

Analysis in $\mathbb{R}^n$ I

Math 20300, Spring 2026

Shuli Chen

based on Boller & Sally and the lecture notes by Yangyang Li

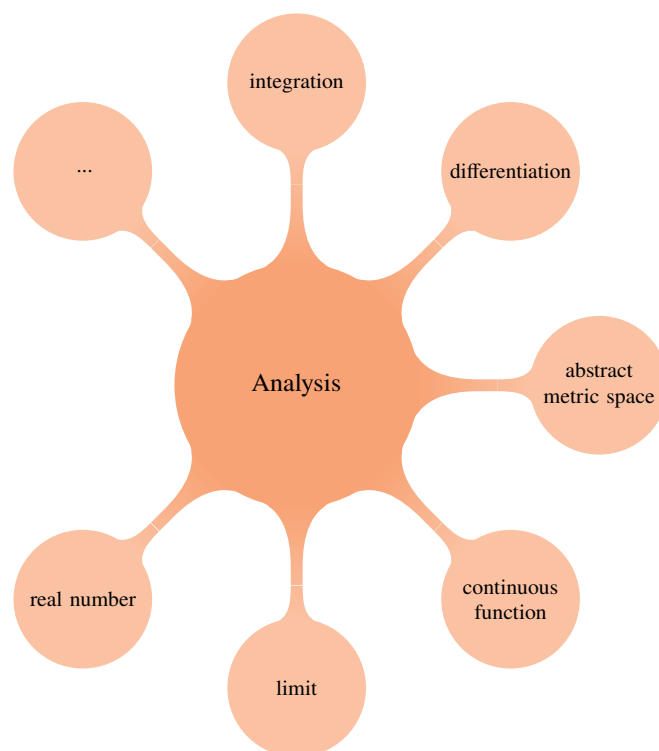
Contents

0	Introduction	5
0.1	What is Analysis?	5
0.2	Why do Analysis?	6
1	The real numbers	9
1.1	Review: the natural numbers, the integers, and the rationals	9
1.2	Cauchy Sequences and the real numbers	10
1.2.1	Operations on Cauchy sequences	12
1.2.2	The construction of the real numbers	12
1.3	Properties of $\mathbb{R}$	13
2	Convergence and limit	19
2.1	Convergence in $\mathbb{R}$	19
2.1.1	More examples	23
2.2	Real Exponentiation	25
2.3	Topology of $\mathbb{R}$	28
2.4	A brief introduction to the geometry of $\mathbb{R}^n$	33
3	Metric spaces	37
3.1	Definitions of metric spaces	37
3.2	Topology of metric spaces	43
3.3	Sequences in metric spaces	49
3.3.1	Sequences in general metric spaces and complete metric spaces	49
3.3.2	The spaces $(\mathbb{R}^n, d_p)$	51

3.3.3	Sequences of functions	53
3.4	Continuous functions	57
3.4.1	Limiting values of functions	57
3.4.2	Interchange of limits	60
3.4.3	Continuous functions	61
3.4.4	When are two spaces the same? (Optional reading)	65
3.5	Compactness	67
3.5.1	Compactness and continuity	70
3.6	Connectedness	72
3.7	The contraction mapping theorem (Optional reading)	75
3.8	The completion of a metric space (Optional reading)	77
3.9	Review	78

0. Introduction

0.1 What is Analysis?



In Analysis I, we will concentrate on the construction of the real numbers, convergence in $\mathbb{R}$, topology of $\mathbb{R}$, the geometry of $\mathbb{R}^n$, and a detailed description of abstract metric spaces. In Analysis II we will deal with differentiation and in Analysis III we will work with integration.

0.2 Why do Analysis?

Let's explore some examples where familiar intuitions fail.

■ **Example 0.1 — Divergent series.** Let's consider the geometric series

$$S = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots.$$

One possible way to sum this series is by multiplying both sides by 2, and we obtain

$$2S = 2 + 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots = 2 + S.$$

Hence, $S = 2$.

However, if we apply the same trick to the series

$$S' = 1 + 2 + 4 + 8 + 16 + \cdots$$

we obtain

$$2S' = 2 + 4 + 8 + 16 + \cdots = S' - 1.$$

This yields $S' = -1$, which is ridiculous. ■

■ **Example 0.2 — Harmonic partial sums.** Consider the sequence $(a_k)_{k=1}^{\infty}$ defined by

$$a_k = \sum_{i=1}^k \frac{1}{i}.$$

Note that

$$a_{k+1} - a_k = \frac{1}{k+1} \rightarrow 0 \quad \text{as } k \rightarrow \infty.$$

Thus, adjacent terms of the sequence become arbitrarily close.

However, as we will learn later, the sequence (a_k) does not converge because it is unbounded.

Each step is small, but there are infinitely many steps, so small errors add up.

Convergence is about total accumulation (a global condition), not just step-by-step behavior (a local condition). ■

■ **Example 0.3 — Interchanging limits.** Let us exam the two limits

$$\lim_{x \rightarrow 0} \lim_{y \rightarrow 0} \frac{x^2}{x^2 + y^2}, \quad \lim_{y \rightarrow 0} \lim_{x \rightarrow 0} \frac{x^2}{x^2 + y^2}$$

At first glance, it may seem that the two limits yield the same result. However, upon closer inspection, we can see that they actually differ:

$$\lim_{x \rightarrow 0} \lim_{y \rightarrow 0} \frac{x^2}{x^2 + y^2} = \lim_{x \rightarrow 0} \frac{x^2}{x^2} = 1,$$

while

$$\lim_{y \rightarrow 0} \lim_{x \rightarrow 0} \frac{x^2}{x^2 + y^2} = 0.$$

■

■ **Example 0.4 — Interchanging integrals.** Let's consider two integrals

$$\int_0^\infty \int_0^1 (e^{-xy} - xye^{-xy}) dy dx, \quad \int_0^\infty \int_0^1 (e^{-xy} - xye^{-xy}) dx dy.$$

A careful computation reveals that they are different.

$$\int_0^\infty \int_0^1 (e^{-xy} - xye^{-xy}) dy dx = \int_0^\infty ye^{-xy} \Big|_{y=0}^{y=1} dx = \int_0^\infty e^{-x} dx = 1,$$

while

$$\int_0^1 \int_0^\infty (e^{-xy} - xye^{-xy}) dx dy = \int_0^1 xe^{-xy} \Big|_{x=0}^{x=\infty} dy = \int_0^1 0 dy = 0,$$

■

These examples underscore the importance of understanding when the computation rules in calculus apply and when they do not. In this course, you will develop an understanding of why mathematical rules work, how to apply them to new situations, and their limitations.

1. The real numbers

1.1 Review: the natural numbers, the integers, and the rationals

The *natural numbers* is typically defined using the five *Peano axioms*. (Basically we introduce the concept of a “successor” for each number, starting from zero.) We *postulate* that such a number system exists, and denote the set of natural numbers by $\mathbb{N}$:

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

There are two binary operations associated with this system, addition $+$ and multiplication $\times$.

The *integers* can be constructed by introducing the binary operation of subtraction $-$. We denote the set of integers by $\mathbb{Z}$ (from the German word *Zahlen*),

$$\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}.$$

The *rational numbers* can be constructed by introducing the binary operation of division $/$. We denote the set of rational numbers by $\mathbb{Q}$.

R $\mathbb{N}^+$ and $\mathbb{Z}^+$ denote the positive integers, and $\mathbb{Q}^+$ denotes the positive rational numbers. Similarly, $\mathbb{Z}^-$ denotes the negative integers, and $\mathbb{Q}^-$ denotes the negative rational numbers.

R If you have not studied the construction of the natural numbers, the integers, and the rational numbers before, you may refer to Chapters 2 and 4 in Terence Tao’s *Analysis I*.

It is clear that there are gaps between the integers. For the rationals, between every two rational numbers there is at least one additional rational.

Proposition 1.1.1 If x and y are two rationals with $x < y$, then there exists a third rational z such that $x < z < y$.

Proof. We can set $z := (x + y)/2$, which satisfies the inequality. ■

Despite this denseness property, there are still “gaps” between the rationals, as suggested by the following instance.

Proposition 1.1.2 There do not exist any rational numbers x for which $x^2 = 2$.

Sketch of proof. Suppose, for the sake of contradiction, that there exists a rational number x for which $x^2 = 2$. Without loss of generality, we may assume that x is positive, so we can write $x = \frac{p}{q}$, where p and q are positive integers that have no common factors (i.e., p and q are coprime).

Then, we have $x^2 = \left(\frac{p}{q}\right)^2 = \frac{p^2}{q^2} = 2$. Multiplying both sides by q^2 , we get $p^2 = 2q^2$.

Since p^2 is even, p must also be even. Let $p = 2k$ for some positive integer k . Substituting into $p^2 = 2q^2$, we get $(2k)^2 = 2q^2$, which simplifies to $4k^2 = 2q^2$. Dividing both sides by 2, we get $2k^2 = q^2$.

Therefore, q^2 is even, which means q must also be even. However, this contradicts our assumption that p and q are coprime. Hence, there cannot be any rational number x such that $x^2 = 2$. ■

Thus, it follows that $\mathbb{Q}$ is not sufficient to represent all the numbers that we encounter in analysis. We need to fill in these gaps and extend $\mathbb{Q}$ to $\mathbb{R}$, called the *real numbers*.

Exercise 1.1 Prove the identity $(a + b)^2 = a^2 + 2ab + b^2$ for all natural numbers a, b using mathematical induction. ■

Exercise 1.2 For every rational number $\varepsilon > 0$, there exists a non-negative rational number x such that $x^2 < 2 < (x + \varepsilon)^2$. ■

1.2 Cauchy Sequences and the real numbers

Intuitively, we would think real numbers as infinite decimals. Look at $\sqrt{2}$ again. What do we know about it? A calculator will tell us $\sqrt{2} = 1.41421\dots$

The key point of this discussion is that when we calculate $\sqrt{2}$ numerically, what we actually obtain is a sequence of approximations to $\sqrt{2}$. We are approximating it by finite decimals

1.4, 1.41, 1.414, 1.4142, 1.41421, ...

But to define real numbers, we want to look beyond finite decimals. So let me define rigorously what a sequence is.

Definition 1.2.1 — sequence. A *sequence* of rational numbers is a function $f : \mathbb{N}^+ \rightarrow \mathbb{Q}$. We often denote such a sequence by $(a_k)_{k=1}^\infty$ or $(a_1, a_2, \dots, a_k, \dots)$, where $a_k = f(k)$.

The aim of this section is to construct the real numbers as limits of sequences of rational numbers. For example, we anticipate that the sequence

1.4, 1.41, 1.414, 1.4142, 1.41421, ...

converges to the limit $\sqrt{2}$.

On the other hand, the following sequence

1, 0, 1, 0, 1, ...

does not seem to have a limit.

So what is a limit?

Definition 1.2.2 — limit. A sequence $(a_k)_{k=1}^{\infty}$ of rational numbers *converges* in $\mathbb{Q}$ to a *limit* $a \in \mathbb{Q}$ provided that for every rational number $r > 0$, there exists an integer $N > 0$ such that for all integers $n \geq N$, we have $|a_n - a| < r$.

In other words, given any prescribed error r , if we go far enough out in the sequence (beyond N) the terms all differ from a by at most r .

R Sometimes, we just say that the sequence $(a_k)_{k=1}^{\infty}$ converges in $\mathbb{Q}$ without mentioning the limit a .

Proposition 1.2.1 The limit a of a convergent sequence is unique.

Exercise 1.3 Prove Proposition 1.2.1. ■

But what if we don't know what the limit is? We need to get the specific limit a out of the definition of limit!

Key observation: on an informal level, if the numbers a_k are getting close to a , they must be getting close to each other.

Definition 1.2.3 — Cauchy sequence. A sequence $(a_k)_{k=1}^{\infty}$ of rational numbers is a *Cauchy sequence* in $\mathbb{Q}$ provided that for every rational number $r > 0$, there exists an integer $N > 0$ such that for all integers $n, m \geq N$, we have $|a_n - a_m| < r$.

But the definition of Cauchy sequence does not involve the limit, as we wanted, so it is not immediately clear that every Cauchy sequence has a limit.

Proposition 1.2.2 If a sequence $(a_k)_{k=1}^{\infty}$ of rational numbers converges in $\mathbb{Q}$, then $(a_k)_{k=1}^{\infty}$ is a Cauchy sequence in $\mathbb{Q}$.

Proof. Suppose that $(a_k)_{k=1}^{\infty}$ converges to $a \in \mathbb{Q}$ and fix an arbitrary $r \in \mathbb{Q}^+$.

By definition, there exists $N \in \mathbb{N}^+$ such that for every integer $n \geq N$,

$$|a_n - a| < \frac{r}{2}.$$

Hence, for all integers $n, m \geq N$,

$$|a_n - a_m| = |a_n - a + a - a_m| \leq |a_n - a| + |a - a_m| < \frac{r}{2} + \frac{r}{2} = r.$$

The arbitrariness of r implies that $(a_k)_{k=1}^{\infty}$ is a Cauchy sequence. ■

Definition 1.2.4 A sequence $(a_k)_{k=1}^{\infty}$ of rational numbers is a *bounded* sequence if the set $\{a_k \mid k \in \mathbb{N}^+\}$ is a bounded set in $\mathbb{Q}$.

R Recall that we say that a set A is bounded if both A is bounded above and A is bounded below.

Lemma 1.2.3 A Cauchy sequence in $\mathbb{Q}$ is a bounded sequence.

Proof. Let $(a_k)_{k=1}^{\infty}$ be a Cauchy sequence in $\mathbb{Q}$. By definition, pick $N \in \mathbb{N}^+$ such that for all $m \geq N$, $|a_m - a_N| < 1$. Therefore, for all $m \geq N$, $|a_m| \leq 1 + |a_N|$.

Let $M = \max\{|a_1|, |a_2|, \dots, |a_{N-1}|, |a_N| + 1\}$. Then $\{a_k \mid k \in \mathbb{N}^+\}$ is bounded above by M . ■

■ **Example 1.1** Let $a_k = \sum_{i=1}^k 1/i$. Then $(a_k)_{k=1}^{\infty}$ is not a bounded sequence (Exercise for you later). ■

1.2.1 Operations on Cauchy sequences

Definition 1.2.5 Given two Cauchy sequences in $\mathbb{Q}$, $(a_k)_{k=1}^\infty$ and $(b_k)_{k=1}^\infty$, we define their *sum* and *product* term-wise,

$$(a_k)_{k=1}^\infty + (b_k)_{k=1}^\infty := (a_k + b_k)_{k=1}^\infty, \quad (1.1)$$

$$(a_k)_{k=1}^\infty (b_k)_{k=1}^\infty := (a_k b_k)_{k=1}^\infty. \quad (1.2)$$

Theorem 1.2.4 The sum and the product of two Cauchy sequences in $\mathbb{Q}$ are Cauchy sequences in $\mathbb{Q}$.

Proof. We only verify the product part, and the sum part is left as an exercise.

Let $(a_k)_{k=1}^\infty$ and $(b_k)_{k=1}^\infty$ be two Cauchy sequences in $\mathbb{Q}$. By Lemma 1.2.3, both sequences are bounded, and hence, we can set $A, B \in \mathbb{Q}^+$ to be upper bounds for $(|a_k|)_{k=1}^\infty$ and $(|b_k|)_{k=1}^\infty$, respectively.

Fix $r \in \mathbb{Q}^+$. Since $(a_k)_{k=1}^\infty$ is a Cauchy sequence, there exists $N_1 \in \mathbb{N}^+$, such that for all integers $n_1, m_1 \geq N_1$, $|a_{n_1} - a_{m_1}| < \frac{r}{2B}$. Similarly, there exists $N_2 \in \mathbb{N}^+$ such that for all integers $n_2, m_2 \geq N_2$, $|b_{n_2} - b_{m_2}| < \frac{r}{2A}$.

Let $N = \max(N_1, N_2)$. For any integers $n, m > N$, we have

$$\begin{aligned} |a_n b_n - a_m b_m| &= |a_n b_n - a_n b_m + a_n b_m - a_m b_m| \\ &\leq |a_n b_n - a_n b_m| + |a_n b_m - a_m b_m| \\ &= |a_n| |b_n - b_m| + |a_n - a_m| |b_m| \\ &< A \frac{r}{2A} + B \frac{r}{2B} \\ &= r. \end{aligned}$$

Hence, $(a_k b_k)_{k=1}^\infty$ is a Cauchy sequence in $\mathbb{Q}$. ■

Exercise 1.4 Prove that the sum and the difference of two Cauchy sequences in $\mathbb{Q}$ are Cauchy sequences in $\mathbb{Q}$. ■

1.2.2 The construction of the real numbers

We want to use Cauchy sequences in $\mathbb{Q}$ to define the real numbers, but this set is too large in the sense that it contains redundant terms, as shown in the following example.

■ **Example 1.2** $(0.9, 0.99, 0.999, 0.9999, 0.99999, \dots)$ and $(1.0, 1.00, 1.000, 1.0000, 1.00000, \dots)$ are two different sequences, but they converge to the same limit 1. Hence, we anticipate that they are equivalent. ■

Let C denote the set of all Cauchy sequences of rational numbers. Let $\mathcal{I}$ be the set of Cauchy sequences in $\mathbb{Q}$ that converge to 0.

Lemma 1.2.5 Suppose that $(a_k)_{k=1}^\infty \in C \setminus \mathcal{I}$. Then there exists a positive rational number r and an integer N such that $|a_n| \geq r$ for all $n \geq N$.

Proof. Since $(a_k)_{k=1}^\infty \notin \mathcal{I}$, there exists a rational number $r > 0$, such that for infinitely many k , $|a_k| \geq 2r$.

Since $(a_k)_{k=1}^\infty \in C$, we can choose a positive integer N such that for all integers $n, m \geq N$, $|a_n - a_m| < r$.

Let $m \geq N$ be an integer such that $|a_m| \geq 2r$. Then for all $n \geq N$, we have

$$|a_n| = |a_n - a_m + a_m| \geq |a_m| - |a_n - a_m| \geq r.$$

We say that a property of rational numbers holds *eventually* for the terms of a sequence $(a_k)_{k=1}^{\infty}$ if there exists some $N \in \mathbb{N}^+$ such that the property holds for a_n whenever $n \geq N$. The lemma above says that for $(a_k)_{k=1}^{\infty} \in C \setminus \mathcal{I}$, $|a_k|$ is greater than some positive rational number eventually.

R $\mathcal{I}$ is expected to represent 0 of the real numbers, and the lemma above confirms that any sequence outside $\mathcal{I}$ is eventually “far away” from 0.

Definition 1.2.6 Given $(a_k)_{k=1}^{\infty}, (b_k)_{k=1}^{\infty} \in C$, we say that $(a_k)_{k=1}^{\infty}$ is *equivalent* to $(b_k)_{k=1}^{\infty}$, denoted by $(a_k)_{k=1}^{\infty} \sim (b_k)_{k=1}^{\infty}$, provided that $(c_k)_{k=1}^{\infty} := (a_k - b_k)_{k=1}^{\infty} \in \mathcal{I}$.

Lemma 1.2.6 $\sim$ defines an equivalence relation on C .

Sketch of proof. One can verify that the relation $\sim$ is

- *reflexive:* $(a_k)_{k=1}^{\infty} \sim (a_k)_{k=1}^{\infty}$;
- *symmetric:* If $(a_k)_{k=1}^{\infty} \sim (b_k)_{k=1}^{\infty}$, then $(b_k)_{k=1}^{\infty} \sim (a_k)_{k=1}^{\infty}$;
- *transitive:* If $(a_k)_{k=1}^{\infty} \sim (b_k)_{k=1}^{\infty}$ and $(b_k)_{k=1}^{\infty} \sim (c_k)_{k=1}^{\infty}$, then $(a_k)_{k=1}^{\infty} \sim (c_k)_{k=1}^{\infty}$.

Therefore, it is an equivalence relation. ■

We denote the set of equivalence classes in C by $\mathbb{R}$, namely, the set of *real numbers*. In particular, $\mathcal{I}$ is 0 in $\mathbb{R}$.

1.3 Properties of $\mathbb{R}$

If $(a_k)_{k=1}^{\infty}$ is a Cauchy sequence in $\mathbb{Q}$, we denote its equivalence class by $[a_k]$. That is,

$$[a_k] = \{(a'_k)_{k=1}^{\infty} \text{ is a Cauchy sequence in } \mathbb{Q} \mid (a_k)_{k=1}^{\infty} \sim (a'_k)_{k=1}^{\infty}\}.$$

One can define the sum and the product of equivalence classes as follows:

$$[a_k] + [b_k] = [a_k + b_k], \tag{1.3}$$

$$[a_k][b_k] = [a_k b_k]. \tag{1.4}$$

Exercise 1.5 Prove that addition and multiplication are well-defined on $\mathbb{R}$, that is, independent of representatives. ■

Let's recall the definition of *field*. Basically, a field is a set on which addition, subtraction, multiplication, and division are defined as we expected them to be.

Definition 1.3.1 A *field* is triple $(F, +, \cdot)$ where F is a set, and $+$ and $\cdot$ are binary operations on F called *addition* and *multiplication* satisfying the following *field axioms*

- *Associativity:* $(a + b) + c = a + (b + c)$, $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.
- *Commutativity:* $a + b = b + a$, $a \cdot b = b \cdot a$.
- *Distributivity:* $a \cdot (b + c) = a \cdot b + a \cdot c$.
- *Identity:* There exist 0 and 1 in F , such that $a + 0 = a$, $a \cdot 1 = a$.
- *Inverses:* For every $a \in F$, there exists $-a \in F$, such that $a + (-a) = 0$; for every $0 \neq a \in F$, there exists $a^{-1} \in F$, such that $a \cdot a^{-1} = 1$.

R For simplicity, we often say that F is a field without mentioning the addition operator and the multiplication operator.

Theorem 1.3.1 $\mathbb{R}$ is a field.

Part of proof. We only show that multiplicative inverses exist for non-zero elements.

Assume that $[a_k] \neq \mathcal{I}$, and $(a_k)_{k=1}^\infty$ is a representative sequence in $\mathbb{Q}$, i.e., $(a_k)_{k=1}^\infty \in [a_k]$. Our goal is to construct $[a_k]^{-1}$. By Lemma 1.2.5, a_k is eventually non-zero, namely, we can choose $M \in \mathbb{N}^+$ and $c > 0$, such that $|a_k| > c$ for all $k \geq M$.

Define a sequence $(b_k)_{k=1}^\infty$ by

$$b_k = \begin{cases} 1 & k < M \\ \frac{1}{a_k} & k \geq M. \end{cases}$$

For any $r \in \mathbb{Q}^+$, there exists an integer $N > M$, such that for all integers $n, m \geq N$, $|a_n - a_m| < c^2 r$. Therefore, for all integers $n, m \geq N$,

$$|b_n - b_m| = \left| \frac{1}{a_n} - \frac{1}{a_m} \right| = \frac{|a_n - a_m|}{|a_n a_m|} \leq \frac{|a_n - a_m|}{c^2} < \frac{c^2 r}{c^2} = r.$$

This shows that $(b_k)_{k=1}^\infty$ is a Cauchy sequence. It is also clear that $[b_k]$ is the multiplicative inverse of $[a_k]$. ■

End of Week 1. To be continued...

Let's recall the definition of an ordering relation.

Definition 1.3.2 — poset, ordering relation. A *partially ordered set* (or *poset*) is a set X , together with a relation $\leq_X$ on X , satisfying the following three properties:

- *Reflexivity*: $\forall x \in X, x \leq_X x$.
- *Antisymmetry*: If $x, y \in X$ are such that $x \leq_X y$ and $y \leq_X x$, then $x = y$.
- *Transitivity*: If $x, y, z \in X$ are such that $x \leq_X y$ and $y \leq_X z$, then $x \leq_X z$.

We call $\leq_X$ the *ordering relation*, and write $x <_X y$ if $x \leq_X y$ and $x \neq y$.

■ **Example 1.3** When X is a collection of sets, we can define $x \leq_X y$ in X when x is a subset of y . For example, $\{2, 7\} < \{2, 5, 7, 8\}$ and $\{5, 8, 9\} \leq \{5, 8, 9\}$. ■

■ **Example 1.4** Let $X = \{1, 2, 3, 4\}$, and let $a \leq_X b$ if b is divisible by a . Then we can verify that X is a poset.

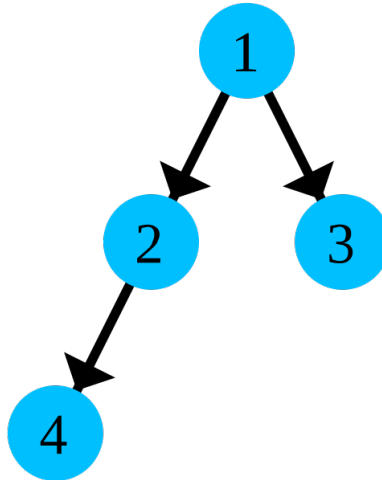


Figure 1.1: Graph of the divisibility of numbers from 1 to 4. Note that 2 and 3 are incomparable here, so this is not a totally ordered set.

For a partially ordered set, some elements might be incomparable. If it turns out that we can compare the order of any two elements, then we have a totally ordered set.

Definition 1.3.3 — totally ordered set. Let X be a poset with $\leq_X$. We call X a *totally ordered set* if $\forall x, y \in X$, we either have $x \leq_X y$ or $y \leq_X x$.

If the totally ordered set turns out to be a field, we also want the order relation to be compatible with the field operations.

Definition 1.3.4 — totally ordered field (also known as ordered field). A field $(F, +, \cdot)$ is a *(totally) ordered field* if

- F is a totally ordered set;
- $\forall x, y, z \in F$, if $x \leq_F y$, then $x + z \leq_F y + z$;
- $\forall x, y \in F$, if $0 \leq_F x$ and $0 \leq_F y$, then $0 \leq_F x \cdot y$.

The next step is to define an order on $\mathbb{R}$.

Definition 1.3.5 — order relation on $\mathbb{R}$. Let $a = [a_k], b = [b_k]$ be *distinct* elements of $\mathbb{R}$, i.e., $a \neq b$. We define $a < b$ if $a_k < b_k$ eventually, and $a > b$ if $a_k > b_k$ eventually.

For $a, b \in \mathbb{R}$, we define $a \leq b$ if $a < b$ or $a = b$.

R Why do we need to define this way? Think about what goes wrong if we don't ask a, b to be distinct, and what goes wrong if we instead define $a \leq b$ if $a_k \leq b_k$ eventually.

Lemma 1.3.2 The order relation on $\mathbb{R}$ defined above is well-defined and makes $\mathbb{R}$ a totally ordered field.

Exercise 1.6 Prove Lemma 1.3.2.

Here is something useful that we'll use a lot. If $x = [x_k]$ is a real number represented by a Cauchy sequence of rationals $(x_k)_{k=1}^\infty$, we want to say that x is the “limit” of the sequence $(x_k)_{k=1}^\infty$. Before we can do that, by looking at the definition of a limit, we need to prove a Cauchy sequence eventually stays in any neighborhood of the real number it represents. (Notice that we haven't defined what absolute values means for real numbers yet.)

Lemma 1.3.3 Let $x = [x_k]$ be a real number represented by a Cauchy sequence of rationals $(x_k)_{k=1}^\infty$.

1. For any rational $r > 0$, we have $x - r \leq x_n \leq x + r$ eventually.
2. Let $a \in \mathbb{R}$. If $a \leq x_k$ eventually, then $a \leq x$. Similarly, if $x_k \leq a$ eventually, then $x \leq a$.

R

- Note that in Part 1, we can also have the strict inequality instead.
- Note that in Part 2, we have the non-strict inequalities everywhere. Is it still true for strict inequalities?

Proof. 1. We'll prove $x_n \leq x + r$ eventually and the other side will be similar. Suppose by contradiction that there exists rational $r > 0$ such that $x_n > x + r$ for infinitely many n .

Since $(x_k)_{k=1}^\infty$ is a Cauchy sequence, there exists $N_1 \in \mathbb{N}^+$ such that $|x_n - x_m| < \frac{r}{2}$ for $m, n \geq N_1$.

Let $(y_k)_{k=1}^\infty$ be another Cauchy sequence of rationals representing x . Since $(x_k)_{k=1}^\infty, (y_k)_{k=1}^\infty$ both represent x , there exists $N_2 \in \mathbb{N}^+$ such that $|x_m - y_m| < \frac{r}{2}$ for $m \geq N_2$.

Let $N = \max\{N_1, N_2\}$. Then by triangle inequality, for $m, n \geq N$,

$$|x_n - y_m| < |x_n - x_m| + |x_m - y_m| < r.$$

On the other hand, since $x_n > x + r$ for infinitely many n , then fixing such an $n \geq N$. By the definition of the order on $\mathbb{R}$, we have that there exists $N' \in \mathbb{N}^+$ such that for all $m \geq N'$, $x_n > y_m + r$.

This contradicts $|x_n - y_m| < r$, so we must have $x_n \leq x + r$ eventually.

2. Exercise. ■

R Since we can choose any representative of $[x]$, we can also simplify this proof by using the same (x_k) instead of (y_k) .

The following theorem shows that $\mathbb{R}$ is an *Archimedean ordered field*.

Theorem 1.3.4 $\mathbb{R}$ has the *Archimedean* property. Namely, if a and b are positive real numbers, then there exists a natural number n such that $na > b$.

R Roughly speaking, the Archimedean property is the property of having no infinitely large or infinitely small elements.

What we will use is that for any $x \in \mathbb{R}^+$, there exists $n \in \mathbb{N}^+$ such that $1/n < x < n$.

Proof. Let $(c_k)_{k=1}^\infty$ be a representation of $a^{-1}b \in \mathbb{R}$. Since $(c_k)_{k=1}^\infty$ is a Cauchy sequence, by Lemma 1.2.3, it has an upper bound $n_0 \in \mathbb{N}$, i.e., $n_0 \geq c_k$. By the lemma above, we have $n_0 \geq a^{-1}b$. Let $n = n_0 + 1$, and then it is clear that $na > b$. ■

Let's recall the definition of upper bound.

Definition 1.3.6 — Upper bound. Let $E \subset \mathbb{R}$ and M be a real number. We say that M is an *upper bound* for E if $\forall x \in E, x \leq M$.

If E has an upper bound, then we say E is bounded from above.

Definition 1.3.7 — Least upper bound. Let $E \subset \mathbb{R}$ and M be a real number. We say that M is a *least upper bound* for E if M is an upper bound for E and any other upper bound M' for E must be larger than or equal to M .

R It follows immediately from the definition that the least upper bound is *unique*.

Theorem 1.3.5 The *least upper bound property* holds in $\mathbb{R}$. Namely, every nonempty subset E of $\mathbb{R}$ that is bounded from above has a least upper bound, denoted by $\sup(E)$ (*supremum* of E).

Proof. Let E be a nonempty subset of $\mathbb{R}$ that is bounded above by $m \in \mathbb{R}^+$ and let $a \in E$. By the Archimedean property, we can choose $M, N \in \mathbb{Z}$ with $M < a$ and $m < N$. (Note: We don't need M to be a lower bound of E ; we just need M to be smaller than some number of E so that M is not an upper bound of E .) [Why do we choose M, N ? This is not to bound E since E could go all the way to $-\infty$, but to bound the upper bounds a_p that we'll construct. We'll only consider upper bounds living between M and N .]

We will define a sequence $(a_p)_{p=1}^\infty$, and show it is a Cauchy sequence that represents a least upper bound for E . To do so, we want to let each a_p be an upper bound and keep pushing it downward.

To be more precise, for $p \in \mathbb{N}^+$, we define a subset of $\mathbb{Q}$, $S_p := \{k2^{-p} \mid k \in \mathbb{Z}, M \leq k2^{-p} \leq N\}$. Note that S_p is nonempty and finite [Think about why this is the case], and $M, N \in S_p$. We set $a_p := \min\{x \in S_p \mid x \text{ is an upper bound for } E\}$. We then have a_p exists because N is a candidate, and $a_p > M$. [Why is $a_p > M$? Because $M < a$, and $a_p \geq a$ since a_p is an upper bound.]

For all $p < q \in \mathbb{N}^+$, since $a_p \in S_q$, we have $a_q \leq a_p$. Since $a_p > M$, we have $a_p - 2^{-p} \geq M$. [Why do we get $a_p - 2^{-p} \geq M$? This is because the distance between two distinct elements in S_p is at least 2^{-p} .] Since $a_p - 2^{-p}$ is not an upper bound for E [Think about why this is the case] while a_q is an upper bound, we also have $a_p - 2^{-p} < a_q$. In summary, for all $p < q$,

$$|a_p - a_q| \leq 2^{-p},$$

which implies that $(a_p)_{p=1}^\infty$ is a decreasing Cauchy sequence.

Let $L = [a_p]$. We claim that L is a least upper bound for E .

We show L is an upper bound of E by contradiction. [Note here how it is easier to construct a proof by contradiction here. To show L is an upper bound directly, we need to check the condition $x \leq L$ for all $x \in E$. In comparison, not being an upper bound gives some $x \in E$ with $x > L$ and it is easier to work with one single x . Also as we have seen in other proofs, strict inequalities are usually easier to deal with.] Suppose $x \in E$ and $x > L$. Then $x - L > 0$. By the Archimedean property again, choose $p \in \mathbb{N}^+$ such that $2^{-p} < x - L$. Since $a_p - 2^{-p} < a_q \leq a_p$ for $p < q$ and $(a_k)_{k=1}^\infty$ is a decreasing Cauchy sequence, by fixing p and applying Lemma 1.3.3 Part 2, these yield

$$a_p - 2^{-p} \leq L \leq a_p.$$

[Think about how this works. We've already seen this useful idea of fixing one subscript and taking another to infinity in the proof of Lemma 1.3.3 Part 1. Also notice that even though the a_q 's satisfy a strict inequality $a_p - 2^{-p} < a_q$, L only satisfies a non-strict one $a_p - 2^{-p} \leq L$.]

In particular, $a_p \leq L + 2^{-p} < x$ which is a contradiction to a_p being an upper bound. Therefore, L is an upper bound for E .

Suppose that H is an upper bound for E but $H < L$. Choose $p \in \mathbb{N}^+$ such that $2^{-p} < L - H$ and $x \in E$ such that $a_p - 2^{-p} < x$. Then, we have

$$a_p < x + 2^{-p} \leq H + 2^{-p} < L.$$

This contradicts the fact that $L \leq a_p$ for all $p \in \mathbb{N}^+$. [Here we used the decreasing condition $a_q \leq a_p$ to get $L \leq a_p$ and then derive a contradiction. If we don't have this decreasing condition, how do we proceed? Thinking about this will be helpful for solving the exercise below.]

In conclusion, L is a least upper bound. ■

■ **Example 1.5** The least upper bound property does not hold for $\mathbb{Q}$. For example, let $E = \{q \in \mathbb{Q} \mid q^2 < 2\}$. Then E is nonempty and bounded above in $\mathbb{Q}$. But its least upper bound would be $\sqrt{2}$, which is not rational. ■

■ **Example 1.6** Let $E = [-1, \frac{1}{6}]$. Let's see how the proof works in this case. We can choose $M = 0$, $N = 1$. Then the first few terms of the sequence (a_p) are

$$a_1 = \frac{1}{2}, a_2 = \frac{1}{4}, a_3 = \frac{1}{4}, a_4 = \frac{3}{16}, a_5 = \frac{3}{16}, a_6 = \frac{11}{64}$$

Draw a picture yourself to see it more clearly. ■

Exercise 1.7 In the above proof, we don't need to choose a_p as dyadic rationals. Here is another way to prove the theorem:

Let $E \subset \mathbb{R}$ be nonempty and bounded above.

1. For each $n \in \mathbb{N}^+$, show we can find $a_n \in \mathbb{Q}$ such that a_n is an upper bound of E but $a_n - \frac{1}{n}$ is not an upper bound.
2. Show that $(a_n)_{n=1}^\infty$ is a Cauchy sequence. (Note in this case it is not necessarily decreasing)
3. Prove that $L = [a_n]$ is a least upper bound of E .

■

Corollary 1.3.6 Every nonempty subset E of $\mathbb{R}$ that is bounded from below has a greatest lower bound, denoted by $\inf(E)$ (*infimum* of E).

R In fact, any two ordered fields with the least upper bound property are order isomorphic. Thus $\mathbb{R}$ can also be defined as the (unique) totally ordered field with the least upper bound property. In Rudin's book, you can find another construction of real numbers using Dedekind cuts.

Further Reading

- Boller & Sally: Section 1.5
- Rudin: Chapter 1
- Tao, *Analysis I*: Chapter 5

2. Convergence and limit

2.1 Convergence in $\mathbb{R}$

In the previous lectures, we've defined convergence in $\mathbb{Q}$, Cauchy sequences in $\mathbb{Q}$, etc. In this section, we will study the similar concepts in $\mathbb{R}$.

Definition 2.1.1 — absolute value. For $x \in \mathbb{R}$, the *absolute value* of x is defined by

$$|x| := \begin{cases} x & x \geq 0, \\ -x & x < 0. \end{cases}$$

Lemma 2.1.1 The absolute value in $\mathbb{R}$ has the following properties:

- $\forall x \in \mathbb{R}, |x| \geq 0$; and $|x| = 0 \iff x = 0$.
- $\forall x, y \in \mathbb{R}, |xy| = |x||y|$.
- $\forall x, y \in \mathbb{R}, |x + y| \leq |x| + |y|$ (triangle inequality).

Definition 2.1.2 — sequence and subsequence. A *sequence* of real numbers is a function $f : \mathbb{N}^+ \rightarrow \mathbb{R}$. We often denote such a sequence by $(a_k)_{k=1}^\infty$ or $(a_1, a_2, \dots, a_k, \dots)$, where $a_k = f(k)$.

Given a sequence $(a_k)_{k=1}^\infty$, a *subsequence* is a sequence of the form $(a_{k_j})_{j=1}^\infty$, where $(k_j)_{j=1}^\infty$ is a strictly monotonic increasing sequence of natural numbers, i.e., $k_{j+1} > k_j$.

In other words, a subsequence of a sequence is obtained by skipping some terms while still keeping infinitely many terms. In particular, $k_j \geq j$. A subsequence contains some, but not necessarily all, terms of the original sequence in their original order.

End of Week 2. To be continued...

Definition 2.1.3 — limit. A sequence $(a_k)_{k=1}^\infty$ of real numbers *converges* to a *limit* $a \in \mathbb{R}$ if for every $\varepsilon > 0$, there exists $N \in \mathbb{N}^+$ such that for every $k \geq N$, we have $|a_k - a| < \varepsilon$. Symbolically,

we write $\lim_{k \rightarrow \infty} a_k = a$.

In this case, we also say that the sequence $(a_k)_{k=1}^{\infty}$ is *convergent*.

If a sequence does not converge, then it is said to *diverge* or to be a *divergent sequence*.

R Note that in the definition of limit, instead of checking the condition “for every $\varepsilon > 0$ ”, it suffices to verify the condition “for every rational $\varepsilon > 0$ ”. Think about why this is the case (recall the Archimedean property).

R In Lemma 1.3.3, we showed that if $x \in \mathbb{R}$ is represented by a Cauchy sequence of rationals $(x_n)_{n=1}^{\infty}$, then $(x_n)_{n=1}^{\infty}$ converges to the limit x .

Exercise 2.1 Use the definition of limit (the ε - N language) to find

$$\lim_{n \rightarrow \infty} \frac{3n^2}{n^2 - 3}.$$

Hint: In this kind of problem that asks you to use the definition directly, you need to first “guess” what the limit is, and then write an ε - N proof. ■

Lemma 2.1.2 The limit a of a convergent sequence is unique.

Proof. Exercise. Cf. Proposition 1.2.1. ■

Exercise 2.2 Let $(a_k)_{k=1}^{\infty}$ be a convergent sequence. Show that any subsequence converges to the same limit. ■

Exercise 2.3 Let $(a_{k_j})_{j=1}^{\infty}$ be a subsequence of $(a_k)_{k=1}^{\infty}$. If

(1) $(a_{k_j})_{j=1}^{\infty}$ converges to a real number a , and

(2) for all $j \in \mathbb{N}^+$ and $k \in \mathbb{N}^+$, if $k_j \leq k \leq k_{j+1}$, then either $a_{k_j} \leq a_k \leq a_{k_{j+1}}$ or $a_{k_{j+1}} \leq a_k \leq a_{k_j}$.

Then $(a_k)_{k=1}^{\infty}$ converges to a .

Also give an example to show that the conclusion can fail if condition (2) is removed. ■

Definition 2.1.4 A sequence $(a_k)_{k=1}^{\infty}$ is *bounded* if the set $\{a_k \mid k \in \mathbb{N}^+\}$ is bounded.

Lemma 2.1.3 Any convergent sequence of real numbers is bounded.

Proof. Exercise. Cf. Proposition 1.2.2 and Lemma 1.2.3. ■

Definition 2.1.5 — monotone sequence. A sequence $(a_k)_{k=1}^{\infty}$ is *increasing* (strictly increasing) if $a_k \leq a_{k+1}$ ($a_k < a_{k+1}$), $\forall k \in \mathbb{N}^+$.

A sequence $(a_k)_{k=1}^{\infty}$ is *decreasing* (strictly decreasing) if $a_k \geq a_{k+1}$ ($a_k > a_{k+1}$), $\forall k \in \mathbb{N}^+$.

A sequence is called *monotone* if it is increasing or decreasing.

Lemma 2.1.4 Let $(a_k)_{k=1}^{\infty}$ be a sequence in $\mathbb{R}$. Then $(a_k)_{k=1}^{\infty}$ has a monotone subsequence.

Proof. Suppose $(a_k)_{k=1}^{\infty}$ does not have an increasing subsequence. [Note that if for all $n \in \mathbb{N}^+$, there exists $k > n$ such that $a_k \geq a_n$, then $(a_k)_{k=1}^{\infty}$ has an increasing subsequence. Think about why this is the case, and form the contrapositive.] Then, there exists $n_1 \in \mathbb{N}^+$ such that $\forall k > n_1$, $a_{n_1} > a_k$. Again, since $(a_k)_{k > n_1}$ does not have an increasing subsequence, there exists $n_2 > n_1$

such that $\forall k > n_2, a_{n_2} > a_k$. Note that $a_{n_1} > a_{n_2}$. Hence, we can inductively construct a strictly decreasing sequence $(a_{n_i})_{i=1}^{\infty}$. ■

R Here the notation $(a_k)_{k>n_1}$ just means the sequence $(a_{n_1+1}, a_{n_1+2}, \dots)$.

You might also see the notation $(a_k)_{k \in \mathbb{N}^+}$ which means the same thing as $(a_k)_{k=1}^{\infty}$.

In general, for a set $S \subset \mathbb{N}$, a notation $(a_k)_{k \in S}$ means the sequence of the a_k 's where the subscript k comes from the set S .

Lemma 2.1.5 Every bounded monotone sequence in $\mathbb{R}$ is convergent.

Proof. Without loss of generality, let $(a_k)_{k=1}^{\infty}$ be a bounded increasing sequence, and let a be the least upper bound of the set $\{a_k \mid k \in \mathbb{N}^+\}$.

For any $\varepsilon > 0$, there exists an N such that $a - \varepsilon < a_N \leq a$ because a is the least upper bound. Since $(a_k)_{k=1}^{\infty}$ is increasing, for all $n \geq N$, we also have $a \geq a_n \geq a_N > a - \varepsilon$ and in particular, $|a_n - a| < \varepsilon$. Hence, $\lim_{n \rightarrow \infty} a_n = a$. ■

Lemma 2.1.6 Every bounded sequence in $\mathbb{R}$ has a convergent subsequence.

Proof. This is a consequence of Lemma 2.1.4 and Lemma 2.1.5. ■

Exercise 2.4 1. Consider the sequence $(a_k)_{k=1}^{\infty}$ of real numbers given by $a_k = \sum_{i=1}^k \frac{1}{i}$. Show it doesn't converge, but for all $\varepsilon > 0$, there exists $N \in \mathbb{N}^+$ such that for all $k \geq N$, we have $|a_{k+1} - a_k| < \varepsilon$.

Remark: This shows for a sequence to converge, it doesn't suffice that adjacent terms get as close as possible.

2. Show the sequence $(b_k)_{k=1}^{\infty}$ of real numbers given by $b_k = \sum_{i=1}^k \frac{1}{i^2}$ is convergent. ■

Definition 2.1.6 — limit supremum and infimum. Let $(a_n)_{n=1}^{\infty}$ be a bounded sequence in $\mathbb{R}$. For each $n \in \mathbb{N}^+$, let $b_n = \sup\{a_k \mid k \geq n\}$ and $c_n = \inf\{a_k \mid k \geq n\}$.

We define the *limit supremum* of $(a_n)_{n=1}^{\infty}$ by $\limsup_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} b_n$.

We define the *limit infimum* of $(a_n)_{n=1}^{\infty}$ by $\liminf_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} c_n$.

R By definition, we can see $\liminf \leq \limsup$ always.

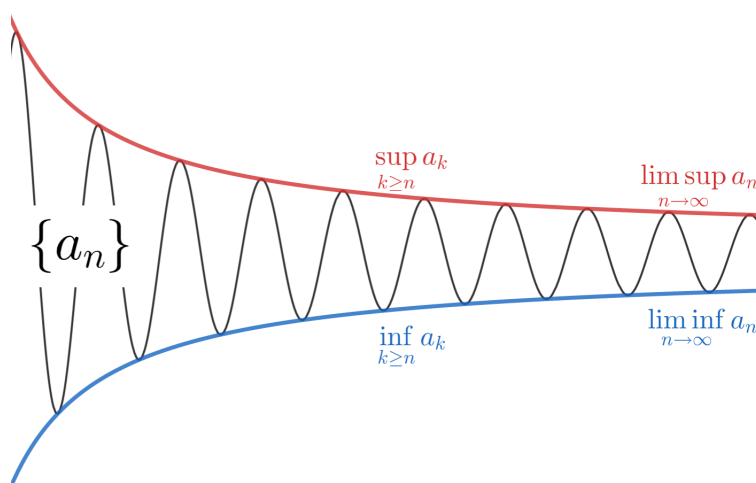


Figure 2.1: Visualization of liminf and limsup

Exercise 2.5 Show that the limit supremum and the limit infimum of any bounded sequence always exist. ■

■ **Example 2.1** Consider a sequence $(a_n)_{n=1}^{\infty}$ given by $a_n = 1 + (-1)^n$. Then $\limsup_{n \rightarrow \infty} a_n = 2$ and $\liminf_{n \rightarrow \infty} a_n = 0$.

Consider another sequence $(b_n)_{n=1}^{\infty}$ given by $b_n = \frac{1}{n} + (-1)^n$. Then $\limsup_{n \rightarrow \infty} b_n = 1$ and $\liminf_{n \rightarrow \infty} b_n = -1$. ■

We can think of the lim sup as “the ceiling the sequence approaches in the long run,” and lim inf is “the floor the sequence approaches in the long run.” The lim sup and lim inf capture the oscillatory behavior. If $\liminf = \limsup$, then we should expect the sequence to have no oscillation, so it should be convergent.

Lemma 2.1.7 A bounded sequence $(a_n)_{n=1}^{\infty}$ is convergent if and only if $\limsup_{n \rightarrow \infty} a_n = \liminf_{n \rightarrow \infty} a_n$. In this case, $\lim_{n \rightarrow \infty} a_n = \limsup_{n \rightarrow \infty} a_n = \liminf_{n \rightarrow \infty} a_n$.

Exercise 2.6 Prove Lemma 2.1.7. ■

Exercise 2.7 Is it true that for a bounded sequence $(a_n)_{n=1}^{\infty}$, there always exists some $m \in \mathbb{N}^+$ such that $\liminf_{n \rightarrow \infty} a_n \leq a_m$? Prove or give a counterexample. ■

Definition 2.1.7 — Cauchy sequence. A sequence $(a_k)_{k=1}^{\infty}$ in $\mathbb{R}$ is a *Cauchy sequence* if for every $\varepsilon > 0$, there exists $N \in \mathbb{N}^+$ such that for all $n, m \geq N$, $|a_n - a_m| < \varepsilon$.

Lemma 2.1.8 Any Cauchy sequence is a bounded sequence.

Proof. It follows from a similar argument as in the proof of Lemma 1.2.3. ■

What’s nice about Cauchy sequences in $\mathbb{R}$ is the following theorem. Recall a Cauchy sequence in $\mathbb{Q}$ might not converge in $\mathbb{Q}$; the limit could be in $\mathbb{R}$ instead. In contrast, every Cauchy sequence in $\mathbb{R}$ does converge in $\mathbb{R}$. There is no larger ordered field in which it could “go further”.

Theorem 2.1.9 — Cauchy criterion. A sequence $(a_k)_{k=1}^{\infty}$ in $\mathbb{R}$ is convergent if and only if it is a Cauchy sequence.

Proof. If $(a_k)_{k=1}^{\infty}$ is convergent, then by the same argument as in the proof of Proposition 1.2.2, we can conclude that it is a Cauchy sequence.

If $(a_k)_{k=1}^{\infty}$ is a Cauchy sequence, by Lemma 2.1.8, we know that it is a bounded sequence. Hence, it has a convergent subsequence $(a_{k_i})_{i=1}^{\infty}$ with the limit a . For every $\varepsilon > 0$, there exists $N \in \mathbb{N}^+$ such that for all $n, m \geq N$, $|a_n - a_m| < \varepsilon/2$, and for all $n \geq N$, $|a_{k_n} - a| < \varepsilon/2$. Since $k_N \geq N$, for all $n \geq k_N$, we have

$$|a_n - a| \leq |a_n - a_{k_N}| + |a_{k_N} - a| < \varepsilon.$$

Hence, $(a_k)_{k=1}^{\infty}$ is a convergent sequence. ■

Exercise 2.8 Prove that if $\lim_{n \rightarrow \infty} a_n = a \neq 0$, then $\lim_{n \rightarrow \infty} \frac{a_{n+1}}{a_n} = 1$. Provide an example that shows this is not true when $a = 0$. ■

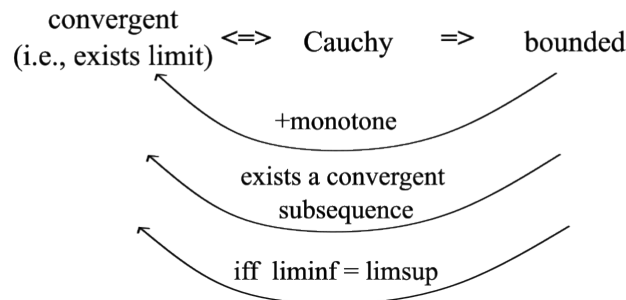
To summarize:

For sequences in $\mathbb{Q}$,

convergent $\implies$ Cauchy $\implies$ bounded

For sequences in $\mathbb{R}$, we have the following picture:

Sequence in $\mathbb{R}$



Proposition 2.1.10 — Limit laws for sequences. Let $(a_k)_{k=1}^{\infty}$ and $(b_k)_{k=1}^{\infty}$ be two sequences in $\mathbb{R}$. If $\lim_{k \rightarrow \infty} a_k = A$ and $\lim_{k \rightarrow \infty} b_k = B$, then

1. $\lim_{k \rightarrow \infty} (a_k + b_k) = A + B$.
2. $\lim_{k \rightarrow \infty} (a_k \cdot b_k) = A \cdot B$.
3. $\lim_{k \rightarrow \infty} (a_k/b_k) = A/B$, if $B \neq 0$, and $b_k \neq 0$.

Exercise 2.9 Suppose that $a_1 = \sqrt{2}$ and $a_{n+1} = \sqrt{2 + a_n}$ for all $n \in \mathbb{N}^+$. Show that $\lim_{n \rightarrow \infty} a_n$ exists, and compute the limit using limit laws. ■

2.1.1 More examples

■ **Example 2.2 — squeeze theorem.** If $a_n \leq b_n \leq c_n$ and $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} c_n = A$, then $\lim_{n \rightarrow \infty} b_n = A$. ■

Proof. You can show it directly using the definition. Alternatively, we can try to use what we learned. Since (a_n) , (c_n) are convergent, they are bounded. Since $a_n \leq b_n \leq c_n$, (b_n) is also

bounded. Then by $a_n \leq b_n \leq c_n$, we have

$$\liminf_{n \rightarrow \infty} a_n \leq \liminf_{n \rightarrow \infty} b_n \leq \limsup_{n \rightarrow \infty} b_n \leq \limsup_{n \rightarrow \infty} c_n.$$

Since $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} c_n = A$, $\liminf_{n \rightarrow \infty} a_n = \limsup_{n \rightarrow \infty} c_n = A$, and thus,

$$\liminf_{n \rightarrow \infty} b_n = \limsup_{n \rightarrow \infty} b_n = A.$$

Consequently, $\lim_{n \rightarrow \infty} b_n$ exists and $\lim_{n \rightarrow \infty} b_n = A$. ■

■ **Example 2.3** If $p > 0$, then $\lim_{n \rightarrow \infty} \frac{1}{n^p} = 0$. ■

Proof. For all $\varepsilon > 0$, take $N > (1/\varepsilon)^{1/p}$, and then for all $n \geq N$, we have $|\frac{1}{n^p} - 0| < \varepsilon$. ■

R We will rigorously define what the exponential n^p means for general real power p in the next section.

■ **Example 2.4** If $p > 0$, then $\lim_{n \rightarrow \infty} \sqrt[p]{p} = 1$. ■

Proof. If $p > 1$, set $x_n = \sqrt[p]{p} - 1 > 0$. By the binomial theorem, $1 + nx_n \leq (1 + x_n)^n = p$, so $0 < x_n \leq \frac{p-1}{n}$. Hence, $\lim_{n \rightarrow \infty} x_n = 0$.

If $p = 1$, then $\lim_{n \rightarrow \infty} \sqrt[p]{p} = \lim_{n \rightarrow \infty} 1 = 1$ holds trivially.

If $p < 1$, then $\lim_{n \rightarrow \infty} \sqrt[p]{p} = \lim_{n \rightarrow \infty} 1/\sqrt[p]{1/p} = 1/\lim_{n \rightarrow \infty} \sqrt[p]{1/p} = 1$ using the quotient law.

[If you try to do this $p < 1$ case directly and set $x_n = 1 - \sqrt[p]{p} > 0$ and compute $p = (1 - x_n)^n$, you see that we have negative terms now and the estimate will be harder. Instead, we don't need to do this harder computation and we only need to use the limit laws.] ■

R Recall the binomial theorem is given by $(a + b)^m = \sum_{k=0}^m \binom{m}{k} a^k b^{m-k}$, where $\binom{m}{k} = \frac{m!}{k!(m-k)!}$ is the binomial coefficient.

R After we've learned the next section, you can also use Lemma 2.2.2 and say that

$$\lim_{n \rightarrow \infty} p^{1/n} = p^{\lim_{n \rightarrow \infty} 1/n} = p^0 = 1.$$

■ **Example 2.5** $\lim_{n \rightarrow \infty} \sqrt[n]{n} = 1$. ■

Proof. Set $x_n = \sqrt[n]{n} - 1 \geq 0$. By the binomial theorem, for $n \geq 2$, $n = (1 + x_n)^n \geq 1 + nx_n + \frac{n(n-1)}{2}x_n^2 \geq \frac{n(n-1)}{2}x_n^2$.

[Why am I considering the $\frac{n(n-1)}{2}x_n^2$ term here? Notice that if I only use the first two terms $1 + nx_n$, I only get $x_n \leq \frac{n-1}{n}$, and $\frac{n-1}{n}$ is converging to 1 instead of 0.]

Hence, for $n \geq 2$, from $n \geq \frac{n(n-1)}{2}x_n^2$ we get

$$0 \leq x_n \leq \sqrt{\frac{2}{n-1}}.$$

Consequently, $\lim_{n \rightarrow \infty} x_n = 0$ and thus, $\lim_{n \rightarrow \infty} \sqrt[n]{n} = 1$. ■

R Using this example and the product law, you can show that $\lim_{n \rightarrow \infty} (n^5)^{1/n} = 1$, etc.

R Can you show $\lim_{n \rightarrow \infty} (1 + n + 2n^2)^{1/n} = 1$? (Hint: you can use squeeze theorem.)

Further Reading

- Boller & Sally: Section 1.6, up to Exercise 1.6.25
- Rudin: Chapter 3, up to “Some special sequences”
- Tao, *Analysis I*: Chapter 6
- Zorich: 3.1.1, 3.1.2, 3.1.3

2.2 Real Exponentiation

In the previous section we’ve already worked with real exponentials in the form a^x for real numbers $a \in \mathbb{R}^+$ and $x \in \mathbb{R}$. In this section, we’ll rigorously define it. Here is the plan: Fix $a > 0$.

- Define over the integers: First we define a^n for natural numbers $n \in \mathbb{N}$, and extend to all integers $n \in \mathbb{Z}$.
- Define over the rationals: Next we define the n -th roots $a^{1/n}$ for $n \in \mathbb{N}^+$, and extend to a^q for all rational q .
- Define over the reals: We define a^x for $x \in \mathbb{R}$ using Cauchy sequences of rationals $(q_n)_{n=1}^\infty$ converging to x and define $a^x = \lim_{n \rightarrow \infty} a^{q_n}$.

You are not responsible for all the details of this construction. We present it because it connects with the general theme: we extend definitions from simpler objects (integers to rationals to reals) using limits, and it reinforces what we learned earlier. Only the basic properties of the exponential function will be used in this course.

Let $a \in \mathbb{R}$, for natural number n , we know how a^n is defined. To raise a to the power 0, we define $a^0 := 1$. To raise a to the power n where n is a positive integer, we have $a^n = a \cdot a \dots a$, where we multiply a for n times.

Now let a be a non-zero real number. Then for any negative integer $-n$, we define $a^{-n} := 1/a^n$.

Next we consider exponentiation to exponents which are not integers. We begin with the notion of an n th root, which we can define using our notion of least upper bound (aka supremum).

Definition 2.2.1 Let $a \geq 0$ be a non-negative real, and let $n \geq 1$ be a positive integer. We define $a^{1/n}$, also known as the n th root of a , by the formula

$$a^{1/n} := \sup\{y \in \mathbb{R} : y \geq 0 \text{ and } y^n \leq a\}.$$

We also write $\sqrt[n]{a}$ for $a^{1/n}$.

R Why do we define it this way? Why don’t we define $a^{1/n}$ as the solution to the equation $y^n = a$? (Do we know that this equation has a nonnegative real solution and the solution is unique? At this point do we have results to guarantee this?)

R You can also define $a^{1/n}$ as $\inf\{y \in \mathbb{R} : y^n \geq a\}$. Then you want to verify the set is nonempty and bounded from below.

To make sure this is well-defined, we need the following lemma:

Lemma 2.2.1 — Existence of n th roots. Let $x \geq 0$ be a non-negative real, and let $n \geq 1$ be a positive integer. Then the set $E := \{y \in \mathbb{R} : y \geq 0 \text{ and } y^n \leq a\}$ is nonempty and is also bounded from above.

Proof. Exercise. ■

Then by Theorem 1.3.5, the least upper bound exists, so $a^{1/n}$ is well-defined. Also from the definition, we can check $(a^{1/n})^n = a$.

R We can use the squeeze theorem to check this. Notice from the definition, $(a^{1/n})^n \leq a \leq (a^{1/n} + \frac{1}{N})^n$ for any integer $N \geq 1$, and $\lim_{N \rightarrow \infty} (a^{1/n} + \frac{1}{N})^n = (\lim_{N \rightarrow \infty} (a^{1/n} + \frac{1}{N}))^n = (a^{1/n})^n$.

R Recall from Example 2.4 that we showed for $a > 0$, we have $\lim_{n \rightarrow \infty} a^{1/n} = 1$.

We can then extend to rational exponents.

Definition 2.2.2 Let $a > 0$ be a positive real number, and let q be a rational number. To define a^q , we write $q = m/n$ for some integer m and positive integer n , and define

$$a^q := (a^{1/n})^m.$$

R We can check that a^q is well-defined for every rational q . Note that this new definition is consistent with our old definition for $a^{1/n}$. We can also check $a^{q_1+q_2} = a^{q_1}a^{q_2}$ for $q_1, q_2 \in \mathbb{Q}$.

Lastly, let's deal with real exponents.

Lemma 2.2.2 — Continuity of exponentiation. Let $a > 0$, and let x be a real number. Let $(q_n)_{n=1}^\infty$ be any sequence of rational numbers converging to x . Then $(a^{q_n})_{n=1}^\infty$ is also a convergent sequence. Furthermore, if $(q'_n)_{n=1}^\infty$ is any other sequence of rational numbers converging to x , then $(a^{q'_n})_{n=1}^\infty$ has the same limit as $(a^{q_n})_{n=1}^\infty$:

$$\lim_{n \rightarrow \infty} a^{q_n} = \lim_{n \rightarrow \infty} a^{q'_n}.$$

Proof. [Reading for you]

There are three cases: $a < 1$, $a = 1$, and $a > 1$. The case $a = 1$ is rather easy (because then $a^q = 1$ for all rational q). We shall just do the case $a > 1$, and leave the case $a < 1$ (which is very similar) to the reader.

Let us first prove that $(a^{q_n})_{n=1}^\infty$ converges. By the Cauchy Criterion (Theorem 2.1.9), it is enough to show that $(a^{q_n})_{n=1}^\infty$ is a Cauchy sequence.

To do this, we need to estimate the distance between a^{q_n} and a^{q_m} ; let us say for the time being that $q_n \geq q_m$, so that $a^{q_n} \geq a^{q_m}$ (since $a > 1$). We have

$$|a^{q_n} - a^{q_m}| = a^{q_n} - a^{q_m} = a^{q_m}(a^{q_n - q_m} - 1).$$

Since $(q_n)_{n=1}^\infty$ is a convergent sequence, it has some upper bound $M \in \mathbb{N}^+$; since $a > 1$, we have $a^{q_m} \leq a^M$. Thus

$$|a^{q_n} - a^{q_m}| \leq a^M(a^{q_n - q_m} - 1).$$

Now let $\varepsilon > 0$. We know by Example 2.4 that, since the sequence $(a^{1/k})_{k=1}^\infty$ converges to 1, there exists some $K \geq 1$ such that

$$|a^{1/K} - 1| < \varepsilon a^{-M}.$$

Now since $(q_n)_{n=1}^\infty$ is convergent, it is a Cauchy sequence, and so there is an $N \geq 1$ such that q_n and q_m are $1/K$ -close for all $n, m \geq N$. Thus we have

$$|a^{q_n} - a^{q_m}| \leq a^M(a^{q_n - q_m} - 1) < a^M(a^{1/K} - 1) < a^M \varepsilon a^{-M} = \varepsilon$$

for every $n, m \geq N$ such that $q_n \geq q_m$. By symmetry we also have this bound when $n, m \geq N$ and $q_n \leq q_m$. Thus the sequence $(a^{q_n})_{n=1}^\infty$ is a Cauchy sequence as desired. This proves the convergence of $(a^{q_n})_{n=1}^\infty$.

Now we prove the second claim. It will suffice to show that

$$\lim_{n \rightarrow \infty} a^{q_n - q'_n} = 1,$$

since the claim would then follow from limit laws (since $a^{q_n} = a^{q_n} a^{q_n - q'_n}$).

Write $r_n := q_n - q'_n$; by limit laws we know that $(r_n)_{n=1}^\infty$ converges to 0. We have to show that for every $\varepsilon > 0$, the sequence $(a^{r_n})_{n=1}^\infty$ is eventually ε -close to 1. But from Example 2.4 we know that the sequence $(a^{1/k})_{k=1}^\infty$ is eventually ε -close to 1. Since by limit laws $\lim_{k \rightarrow \infty} a^{-1/k}$ is also equal to 1, we know that $(a^{-1/k})_{k=1}^\infty$ is also eventually ε -close to 1. Thus we can find a K such that $a^{1/K}$ and $a^{-1/K}$ are both ε -close to 1. But since $(r_n)_{n=1}^\infty$ is convergent to 0, it is eventually $1/K$ -close to 0, so that eventually $-1/K \leq r_n \leq 1/K$, and thus $a^{-1/K} \leq a^{r_n} \leq a^{1/K}$. In particular a^{r_n} is also eventually ε -close to 1, as desired. ■

R Here we are using “ a and b are ε -close” to mean $|a - b| \leq \varepsilon$.

We may now make the following definition.

Definition 2.2.3 — Exponentiation to a real exponent. Let $a > 0$ be real, and let x be a real number. We define a^x by the formula $a^x = \lim_{n \rightarrow \infty} a^{q_n}$, where $(q_n)_{n=1}^\infty$ is any sequence of rational numbers converging to x .

Lemma 2.2.3 — Properties of real exponentiation. Let $a, b > 0$ be positive reals, and let q, r be real numbers. Then

1. a^q is a positive real.
2. $a^{q+r} = a^q a^r$ and $(a^q)^r = a^{qr}$.
3. $a^{-q} = 1/a^q$.
4. If $q > 0$, then $a > b$ if and only if $a^q > b^q$.
5. If $a > 1$, then $a^q > a^r$ if and only if $q > r$. If $a < 1$, then $a^q > a^r$ if and only if $q < r$.
6. $(ab)^q = a^q b^q$.

Exercise 2.10 Another common way to define the real exponential is via the Euler’s number e .

For a real number x , we define $E_n(x) = \sum_{k=0}^n \frac{x^k}{k!}$.

We then define $\exp(x) := \lim_{n \rightarrow \infty} E_n(x)$, and $e := \exp(1)$.

1. For any $x \in \mathbb{R}$, show the limit $\lim_{n \rightarrow \infty} E_n(x)$ exists via the Cauchy criterion.

Hint: We want to bound $|E_m(x) - E_n(x)| = \left| \sum_{k=n+1}^m \frac{x^k}{k!} \right|$. Fix a positive integer $N_0 > 2|x|$.

For integer $k > N_0$, we have $\frac{|x|}{k} < \frac{1}{2}$, so

$$\frac{|x|^k}{k!} < \frac{1}{2} \frac{|x|^{k-1}}{(k-1)!}.$$

Use this inequality repeatedly to compare $\frac{|x|^k}{k!}$ with $\frac{|x|^{N_0}}{(N_0)!}$.

2. Show $\exp(x + y) = \exp(x) \cdot \exp(y)$.

Hint: You can use the binomial theorem to show

$$|E_n(x)E_n(y) - E_n(x+y)| \leq \sum_{m=n+1}^{2n} \frac{(|x| + |y|)^m}{m!},$$

and notice that from part (1), by taking n large, the right hand side can be made arbitrarily small.

Recall the binomial theorem is $(a + b)^m = \sum_{k=0}^m \binom{m}{k} a^k b^{m-k}$, where $\binom{m}{k} = \frac{m!}{k!(m-k)!}$ is the binomial coefficient.

3. Deduce from part (2) that $\exp(x) > 0$ always, and $\exp(x) < \exp(y)$ for $x < y$.
4. Using part (2) and (3), show the new definition agrees with the real exponential defined in class. That is, verify $\exp(x) = e^x$ for all x . You don't need to write all the details, but you need to include the key steps and results you used. (For example, if you use induction to show something, you don't need to write the induction in detail and you can just directly say you get this thing by induction instead.)

Hint: For part (4), you don't need to redo those computations about convergence. All you need to do is to use the very powerful formulas you proved in part (2) and (3).

In the previous exercise we see that we can define real exponential e^x as $\exp(x)$. Once we've done that, then using the natural log function $\log$ (we'll talk about how to define it as the inverse function of $\exp(x)$ later), for every $a > 0$, we can define $a^x = \exp(x \log a)$.

Further Reading

- Tao, *Analysis I*: Sections 5.6, 6.7

End of Week 3. To be continued...

2.3 Topology of $\mathbb{R}$

You can think of topology as the study of the properties and spatial relationships of things that remain unchanged even when they are bent, stretched, or otherwise distorted. In comparison, geometry studies more rigid things like lengths and distances, which would be changed under such distortions.

Definition 2.3.1 — accumulation point. Let S be a subset of $\mathbb{R}$. Then $x \in \mathbb{R}$ is an *accumulation point* of S if, for all $\varepsilon > 0$, $((x - \varepsilon, x + \varepsilon) \setminus \{x\}) \cap S \neq \emptyset$. In other words, for all $\varepsilon > 0$, there exists $y \in S$ such that $y \neq x$ and $|x - y| < \varepsilon$.

R In some other textbooks, an accumulation point is also called a limit point, or a cluster point.

Thus, x is an accumulation point of S if every interval around x contains points of S other than x . Of course, x does not have to be an element of S in order to be an accumulation point of S .

- **Example 2.6** The set of accumulation points of $(0, 1)$ is $[0, 1]$. ■
- **Example 2.7** The set of accumulation points of $\mathbb{Z}$ is $\emptyset$. ■
- **Example 2.8** The set of accumulation points of $\mathbb{Q}$ is $\mathbb{R}$. ■
- **Example 2.9** The set of accumulation points of $\mathbb{R} \setminus \mathbb{Q}$ is also $\mathbb{R}$. ■
- **Example 2.10** The set of accumulation points of $\{\frac{1}{n} \mid n \in \mathbb{N}^+\}$ is $\{0\}$. ■
- **Example 2.11** The set of accumulation points of $\{(-1)^n + \frac{1}{n} \mid n \in \mathbb{N}^+\}$ is $\{-1, 1\}$. ■
- **Example 2.12** Any finite subset of $\mathbb{R}$ has no accumulation points. ■

Theorem 2.3.1 — Bolzano-Weierstrass Theorem. Let S be a bounded, infinite subset of $\mathbb{R}$. Then S has an accumulation point in $\mathbb{R}$.

Proof. Pick an infinite sequence $(a_k)_{k=1}^{\infty}$ of distinct elements of S . Since S is bounded, (a_n) is also bounded, so by Lemma 2.1.6, it has a convergent subsequence $(a_{k_j})_{j=1}^{\infty}$.

We set $a = \lim_{j \rightarrow \infty} a_{k_j}$. By definition, for every $\varepsilon > 0$, there exist two distinct a_{k_j} and $a_{k_j'}$ in $(a - \varepsilon, a + \varepsilon)$ and thus, at least one of them is different from a . It follows that a is an accumulation point of S . ■

So basically an accumulation point of a set is a generalization of a limit of a sequence. The only difference is that in a sequence we allow repeated terms but in a set we don't have repeated elements. so for a constant sequence it has a limit but viewed as a set it has no accumulation points.

The next exercise shows an accumulation point of a sequence is essentially the limit of some subsequence.

Exercise 2.11 Let $(x_n)_{n=1}^{\infty}$ be a bounded sequence. Let $S = \{x_n \mid n \in \mathbb{N}^+\}$.

1. Show a is an accumulation point of S if and only if there exists a subsequence $(x_{n_k})_{k=1}^{\infty}$ of $(x_n)_{n=1}^{\infty}$ with all terms distinct that converges to a .
2. Suppose all terms of $(x_n)_{n=1}^{\infty}$ are distinct. Show $\liminf_{n \rightarrow \infty} x_n$ and $\limsup_{n \rightarrow \infty} x_n$ are accumulation points of S .
3. Let a be an accumulation point of S . Show $\liminf_{n \rightarrow \infty} x_n \leq a \leq \limsup_{n \rightarrow \infty} x_n$. ■

Definition 2.3.2 — open set. Let S be a subset of $\mathbb{R}$. We say that S is an *open set* in $\mathbb{R}$ if, for each point $x \in S$, there is an $\varepsilon > 0$ such that $(x - \varepsilon, x + \varepsilon) \subset S$.

- **Example 2.13** $\emptyset, \mathbb{R}$ are open sets in $\mathbb{R}$. ■
- **Example 2.14** $\mathbb{Z}$ is not open in $\mathbb{R}$. But the complement $Z^c = \mathbb{R} \setminus \mathbb{Z}$ is open in $\mathbb{R}$. ■
- **Example 2.15** $\mathbb{Q}$ is not open in $\mathbb{R}$. The complement $\mathbb{Q}^c = \mathbb{R} \setminus \mathbb{Q}$ is not open in $\mathbb{R}$ either. ■
- **Example 2.16** An open interval (a, b) is an open set in $\mathbb{R}$, where $a < b$. ■
- **Example 2.17** A union of open intervals like $(a_1, b_2) \cup (a_2, b_2)$ is an open set in $\mathbb{R}$. ■

Definition 2.3.3 — closed set. Let S be a subset of $\mathbb{R}$. We say that S is a *closed set* in $\mathbb{R}$ if the complement of S is an open set in $\mathbb{R}$.

- **Example 2.18** $\emptyset, \mathbb{R}$ are closed sets in $\mathbb{R}$. ■
- **Example 2.19** $\mathbb{Z}$ is closed in $\mathbb{R}$. The complement $Z^c = \mathbb{R} \setminus \mathbb{Z}$ is not closed in $\mathbb{R}$. ■
- **Example 2.20** $\mathbb{Q}$ is not closed in $\mathbb{R}$. And $\mathbb{R} \setminus \mathbb{Q}$ is not closed in $\mathbb{R}$ either. ■
- **Example 2.21** A closed interval $[a, b]$ is a closed set in $\mathbb{R}$, where $a < b$. ■
- **Example 2.22** An half-open, half-closed interval like $[a, b)$ is neither closed nor open in $\mathbb{R}$. ■

So we see in general, a set could be neither open nor closed, and $\emptyset$ and $\mathbb{R}$ are both open and closed.

In fact, we will show later that in $\mathbb{R}$, the only sets that are both open and closed are $\emptyset$ and $\mathbb{R}$.

We want a better way to characterize closed sets.

Lemma 2.3.2 A subset of $\mathbb{R}$ is closed if and only if it contains all its accumulation points.

Proof. Suppose A is a closed subset and let a be an arbitrary accumulation point of A . Assume, for the sake of contradiction, that $a \notin A$. By definition, the complement of A , denoted $A^c = \mathbb{R} \setminus A$, is

an open subset, so there exists $\varepsilon > 0$ such that $(a - \varepsilon, a + \varepsilon) \subseteq A^c$. Hence, $(a - \varepsilon, a + \varepsilon) \cap A = \emptyset$, which contradicts the assumption that a is an accumulation point of A . Since a was arbitrary, this shows that A contains all of its accumulation points.

Now suppose that A contains all of its accumulation points. To show that A is closed, it suffices to prove that A^c is open. Let $a \in A^c$ be arbitrary. We aim to show that there exists $\varepsilon > 0$ such that $(a - \varepsilon, a + \varepsilon) \subseteq A^c$. Suppose, for the sake of contradiction, that for all $\varepsilon > 0$, $(a - \varepsilon, a + \varepsilon) \not\subseteq A^c$, that is, $(a - \varepsilon, a + \varepsilon) \cap A \neq \emptyset$. Since $a \notin A$, for all $\varepsilon > 0$, $((a - \varepsilon, a + \varepsilon) \setminus \{a\}) \cap A \neq \emptyset$, and thus, a is an accumulation point of A , contradicting the assumption that $a \notin A$. Therefore, A^c is open. ■

Exercise 2.12 Show that

1. An *arbitrary union* of open sets in $\mathbb{R}$ is open in $\mathbb{R}$. As a consequence, an *arbitrary intersection* of closed sets in $\mathbb{R}$ is closed in $\mathbb{R}$.
2. A *finite intersection* of open sets in $\mathbb{R}$ is open in $\mathbb{R}$. As a consequence, a *finite union* of closed sets in $\mathbb{R}$ is closed in $\mathbb{R}$.
3. An infinite intersection of open sets is not necessarily open by providing an example. As a consequence, an infinite union of closed sets is not necessarily closed.

R It follows from definition that every open set in $\mathbb{R}$ can be written as a union of open intervals. It can be further shown that every open set in $\mathbb{R}$ can be written as a countable union of disjoint intervals.

- Exercise 2.13**
1. Give an example of a bounded subset A of $\mathbb{R}$ with infinitely many points such that neither $\inf A$ nor $\sup A$ is an accumulation point of A .
 2. Let $A \subset \mathbb{R}$ be a set, and let S be the set of all accumulation points of A . Show S is closed.

Definition 2.3.4 — open covering. Let A be a subset of $\mathbb{R}$. An *open covering* (also called open cover) of A is a collection of open sets $\{U_i\}_{i \in I}$ such that $A \subset \bigcup_{i \in I} U_i$.

R Note here we are only asking A is contained in $\bigcup_{i \in I} U_i$, not equal. Recall that an arbitrary union of open sets is open, so $\bigcup_{i \in I} U_i$ is open. So if A is not open, it is impossible to have $A = \bigcup_{i \in I} U_i$.

R Being an open covering of a set is very flexible. Note that $\{\mathbb{R}\}$ is an opening covering of any set in $\mathbb{R}$.

Definition 2.3.5 — compact set. A subset A of $\mathbb{R}$ is a *compact set* if every open covering of A has a *finite subcovering*. Namely, if $\{U_i\}_{i \in I}$ is an open covering, then there exists a finite subcollection $U_1, \dots, U_n$ such that $A \subset \bigcup_{k=1}^n U_k$.

Intuitively, what compactness actually gives you is: “finite control from infinite data.”

■ **Example 2.23** Consider the finite set $A = \{1, 2, \dots, n\}$. We show it is compact. Take any open covering $\{U_i\}_{i \in I}$ of A . For each $k \in A$, there exists some U_{i_k} such that $k \in U_{i_k}$. Then $\{U_{i_1}, U_{i_2}, \dots, U_{i_n}\}$ is a finite subcovering of A . This shows A is compact. ■

■ **Example 2.24** Consider the unbounded set $\mathbb{Z}$. For each $i \in \mathbb{Z}$ take $U_i = (i - \frac{1}{2}, i + \frac{1}{2})$. These U_i 's form an open covering of $\mathbb{Z}$. But it has no finite subcovering, showing this unbounded set $\mathbb{Z}$ cannot be compact. ■

■ **Example 2.25** Consider the open set $(0, 1)$. For each $i \in \mathbb{N}$ take $U_i = (\frac{1}{i}, 2)$. These U_i 's form an open covering of $(0, 1)$. But it has no finite subcovering, showing this non-closed set $(0, 1)$ cannot be compact. ■

The next theorem, the Heine-Borel theorem for $\mathbb{R}$, is the second of the two basic topological theorems for the real numbers, the first of which is the Bolzano-Weierstrass theorem. We shall see versions of these theorems again in Chapter 3.

Theorem 2.3.3 — Heine-Borel Theorem. Let S be a subset of $\mathbb{R}$. Then S is compact if and only if it is closed and bounded.

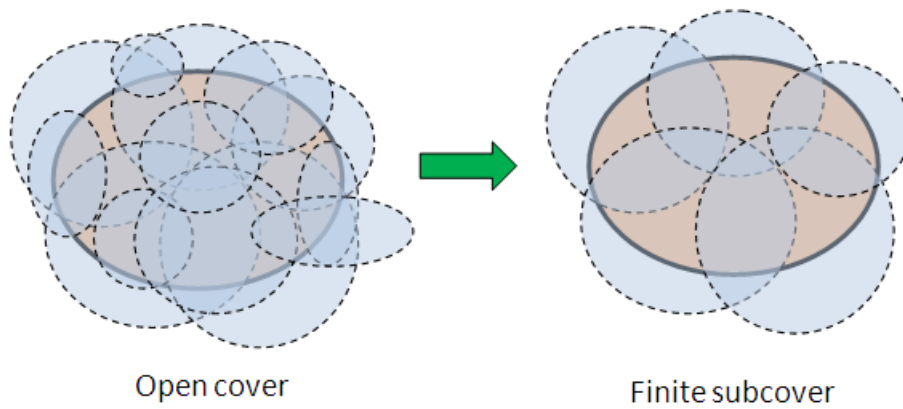


Figure 2.2: Visualization of the Heine-Borel Theorem

Proof. The part “compact $\implies$ closed and bounded” is exercise. We’ll do “closed and bounded $\implies$ compact”.

Step 1: Set up the problem.

[We want to show every open covering has a finite subcovering. This is a global statement about all coverings, and it is hard to prove directly. So we argue by contradiction.]

Let S be a closed bounded set in $\mathbb{R}$. Assume, for the sake of contradiction, that there exists an open covering $\{U_i\}_{i \in I}$ of S such that S can not be covered by finitely many U_i . In particular, S is nonempty.

Step 2: Localize the failure by repeated bisection. Use boundedness here.

[No matter which finitely many sets you pick, they will always miss some point of S . There is always some “uncovered bad part” left over. Our goal is to locate this leftover bad part by “zooming in”. The idea is that the failure of finite cover must persist in some smaller region.]

Since S is bounded, there exists $a_0, b_0 \in \mathbb{R}$ such that $S \subset [a_0, b_0]$. By our assumption, either $S \cap [a_0, \frac{a_0+b_0}{2}]$ or $S \cap [\frac{a_0+b_0}{2}, b_0]$ cannot be covered by finitely many U_i . We will denote this new closed interval by $[a_1, b_1]$. We repeat the argument to construct a sequence of intervals

$$[a_0, b_0] \supset [a_1, b_1] \supset [a_2, b_2] \supset \cdots,$$

such that for any $j \in \mathbb{N}$, $S \cap [a_j, b_j]$ cannot be covered by finitely many U_i . Moreover,

$$b_j - a_j = \frac{b_0 - a_0}{2^j} \rightarrow 0, \quad \text{as } j \rightarrow \infty. \quad (2.1)$$

Step 3: Shrink to a point $c \in S$. Use closedness and Cauchy criterion here.

[Intuitively, shrinking intervals force us to a single point.]

We consider a sequence $(c_k)_{k=1}^{\infty}$ where $c_k \in S \cap [a_k, b_k]$. By (2.1), we know that c_k is a Cauchy sequence, and thus converges to some $c \in \mathbb{R}$. Hence, c is an accumulation point of S , or $c \in S$. By the closedness of S , the first case also implies that $c \in S$. Furthermore, for every $j \in \mathbb{N}$, $c \in S \cap [a_j, b_j]$.

Step 4: Derive contradiction.

There exists an open set U_{i_0} such that $c \in U_{i_0}$, and there exists $\varepsilon > 0$, $(c - \varepsilon, c + \varepsilon) \subset U_{i_0}$. For sufficiently large $j \in \mathbb{N}$, $b_j - a_j = \frac{b_0 - a_0}{2^j} < \varepsilon$, and thus, $[a_j, b_j] \subset (c - \varepsilon, c + \varepsilon) \subset U_{i_0}$. This contradicts our assumption that $S \cap [a_j, b_j]$ cannot be covered by finitely many U_{i_0} .

Therefore, S can be covered by finitely many U_i . ■

R In the proof we see how both bounded and closed are used. If the set is not bounded, say $\mathbb{R}$, then we see Step 2 doesn't work: no matter how many times we subdivide, it is still possible that the subinterval is still unbounded. If the set is not closed, then we see Step 3 doesn't work: the point c we get in Step 3 doesn't need to be in S , so it doesn't need to be contained in any of the U_i 's, and we can't get a contradiction in Step 4.

So now we know in $\mathbb{R}$, a compact set is closed and bounded. Then we do we need this abstract definition of compactness? It's because when we move on to general metric spaces, we will see that being closed and bounded doesn't imply being compact (for examples, even in $\mathbb{Q}$). To get the finite control over infinite data, we actually need the compactness.

As we have seen, it is often easier to work with sequences than sets. Hence we introduce a related notion.

Definition 2.3.6 — sequentially compact set. A subset A of $\mathbb{R}$ is *sequentially compact* if every sequence in A has a subsequence that converges in A .

You can think that a sequentially compact set is a set where every infinite “scatter of points” must contain a cluster. There is no escape (it must have a convergent subsequence) and no missing limit (the subsequential limit is in it).

What's nice is that being sequentially compact is actually equivalent to being compact.

Proposition 2.3.4 Let A be a subset of $\mathbb{R}$. The following three statements are equivalent.

1. A is bounded and closed.
2. A is compact.
3. A is sequentially compact.

Exercise 2.14 Prove Proposition 2.3.4. ■

Proposition 2.3.5 — Cantor's intersection theorem. Let $I_1, I_2, \dots$ be nonempty compact sets such that $I_1 \supset I_2 \supset I_3 \supset \dots$. Then $P := \bigcap_{n=1}^{\infty} I_n$ is also nonempty.

Exercise 2.15 Prove Cantor's intersection theorem. (Hint: You can either use the definition directly, or use sequential compactness.) ■

In other words, you can't squeeze a compact set down to nothing.

■ **Example 2.26** If the I_n 's in the previous theorem are not compact, then the intersection could be empty. For example, take $I_n = (0, \frac{1}{n})$. Then $\bigcap_{n=1}^{\infty} I_n = \emptyset$. ■

Further Reading

- Boller & Sally: Section 1.6
- Rudin: Chapter 2
- Tao, *Analysis I*: Section 9.1

End of Week 4. To be continued...

2.4 A brief introduction to the geometry of $\mathbb{R}^n$

In this section, we will investigate the *n-dimensional Euclidean space*, *n-dimensional real space* or *real n-space* $\mathbb{R}^n$, which is the product space of n spaces $\mathbb{R}$, i.e. $\mathbb{R} \times \cdots \times \mathbb{R}$. A point in $\mathbb{R}^n$ is an n -tuple of real numbers, and is written as

$$\mathbf{x} = (x_1, x_2, \dots, x_n)$$

where $x_i \in \mathbb{R}$.

We also want to define on the set $\mathbb{R}^n$ various “structures”: vector space, inner product space, normed space, and metric space. Each of these structures can be defined by rather simple formulas; however, in general we will also be interested in a more abstract description of these structures. In this course, we are interested in the metric space structure. To that end, we will first describe the Euclidean version in this section, then present some of the basic properties of the structure, and finally use these properties to define a general notion of the metric structure in Chapter 3.

We begin with the vector space structure of $\mathbb{R}^n$. The study of vector spaces is called *linear algebra* and is the topic for another course. Apart from thinking n -tuples of real numbers as points, we can also consider n -tuples of real numbers as n -dimensional vectors giving the data of a direction and a magnitude.

Definition 2.4.1 Given $\mathbf{x} = (x_1, x_2, \dots, x_n)$, $\mathbf{y} = (y_1, y_2, \dots, y_n)$ and $c \in \mathbb{R}$, we define

$$\begin{aligned}\mathbf{x} + \mathbf{y} &:= (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n) \\ c\mathbf{x} &:= (c \cdot x_1, c \cdot x_2, \dots, c \cdot x_n).\end{aligned}$$

The vector space structure of $\mathbb{R}^n$ is not enough—on its own—to allow us to express geometric concepts like length, angle, or distance. We are thus led to consider more structures.

Definition 2.4.2 — dot product (also called scalar product). Let $\mathbf{v} = (v_1, \dots, v_n)$ and $\mathbf{w} = (w_1, \dots, w_n)$ be two vectors in $\mathbb{R}^n$. The *dot product* of $\mathbf{v}$ and $\mathbf{w}$ is

$$\mathbf{v} \cdot \mathbf{w} := \sum_{i=1}^n v_i w_i.$$

The dot product is also denoted by $\langle \mathbf{v}, \mathbf{w} \rangle$.

Lemma 2.4.1 For all $\mathbf{v}, \mathbf{v}_1, \mathbf{v}_2, \mathbf{w} \in \mathbb{R}^n$, and $\forall \alpha \in \mathbb{R}$,

1. $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$.
2. $\langle \mathbf{v}_1 + \alpha \mathbf{v}_2, \mathbf{w} \rangle = \langle \mathbf{v}_1, \mathbf{w} \rangle + \alpha \langle \mathbf{v}_2, \mathbf{w} \rangle$.

Definition 2.4.3 — norm. If $\mathbf{v} = (v_1, \dots, v_n) \in \mathbb{R}^n$, then the *norm* or *length* of $\mathbf{v}$ is defined by

$$\|\mathbf{v}\| := \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} = (v_1^2 + v_2^2 + \cdots + v_n^2)^{1/2}.$$

Lemma 2.4.2 The norm satisfies the following properties.

1. *positive definite*: $\forall \mathbf{v} \in \mathbb{R}^n, \|\mathbf{v}\| \geq 0$; $\|\mathbf{v}\| = 0$ iff $\mathbf{v} = 0$;
2. *positive homogeneous*: $\forall \mathbf{v} \in \mathbb{R}^n, \alpha \in \mathbb{R}, \|\alpha \mathbf{v}\| = |\alpha| \|\mathbf{v}\|$;
3. *triangle inequality*: $\forall \mathbf{v}, \mathbf{w} \in \mathbb{R}^n, \|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$.

Exercise 2.16 Prove Lemma 2.4.2. Hint for 3: Use Cauchy-Schwarz inequality or AM-GM inequality. ■

The definition of the norm leads to the usual definition of Euclidean distance between two points in $\mathbb{R}^n$.

Definition 2.4.4 Given two points $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, the *Euclidean distance* between $\mathbf{x}$ and $\mathbf{y}$ is given by

$$d(\mathbf{x}, \mathbf{y}) := \|\mathbf{x} - \mathbf{y}\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}.$$

■ **Example 2.27** The Euclidean distance between $(0, 0, 0)$ and $(1, 1, 1)$ is $\sqrt{3}$. ■

Lemma 2.4.3 This Euclidean distance d satisfies three basic conditions:

1. *positive definite*: $d(x, y) \geq 0$ with equality if and only if $x = y$.
2. *symmetric*: $d(x, y) = d(y, x)$.
3. *triangle inequality*: $d(x, z) \leq d(x, y) + d(y, z)$.

Proof. For triangle inequality, set $v := x - y$ and $w := y - z$. Then apply item 3 of Lemma 2.4.2. ■

We mentioned that to show the triangle inequality we might use the AM-GM inequality.

The arithmetic mean, or less precisely the average, of a list of n numbers $x_1, x_2, \dots, x_n$ is the sum of the numbers divided by n :

$$\frac{x_1 + x_2 + \dots + x_n}{n}$$

The geometric mean is similar, except that it is only defined for a list of nonnegative real numbers, and uses multiplication and a root in place of addition and division:

$$\sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$$

The inequality of arithmetic and geometric means, or more briefly the AM–GM inequality, states that for any list of n nonnegative real numbers $x_1, x_2, \dots, x_n$, we have

$$\frac{x_1 + x_2 + \dots + x_n}{n} \geq \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n},$$

and that equality holds if and only if $x_1 = x_2 = \dots = x_n$.

For $n = 2$ this is the familiar inequality $\frac{x+y}{2} \geq \sqrt{xy}$. (Square it and rearrange, we see this is just $(x - y)^2 \geq 0$.)

R Here are some direct applications of the AM-GM inequality:

- Show that among all rectangles with a given perimeter, the square has the largest area.
- Show that among all boxes with a given surface area, the cube has the largest volume.

Theorem 2.4.4 — Cauchy-Schwarz inequality. For all $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, we have $|\langle \mathbf{v}, \mathbf{w} \rangle| \leq \|\mathbf{v}\| \|\mathbf{w}\|$.

Proof. Let λ be a real number. Then

$$\begin{aligned} 0 &\leq \langle \mathbf{v} - \lambda \mathbf{w}, \mathbf{v} - \lambda \mathbf{w} \rangle \\ &= \langle \mathbf{v}, \mathbf{v} \rangle - 2\lambda \langle \mathbf{v}, \mathbf{w} \rangle + \lambda^2 \langle \mathbf{w}, \mathbf{w} \rangle \\ &= \|\mathbf{v}\|^2 - 2\lambda \langle \mathbf{v}, \mathbf{w} \rangle + \lambda^2 \|\mathbf{w}\|^2. \end{aligned}$$

This is a quadratic polynomial in λ which is always greater than or equal to 0. For this inequality to hold, the discriminant of this quadratic must be nonpositive, namely,

$$0 \geq b^2 - 4ac = (2\langle \mathbf{v}, \mathbf{w} \rangle)^2 - 4\|\mathbf{v}\|^2\|\mathbf{w}\|^2.$$

Therefore, $|\langle \mathbf{v}, \mathbf{w} \rangle| \leq \|\mathbf{v}\|\|\mathbf{w}\|$. ■

 If we write it out, then the Cauchy-Schwarz inequality states that for any real numbers x_i, y_i we have

$$(x_1 y_1 + \cdots + x_n y_n)^2 \leq (x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2)$$

“The square of the sum of products $\leq$ the product of the sum of squares.”

■ **Example 2.28** Take all $y_i = 1$. Then

$$(x_1 + \cdots + x_n)^2 \leq n(x_1^2 + \cdots + x_n^2).$$

Rearranging yields

$$\frac{x_1 + \cdots + x_n}{n} \leq \sqrt{\frac{x_1^2 + \cdots + x_n^2}{n}},$$

or in words, that the mean is bounded above by the root mean square. This already hints at a connection to probability and statistics. Inspired by this, let's impose the additional condition $x_1 + \cdots + x_n = 1$ (the sum of probabilities of all possible events is 1). If this is the case, then

$$x_1^2 + \cdots + x_n^2 \geq 1/n.$$

For example, if we take $n = 6$ and let x_i be the probability getting number i for a dice. This shows that if I roll a 6-sided dice twice, the probability that both rolls produce the same number is at least $1/6$ - no matter whether or not the dice is fairly weighted! ■

■ **Example 2.29 — angle.** Let θ be the angle between $\mathbf{v}$ and $\mathbf{w}$, then $\langle \mathbf{v}, \mathbf{w} \rangle = \|\mathbf{v}\|\|\mathbf{w}\| \cos \theta$. In particular, for two non-zero vectors $\mathbf{v}$ and $\mathbf{w}$, $\mathbf{v} \perp \mathbf{w}$ (It reads $\mathbf{v}$ is *perpendicular* to $\mathbf{w}$), i.e., $\theta = \frac{\pi}{2}$, if and only if $\langle \mathbf{v}, \mathbf{w} \rangle = 0$. ■

Further Reading

- *Rudin*: Chapter 3
- *Zorich*: Sections 7.1, 8.1

3. Metric spaces

3.1 Definitions of metric spaces

We have already encountered the notion of distance in n -dimensional Euclidean space. In this chapter, we generalize the notion of distance in $\mathbb{R}^n$ to that of a metric on any set X . This new notion of distance will allow us to define open sets, closed sets, and compact sets as we did for the real numbers.

Definition 3.1.1 — metric space. A *metric space* is a pair (X, d) where X is a set and $d : X \times X \rightarrow \mathbb{R}$ is a map satisfying the following properties.

1. *positive definite*: For all $x_1, x_2 \in X$, $d(x_1, x_2) \geq 0$, and $d(x_1, x_2) = 0$ iff $x_1 = x_2$.
2. *symmetric*: For all $x_1, x_2 \in X$, $d(x_1, x_2) = d(x_2, x_1)$.
3. *triangle inequality*: For all $x_1, x_2, x_3 \in X$,

$$d(x_1, x_2) \leq d(x_1, x_3) + d(x_3, x_2).$$

R For the distance function d , because it measures the distance between two points, we naturally need two arguments.

We need the triangle inequality not only because this is what the Euclidean distance satisfies, but also because we want “nearness” to be a transitive property: if x_1 is close to x_3 and x_3 is close to x_2 , we naturally want x_1 to be close to x_2 .

Any space can be turned into a metric space via the discrete metric:

■ **Example 3.1 — discrete metric space.** Let X be any nonempty set and for $x_1, x_2 \in X$, we define a *discrete metric*

$$d(x_1, x_2) = \begin{cases} 0 & x_1 = x_2 \\ 1 & x_1 \neq x_2. \end{cases}$$

■

- R** If X has finitely many points (say N points), you can visualize the discrete metric space as sitting in Euclidean $\mathbb{R}^{N-1}$: for $N = 3$ this space is the 3 vertices of an equilateral triangle in $\mathbb{R}^2$. for $N = 4$ this space is the 4 vertices of an ideal tetrahedron in $\mathbb{R}^3$. For general N this is an ideal N -simplex in $\mathbb{R}^{N-1}$.

A related notion is the norm.

Definition 3.1.2 — norm. Given a vector space X , a norm on X is a real-valued function $\|\cdot\| : X \rightarrow \mathbb{R}$ with the following properties

1. *positive definiteness*: $\forall \mathbf{v} \in X, \|\mathbf{v}\| \geq 0$; $\|\mathbf{v}\| = 0$ iff $\mathbf{v} = 0$;
2. *positive homogeneity*: $\forall \mathbf{v} \in X, \alpha \in \mathbb{R}, \|\alpha \mathbf{v}\| = |\alpha| \|\mathbf{v}\|$;
3. *triangle inequality*: $\forall \mathbf{v}, \mathbf{w} \in X, \|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$.

- R** Note that in order to define a norm, we need X to be a *vector space* instead of any set. This is because we need to be able to add two vectors (for property 3) and multiply one vector by a scalar (for property 2).

From the definitions, we see directly that

Lemma 3.1.1 Given a norm $\|\cdot\|$ on a vector space X , we can get an induced metric d by $d(x, y) = \|x - y\|$.

■ **Example 3.2 — usual Euclidean norm and metric on $\mathbb{R}^n$.** For $x = (x_1, \dots, x_n), y = (y_1, \dots, y_n) \in \mathbb{R}^n$, the usual Euclidean norm is given by

$$\|x\| := \sqrt{\sum_{i=1}^n x_i^2},$$

and then the usual Euclidean metric is given by

$$d(x, y) := \sqrt{\sum_{i=1}^n (x_i - y_i)^2}.$$

■

■ **Example 3.3 — taxicab distance/Manhattan distance.** The usual Euclidean metric is not the only way to define the distance between two points in $\mathbb{R}^n$. Imagine a city laid out in a perfect grid (streets run north–south and east–west). A taxi cannot drive diagonally through buildings. We want to measure the actual distance the taxi needs to travel between two places.

Motivated by this, for $x = (x_1, \dots, x_n), y = (y_1, \dots, y_n) \in \mathbb{R}^n$, we can define the *taxicab norm* by

$$\|x\|_1 := \sum_{i=1}^n |x_i|,$$

and then get the *taxicab metric* by

$$d_1(x, y) := \sum_{i=1}^n |x_i - y_i|.$$

■

■ **Example 3.4 — ℓ_p norm and d_p metric.** More generally, let $p \geq 1$ (p can be any real number ≥ 1 , not just integers!). For $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, we define the ℓ_p norm (also called p -norm)

$$\|x\|_p = \left(\sum_{j=1}^n |x_j|^p \right)^{1/p}.$$

As before, for $x = (x_1, \dots, x_n), y = (y_1, \dots, y_n) \in \mathbb{R}^n$, we define a metric by

$$d_p(x, y) := \|x - y\|_p.$$

We'll check later that $\|x\|_p$ is indeed a norm (so d_p is indeed metric) on $\mathbb{R}^n$. Note that if $p = 1$, this is the taxicab metric. Note further that if $p = 2$, this is the usual Euclidean distance on $\mathbb{R}^n$. ■

R Why do we need to take the p -th power and then the p -th root? Why not the p -th power and then the q -th root? The reason is that we want the positive homogeneity to be satisfied.

As p increases, we can see that small coordinate differences matter less, and large coordinate differences matter more. So when p goes to infinity, only the largest coordinate differences matter. Therefore we can define the d_∞ metric.

■ **Example 3.5 — d_∞ metric.** For $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, we define the ℓ_p norm

$$\|x\|_\infty = \max_{1 \leq j \leq n} |x_j|,$$

and for $x = (x_1, \dots, x_n), y = (y_1, \dots, y_n) \in \mathbb{R}^n$, we define a metric by

$$d_\infty(x, y) := \|x - y\|_\infty.$$

Again, we'll check later that this is indeed a metric on $\mathbb{R}^n$. ■

Exercise 3.1 Let $x, y \in \mathbb{R}^n$. Show that

$$d_\infty(x, y) = \lim_{p \rightarrow \infty} d_p(x, y).$$

Hint: You can use $\lim_{p \rightarrow \infty} n^{1/p} = 1$ and the squeeze theorem for limits. ■

Exercise 3.2 On $\mathbb{R}$, we define a map $d : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ by the following formula

$$d(x, y) = \sqrt{|x - y|},$$

where $x, y \in \mathbb{R}$. Show that d is a metric on $\mathbb{R}$, but $\sqrt{|x|}$ doesn't define a norm on $\mathbb{R}$. ■

■ **Example 3.6 — What about $0 < p < 1$?** Now consider the case $0 < p < 1$. For $n \geq 2, x \in \mathbb{R}^n$, if we define

$$\|x\|_p = \left(\sum_{j=1}^n |x_j|^p \right)^{1/p}.$$

then it *isn't* a norm! We can check that the triangle inequality is violated: For instance, if $x = (1, 0, \dots, 0)$ and $y = (0, 1, 0, \dots, 0)$ then $\|x\|_p = 1, \|y\|_p = 1$, yet $\|x + y\|_p = 2^{1/p} > 2$ so that

$$\|x\|_p + \|y\|_p = 2 < 2^{1/p} = \|x + y\|_p.$$

Thus, if we take $d(x, y) = \|x - y\|_p$ this is also *not* a metric; the triangle inequality is again violated by considering $x = (1, 0, \dots, 0), y = (0, 1, 0, \dots, 0)$, and $z = (0, 0, \dots, 0)$. ■

Here we give some applications of the d_p metric.

■ **Example 3.7 — use of d_p metric.** Consider a scenario where a model predicts outcomes $\hat{x}_i$ for individuals $i = 1, \dots, n$, and the true outcomes are x_i . We aim to minimize the prediction error. The prediction error can be thought as the “distance” from our prediction to the actual outcome. Naturally, we can use different d_p distances to define the prediction errors.

Let $\hat{x} = (\hat{x}_1, \dots, \hat{x}_n)$ and let $x = (x_1, \dots, x_n)$. Then recall

$$d_p(\hat{x}, x) = \left(\sum |\hat{x}_i - x_i|^p \right)^{1/p}.$$

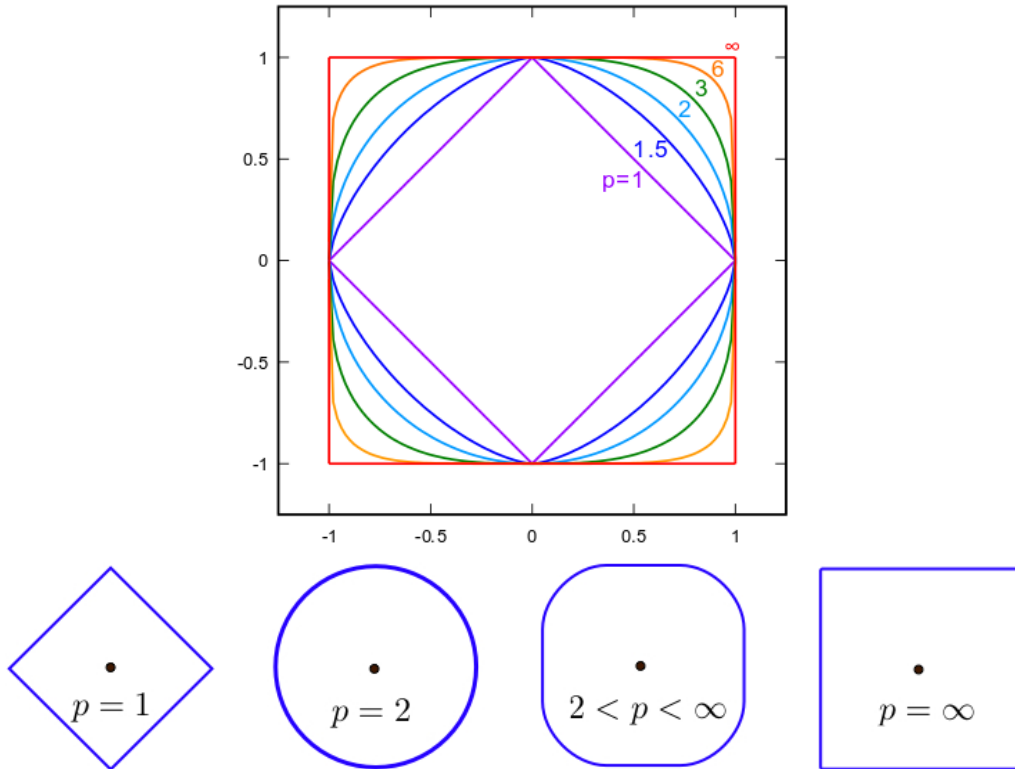
Which p should we choose to measure the error?

For $p = 1$, this is known as the mean absolute error. The d_1 metric is robust; it gives equal weight to all deviations, making it less sensitive to large outliers compared to higher p . The problem with d_1 metric is that d_1 is non-differentiable at zero, which can complicate optimization in gradient-based algorithms. Thus if you want both the robustness and smoothness, you may consider some $1 < p < 2$.

For $p = 2$, this is known as the mean squared error. It agrees with the intuitive geometric distance in Euclidean space, and is well-suited if the data follows a Gaussian distribution.

When p is larger, we focus more on the largest deviations. The p -th power term will blow up the error difference of the outliers. So if you want to emphasize large errors you can choose larger p .

When $p = \infty$, it only focuses on the largest error. This makes it ideal in applications where the worst-case scenario or maximum error is critical to evaluate or minimize. But it also leads to non-smooth or discontinuous optimization landscapes, requiring special methods to handle. ■



The unit sphere for different metrics: $\|x\|_{l_p} = 1$ in $\mathbb{R}^2$.

Figure 3.1: The unit ball for different ℓ_p norms

Definition 3.1.3 — $\ell_n^p(\mathbb{R})$. For $1 \leq p \leq \infty$, the metric space $(\mathbb{R}^n, d_p)$ is also denoted by $\ell_n^p(\mathbb{R})$.

R Note that the d_p metrics are all the same on $\mathbb{R}$, and all equal the usual Euclidean metric on $\mathbb{R}$. But of course there could also be other metrics on $\mathbb{R}$, it is just that they do not come from ℓ_p norms.

Exercise 3.3 — new metric spaces from old ones.

1. Let (X, d_X) and (Y, d_Y) be metric spaces. Show that $(X \times Y, d)$ with $d((x_1, y_1), (x_2, y_2)) := d_X(x_1, x_2) + d_Y(y_1, y_2)$ is a metric space. Note that the taxicab metric is the product metric on $\mathbb{R}^n$.
2. Let X be a set and let d_1, d_2 be two metrics on X . Show that (X, d) with $d(x, y) := \max\{d_1(x, y), d_2(x, y)\}$ is a metric space.
3. Let X be a set and let d_1, d_2 be two metrics on X . Give an example to show that $(X \times Y, d)$ with $d(x, y) := \min\{d_1(x, y), d_2(x, y)\}$ need NOT be a metric space.

End of Week 5. To be continued...

To show that $d_p, 1 \leq p \leq \infty$ is indeed a metric on $\mathbb{R}^n$ (or to check $\|\cdot\|_p$ is indeed a norm on $\mathbb{R}^n$), we need to verify the triangle inequality. For $1 < p < \infty$, we need the following inequality:

Theorem 3.1.2 — **Hölder's inequality**. Suppose that $p, q > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$. For $x = (x_1, \dots, x_n), y = (y_1, \dots, y_n) \in \mathbb{R}^n$, we have

$$\sum_{j=1}^n |x_j y_j| \leq \|x\|_p \|y\|_q.$$

R $\frac{1}{p} + \frac{1}{q} = 1 \implies p + q = pq \implies (p-1)(q-1) = 1 \implies q = \frac{p}{p-1}.$

The Cauchy-Schwarz inequality (Theorem 2.4.4) is a special case of Hölder's inequality.

Corollary 3.1.3 — **Cauchy-Schwarz inequality**. For $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, we have $|\langle \mathbf{v}, \mathbf{w} \rangle| \leq \|\mathbf{v}\| \|\mathbf{w}\|$.

Proof. This follows from the Hölder's inequality with $p = q = 2$, and $|\langle \mathbf{v}, \mathbf{w} \rangle| = |\sum_{j=1}^n v_j w_j| \leq \sum_{j=1}^n |v_j w_j|$. ■

Proof of Hölder's inequality. The proof is based on the following Young's inequality.

$$\sum_{k=1}^n \frac{|x_k|}{\|x\|_p} \frac{|y_k|}{\|y\|_q} \leq \sum_{k=1}^n \left(\frac{|x_k|^p}{p \|x\|_p^p} + \frac{|y_k|^q}{q \|y\|_q^q} \right) = \frac{1}{p} + \frac{1}{q} = 1.$$

Lemma 3.1.4 — **Young's inequality**. For all $a, b \in \mathbb{R}^+$, and $p, q > 1$ with $\frac{1}{p} + \frac{1}{q} = 1$, we have

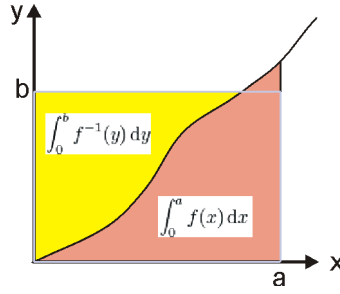
$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}.$$

Not rigorous proof. Consider the function $y = f(x) = x^{p-1}$ for $x \geq 0$. It has an inverse function $x = f^{-1}(y) = y^{q-1}$ for $y \geq 0$. (Recall $(p-1)(q-1) = 1$.)

We have

$$\int_0^a x^{p-1} dx + \int_0^b y^{q-1} dy = \frac{a^p}{p} + \frac{b^q}{q},$$

and the inequality follows from the picture below (which is the case when $b \leq a^{p-1}$; the case $b > a^{p-1}$ is similar), and equality holds if and only if $b = a^{p-1}$, which is equivalent to saying $b^q = a^p$.



■

Rigorous proof. [Reading homework for you]

Using the AM-GM inequality, the inequality holds for $p = q = 2$, i.e. $ab \leq \frac{a^2+b^2}{2}$. By induction on $k \in \mathbb{N}^+$, we can (Exercise for you in Homework) prove that for $(a_i)_{i=1}^{2^k}$ in $\mathbb{R}^+$, we have

$$\prod_{i=1}^{2^k} a_i \leq \frac{\sum_{i=1}^{2^k} (a_i)^{2^k}}{2^k}. \quad (3.1)$$

If $p = \frac{2^n}{p'}$ and $q = \frac{2^n}{q'}$ for $p', q' \in \mathbb{N}^+$ and $p' + q' = 2^n$, Young's inequality is

$$ab \leq \frac{p'a^{(2^n/p')} + q'b^{(2^n/q')}}{2^n}.$$

Setting $\bar{a} = a^{1/p'}$ and $\bar{b} = b^{1/q'}$, the inequality becomes

$$\bar{a}^{p'} \bar{b}^{q'} \leq \frac{p'\bar{a}^{2^n} + q'\bar{b}^{2^n}}{2^n}.$$

This is an immediate consequence of (3.1).

For general positive real numbers p and q , we can (Exercise for you) find two sequences $\left(\frac{2^n}{p_n}\right)_{n=1}^{\infty}$ and $q = \left(\frac{2^n}{q_n}\right)_{n=1}^{\infty}$, such that for all $n \in \mathbb{N}^+$, $p_n, q_n \in \mathbb{N}^+$ with $p_n + q_n = 2^n$, and

$$p = \lim_{n \rightarrow \infty} \frac{2^n}{p_n}, \quad q = \lim_{n \rightarrow \infty} \frac{2^n}{q_n}.$$

Then we have

$$ab \leq \lim_{n \rightarrow \infty} \frac{a^{(2^n/p_n)}}{2^n/p_n} + \frac{b^{(2^n/q_n)}}{2^n/q_n} = \frac{a^p}{p} + \frac{b^q}{q}.$$

This completes the proof. ■

Using the Hölder's inequality, we can then show:

Exercise 3.4 For $1 \leq p \leq \infty$, show that $(\mathbb{R}^n, d_p)$ is a metric space.

Hint: You just need to show $\|\cdot\|_p$ is a norm on $\mathbb{R}^n$, so essentially you just need to check the triangle inequality for $\|\cdot\|_p$. Consider the cases $p = 1$, $p \in (1, \infty)$, and $p = \infty$ separately. For $1 < p < \infty$, apply Hölder's inequality and

$$\|x + y\|_p^p \leq \sum_i |x_i + y_i|^p \leq \sum_i |x_i + y_i|^{p-1} |x_i| + \sum_i |x_i + y_i|^{p-1} |y_i|$$

to prove the triangle inequality. ■

Not all metrics come from norms. The discrete metric is one example. Here is another one.

■ **Example 3.8 — strings with Hamming distance.** We let the set X consists of strings of equal length, e.g., $X = \{\text{"cat"}, \text{"bat"}, \text{"car"}, \text{"bot"}\}$. The metric d is the number of positions where the characters differ in the two strings. For example:

$d(\text{"cat"}, \text{"bat"}) = 1$ (only the first letter differs).

$d(\text{"bat"}, \text{"car"}) = 2$ (the first and third letters differ).

$d(\text{"car"}, \text{"bot"}) = 3$ (all letters differ). ■

Exercise 3.5 On $\mathbb{R}$, we define a map $d : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ by the following formula

$$d(x, y) = \frac{|x - y|}{1 + |x - y|},$$

where $x, y \in \mathbb{R}$. Then show that $(\mathbb{R}, d)$ is a metric space that is not normed. ■

Further Reading

- Boller & Sally: Chapter 3
- Rudin: Chapter 2, “Metric Spaces”
- Zorich: Mathematical Analysis II, 9.1
- Tao: Analysis II, 1.1

3.2 Topology of metric spaces

Definition 3.2.1 — subspace. Let (X, d) be a metric space, and let Y be a subset of X . Let $d|_{Y \times Y} : Y \times Y \rightarrow [0, \infty)$, called the *inherited metric*, be the restriction of d to the subset $Y \times Y \subset X \times X$. Then $(Y, d|_{Y \times Y})$ (or simply denoted by (Y, d)) is a *subspace* of (X, d) .

 The subspace (Y, d) is also a metric space.

■ **Example 3.9** $([0, 1], d_2)$, $((0, 1), d_2)$ and $(\{0, 1\}, d_2)$ are all subspaces of $(\mathbb{R}, d_2)$. ■

■ **Example 3.10** $(\mathbb{Z}, d_2)$ and $(\mathbb{Q}, d_2)$ are both subspaces of $(\mathbb{R}, d_2)$. ■

■ **Example 3.11** Sometimes the inherited metric for a subspace may not be the most natural one to consider. For example, the unit circle in the plane as a subspace of $\mathbb{R}^2$ has a chord-length distance function (the straight line), as illustrated in Figure 3.2. It is not difficult to verify that arc-length distance makes the circle a metric space—a different but perhaps more natural metric space. ■

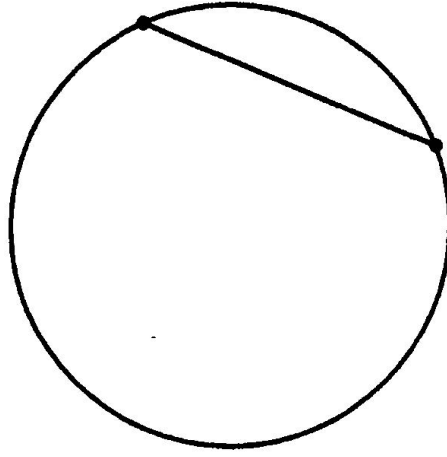


Figure 3.2: chord-length vs arc-length

Definition 3.2.2 — open and closed balls. Let (X, d) be a metric space, let $x_0 \in X$ and let $r \in \mathbb{R}^+$. We define the (*open*) ball $B_{(X,d)}(x_0, r)$ in X , centered at x_0 , and with radius r , in the metric d , to be the set

$$B_{(X,d)}(x_0, r) := \{x \in X \mid d(x, x_0) < r\}.$$

When it is clear what (X, d) is, we abbreviate $B_{(X,d)}(x_0, r)$ as $B(x_0, r)$ or $B_r(x_0)$.

The closed ball centered at x_0 and with radius r is the subset of X defined by

$$\bar{B}_{(X,d)}(x_0, r) := \{x \in X \mid d(x, x_0) \leq r\}.$$

When it is clear what (X, d) is, we abbreviate $\bar{B}_{(X,d)}(x_0, r)$ as $\bar{B}(x_0, r)$ or $\bar{B}_r(x_0)$.

R If $r = 1$, then the ball is called a *unit ball*.

- **Example 3.12** $B_{(\mathbb{R}^2, d_2)}((0, 0), 1) = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1\}$ is an open disc. ■
- **Example 3.13** $B_{(\mathbb{R}^2, d_1)}((0, 0), 1) = \{(x, y) \in \mathbb{R}^2 \mid |x| + |y| < 1\}$ is an open diamond. See Figure 3.1. ■
- **Example 3.14** For the discrete metric d_{disc} on a nonempty set X ,

$$B_{(X, d_{\text{disc}})}(x_0, r) = \begin{cases} \{x_0\} & r \in (0, 1] \\ X & r > 1, \end{cases}$$

and

$$\bar{B}_{(X, d_{\text{disc}})}(x_0, r) = \begin{cases} \{x_0\} & r \in (0, 1) \\ X & r \geq 1. \end{cases}$$

- **Example 3.15** In $[0, 3]$, $B(0, 1) = [0, 1)$ and $B(1, 1) = (0, 2)$. ■

Exercise 3.6 For $1 \leq p < q$, show that the unit ball $B_{(\mathbb{R}^n, d_p)}(\mathbf{0}, 1)$ is contained in the unit ball $B_{(\mathbb{R}^n, d_q)}(\mathbf{0}, 1)$. ■

Definition 3.2.3 — open and closed sets. Let (X, d) be a metric space and let $A \subset X$.

The set A is an *open set* in X if, for each $a \in A$, there is an $r > 0$ such that $B(a, r) \subset A$.

The set A is a *closed set* in X if $X \setminus A$ is an open set in X .

■ **Example 3.16** The open ball $B((0, 0), 1)$ is an open set in $(\mathbb{R}^2, d_2)$.

Indeed, for any $(x, y) \in B((0, 0), 1)$, let $r = 1 - d_2((x, y), (0, 0)) > 0$, and then we see that $B((x, y), r) \subset B((0, 0), 1)$. ■

■ **Example 3.17** The *closed ball* $\bar{B}((0, 0), 1) = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$ is a closed set in $(\mathbb{R}^2, d_2)$.

Indeed, for any $(x, y) \notin \bar{B}((0, 0), 1)$, let $r = d_2((x, y), (0, 0)) - 1 > 0$ and then we see that $B((x, y), r) \subset \mathbb{R}^2 \setminus \bar{B}((0, 0), 1)$. ■

■ **Example 3.18** $[0, 1]$ is an open set in $[0, 3]$. ■

■ **Example 3.19** $\mathbb{Q}$ as a subset of $\mathbb{R}$ with the usual metric is neither open nor closed in $\mathbb{R}$. Of course, if the metric space is simply $\mathbb{Q}$ with the usual metric, then $\mathbb{Q}$ is both open and closed in $\mathbb{Q}$. ■

Exercise 3.7 Use the definition of closed set to prove that the following are closed sets.

1. The x -axis in $\mathbb{R}^2$ with the usual metric d_2 .
2. Any subset of a discrete metric space.
3. $(0, 1]$ in $X = (0, 5)$ with the metric inherited from $\mathbb{R}$.

Similar to the situation in $\mathbb{R}$, open and closed sets behave nicely under certain set-theoretic operations.

Theorem 3.2.1 Let (X, d) be a metric space.

- (1) $\emptyset, X$ are both open and closed in X .
- (2) For any $x_0 \in X, r \in \mathbb{R}^+, B(x_0, r)$ is open in X , and the *closed ball* $\bar{B}(x_0, r) = \{y \in X \mid d(x_0, y) \leq r\}$ is closed in X .
- (3) For any $x_0 \in X$, the *singleton set* $\{x_0\} \subset X$ is closed in X .
- (4) If $\{E_\alpha\}_{\alpha \in I}$ is a collection of open sets in X , then $\bigcup_{\alpha \in I} E_\alpha$ is also open in X . If $\{E_\alpha\}_{\alpha \in I}$ is a collection of closed sets in X , then $\bigcap_{\alpha \in I} E_\alpha$ is also closed in X .
i.e., an *arbitrary union* of open sets is open. As a consequence, an *arbitrary intersection* of closed sets is closed.
- (5) If $E_1, \dots, E_n$ is a finite collection of open sets in X , then $\bigcap_{i=1}^n E_i$ is also open in X . If $E_1, \dots, E_n$ is a finite collection of closed sets in X , then $\bigcup_{i=1}^n E_i$ is also closed in X .
i.e., a *finite intersection* of open sets is open. As a consequence, a *finite union* of closed sets is closed.



$\emptyset, X$ are both open and closed in X , but they are not necessarily the only sets that are both open and closed! For example, if we consider $\mathbb{Z}$ with the usual metric, then every subset is both open and closed.

■ **Example 3.20** Similar to the situation in $\mathbb{R}$, there can be problems with infinite intersections of open sets in general. For example, let $A_n = B_{1/n}((0, 0))$ in $\mathbb{R}^2$ with the usual metric. Then

$$\bigcap_{n=1}^{\infty} A_n = \{(0, 0)\}$$

is not open. ■

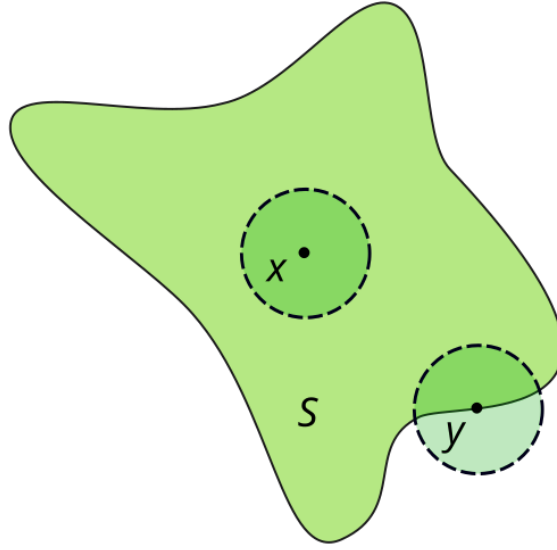


Figure 3.3: Here, x is an interior point of S and y is a boundary point of S

Exercise 3.8 Let (X, d) be a metric space, let A be an open set in X , and let B be a closed set in X . Show that $A \setminus B$ is open and $B \setminus A$ is closed. ■

Definition 3.2.4 — interior, exterior, boundary. Let (X, d) be a metric space, let $A \subset X$.

- A point $x_0 \in X$ is an *interior point* of A if there exists $r > 0$ such that $B(x_0, r) \subset A$; The set of all interior points of A is called the *interior* of A , and is denoted by $\text{int}(A)$ or A° .
- A point $x_0 \in X$ is an *exterior point* of A if there exists $r > 0$ such that $B(x_0, r) \cap A = \emptyset$; The set of all exterior points of A is called the *exterior* of A and is denoted by $\text{ext}(A)$.
- A point $x_0 \in X$ is a *boundary point* of A if it is neither an interior point nor an exterior point of A ; The set of all boundary points of A is called the *boundary* of A and is denoted by ∂A .

So we have $X = \text{int}(A) \cup \text{ext}(A) \cup \partial A$.

- **Example 3.21** For $[0, 1] \subset \mathbb{R}$, its interior is $(0, 1)$ while its boundary is $\{0, 1\}$. ■
- **Example 3.22** For $(2, 3) \cup \{4\} \subset \mathbb{R}$, its interior is $(2, 3)$ while its boundary is $\{2, 3, 4\}$. ■
- **Example 3.23** In the Euclidean space $\mathbb{R}^2$, both $B((0, 0), 1)$ and $\bar{B}((0, 0), 1)$ have interior $B((0, 0), 1)$ and boundary $\partial B((0, 0), 1) = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$. The boundary set is called a *circle*. ■
- **Example 3.24** In the n -dimensional Euclidean space $\mathbb{R}^n$,

$$S^{n-1} = \partial B(\mathbf{0}, 1) = \{x \in \mathbb{R}^n \mid \sum_{i=1}^n x_i^2 = 1\}$$

is called an $(n - 1)$ -*sphere*. ■

- **Example 3.25** $\mathbb{Z} \subset \mathbb{R}$ has no interior point. The exterior is $\mathbb{R} \setminus \mathbb{Z}$ and the boundary is $\mathbb{Z}$ itself. ■
- **Example 3.26** $\mathbb{Q} \subset \mathbb{R}$ has no interior or exterior point. The boundary of $\mathbb{Q}$ is $\mathbb{R}$. ■
- **Example 3.27** Let $A = \{\frac{1}{n} \mid n \in \mathbb{N}^+\} \subset \mathbb{R}$ with the usual metric. Then the boundary points are $A \cup \{0\}$. The set A is neither open nor closed. ■

Lemma 3.2.2 Let A be a subset of a metric space (X, d) .

- (1) $\text{int}(A)$ is the union of all open sets which are contained in A . Thus, the interior of A is the “largest” open set contained in A .
- (2) $\text{ext}(A)$ is open.
- (3) ∂A is closed.

Proof. Since $\partial A = X \setminus (\text{int}(A) \cup \text{ext}(A))$, it suffices to prove (1) and (2).

1. Let $U \subset A$ be an open set. For $x \in U$, there exists $r > 0$ such that $B(x, r) \subset U \subset A$, so $x \in \text{int}(A)$. Thus $U \subset \text{int}(A)$. So $\text{int}(A)$ contains all open sets contained in A .
For $x \in \text{int}(A)$, there exists $r > 0$ such that $B(x, r) \subset A$. For any $y \in B(x, r)$, let $r_y = r - d(x, y) > 0$, and then $B(y, r_y) \subset B(x, r) \subset A$, so $y \in \text{int}(A)$. Consequently, $B(x, r) \subset \text{int}(A)$ and $\text{int}(A)$ is open. Since $\text{int}(A)$ is open, $\text{int}(A)$ is the union of all open sets which are contained in A .
2. For $x \in \text{ext}(A)$, there exists $r > 0$ such that $B(x, r) \subset X \setminus A$. For any $y \in B(x, r)$, let $r_y = r - d(x, y) > 0$, and then $B(y, r_y) \subset B(x, r) \subset X \setminus A$, so $y \in \text{ext}(A)$. Consequently, $B(x, r) \subset \text{ext}(A)$ and $\text{ext}(A)$ is open. ■

Corollary 3.2.3 Let A be a subset of a metric space (X, d) . Then A is open if and only if $A = \text{int}(A)$.

End of Week 6. To be continued...

Definition 3.2.5 — accumulation, isolated and adherent points. Let (X, d) be a metric space and let $A \subset X$.

- A point $x_0 \in X$ is an *accumulation point* of A if, for every $r > 0$, we have $(B(x_0, r) \setminus \{x_0\}) \cap A \neq \emptyset$.
- A point $x_0 \in X$ is an *isolated point* of A if there exists $r > 0$ such that $B(x_0, r) \cap A = \{x_0\}$. In particular, $x_0 \in A$.
- A point $x_0 \in X$ is an *adherent point* of A if for every $r > 0$, we have $B(x_0, r) \cap A \neq \emptyset$.

R If x_0 is an accumulation point of A , there are points of A (other than x_0) that are arbitrarily close to x_0 . Note that, x_0 may or may not be an element of A .

Lemma 3.2.4 A point $x_0 \in X$ is an adherent point of A if and only if it is either an accumulation point or an isolated point.

Proof. [Reading for you]

Suppose $x_0 \in X$ is an adherent point of A , and x_0 is not an isolated point of A . Then for every $r > 0$, we have $B(x_0, r) \cap A \neq \emptyset$, and $B(x_0, r) \cap A \neq \{x_0\}$. Thus $(B(x_0, r) \setminus \{x_0\}) \cap A \neq \emptyset$, so x_0 is an accumulation point.

Conversely, suppose $x_0 \in X$ is an accumulation point. Then for every $r > 0$, we have $(B(x_0, r) \setminus \{x_0\}) \cap A \neq \emptyset$, so in particular $B(x_0, r) \cap A \neq \emptyset$, showing x_0 is an adherent point. Lastly, suppose x_0 is an isolated point. Since $x_0 \in A$, we have $B(x_0, r) \cap A \neq \emptyset$, so x_0 is an adherent point. ■

■ **Example 3.28** $\mathbb{Z} \subset \mathbb{R}$ has no accumulation point, and every point of $\mathbb{Z}$ is an isolated point. ■

■ **Example 3.29** For $\mathbb{Q} \subset \mathbb{R}$, every point in $\mathbb{R}$ is an accumulation point of $\mathbb{Q}$, and $\mathbb{Q}$ has no isolated points. ■

■ **Example 3.30** Let $A = \{\frac{1}{n} \mid n \in \mathbb{N}^+\} \subset \mathbb{R}$ with the usual metric. Then every point of A is an isolated point, and the point 0 is the only accumulation point of the set. ■

Lemma 3.2.5 Let (X, d) be a metric space and let $A \subset X$. Then any $x_0 \in A$ is an adherent point of A .

Proof. This follows from the definition. ■

R Note that $\{\text{adherent points of } A\} = \{\text{accumulation points of } A\} \cup \{\text{isolated points of } A\}$ as a disjoint union, and $\{\text{adherent points of } A\} = A \cup \{\text{accumulation points of } A\}$ as a (not necessarily disjoint) union.

Recall that we proved a subset of $\mathbb{R}$ is closed if and only if it contains all its accumulation points. This continues to be the case in a general metric space.

Lemma 3.2.6 Let (X, d) be a metric space and let $A \subset X$. The following statements are logically equivalent:

- (1) A is closed in X .
- (2) A contains all of its accumulation points.
- (3) A contains all of its adherent points.
- (4) A is the set of all its adherent points.

Proof. The proof of (1) $\iff$ (2) is essentially the same as the proof of Lemma 2.3.2.

The proof of (2) $\iff$ (3) follows from the fact that isolated points are contained in A .

The proof of (3) $\iff$ (4) follows from the previous lemma. ■

If a set A in a metric space X is not closed and we wish that it were, then we can do something about it.

Definition 3.2.6 — closure. Let (X, d) be a metric space and let $A \subset X$. The *closure* of A is the set of all adherent points of A , and is denoted by $\bar{A}$ or $\text{clos}(A)$. In particular, $A \subset \bar{A}$.

■ **Example 3.31** For $[0, 1) \subset \mathbb{R}$, the closure is $[0, 1]$. ■

■ **Example 3.32** For $\mathbb{Z} \subset \mathbb{R}$, the closure is itself. i.e., $\bar{\mathbb{Z}} = \mathbb{Z}$. ■

■ **Example 3.33** $\bar{\mathbb{Q}} = \mathbb{R}$. ■

■ **Example 3.34** Let $A = \{\frac{1}{n} \mid n \in \mathbb{N}^+\} \subset \mathbb{R}$ with the usual metric. Then $\bar{A} = A \cup \{0\}$. ■

Now we present another characterization of adherent points.

Lemma 3.2.7 Let (X, d) be a metric space, let $A \subset X$, and let $x_0 \in X$. Then x_0 is an adherent point of A if and only if x_0 is either an interior point or a boundary point of A .

Proof. By definition, x_0 is an adherent point of A if and only if for all $r > 0$, $B_r(x_0) \cap A \neq \emptyset$. This is equivalent to $x_0 \notin \text{ext}(A)$, namely, x_0 is either an interior point or a boundary point of A . ■

Corollary 3.2.8 Let (X, d) be a metric space and let $A \subset X$.

$$\bar{A} = \text{int}(A) \cup \partial A = X \setminus \text{ext}(A).$$

In particular, $\bar{A}$ is a closed set.

There is another way to characterize the closure of a set A .

Corollary 3.2.9 Let (X, d) be a metric space and let $A \subset X$. Then $\bar{A}$ is the intersection of all closed sets in X which contain A , hence closed. Thus the closure of A is the “smallest” closed set that contains A .

Proof. Since $\bar{A}$ is a closed set and $A \subset \bar{A}$, it suffices to show that for all closed sets B in X with $A \subset B$, we have $\bar{A} \subset B$. Indeed, by definition, $\text{ext}(A) \supset \text{ext}(B)$ since $A \subset B$. Hence, by the previous corollary, we have

$$\bar{A} = X \setminus \text{ext}(A) \subset X \setminus \text{ext}(B) = \bar{B} = B,$$

where the last equality follows from the assumption that B is closed. ■

Exercise 3.9 Give an example of a metric space (X, d) , $x_0 \in X$ and $r > 0$, such that

$$\overline{B(x_0, r)} \neq \{y \in X \mid d(x_0, y) \leq r\}.$$

Here, $\overline{B(x_0, r)}$ is the closure of the open ball $B(x_0, r)$. ■

Exercise 3.10 Let A be a subset of (X, d) . Show that $\partial A = \bar{A} \cap \overline{A^c}$. Here, $A^c := X \setminus A$. ■

So we have the following characterizations of the closure:

$$\begin{aligned} \bar{A} &= \{\text{adherent points of } A\} \\ &= \{\text{accumulation points of } A\} \cup \{\text{isolated points of } A\} \\ &= \{\text{accumulation points of } A\} \cup A \\ &= \text{int}(A) \cup \partial A \\ &= X \setminus \text{ext}(A) \\ &= \text{the intersection of all closed sets in } X \text{ which contain } A. \end{aligned}$$

Definition 3.2.7 Let (X, d) be a metric space. A subset $A \subset X$ is said to be dense in X if $\bar{A} = X$.

R The condition $\bar{A} = X$ is equivalent to saying every point in X is an adherent point of A . Thus, A is dense if and only if for any $x \in X$, for any $\varepsilon > 0$, there exists $y \in A$ such that $d(x, y) < \varepsilon$.

- **Example 3.35** In the usual metric, $\mathbb{Q}$ is dense in $\mathbb{R}$. ■
- **Example 3.36** In any metric space X , X is dense in X . ■
- **Example 3.37** In any discrete metric space X , the only dense subset of X is X itself. This is because in a discrete metric space X , any subset is closed so the closure is itself. ■

Exercise 3.11 Show the “dyadic numbers,” that is, the set $D = \{\frac{a}{2^n} \mid a \in \mathbb{Z}, n \in \mathbb{N}\}$, are dense in $\mathbb{R}$ in the usual metric. ■

3.3 Sequences in metric spaces

3.3.1 Sequences in general metric spaces and complete metric spaces

In a metric space, we can also consider sequences and subsequences.

Definition 3.3.1 — sequence and subsequence. A *sequence* in a metric space (X, d) is a function $f : \mathbb{N}^+ \rightarrow X$. We often denote such a sequence by $(a_k)_{k=1}^\infty$ or $(a_1, a_2, \dots, a_k, \dots)$, where $a_k = f(k)$.

Given a sequence $(a_k)_{k=1}^\infty$, a *subsequence* is a sequence of the form $(a_{k_j})_{j=1}^\infty$, where $(k_j)_{j \in \mathbb{N}^+}$ is a strictly monotonic increasing sequence of natural numbers, i.e., $k_{j+1} > k_j$.

Definition 3.3.2 — convergence and limit. Let $(a_n)_{n=1}^\infty$ is a sequence in a metric space (X, d) , and let $a \in X$. We say that $(a_n)_{n=1}^\infty$ *converges* to a (with respect to the metric d) provided that $\lim_{n \rightarrow \infty} d(a_n, a) = 0$. In other words, for all $\varepsilon > 0$, there exists an $N \in \mathbb{N}^+$ such that for all $n \geq N$, $d(a_n, a) < \varepsilon$.

a is called the *limit* and we write $\lim_{n \rightarrow \infty} a_n = a$.

Lemma 3.3.1 — uniqueness of limit. Every convergent sequence in a metric space has a unique limit.

Proof. [Basically the same proof as before, reading exercise for you] Assume that (x_n) converges to $x, y \in X$. For any $\varepsilon > 0$ there exist $N, N' \in \mathbb{N}^+$ such that $d(x_n, x) < \varepsilon$ for $n \geq N$ and $d(x_n, y) < \varepsilon$ for $n \geq N'$. Then for any $n \geq \max\{N, N'\}$ we have $d(x, y) \leq d(x, x_n) + d(x_n, y) < \varepsilon + \varepsilon = 2\varepsilon$. This inequality is true for any $\varepsilon > 0$, so we must have $d(x, y) = 0$, which implies $x = y$. ■

Sequences are useful for the characterization of accumulation points of subsets.

Lemma 3.3.2 Let (X, d) be a metric space. Let $A \subset X$ be a subset. A point $x \in X$ is an accumulation point of $A \subset X$ if and only if there exists a sequence in $A \setminus \{x\}$ convergent to x .

Proof. [Basically the same situation as before, reading exercise for you]

($\Rightarrow$) : Let $x \in X$ be an accumulation point of A . Then, for any $n \geq 1$ there exists $x_n \in A \setminus \{x\}$ such that $d(x_n, x) < 1/n$. This implies that the sequence $(x_n)_{n=1}^\infty$ is contained in $A \setminus \{x\}$ and converges to x .

($\Leftarrow$) : Assume that there exists a sequence $(x_n)_{n=1}^\infty$ in $A \setminus \{x\}$ convergent to x . Then, for any $\varepsilon > 0$ there exists $n \geq 1$ such that $d(x_n, x) < \varepsilon$. This implies that x is an accumulation point of A . ■

R Similarly, a point $x \in X$ is an adherent point of $A \subset X$ if and only if there exists a sequence in A convergent to x .

Corollary 3.3.3 A subset $A \subset X$ is closed if and only if the limit of any convergent sequence of points in A is contained in A .

Definition 3.3.3 — bounded set and bounded sequence. A subset A in a metric space (X, d) is *bounded* if $X = \emptyset$ or there exists a point $x \in X$ and $r > 0$ such that $A \subset B(x, r)$.

A sequence $(a_n)_{n=1}^\infty$ in (X, d) is *bounded* if the set $\{a_n \mid n \in \mathbb{N}^+\}$ is bounded.

Definition 3.3.4 — Cauchy sequence. A sequence $(a_n)_{n=1}^\infty$ in a metric space (X, d) is a *Cauchy sequence* provided that for all $\varepsilon > 0$, there exists $N \in \mathbb{N}^+$, for all $n, m \geq N$, $d(a_n, a_m) < \varepsilon$.

Lemma 3.3.4 In any metric space (X, d) , a convergent sequence is a Cauchy sequence.

Proof. This follows from the definitions and the triangle inequality. ■

Lemma 3.3.5 In any metric space (X, d) , a Cauchy sequence is a bounded sequence.

Proof. This follows from the definitions and the same proof as the rational case. ■

To summarize, in general metric space, we still have

$$\text{convergent} \implies \text{Cauchy} \implies \text{bounded}$$

On the other hand, it may be that a sequence in a metric space is a Cauchy sequence even though it does not converge. For example, as we observed in Chapter 1, Cauchy sequences in $\mathbb{Q}$ with the usual metric do not necessarily converge in $\mathbb{Q}$.

Definition 3.3.5 — complete metric space. A metric space (X, d) is *complete* if every Cauchy sequence in X converges.

That is, for complete metric spaces, we have

$$\text{convergent seq in complete metric space} \Leftrightarrow \text{Cauchy seq in complete metric space} \Rightarrow \text{bounded}$$

■ **Example 3.38** The following metric spaces are complete.

1. $\mathbb{R}$ with the usual metric;
2. $\mathbb{Z}$ with the usual metric;
3. any discrete metric space;

Indeed, a sequence $(a_n)_{n=1}^{\infty}$ in a discrete metric space (X, d_{disc}) is a Cauchy sequence if and only if there exists $N \in \mathbb{N}^+$ and $a \in X$, such that for all $n \geq N$, $a_n = a$. Such a sequence is called *eventually constant*. In particular, $\lim_{n \rightarrow \infty} a_n = a$ and thus (X, d_{disc}) is a complete metric space.

4. any metric space (X, d) where X has only finitely many points.

■ **Example 3.39** The following metric spaces are *not* complete.

1. $\mathbb{Q}$ with the usual metric;
2. Any proper nonempty open subset of $\mathbb{R}$ with the inherited metric.

So what about closed subsets?

Lemma 3.3.6 Let (X, d) be a complete metric space. Then a subset Y is also a complete metric space with the inherited metric if and only if Y is closed as a subset of X .

Proof. Assume that Y is closed. Assume that $(x_n)_{n=1}^{\infty}$ is a Cauchy sequence in Y . Then it is Cauchy in X , and therefore it is convergent to some $x \in X$ by the completeness of X . As Y is closed, we conclude that $x \in Y$. This means that $(x_n)_{n=1}^{\infty}$ is convergent to a point in Y . Therefore Y is complete.

Assume that Y is complete. Let $x \in X$ be an accumulation point of Y . Then there exists a sequence $(x_n)_{n=1}^{\infty}$ in Y convergent to x . We have to show that $x \in Y$. As $(x_n)_{n=1}^{\infty}$ is a convergent sequence, it is a Cauchy sequence. Since Y is complete, it converges to some $y \in Y$. We conclude from the uniqueness of limits that $x = y \in Y$. ■

3.3.2 The spaces $(\mathbb{R}^n, d_p)$

We now turn to an investigation of convergence in the spaces $\mathbb{R}^n$ with the d_p metrics.

■ **Example 3.40** Let $((x_n, y_n))_{n=1}^{\infty}$ be a sequence in $(\mathbb{R}^2, d_2)$. Then $\lim_{n \rightarrow \infty} (x_n, y_n) = (x, y) \in \mathbb{R}^2$ if and only if

$$\lim_{n \rightarrow \infty} x_n = x, \quad \lim_{n \rightarrow \infty} y_n = y.$$

■

More generally, we have that a sequence converges in $(\mathbb{R}^n, d_p)$ if and only if every coordinate converges.

Lemma 3.3.7 Let $(x_k)_{k=1}^\infty$ be a sequence in $\mathbb{R}^n$. We write $x_k = (x_{k,1}, x_{k,2}, \dots, x_{k,n})$. Let $y = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$. Let $p \in [1, \infty]$. Then the following statements are equivalent:

- (1) $(x_k)_{k=1}^\infty$ converges to y with respect to the metric d_2 .
- (2) $(x_k)_{k=1}^\infty$ converges to y with respect to the metric d_p .
- (3) For every $j \in \{1, \dots, n\}$, $\lim_{k \rightarrow \infty} x_{k,j} = y_j$.

Proof. Exercise. ■

So even though the d_p metrics are different, the notion of convergence is the same in all these spaces.

In $(\mathbb{R}^n, d_p)$ we also have the Cauchy criterion:

Proposition 3.3.8 For $1 \leq p \leq \infty$, the metric space $(\mathbb{R}^n, d_p)$ is a complete metric space.

Proof. Exercise. ■

And in $(\mathbb{R}^n, d_p)$ we can also partially go back from being bounded to being convergent via a subsequence:

Proposition 3.3.9 For any $n \in \mathbb{N}^+$ and $p \in [1, \infty]$, every bounded sequence in $(\mathbb{R}^n, d_p)$ has a convergent subsequence.

R For the proof, we need to choose the convergent subsequence in some way. For $n = 1$ we know how to do it, and we know a subsequence converges if and only if each coordinate converges. So the idea is to look at each coordinate, and try to piece them together. We can try to first figure out the 2-dimensional case and generalize to n -dimension using induction.

Proof. We prove the lemma by induction.

The base case $n = 1$ follows from Lemma 2.1.6.

Suppose that the lemma holds for $n \in \mathbb{N}^+$. Now we consider the case for $n + 1$. Let $(x_k)_{k=1}^\infty$ be a bounded sequence in $(\mathbb{R}^{n+1}, d_p)$. We write $x_k = (x_{k,1}, x_{k,2}, \dots, x_{k,n+1})$ and $x'_k = (x_{k,1}, x_{k,2}, \dots, x_{k,n})$. Then $(x'_k)_k$ is a bounded sequence in $(\mathbb{R}^n, d_p)$, and $(x_{k,n+1})_k$ is a bounded sequence in $\mathbb{R}$.

Since $(x'_{k_j})_{j=1}^\infty$ is a bounded sequence in $(\mathbb{R}^n, d_p)$, by the inductive assumption, it has a convergent subsequence $(x'_{k_{j_s}})_{s=1}^\infty$.

Since $(x_{k,n+1})_{k=1}^\infty$ is a bounded sequence in $\mathbb{R}$, the subsequence $(x_{k_{j_s},n+1})_{s=1}^\infty$ is also bounded. So by Lemma 2.1.6 again, the subsequence $(x_{k_{j_s},n+1})_{s=1}^\infty$ has a convergent subsequence $(x_{k_{j_{s_t}},n+1})_{t=1}^\infty$.

Now we look at the further subsequence $(x'_{k_{j_{s_t}}})_{t=1}^\infty$ of the subsequence $(x'_{k_{j_s}})_{s=1}^\infty$. It is a subsequence of a convergent sequence, hence also convergent. By Lemma 3.3.7, each component of it is convergent.

Thus we can form the subsequence $(x_{k_{j_{s_t}}})_{t=1}^\infty$ of $(x_k)_{k=1}^\infty$ in $(\mathbb{R}^{n+1}, d_p)$. By Lemma 3.3.7, this is convergent since every component is convergent. We have thus found the desired convergent subsequence.

In summary, the lemma holds for all $n \in \mathbb{N}^+$. ■

R Why in the proof, we need to get the subsequence of a subsequence? What goes wrong if I just choose a subsequence of x'_k and a subsequence of $x_{k,n+1}$ independently?

■ **Example 3.41** Let (X, d) be a discrete metric space such that X is an infinite set and let $(x_n)_{n=1}^{\infty}$ be a sequence in X consisting of distinct elements from X . Then $(x_n)_{n=1}^{\infty}$ is a bounded sequence, but it doesn't have a convergent subsequence.

Therefore, the previous lemma *doesn't* hold for all complete metric spaces. ■

Theorem 3.3.10 — Bolzano-Weierstrass theorem. If A is a bounded infinite subset in $\mathbb{R}^n$ with the usual Euclidean metric (or more generally any d_p metric for $p \in [1, \infty]$), then A has an accumulation point.

■ **Example 3.42** Again, this theorem *doesn't* hold for all complete metric spaces. The discrete metric space (X, d) is bounded, but has no accumulation point. ■

Proof. Let $(x_k)_{k=1}^{\infty}$ be a sequence in A such that $x_k \neq x_j$ if $k \neq j$. Since A is a bounded set, $(x_k)_{k=1}^{\infty}$ is a bounded sequence and by Lemma 3.3.9, has a convergent subsequence $(x_{k_j})_{j=1}^{\infty}$. Then $x_0 = \lim_{j \rightarrow \infty} x_{k_j}$ is an accumulation point of A . ■

End of Week 7. To be continued...

3.3.3 Sequences of functions

Now we discuss the convergence of sequences of functions $(f_n)_{n=1}^{\infty}$. Here each f_n is a function. There is more than one notion of what it means for a sequence of functions to converge. The two of the most important of these notions are *pointwise convergence* and *uniform convergence*. We do this in the case of sequences of bounded functions from a set X to $\mathbb{R}$.

Definition 3.3.6 — the space of bounded functions $\mathcal{B}(X, \mathbb{R})$. Let X be a set and $\mathcal{B}(X, \mathbb{R})$ denote the set of all bounded functions from X to $\mathbb{R}$. Namely, $f \in \mathcal{B}(X, \mathbb{R})$ if and only if there exists $M > 0$, such that for all $x \in X$, $|f(x)| \leq M$.

R Note that we can add two bounded functions and multiply one bounded function by a scalar and still get a bounded function. So $\mathcal{B}(X, \mathbb{R})$ is a vector space.

Definition 3.3.7 — pointwise convergence. Let X be a set, and let $(f_n)_{n=1}^{\infty}$ be a sequence in $\mathcal{B}(X, \mathbb{R})$. We say that $(f_n)_{n=1}^{\infty}$ *converges pointwise* to a function $f : X \rightarrow \mathbb{R}$ (not necessarily bounded) if for every $x \in X$, $\lim_{n \rightarrow \infty} f_n(x) = f(x)$. In this case, f is called the *pointwise limit* of $(f_n)_{n=1}^{\infty}$.

R The pointwise limit is unique.

■ **Example 3.43** Let $f_n : [0, 1] \rightarrow \mathbb{R}$ be given by

$$f_n(x) = \begin{cases} 0 & 0 \leq x \leq 1 - \frac{1}{n} \\ nx - (n-1) & 1 - \frac{1}{n} < x \leq 1. \end{cases}$$

Note that for all $x < 1$, $(f_n(x))_{n=1}^{\infty}$ is eventually the constant 0, so $\lim_{n \rightarrow \infty} f_n(x) = 0$. When $x = 1$, $f_n(x) = 1$ for all $n \in \mathbb{N}^+$, so $\lim_{n \rightarrow \infty} f_n(1) = 1$. Hence, $(f_n)_{n=1}^{\infty}$ converges pointwise to

$$f(x) = \begin{cases} 0 & 0 \leq x < 1 \\ 1 & x = 1. \end{cases}$$

Notice that there is a jump! ■

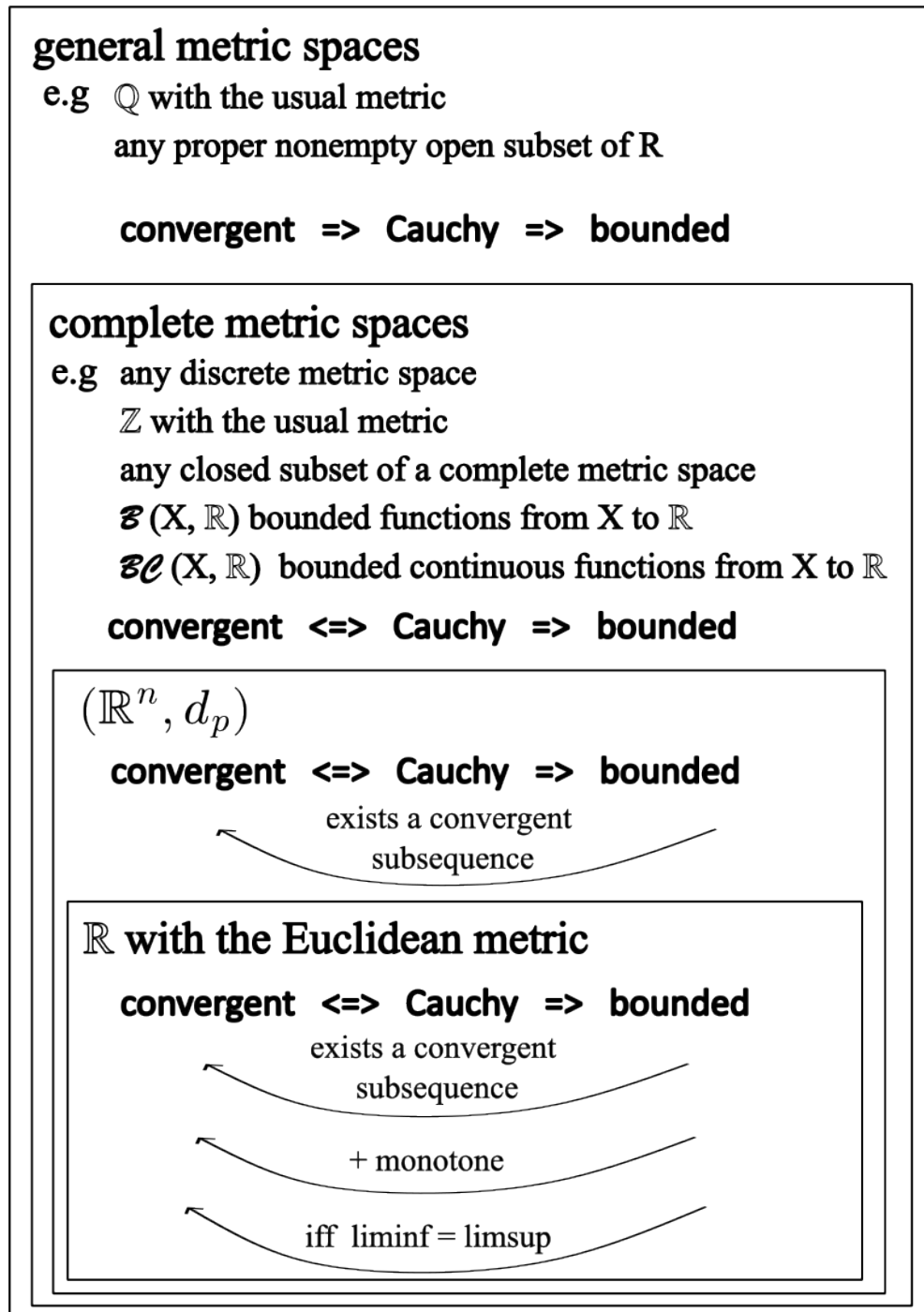


Figure 3.4: Sequences in metric spaces.

■ **Example 3.44** Let $f_n : [0, 1] \rightarrow \mathbb{R}$ be given by

$$f_n(x) = \begin{cases} 0 & 0 \leq x < \frac{1}{n} \\ \frac{1}{x} & \frac{1}{n} \leq x \leq 1. \end{cases}$$

They are bounded functions, but their pointwise limit is an unbounded function

$$f(x) = \begin{cases} \frac{1}{x} & 0 < x \leq 1 \\ 0 & x = 0. \end{cases}$$

■

The other major notion of convergence for sequences of functions that we will discuss is uniform convergence. This notion of convergence utilizes the sup metric on $\mathcal{B}(X, \mathbb{R})$ defined below:

Definition 3.3.8 — sup-norm and sup-metric. For $f \in \mathcal{B}(X, \mathbb{R})$, we define the sup-norm of f by $\|f\| = \sup_{x \in X} |f(x)|$. (Check that it is indeed a norm!)

Then we can define the distance function d as $d(f, g) = \|f - g\| = \sup_{x \in X} |f(x) - g(x)|$ for $f, g \in \mathcal{B}(X, \mathbb{R})$. It follows that $(\mathcal{B}(X, \mathbb{R}), d)$ is a metric space.

Definition 3.3.9 — uniform convergence. Let X be a set, and let $(f_n)_{n=1}^\infty$ be a sequence in $\mathcal{B}(X, \mathbb{R})$. We say that $(f_n)_{n=1}^\infty$ *converges uniformly* to a function f if for all $\varepsilon > 0$, $\exists N \in \mathbb{N}^+$, such that for $n \geq N$,

$$\sup_{x \in X} |f_n(x) - f(x)| < \varepsilon.$$

R If the sequence $(f_n)_{n=1}^\infty$ converges to f uniformly, then $f \in \mathcal{B}(X, \mathbb{R})$ as well.

R We can see from the definition that uniform convergence is equivalent to the convergence with respect to the sup metric d we defines on $\mathcal{B}(X, \mathbb{R})$. Thus by Lemma 3.3.1, the uniform limit is unique.

Uniform convergence is a stronger notion:

Lemma 3.3.11 If a sequence $(f_n)_{n=1}^\infty$ in $\mathcal{B}(X, \mathbb{R})$ converges uniformly to a function f , then it converges pointwise to the function f .

Exercise 3.12 Let $f_n(x) = x^n$ for $n \in \mathbb{N}^+$.

- (1) Show that the sequence $(f_n)_{n=1}^\infty$ converges pointwise to the function $f(x) = 0$ on the interval $(-1, 1)$.
- (2) Show that the sequence $(f_n)_{n=1}^\infty$ converges uniformly to the function $f(x) = 0$ on the interval $[-\frac{1}{2}, \frac{1}{2}]$.
- (3) Show that the sequence $(f_n)_{n=1}^\infty$ doesn't converges uniformly on the interval $(-1, 1)$.

■

Recall that being complete means all the Cauchy sequences converge.

Theorem 3.3.12 The space $\mathcal{B}(X, \mathbb{R})$ is a complete metric space. Here, recall that the metric d is

the “sup metric”, i.e., for every $f, g \in \mathcal{B}(X, \mathbb{R})$,

$$d(f, g) = \sup_{x \in X} |f(x) - g(x)|.$$

Proof. Suppose that $(f_n)_{n=1}^\infty$ is a Cauchy sequence in $\mathcal{B}(X, \mathbb{R})$. Then, for every $x \in X$, the sequence $(f_n(x))_{n=1}^\infty$ is a Cauchy sequence since

$$|f_n(x) - f_m(x)| \leq \sup_{y \in X} |f_n(y) - f_m(y)|.$$

Thus, $(f_n)_{n=1}^\infty$ pointwise converges to a limit $f : X \rightarrow \mathbb{R}$.

We want to show that $(f_n)_{n=1}^\infty$ converges uniformly to f . For all $\varepsilon > 0$, there exists $N \in \mathbb{N}^+$, such that for all $n, m \geq N$,

$$\sup_{y \in X} |f_n(y) - f_m(y)| < \frac{\varepsilon}{2}.$$

Then for each x , there exists $N(x) > N$, such that

$$|f_{N(x)}(x) - f(x)| < \frac{\varepsilon}{2},$$

and thus, for all $n \geq N$,

$$|f_n(x) - f(x)| \leq |f_n(x) - f_{N(x)}(x)| + |f_{N(x)}(x) - f(x)| < \varepsilon.$$

To complete the proof, it suffices to show that $f \in \mathcal{B}(X, \mathbb{R})$. Indeed, this follows immediately from the uniform convergence. Namely, for all $\varepsilon > 0$, there exists $N \in \mathbb{N}^+$, such that for all $x \in X$,

$$|f(x)| < |f_N(x)| + \varepsilon.$$

■

■ **Example 3.45** Let's revisit this example: Let $f_n : [0, 1] \rightarrow \mathbb{R}$ be given by

$$f_n(x) = \begin{cases} 0 & 0 \leq x \leq 1 - \frac{1}{n} \\ nx - (n-1) & 1 - \frac{1}{n} < x \leq 1. \end{cases}$$

We find that $(f_n)_{n=1}^\infty$ converges pointwise to

$$f(x) = \begin{cases} 0 & 0 \leq x < 1 \\ 1 & x = 1. \end{cases}$$

However, this convergence is *not* uniform: we find for all $n \geq 1$,

$$\sup_{x \in [0, 1]} |f_n(x) - f(x)| = 1.$$

And the sequence $(f_n)_{n=1}^\infty$ is not Cauchy in the metric space $(\mathcal{B}(X, \mathbb{R}), d)$; we can see this either by directly computations or by applying Theorem 3.3.12, which shows a Cauchy sequence must converge uniformly. ■

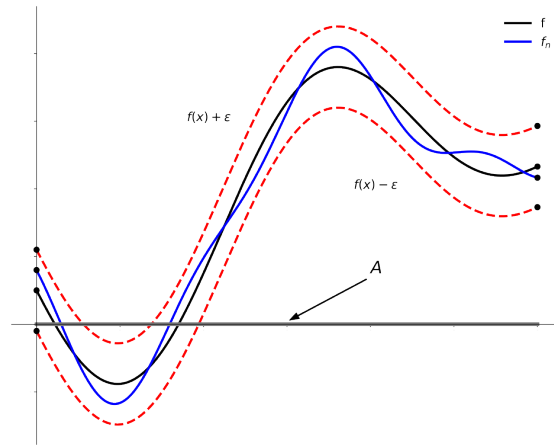


Figure 3.5: uniform convergence visualization; all the f_n are eventually within distance ε from the uniform limit f

3.4 Continuous functions

3.4.1 Limiting values of functions

So far, we have discussed convergence of a sequence in a metric space. In fact, we can generalize this concept to a function defined on a metric space.

In calculus, you've probably seen notations like $\lim_{x \rightarrow x_0} f(x)$. Here we generalize that notion to general metric space. We consider functions that maps to some metric space, not just $\mathbb{R}$. This is because we want to form composition of functions, and functions to spaces other than $\mathbb{R}$ (say $\mathbb{R}^n$) is also very common.

Definition 3.4.1 — convergence of a function at a point. Let (X, d_X) and (Y, d_Y) be metric spaces, let $f : E \rightarrow Y$ be a function, let $E \subset X$, let $x_0 \in X$ be an accumulation point of E , and let $y_0 \in Y$. We say that f *converges to y_0 at x_0* and write

$$\lim_{E \ni x \rightarrow x_0} f(x) = y_0,$$

if $\forall \varepsilon > 0, \exists \delta > 0$, such that $\forall x \in E$ with $0 < d_X(x, x_0) < \delta, d_Y(f(x), y_0) < \varepsilon$.

If f does not converge to any $y \in Y$ at x_0 , we say that f *diverges at x_0* .

R The definition is local, namely, we can replace E by $E \cap B(x_0, r)$ for any $r > 0$ in the definition.

R If $E = X$ or $E \supset B_{(X,d)}(x_0, r)$ for some $r > 0$, by convention, we can simply write $\lim_{x \rightarrow x_0} f(x) = y_0$.

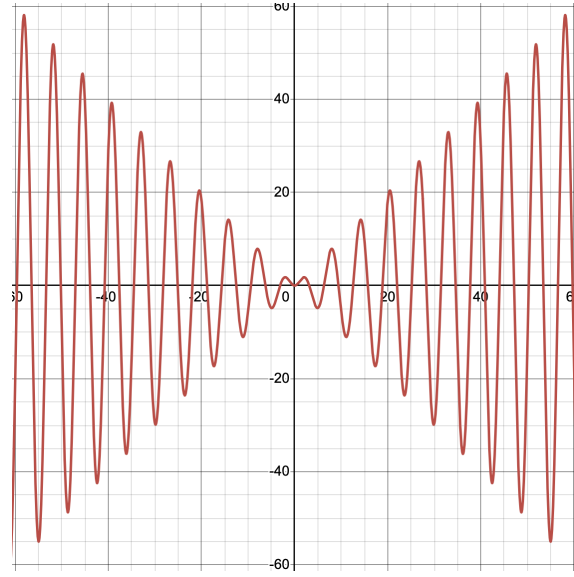
■ **Example 3.46** Let $E = \mathbb{R} \setminus \{0\}$, and $f(x) = x \sin \frac{1}{x}$. Then we have

$$\lim_{E \ni x \rightarrow 0} x \sin \frac{1}{x} = 0.$$

Indeed, $\forall \varepsilon > 0$, let $\delta = \varepsilon$. Then for $0 < |x| < \delta$,

$$\left| x \sin \frac{1}{x} \right| \leq |x| < \delta = \varepsilon.$$

■

Figure 3.6: The graph of $x \sin(1/x)$

■ **Example 3.47** The sign function $\text{sgn} : \mathbb{R} \rightarrow \{-1, 0, 1\}$ is defined by

$$\text{sgn}(x) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0. \end{cases}$$

Then $\lim_{x \rightarrow 0} \text{sgn}(x)$ does not exist, but

$$\lim_{x \rightarrow 0^-} \text{sgn}(x) \equiv \lim_{\mathbb{R}^- \ni x \rightarrow 0} \text{sgn}(x) = -1, \quad \lim_{x \rightarrow 0^+} \text{sgn}(x) \equiv \lim_{\mathbb{R}^+ \ni x \rightarrow 0} \text{sgn}(x) = 1.$$

■

We can also characterize convergence using sequences.

Lemma 3.4.1 Let (X, d_X) and (Y, d_Y) be metric spaces, let $f : E \rightarrow Y$ be a function, let $E \subset X$, let $x_0 \in X$ be an accumulation point of E , and let $y_0 \in Y$. Then $\lim_{E \ni x \rightarrow x_0} f(x) = y_0$ if and only if for every sequence $(x_n)_{n=1}^\infty$ in $E \setminus \{x_0\}$ satisfying $\lim_{n \rightarrow \infty} x_n = x_0$, we have $\lim_{n \rightarrow \infty} f(x_n) = y_0$.

Proof. [Reading for you]

If $\lim_{E \ni x \rightarrow x_0} f(x) = y_0$, then for every $\varepsilon > 0$, there exists $\delta > 0$, such that whenever $x \in E$ with $0 < d_X(x, x_0) < \delta$, $d_Y(f(x), y_0) < \varepsilon$. Since $\lim_{n \rightarrow \infty} x_n = x_0$, there exists $N \in \mathbb{N}^+$, such that for all $n \geq N$, $0 < d_X(x_n, x_0) < \delta$, and thus, $d_Y(f(x_n), y_0) < \varepsilon$. This implies $\lim_{n \rightarrow \infty} f(x_n) = y_0$.

We now prove the converse. If $\lim_{E \ni x \rightarrow x_0} f(x) \neq y_0$, then $\exists \varepsilon_0$, for each $\delta_n = \frac{1}{n}$, there exists $x_n \in E$ satisfying $0 < d_X(x_n, x_0) < \delta_n$ and $d_Y(f(x_n), y_0) \geq \varepsilon_0$. Hence, we obtain a sequence $(x_n)_{n=1}^\infty$ in $E \setminus \{x_0\}$ but $\lim_{n \rightarrow \infty} f(x_n) \neq y_0$, giving a contradiction. ■

Recall that if $f : E \rightarrow \mathbb{R}$ and $g : E \rightarrow \mathbb{R}$ have a common domain of definition E , we can define functions on the same set by:

$$\begin{aligned} (f + g)(x) &:= f(x) + g(x), \\ (f \cdot g)(x) &:= f(x) \cdot g(x), \\ \left(\frac{f}{g}\right)(x) &:= \frac{f(x)}{g(x)}, \quad \text{if } g(x) \neq 0 \text{ for } x \in E. \end{aligned}$$

The following proposition follows from the limit laws for sequences.

Proposition 3.4.2 — limit laws for functions. Let $f : E \rightarrow \mathbb{R}$ and $g : E \rightarrow \mathbb{R}$ be two functions. If $\lim_{E \ni x \rightarrow x_0} f(x) = A$ and $\lim_{E \ni x \rightarrow x_0} g(x) = B$, then

1. $\lim_{E \ni x \rightarrow x_0} (f + g)(x) = A + B$;
2. $\lim_{E \ni x \rightarrow x_0} (f \cdot g)(x) = A \cdot B$;
3. $\lim_{E \ni x \rightarrow x_0} \left(\frac{f}{g}\right)(x) = \frac{A}{B}$, if $B \neq 0$ and $g(x) \neq 0$ for $x \in E$.

Proposition 3.4.3 — limit law for compositions. Let (X, d_X) , (Y, d_Y) and (Z, d_Z) be metric spaces, let $x_0 \in X$, $y_0 \in Y$ be accumulation points, $z_0 \in Z$, and let $f : X \rightarrow Y \setminus \{y_0\}$, $g : Y \rightarrow Z$ be two functions. If $\lim_{x \rightarrow x_0} f(x) = y_0$ and $\lim_{y \rightarrow y_0} g(y) = z_0$, then

$$\lim_{x \rightarrow x_0} g \circ f(x) = z_0.$$

Exercise 3.13 Prove Proposition 3.4.3. ■

■ **Example 3.48** $\lim_{x \rightarrow 0} \frac{\sin x}{x} = 1$.

[The proof is reading exercise for you]

It suffices to show that for $0 < |x| < \frac{\pi}{2}$,

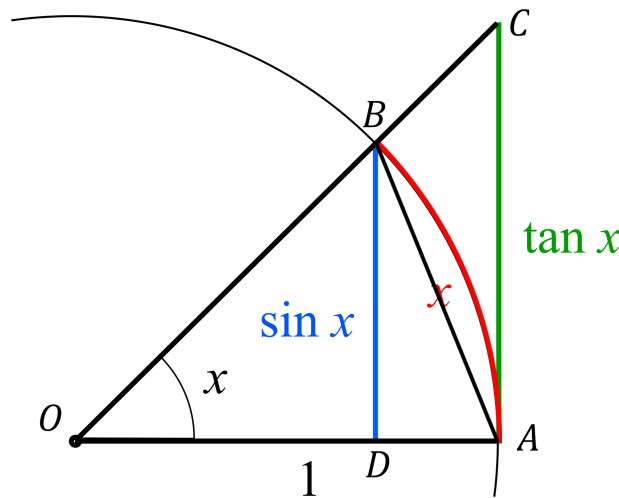
$$\cos x < \frac{\sin x}{x} < 1. \quad (3.2)$$

Indeed, if the inequality (3.2) holds, then we have $|\sin(x)| \leq |x|$, and thus $\lim_{x \rightarrow 0} \sin(x) = 0$. Since (3.2) can also be written as

$$\sqrt{1 - \sin^2 x} < \frac{\sin x}{x} < 1.$$

By the squeeze theorem (Example 2.2), $\lim_{x \rightarrow 0} \frac{\sin x}{x} = 1$ holds.

To see (3.2), we appeal to the high school definition of $\sin$ and $\tan$ and look at the picture below, where we have a circle of radius 1, and x is the angle in radian. Then $\sin x = \frac{BD}{OB} = BD$, and $\tan x = \frac{AC}{OA} = AC$.



From the picture, we know that $\text{area}(\triangle OAB) < \text{area}(\text{sector } OAB) < \text{area}(\triangle OAC)$, so $\frac{1 \cdot \sin x}{2} < \frac{x}{2\pi} \pi < \frac{1 \cdot \tan x}{2}$. Simplifying, we get $\sin x < x < \tan x$, and this implies $\cos x < \frac{\sin x}{x} < 1$. ■

3.4.2 Interchange of limits

Recall that in $\mathcal{B}(X, \mathbb{R})$, a sequence $(f_n)_{n=1}^\infty$ converges pointwise to $f : X \rightarrow \mathbb{R}$ if $\forall x \in X$,

$$\lim_{n \rightarrow \infty} |f_n(x) - f(x)| = 0,$$

and the sequence converges uniformly to f if

$$\lim_{n \rightarrow \infty} (\sup_{x \in X} |f_n(x) - f(x)|) = 0.$$

In this section, we have investigated convergence of a function at a fixed point $x_0 \in X$ that is an accumulation point. We want to determine when the order of limits can be interchanged, i.e., when

$$\lim_{x \rightarrow x_0} \lim_{n \rightarrow \infty} f_n(x) = \lim_{n \rightarrow \infty} \lim_{x \rightarrow x_0} f_n(x),$$

provided the limits exist.

The following example implies that pointwise convergence is not sufficient.

■ **Example 3.49** Let $f_n(x) = x^n$ defined on $[0, 1]$, and the sequence converges pointwise to

$$f(x) = \begin{cases} 0 & x \in [0, 1) \\ 1 & x = 1 \end{cases}.$$

We have $\lim_{n \rightarrow \infty} \lim_{x \rightarrow 1} f_n(x) = 1 \neq 0 = \lim_{x \rightarrow 1} \lim_{n \rightarrow \infty} f_n(x)$. ■

On the other hand, for uniform convergence we can interchange the limit.

■ **Example 3.50** Let $f_n(x) = \frac{\sin(x/n)}{x/n}$ be defined on $(0, 1]$, and the sequence converges pointwise to

$$f(x) = 1.$$

(We can check that this is actually uniform convergence.)

We have $\lim_{n \rightarrow \infty} \lim_{x \rightarrow 0} f_n(x) = 1 = \lim_{x \rightarrow 0} \lim_{n \rightarrow \infty} f_n(x)$. ■

Proposition 3.4.4 — Interchange of limits and uniform limits. Let (X, d_X) be a metric space, and let $E \subset X$. Let $(f_n)_{n=1}^\infty$ be a sequence of functions from $E \rightarrow \mathbb{R}$ which converges *uniformly* in E to some function $f : E \rightarrow \mathbb{R}$. Let $x_0 \in X$ be an accumulation point of E , and suppose that for each n the limit $\lim_{E \ni x \rightarrow x_0} f_n(x)$ exists.

Then $\lim_{E \ni x \rightarrow x_0} f(x)$ also exists and is equal to $\lim_{n \rightarrow \infty} \lim_{E \ni x \rightarrow x_0} f_n(x)$. In other words,

$$\lim_{n \rightarrow \infty} \lim_{E \ni x \rightarrow x_0} f_n(x) = \lim_{E \ni x \rightarrow x_0} \lim_{n \rightarrow \infty} f_n(x).$$

Proof. [Reading exercise for you]

For this, we use Lemma 3.4.1. Then to show $\lim_{E \ni x \rightarrow x_0} f(x)$ exists, we just need to show that for every $(x_n)_{n=1}^\infty$ in $E \setminus \{x_0\}$ with $\lim_{n \rightarrow \infty} x_n = x_0$, the sequence $(f(x_n))_{n=1}^\infty$ converges and different such sequences all converge to the same limit.

First, we show that for every $(x_n)_{n=1}^\infty$ in $E \setminus \{x_0\}$ with $\lim_{n \rightarrow \infty} x_n = x_0$, the sequence $(f(x_n))_{n=1}^\infty$ is a Cauchy sequence. Consequently, since $\mathbb{R}$ is complete, $\lim_{n \rightarrow \infty} f(x_n)$ exists.

$\forall \varepsilon > 0$, by the definition of uniform convergence, $\exists N \in \mathbb{N}^+$, $\forall n \geq N$, $\sup_{x \in E} |f(x) - f_n(x)| < \varepsilon$, and in particular, $\forall k \in \mathbb{N}^+$, $|f(x_k) - f_n(x_k)| < \varepsilon$. Since the limit $\lim_{E \ni x \rightarrow x_0} f_n(x)$ exists, by Lemma 3.4.1, the sequence $(f_n(x_k))_{k=1}^\infty$ is a Cauchy sequence, and thus, $\exists N_1 \in \mathbb{N}^+$, $\forall k, l \geq N_1$, $|f_n(x_k) - f_n(x_l)| < \varepsilon$. These imply that

$$|f(x_k) - f(x_l)| \leq |f(x_k) - f_n(x_k)| + |f_n(x_k) - f_n(x_l)| + |f_n(x_l) - f(x_l)| < 3\varepsilon,$$

so $(f(x_n))_{n=1}^{\infty}$ is a Cauchy sequence.

Second, if $(x_n)_{n=1}^{\infty}$ and $(x'_n)_{n=1}^{\infty}$ are two sequences in $E \setminus \{x_0\}$ with $\lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} x'_n = x_0$, then $(y_n)_{n=1}^{\infty}$ defined by

$$y_n = \begin{cases} x_k & n = 2k \\ x'_k & n = 2k + 1 \end{cases}$$

also converges to x_0 . Hence, $\lim_{n \rightarrow \infty} f(x_n) = \lim_{n \rightarrow \infty} f(y_n) = \lim_{n \rightarrow \infty} f(x'_n)$. By Lemma 3.4.1 again, $L = \lim_{E \ni x \rightarrow x_0} f(x)$ exists.

Finally, $\forall \varepsilon > 0$, we choose $N \in \mathbb{N}^+$ such that $\forall n \geq N$, $\sup_{x \in E} |f(x) - f_n(x)| < \varepsilon$. For each $n \geq N$, we choose $\delta_n > 0$, such that $\forall x \in E$ with $0 < d_X(x, x_0) < \delta_n$, $|f_n(x) - \lim_{E \ni x \rightarrow x_0} f_n(x)| < \varepsilon$, and $|f(x) - L| < \varepsilon$. Then, we fix a $x_1 \in E$ with $0 < d_X(x, x_0) < \delta_n$, and we obtain

$$|\lim_{E \ni x \rightarrow x_0} f_n(x) - L| \leq |\lim_{E \ni x \rightarrow x_0} f_n(x) - f_n(x_1)| + |f_n(x_1) - f(x_1)| + |f(x_1) - L| < 3\varepsilon.$$

In conclusion, $\lim_{n \rightarrow \infty} \lim_{E \ni x \rightarrow x_0} f_n(x) = \lim_{E \ni x \rightarrow x_0} \lim_{n \rightarrow \infty} f_n(x)$. ■

3.4.3 Continuous functions

Definition 3.4.2 — continuity. Let (X, d_X) and (Y, d_Y) be metric spaces, and let $f : X \rightarrow Y$ be a function. For $x_0 \in X$, we say that f is *continuous* at x_0 if $\forall \varepsilon > 0$, $\exists \delta > 0$, such that whenever $d_X(x, x_0) < \delta$, $d_Y(f(x), f(x_0)) < \varepsilon$.

We say that f is *continuous* if and only if f is continuous at every $x \in X$.

R If x_0 is an accumulation point, the ε - δ definition is simply $\lim_{x \rightarrow x_0} f(x) = f(x_0)$.

R If x_0 is an isolated point, then f is automatically continuous at x_0 .

The continuity is also a *local* property; a function may be continuous at some points, but not at others. We often deal with functions $f : X \rightarrow Y$ which are continuous at every point of X . In this case, we simply say that f is continuous on X without reference to any particular point.

■ **Example 3.51** If $f : X \rightarrow Y$ is a constant function (i.e., there exists $y \in Y$ such that $f(x) = y$ for all $x \in X$), then f is continuous on X . ■

■ **Example 3.52** Let the domain and target both be (X, d_X) and let $f : X \rightarrow X$ be the identity function $f(x) = x$. Then f is continuous on X . ■

■ **Example 3.53** The *Dirichlet function* on $\mathbb{R}$ is the indicator function of $\mathbb{Q}$, i.e.

$$f(x) = \begin{cases} 0 & x \notin \mathbb{Q} \\ 1 & x \in \mathbb{Q} \end{cases}.$$

It is nowhere continuous on $\mathbb{R}$. ■

■ **Example 3.54** Suppose X is a discrete metric space, and let f be any function from X to Y where Y is any other metric space. Then f is continuous on X . ■

■ **Example 3.55 — Thomae's function.** Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be

$$f(x) = \begin{cases} 0 & x \notin \mathbb{Q}, \text{ or } x = 0 \\ \frac{1}{q} & x = p/q, \text{ and } p, q \text{ coprime, } q > 0 \end{cases}.$$

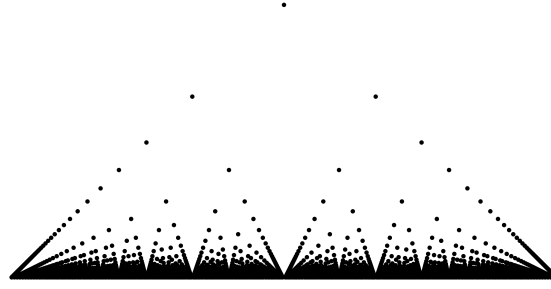


Figure 3.7: Point plot of the Thomae's function on the interval $(0, 1)$. The topmost point in the middle shows $f(1/2) = 1/2$.

Then f is continuous at 0 and irrational numbers, and discontinuous at non-zero rational numbers.

To see that f is continuous at irrational number x_0 , for every $\varepsilon > 0$, let N be a positive integer such that $\frac{1}{N} < \varepsilon$. We define a set of rational numbers

$$S_N := \left\{ \frac{p}{q} \mid 1 \leq q \leq N, p \in \mathbb{Z}, \frac{p}{q} \in [x_0 - 1, x_0 + 1] \right\},$$

which is a finite set. Let $\delta = \min_{x \in S_N} |x - x_0|$. This is positive since x_0 is irrational, and we also see that $B(x, \delta) \cap S_N = \emptyset$. Now for all $x \in B(x_0, \delta)$, either

- $x \notin \mathbb{Q}$, and thus, $f(x) = 0$; Or
- $x = \frac{p}{q} \in \mathbb{Q}$ where $q > N$, and thus, $f(x) = \frac{1}{q} < \frac{1}{N} < \varepsilon$.

In either case, we have $|f(x)| < \varepsilon$. Therefore, $\lim_{x \rightarrow x_0} f(x) = 0 = f(x_0)$. ■

End of Week 8. To be continued...

Theorem 3.4.5 — continuity preserves convergence. Suppose (X, d_X) and (Y, d_Y) are two metric space. Let $f : X \rightarrow Y$ be a function and let $x_0 \in X$. Then the following are equivalent:

1. f is continuous at x_0 .
2. Whenever $(x_n)_{n=1}^\infty$ is a sequence in X with $\lim_{n \rightarrow \infty} x_n = x_0$, we have

$$\lim_{n \rightarrow \infty} f(x_n) = f(x_0).$$

3. For every open set $V \subset Y$ that contains $f(x_0)$, there exists an open set $U \subset X$ containing x_0 such that

$$f(U) \subset V.$$

Proof. 1 $\iff$ 2: If x_0 is an isolated point, then both statements are trivial. If x_0 is an accumulation point, then this follows from Lemma 3.4.1.

2 $\implies$ 3: Assume that 3 is false. Then, there exists an open set $V_0 \ni f(x_0)$, and for every $n \in \mathbb{N}^+$, there exists $x_n \in B(x_0, \frac{1}{n})$ such that $f(x_n) \notin V_0$. As V_0 is open, there exists $r > 0$ such that $B(f(x_0), r) \subset V_0$, and thus, $\forall n \in \mathbb{N}^+$, $d_Y(f(x_0), f(x_n)) \geq r$. However, by 2, since $\lim_{n \rightarrow \infty} x_n = x_0$, we have $\lim_{n \rightarrow \infty} f(x_n) = f(x_0)$ which gives a contradiction.

3 $\implies$ 1: For any $\varepsilon > 0$, let $V = B(f(x_0), \varepsilon)$. Then, by 3, there exists an open set $U \ni x_0$ such that $f(U) \subset V$. As U is open, there exists $\delta > 0$ such that $B(x_0, \delta) \subset U$, and thus, $f(B(x_0, \delta)) \subset B(f(x_0), \varepsilon)$. In particular, for all $x \in B(x_0, \delta)$, $d_Y(f(x), f(x_0)) < \varepsilon$, and therefore, f is continuous at x_0 . ■

Theorem 3.4.6 — topological characterization of continuity. Let (X, d_X) and (Y, d_Y) be two metric space, and let $f : X \rightarrow Y$ be a function. The following statements are equivalent:

1. f is continuous.
2. For every open set $V \subset Y$, $f^{-1}(V) := \{x \in X : f(x) \in V\}$ is open in X .
3. For every closed set $F \subset Y$, $f^{-1}(F) := \{x \in X : f(x) \in F\}$ is closed in X .

Proof. $1 \iff 2$: Since $f(U) \subset V$ if and only if $U \subset f^{-1}(V)$, this follows from the equivalence $1 \iff 3$ in Theorem 3.4.5.

$2 \iff 3$: If 2 holds, then for every closed F , $f^{-1}(F) = X \setminus f^{-1}(Y \setminus F)$ is closed since $Y \setminus F$ is open. Similarly, if 3 holds, then for every open V , $f^{-1}(V) = X \setminus f^{-1}(Y \setminus V)$ is open since $Y \setminus V$ is closed. ■

Corollary 3.4.7 — continuity preserved by composition. Let (X, d_X) , (Y, d_Y) and (Z, d_Z) be metric spaces, and let $f : X \rightarrow Y$, $g : Y \rightarrow Z$ be functions.

1. If f is continuous at $x_0 \in X$ and g is continuous at $f(x_0)$, then $g \circ f$ is continuous at x_0 .
2. If f is continuous, and g is continuous, then $g \circ f$ is continuous.

Definition 3.4.3 Given two functions $f : X \rightarrow Y$ and $g : X \rightarrow Z$, we define their *Cartesian product* $(f, g) : X \rightarrow Y \times Z$ by

$$(f, g)(x) = (f(x), g(x)).$$

Lemma 3.4.8 Let $f : X \rightarrow \mathbb{R}$ and $g : X \rightarrow \mathbb{R}$ be functions and let $(f, g) : X \rightarrow \mathbb{R}^2$ be their Cartesian product. Suppose that $\mathbb{R}^2$ is given the Euclidean metric.

1. For $x_0 \in X$, (f, g) is continuous at x_0 if and only if both f and g are continuous at x_0 .
2. (f, g) is continuous if and only if both f and g are continuous.

Lemma 3.4.9 The addition function $+$: $\mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ given by $(x, y) \mapsto x + y$, the subtraction function $(x, y) \mapsto x - y$, the multiplication function $(x, y) \mapsto xy$, the maximum function $(x, y) \mapsto \max(x, y)$, and the minimum function $(x, y) \mapsto \min(x, y)$ are all continuous functions.

Proof. Here we only verify the continuity of the maximum function and the minimum function. Indeed, we have

$$\max(x, y) = \frac{x + y + |x - y|}{2}, \quad \min(x, y) = \frac{x + y - |x - y|}{2},$$

and since $x \mapsto |x|$ is continuous, both the maximum function and the minimum function are continuous. ■

Corollary 3.4.10 — algebraic operations preserve continuity. Let (X, d) be a metric space, and let $f : X \rightarrow \mathbb{R}$ and $g : X \rightarrow \mathbb{R}$ be functions. Let c be a real number.

1. If $x_0 \in X$ and f and g are continuous at x_0 , then the functions $f + g : X \rightarrow \mathbb{R}$, $f - g : X \rightarrow \mathbb{R}$, $fg : X \rightarrow \mathbb{R}$, $\max(f, g) : X \rightarrow \mathbb{R}$, $\min(f, g) : X \rightarrow \mathbb{R}$, and $cf : X \rightarrow \mathbb{R}$ are also continuous at x_0 . If $g(x) \neq 0$ for all $x \in X$, then $f/g : X \rightarrow \mathbb{R}$ is also continuous at x_0 .
2. If f and g are continuous, then the functions $f + g : X \rightarrow \mathbb{R}$, $f - g : X \rightarrow \mathbb{R}$, $fg : X \rightarrow \mathbb{R}$, $\max(f, g) : X \rightarrow \mathbb{R}$, $\min(f, g) : X \rightarrow \mathbb{R}$, and $cf : X \rightarrow \mathbb{R}$ are also continuous. If $g(x) \neq 0$ for all $x \in X$, then $f/g : X \rightarrow \mathbb{R}$ is also continuous.

■ **Example 3.56 — polynomials on $\mathbb{R}$ are continuous.** On $\mathbb{R}$ with the usual metric, $f(x) = x^2 + 2x + 3$ is continuous. More generally, using the corollary, we can show any polynomial function on $\mathbb{R}$ is continuous. ■

■ **Example 3.57 — continuity for functions on $\mathbb{R}^2$.** Now for functions $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ where we equip $\mathbb{R}^2$ with the Euclidean metric d_2 , how do we determine if they are continuous? For the simplest example, let's say we want to determine if the function is continuous at the origin $(0, 0)$. The method to use *polar coordinates* $x = r \cos(\theta)$, $y = r \sin(\theta)$. Notice that $(x, y) \rightarrow (0, 0)$ is equal to saying the Euclidean norm $\|(x, y)\| \rightarrow 0$ (think about why this is true; this is essentially the definition), and the Euclidean norm $\|(x, y)\|$ is exactly r . Thus taking $(x, y) \rightarrow 0$ is exactly the same as taking $r \rightarrow 0$, and we get no control on θ (how θ is moving is encoding the path we take to get to $(0, 0)$). ■

Exercise 3.14 Let $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ be a function.

1. If f is continuous, show that for each $x \in \mathbb{R}$, the function $y \mapsto f(x, y)$ is continuous on $\mathbb{R}$; for each $y \in \mathbb{R}$, the function $x \mapsto f(x, y)$ is continuous on $\mathbb{R}$.
2. Suppose $f(x, y) = \frac{x^2}{y}$ when $y \neq 0$ and $f(x, y) = 0$ when $y = 0$. Show that for each $(x, y) \in \mathbb{R}^2$, $\lim_{t \rightarrow 0} f(tx, ty) = 0$, but f is not continuous at $(0, 0)$. Thus *being continuous on every line through the origin is not enough to guarantee continuity at the origin!*
3. Suppose $f(x, y) = \frac{x^3}{x^2 + y^2}$ when $(x, y) \neq (0, 0)$ and $f(x, y) = 0$ when $(x, y) = (0, 0)$. Then is f continuous at $(0, 0)$? Prove or disprove. ■

For metric spaces (X, d_X) and (Y, d_Y) , we use $\mathcal{BC}(X, Y)$ to denote the subspace of $\mathcal{B}(X, Y)$ consisting of bounded continuous functions. This is also a metric space. Recall in Theorem 3.3.12, we showed that $\mathcal{B}(X, \mathbb{R})$ is a complete metric space. (In fact, the same proof implies that $\mathcal{B}(X, Y)$ is complete provided that Y is complete.)

Theorem 3.4.11 $\mathcal{BC}(X, \mathbb{R})$ is a complete metric space. In particular, the uniform limit of bounded continuous functions is also a bounded continuous function.

Proof. Given a Cauchy sequence $(f_n)_{n=1}^\infty$ in $\mathcal{BC}(X, \mathbb{R})$, this is also a Cauchy sequence in the complete metric space $\mathcal{B}(X, \mathbb{R})$, so it converges uniformly to $f \in \mathcal{B}(X, \mathbb{R})$. It suffices to show that $f \in \mathcal{BC}(X, \mathbb{R})$.

In other words, for every $x_0 \in X$, we would like to show that $\lim_{x \rightarrow x_0} f(x) = f(x_0)$. Indeed, this follows from the interchange of limits and uniform limits, namely,

$$\lim_{x \rightarrow x_0} f(x) = \lim_{x \rightarrow x_0} \lim_{n \rightarrow \infty} f_n(x) = \lim_{n \rightarrow \infty} \lim_{x \rightarrow x_0} f_n(x) = \lim_{n \rightarrow \infty} f_n(x_0) = f(x_0).$$

R More generally, $\mathcal{BC}(X, Y)$ is a complete metric space provided that Y is complete, and the proof for this is basically the same as above.

■ **Example 3.58 — exponential function is continuous.** Let $f_n(x) = \sum_{k=0}^n \frac{x^k}{k!}$ defined on $[-1, 1]$. This is continuous because it is a polynomial. Observe that $(f_n)_n$ is a Cauchy sequence in $\mathcal{BC}([-1, 1], \mathbb{R})$ (recall we showed in homework that for each x , $(f_n(x))_n$ is a Cauchy sequence, and we can do a similar proof here), and thus it has a limit

$$f(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!} \in \mathcal{BC}([-1, 1], \mathbb{R}).$$

As we've seen in homework, this is exactly the function e^x . Thus this example shows e^x is continuous. ■

There is an additional property of continuous functions which is important for future applications.

Definition 3.4.4 — uniform continuity. Let (X, d) and (X', d') be metric spaces, and let f be a continuous function from X to X' . We say that f is *uniformly continuous* if, given $\varepsilon > 0$, there exists $\delta > 0$ such that, for any pair $x, y \in X$, we have $d'(f(x), f(y)) < \varepsilon$ whenever $d(x, y) < \delta$.

So, f is uniformly continuous if it is continuous at every point and, for a given $\varepsilon > 0$, we can find a corresponding δ that is independent of the point. In general this is a stronger notion than being continuous. This gives a form of continuity that is global rather than local.

■ **Example 3.59** The function $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = x + 2$ is uniformly continuous on $\mathbb{R}$. ■

■ **Example 3.60** Consider the function $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = x^2$. Then f is not uniformly continuous on $\mathbb{R}$. However, if we only consider the function f on $[0, 1]$, then it is uniformly continuous on $[0, 1]$. ■

Exercise 3.15 1. Show that $f(x) = \frac{1}{x}$ is not uniformly continuous on $(0, 2)$.

2. Show that $f(x) = \sin(x)$ is uniformly continuous on $\mathbb{R}$.

Hint. You can use $\sin(x) - \sin(y) = 2 \sin\left(\frac{x-y}{2}\right) \cos\left(\frac{x+y}{2}\right)$, and recall $|\sin x| \leq |x|$.

3. Show that $f(x) = \frac{x}{x+2} + x^2$ is uniformly continuous on $[0, 1]$.

Hint. Of course you can show this directly using the definition, but you can also try to apply Theorem 3.5.11. ■

R Notice that $\cos(x) = \sin\left(\frac{\pi}{2} - x\right)$. So it is the composition of two continuous functions $\sin x$ and $\frac{\pi}{2} - x$, so it is continuous.

As a summary, we've learned

- How continuity interacts with topology.
- On $\mathbb{R}$: we've shown all polynomials, e^x , $\sin x$, and $\cos x$, and their sum, product, composition, etc. are continuous.
- On $\mathbb{R}^2$: we can change to polar coordinates to test continuity.
- The space $\mathcal{BC}(X, \mathbb{R})$
- A stronger notion: uniform continuity

3.4.4 When are two spaces the same? (Optional reading)

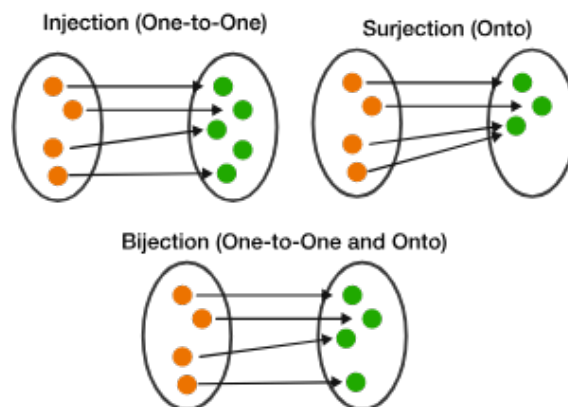
In math, we often like to talk about when two spaces are “the same”. Depending on the spaces and the structures on the spaces, there are many different notions.

Recall that if we consider sets X, Y with no additional structure, then there is the notion of a bijection, which is a one-to-one correspondence between the elements of the two sets.

Definition 3.4.5 — bijection. A function $f : X \rightarrow Y$ is a bijection if it is both injective (one-to-one) and surjective (onto). It means that for every element $y \in Y$, there is exactly one element $x \in X$ such that $f(x) = y$.

R Recall

- Being injective means $\forall x, x' \in X$, if $x \neq x'$, then $f(x) \neq f(x')$.
- Being surjective means $\forall y \in Y$, there exists $x \in X$ such that $y = f(x)$.



Definition 3.4.6 — homeomorphism. Let (X, d) and (X', d') be metric spaces. A function $f : X \rightarrow X'$ is a *homeomorphism* if:

- f is a bijection.
- f is continuous.
- f^{-1} is continuous.

We say (X, d) and (X', d') are *homeomorphic* if there exists a homeomorphism $f : X \rightarrow X'$.

R By Theorem 3.4.6, a homeomorphism maps the set of open sets of X to the set of open sets of X' .

The word “homeomorphism” is from Greek roots meaning “similar shape.” A homeomorphism is a mapping that preserves all the topological properties of a given space.

- **Example 3.61** 1. $f : \mathbb{R} \rightarrow \mathbb{R}$ given by $f(x) = 2x + 1$ is a homeomorphism.
- 2. $f : \mathbb{R} \rightarrow \mathbb{R}$ given by $f(x) = x^2$ is not a homeomorphism because it is not a bijection.
- 3. $f : \mathbb{R} \rightarrow \mathbb{R}$ given by $f(x) = x^3$ is a homeomorphism.

■ **Example 3.62** Consider the function $f : [0, 2\pi) \rightarrow S^1$ (the unit circle in $\mathbb{R}^2$) defined by $f(\varphi) = (\cos \varphi, \sin \varphi)$. This function is bijective and continuous, but not a homeomorphism. The function f^{-1} is not continuous at the point $(1, 0)$.

■ **Example 3.63** $\tan : (-\pi/2, \pi/2) \rightarrow \mathbb{R}$ is a homeomorphism with inverse $\tan^{-1} : \mathbb{R} \rightarrow (-\pi/2, \pi/2)$.

Exercise 3.16 Let $X = \mathbb{R}$ with the discrete metric, and $X' = \mathbb{R}$ with the usual metric. Show that $I : X \rightarrow X'$, $I(x) = x$ is a bijection but not a homeomorphism.

Theorem 3.4.12 For $1 \leq p < q \leq \infty$, the identity map $I(x) = x$ is a homeomorphism from $(\mathbb{R}^n, d_p)$ to $(\mathbb{R}^n, d_q)$ is a homeomorphism.

Proof. Apparently, I is a bijection.

Using a homework exercise, we have that a set A is open under the d_p metric iff it is open under the d_q metric.

Thus, let $A \subset \mathbb{R}^n$ be an open set in $(\mathbb{R}^n, d_q)$. then it is also open in $(\mathbb{R}^n, d_p)$. This shows f is continuous.

Conversely, let $A \subset \mathbb{R}^n$ be an open set in $(\mathbb{R}^n, d_p)$. then it is also open in $(\mathbb{R}^n, d_q)$. This shows f^{-1} is continuous.

In summary, I is a homeomorphism. ■

Definition 3.4.7 — isometry. Let (X, d) and (X', d') be metric spaces. A bijection $f : X \rightarrow X'$ is an *isometry* if $\forall x_1, x_2 \in X$,

$$d_X(x_1, x_2) = d_{X'}(f(x_1), f(x_2)).$$

We say (X, d) and (X', d') are *isometric* if there exists an isometry $f : X \rightarrow X'$.

Lemma 3.4.13 An isometry is a homeomorphism.

Proof. Exercise. ■

Exercise 3.17 Let $\mathbb{R}^2$ be equipped with the usual metric. Show that the following functions are isometries from $\mathbb{R}^2$ to itself.

1. Translation by a vector $(a, b) \in \mathbb{R}^2$: $T_{(a,b)}(x, y) = (x + a, y + b)$.
2. Counterclockwise rotation about the origin by an angle $\theta \in \mathbb{R}$: $R_\theta(x, y) = (x \cos \theta - y \sin \theta, x \sin \theta + y \cos \theta)$.
3. Reflection over a line through the origin making an angle θ with the x -axis: $S_\theta(x, y) = (x \cos 2\theta + y \sin 2\theta, x \sin 2\theta - y \cos 2\theta)$. ■

3.5 Compactness

Definition 3.5.1 — open covering. Let (X, d_X) be a metric space and let $A \subset X$. We say that a family of open subsets $\{U_j\}_{j \in I}$ is an *open covering* of A if A is contained in the union of the U_j , i.e.,

$$A \subset \bigcup_{j \in I} U_j.$$

A *finite subcovering* of $\{U_j\}_{j \in I}$ is a finite subcollection of open sets $\{U_{j_1}, \dots, U_{j_n}\}$ such that

$$A \subset \bigcup_{k=1}^n U_{j_k}.$$

Definition 3.5.2 — compact set. Let A be a subset of a metric space X . We say that A is *compact* if every open covering of A has a finite subcovering.

Recall that the Heine-Borel Theorem in $\mathbb{R}$ states that a subset of $\mathbb{R}$ with the usual metric is compact if and only if it is closed and bounded. The statement of the Heine-Borel theorem is certainly not true in a general metric space. For example, take $\mathbb{R}$ with the discrete metric. Then, the whole space $\mathbb{R}$ is closed and bounded in this metric. However, take an open covering consisting of the individual points in $\mathbb{R}$. This covering does not have a finite subcovering.

What we have in general is one direction of the implication:

Theorem 3.5.1 If a subset A of a metric space X is compact, then A is both bounded and closed.

Proof. [Similar to the situation in $\mathbb{R}$, reading for you]

If A is unbounded, then for $x_0 \in X$, the open covering

$$\{B(x_0, n) \mid n \in \mathbb{N}^+\},$$

has no finite subcovering. This contradicts the assumption that A is compact. Therefore, A must be bounded.

If A is not closed, let $x_0 \in X$ be an accumulation point of A that is not in A . For each $n \in \mathbb{N}^+$, let $D_n = \{x \in X \mid d_X(x, x_0) > \frac{1}{n}\}$. Then the open covering

$$\{D_n \mid n \in \mathbb{N}^+\}$$

has no finite subcovering. This contradicts the assumption that A is compact. Therefore, A must be closed. ■

Corollary 3.5.2 If A is a compact set in a metric space X , then for every infinite subset C of A , C has an accumulation point which lies in A .

Proof. [Similar to the situation in $\mathbb{R}$, reading for you]

Suppose that C is an infinite subset of A which does not have an accumulation point in A . Then for every $a \in A$, we can find an open ball $B(a, r_a)$ centered at a such that $B(a, r_a) \cap C \subset \{a\}$. In other words, the intersection is either empty or a singleton set.

We can then construct an open covering of A by taking the union of all these balls: $B(a, r_a) : a \in A$. Since A is compact, this covering has a finite subcovering, say $B(a_1, r_{a_1}), \dots, B(a_n, r_{a_n})$. Thus,

$$C \subset A \subset B_{a_1} \cup \dots \cup B_{a_n},$$

and each B_{a_i} contains at most one element of C . This implies that C has at most n elements, which gives a contradiction. ■

■ **Example 3.64** Consider the set $X = [0, 1] \cap \mathbb{Q}$, equipped with the inherited metric. Then as a metric space, it is closed and bounded, but it is not compact.

We can also verify this directly using the definition: pick any irrational $\xi \in (0, 1)$, and consider the open covering $\{U_n\}_{n=n_0}^\infty$ given by

$$U_n = \left[0, \xi - \frac{1}{n}\right) \cup \left(\xi + \frac{1}{n}, 1\right],$$

where n_0 is chosen large enough that $\xi - \frac{1}{n_0} > 0$ and $\xi + \frac{1}{n_0} < 1$.

This is the same idea as the proof showing compact sets must be closed. ■

Lemma 3.5.3 A closed subset of a compact set is compact.

Proof. Let A be a closed subset of a compact subset $B \subset X$. Let $\{U_j\}_{j \in I}$ be an open cover of A . Then $X \setminus A$ is open, and $\{U_i\}_{i \in I} \cup (X \setminus A)$ is an open cover of B . Since B is compact, we can extract a finite subcover $\{U_{j_1}, U_{j_2}, \dots, U_{j_n}, X \setminus A\}$. Since $A \cap (X \setminus A) = \emptyset$, we have that $\{U_{j_1}, U_{j_2}, \dots, U_{j_n}\}$ is a subcover of $\{U_j\}$. ■

Corollary 3.5.4 Arbitrary intersection of compact sets is compact.

Exercise 3.18 1. Show that a finite union of compact sets is compact.
2. Give an example of a countable union of compact sets that is not compact. ■

Recall in $\mathbb{R}$ with the Euclidean metric, we have *sequentially compact* $\iff$ *compact* $\iff$ *closed and bounded*. In a general metric space, we also have the notion of sequential compactness.

Definition 3.5.3 — sequentially compact set. Let A be a subset of a metric space X . We say that A is *sequentially compact* if every sequence in A has a subsequence that converges to an element of A .

Theorem 3.5.5 Let A be a subset of a metric space X . A is compact if and only if A is sequentially compact.

Proof. [Reading exercise for you; recall we proved this for $\mathbb{R}$ in homework]

Suppose that A is compact, and $(a_n)_{n=1}^\infty$ is a sequence in A . If $S = \{a_n \mid n \in \mathbb{N}^+\}$ is a finite set, then $(a_n)_{n=1}^\infty$ has a constant subsequence, and thus converges to an element of A . Otherwise, by Corollary 3.5.2, S has an accumulation point in A , and thus, $(a_n)_{n=1}^\infty$ has a subsequence that converges to this element. Therefore, A is sequentially compact.

Suppose that A is sequentially compact, and $\{U_i\}_{i \in I}$ is an open covering. For each $a \in A$, we define

$$r(a) := \sup\{r \in (0, \infty) \mid B(a, r) \subset U_i \text{ for some } i \in I\}.$$

Since U_i 's are open, $r(a)$ is always positive.

We claim that there exists $\delta > 0$ (this is called a *Lebesgue number*) such that $\forall a \in A, r(a) > \delta$. Otherwise, there exists a sequence $(a_n)_{n=1}^\infty$ in A with $\lim_{n \rightarrow \infty} r(a_n) = 0$. Since A is sequentially compact, it has a convergent subsequence, denoted by $(a_{n_j})_{j=1}^\infty$, with limit $a_\infty = \lim_{j \rightarrow \infty} a_{n_j} \in A$. Then there exists a $U_{i_0} \supset B(a_\infty, r(a_\infty)/2)$. Moreover, $\exists N \in \mathbb{N}^+, \forall j \geq N$,

$$d(a_{n_j}, a_\infty) < r(a_\infty)/4, \quad r(a_{n_j}) < r(a_\infty)/4.$$

However, $B(a_{n_j}, r(a_\infty)/4) \subset B(a_\infty, r(a_\infty)/2) \subset U_{i_0}$, which contradicts the definition of $r(a_{n_j})$.

We now construct the finite subcovering by the following recursive procedure. We let a_1 be any point in A and let $U_{i_1} \supset B(a_1, \delta)$. Suppose that for $l \in \mathbb{N}^+$, $(a_j)_{j=1}^l$ and $(U_{i_j})_{j=1}^l$ are chosen, but $A \not\subset \bigcup_{j=1}^l U_{i_j}$. Then we can let a_{l+1} be any point in $A \setminus \bigcup_{j=1}^l U_{i_j}$ and $U_{i_{l+1}} \supset B(a_{l+1}, \delta)$. We observe that for $j \neq k$, $d(a_j, a_k) > \delta$. Therefore, the procedure stops in finitely many steps, since otherwise we obtain a non-Cauchy sequence $(a_j)_{j=1}^\infty$ in A contradicting to the sequential compactness. In other words, there exists a finite subcollection $U_{i_1}, \dots, U_{i_n}$ such that

$$A \subset \bigcup_{j=1}^n U_{i_j}.$$

■

Corollary 3.5.6 A compact metric space (X, d) is complete.



Recall the examples of complete metric spaces we have so far:

- $\mathbb{R}$
- $\mathbb{Z}$
- Any discrete metric space (X, d_{disc})
- $(\mathbb{R}^n, d_p)$
- Any closed subset of a complete metric space
- Any compact metric space

- $\mathcal{B}(X, \mathbb{R})$
- $\mathcal{BC}(X, \mathbb{R})$

Using sequential compactness, we can also relate the Heine-Borel property to property about sequences:

Proposition 3.5.7 The following two statements are equivalent for a metric space (X, d) :

1. Any bounded sequence in (X, d) has a convergent subsequence.
2. (Heine-Borel property) Any closed and bounded set in (X, d) is compact.

Proof. Exercise. ■

As a corollary, we get

Theorem 3.5.8 — Heine-Borel for $\mathbb{R}^n$. A nonempty subset A of $\mathbb{R}^n$ with the Euclidean metric (or more generally, any d_p metric for $p \in [1, \infty]$) is compact if and only if it is closed and bounded.

Proof. This is a direct consequence of the previous proposition and Proposition 3.3.9. ■



As a summary,

1. For $\mathbb{R}^n$ with the Euclidean metric (or more generally, any d_p metric for $p \in [1, \infty]$), we have *sequentially compact* $\iff$ *compact* $\iff$ *closed and bounded*.
2. For general metric spaces, we only have *sequentially compact* $\iff$ *compact* $\implies$ *closed and bounded*.

3.5.1 Compactness and continuity

Recall that the notion of compactness is to have finite control over infinite data, so compactness is exactly the condition that lets local control become global control.

Recall that continuity is a local condition: near each point, things behave nicely. Compactness precisely lets you upgrade this to global statements and we get uniform continuity.

And we can see, being closed and bounded is not enough: For example, in a discrete metric space, every set is closed and bounded, and any function is continuous, so I get no control on continuous functions.

Theorem 3.5.9 — continuity preserves compactness. Let $f : X \rightarrow X'$ be a continuous function of metric spaces. If $A \subset X$ is compact, then $f(A) \subset X'$ is compact.

Proof. Exercise. ■

Proposition 3.5.10 Suppose that X is a compact metric space. A continuous function $f : X \rightarrow \mathbb{R}$ attains a maximum and minimum value on X . (i.e., there exist points $x_{\max}, x_{\min} \in X$ such that $f(x_{\max}) \geq f(x)$ and $f(x_{\min}) \leq f(x)$ for every $x \in X$.)

Proof. Exercise. ■

- Exercise 3.19**
1. Let $f : X \rightarrow X'$ be a continuous map of metric spaces. Show that if $A \subset X$ is compact, then $f(A) \subset X'$ is compact.
 2. Suppose that X is a compact metric space. Using (1), deduce that a continuous function $f : X \rightarrow \mathbb{R}$ is bounded.
 3. Suppose that X is a compact metric space. Show that a continuous function $f : X \rightarrow \mathbb{R}$

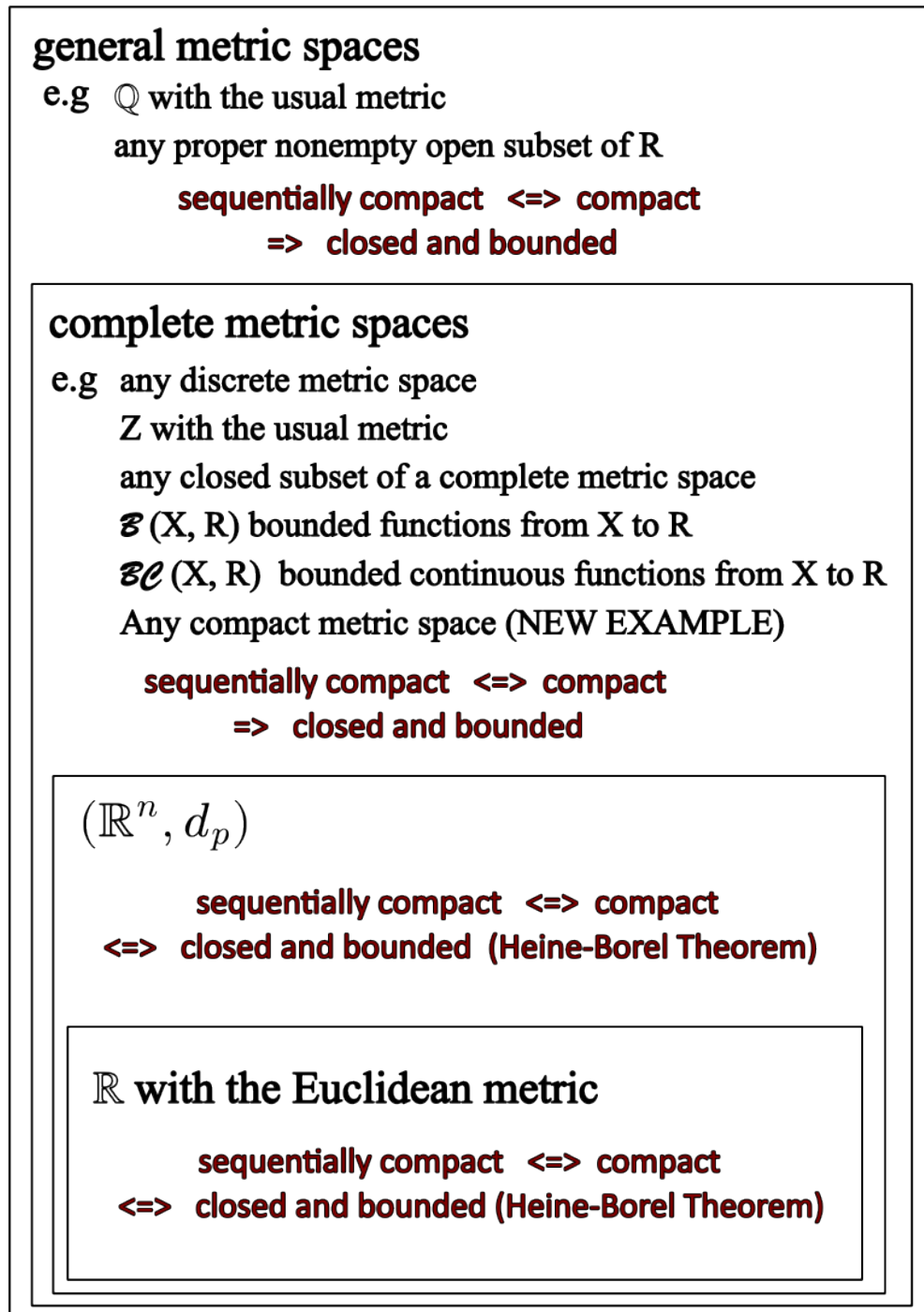


Figure 3.8: Compactness in metric spaces

attains a maximum and minimum value on X .

Recall the definition of uniform continuity. In general this is a stronger notion than continuity, but on compact domains, they are equivalent. This is a very useful way to show some function is uniformly continuous.

Theorem 3.5.11 — continuity implies uniform continuity on compact sets. Suppose that X and X' are metric spaces with X compact, and $f : X \rightarrow X'$ is continuous. Then f is uniformly continuous on X .

Proof. For any fixed $\varepsilon > 0$, since f is continuous, for each $x \in X$, there exists an $r(x) > 0$ such that for all $y \in B(x, r(x))$, $d_{X'}(f(x), f(y)) < \varepsilon/2$. This gives an open covering $\{B(x, r(x)/2)\}_{x \in X}$ of X , and by the compactness of X , it has a finite subcovering $\{B(x_i, r(x_i)/2)\}_{i=1}^n$.

Let $\delta = \min_i r(x_i)/2$. $\forall x, y \in X$ with $d_X(x, y) < \delta$, choose an i_0 such that $x \in B(x_{i_0}, r(x_{i_0})/2)$. Then both x and y lie in $B(x_{i_0}, r(x_{i_0}))$, so

$$d_{X'}(f(x), f(y)) \leq d_{X'}(f(x), f(x_{i_0})) + d_{X'}(f(y), f(x_{i_0})) < \varepsilon.$$

Therefore, f is uniformly continuous on X . ■

Lastly, here is a theorem that relates to Section 3.4.4 if you've read that section:

Theorem 3.5.12 Suppose that X and X' are metric spaces with X compact, and $f : X \rightarrow X'$ is a continuous bijection. Then f^{-1} is also continuous, so f is a homeomorphism.

Proof. It suffices to show that f^{-1} is continuous. In other words, we need to prove that for any closed set $A \subset X$, $f(A)$ is closed.

Since X is compact, by Lemma 3.5.3, for a closed $A \subset X$, A is also compact. By Theorem 3.5.9, $f(A)$ is a compact subset of X' and thus, a closed set. ■

R Recall Example 3.62, where we have the function $f : [0, 2\pi) \rightarrow S^1$ which is a continuous bijection but is not a homeomorphism. The problem is that $X = [0, 2\pi)$ is not compact.

3.6 Connectedness

Intuitively, we think about connected space if the space comes in one piece. How do we say it is only one piece? Well; that means you can't split it into two separate chunks without tearing. To make it rigorous, we use open sets to encode chunks.

Definition 3.6.1 — connected spaces. Let (X, d) be a metric space. We say that X is *disconnected* (or *not connected*) if there exist disjoint (i.e. $U \cap V = \emptyset$) non-empty open sets $U, V \subset X$ such that $X = U \cup V$. We say that X is *connected* if no such U and V exist.

■ **Example 3.65** $X = [1, 2] \cup [3, 4]$ is disconnected. ■

Definition 3.6.2 — connected sets. Let (X, d) be a metric space and $Y \subset X$. We say that Y is *disconnected* if (Y, d) is disconnected. We say that Y is *connected* if (Y, d) is connected.

(Equivalently, we say that Y is *disconnected* if there exist disjoint (i.e. $U \cap V = \emptyset$) open sets $U, V \subset X$ such that

- $Y \cap U \neq \emptyset$ and $Y \cap V \neq \emptyset$,
- $Y \subset U \cup V$.

We say Y is *connected* if no such U and V exist.)

If a metric space X is not connected, then it can be decomposed into subsets called connected components.

Definition 3.6.3 — connected component. If X is a metric space and x_0 is in X , then the *connected component* of x_0 in X is the union of the connected sets that contain x_0 .

Exercise 3.20 1. Let (X, d) be a metric space, and let $(E_\alpha)_{\alpha \in I}$ be a collection of connected sets in X with I non-empty. Suppose also that $\bigcap_{\alpha \in I} E_\alpha$ is non-empty. Show that $\bigcup_{\alpha \in I} E_\alpha$ is connected.

In particular, this shows for $x_0 \in X$, the connected component of x_0 is a connected set in X .

2. Show that if A is a connected subset of a metric space, then $\bar{A}$ is connected. Deduce that connected components are closed.

On the real line, connected sets are easy to describe.

Theorem 3.6.1 — connected sets in $\mathbb{R}$. Let X be a non-empty subset of $\mathbb{R}$. The following are equivalent:

- (1) X is connected;
- (2) Whenever $x, y \in X$ and $x < y$, the interval $[x, y] \subset X$;
- (3) X is an interval.

R Here by interval we mean any of the following: $[a, b]$, (a, b) , $[a, b)$, $(a, b]$, $(-\infty, b]$, $(-\infty, b)$, $[a, \infty)$, (a, ∞) , $\mathbb{R} = (-\infty, \infty)$.

Proof of (1) $\iff$ (2). [Reading for you]

(1) $\implies$ (2): Suppose, for the sake of contradiction, that there exist $x < z < y$, $x, y \in X$ but $z \notin X$. Then we can define $U = (-\infty, z) \cap X$ and $V = (z, \infty) \cap X$, which are disjoint non-empty open sets in X and $X = U \cup V$. This contradicts to the connectedness of X .

(2) $\implies$ (1): Suppose, for the sake of contradiction, that there exist disjoint non-empty open sets U and V in X such that $X = U \cup V$. Choose $u \in U$ and $v \in V$, and without loss of generality, we assume that $u < v$ and thus, $[u, v] \subset X$. Let $z = \sup(U \cap [u, v])$. It is clear that $z \in [u, v] \subset X$, so either $z \in U$ or $z \in V$. If $z \in U$, then $z \neq v \in V$, and there exists $r > 0$ such that $[z, z + r) \subset [u, v] \cap U$. However, by the definition of z , we know that $(z, z + r) \not\subset U$, which gives a contradiction. Otherwise, if $z \in V$, then $z \neq u \in U$, and there exists $r > 0$ such that $(z - r, z] \subset [u, v] \cap V$. However, again by the definition of z , $(z - r, z] \cap U \neq \emptyset$, which contradicts the assumption that $U \cap V = \emptyset$. ■

Exercise 3.21 Show that (2) $\iff$ (3) in the previous theorem. ■

Theorem 3.6.2 — Continuity preserves connectedness. Let $f : X \rightarrow Y$ be a continuous map from one metric space (X, d_X) to another (Y, d_Y) . Let E be any connected subset of X . Then $f(E)$ is also connected.

Proof. We can restrict f to E , and $f : E \rightarrow f(E)$ is also a continuous function. Suppose that $f(E)$ is disconnected, then there exist two non-empty disjoint open subsets U, V in $f(E)$ such that $f(E) = U \cup V$. Because U, V are nonempty in $f(E)$, we have $f^{-1}(U)$ and $f^{-1}(V)$ are

nonempty. Because $f(E) = U \cup V$, we have $E = f^{-1}(U) \cup f^{-1}(V)$. Because $U \cap V = \emptyset$, we have $f^{-1}(U) \cap f^{-1}(V) = \emptyset$. Since f is continuous, both $f^{-1}(U)$ and $f^{-1}(V)$ are open, which gives a contradiction. ■

The focus of this section is the following theorem. It shows existence of solutions without solving explicitly.

The intermediate value theorem is also the foundation of numerical methods: The bisection method is basically IVT made algorithmic. If signs differ at endpoints, repeatedly halve the interval. IVT guarantees the root never disappears.

Corollary 3.6.3 — Intermediate value theorem. Let $f : X \rightarrow \mathbb{R}$ be a continuous function. Let E be any connected subset of X , and let a, b be any two elements of E . Let y be a real number between $f(a)$ and $f(b)$, i.e., either $f(a) \leq y \leq f(b)$ or $f(a) \geq y \geq f(b)$. Then there exists $c \in E$ such that $f(c) = y$.

■ **Example 3.66 — $\log x$ is continuous.** Recall that we've shown in homework that the function $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = e^x$ is injective (actually strictly increasing), and we've verified that e^x is continuous.

What is the image $f(\mathbb{R})$? For any x , $e^x > 0$, so $f(\mathbb{R}) \subset (0, \infty)$. Since $\lim_{x \rightarrow -\infty} e^x = 0$ and $\lim_{x \rightarrow \infty} e^x = \infty$, using the intermediate value theorem, we can show that $f(\mathbb{R}) = (0, \infty]$ (think about why this is true).

Thus, $f : \mathbb{R} \rightarrow (0, \infty)$ is a bijection, and we can consider its inverse function f^{-1} , which is the natural logarithmic function $\log : (0, \infty) \rightarrow \mathbb{R}$.

We can show $\log$ is continuous: Consider any closed interval $[a, b] \subset (0, \infty)$. Then $\log([a, b]) = f^{-1}([a, b])$ is closed because $f = e^x$ is continuous, and it is bounded which we can verify using e^x is a bijection and strictly increasing (think about why this is true).

Thus $\log([a, b])$ is a compact set, and then $f : e^x : \log([a, b]) \rightarrow [a, b]$ is a continuous bijection on a compact set, so the inverse function $f^{-1} = \log$ is continuous by Theorem 3.5.12.

Since $\log$ is compact on any closed interval on $(0, \infty]$, it is continuous everywhere. ■

■ **Example 3.67 — x^p is continuous on $(0, \infty)$.** Now can show for any $p \in \mathbb{R}$, the function $f(x) = x^p$ is continuous on $(0, \infty)$. This is because $x^p = \exp(p \log x)$ and we've verified both $\exp$ and $\log$ are continuous and x^p is just their composition. ■

Exercise 3.22 1. Let (X, d_{disc}) be a metric space with the discrete metric. Let E be a subset of X which contains at least two elements. Show that E is disconnected.
2. Let $f : X \rightarrow Y$ be a function from a connected metric space (X, d) to a metric space (Y, d_{disc}) with the discrete metric. Show that f is continuous if and only if it is constant. ■

Exercise 3.23 Take $\mathbb{R}$ with the usual metric, and let $f : \mathbb{R} \rightarrow \mathbb{R}$ be given by $f(x) = x^n$ for $n \in \mathbb{N}^+$. If b is a positive real number, show that there exists a unique positive real number a such that $a^n = b$ via the intermediate value theorem.

(Recall that earlier we find such a and defined it as the n -th root using supremum.) ■

A related notion is path-connectedness, which means you can move from a point to another along a continuous path inside the space.

Definition 3.6.4 — path-connected set. Let (X, d) be a metric space, and let E be a subset of X . We say that E is *path-connected* if, for every $x, y \in E$, there exists a continuous function

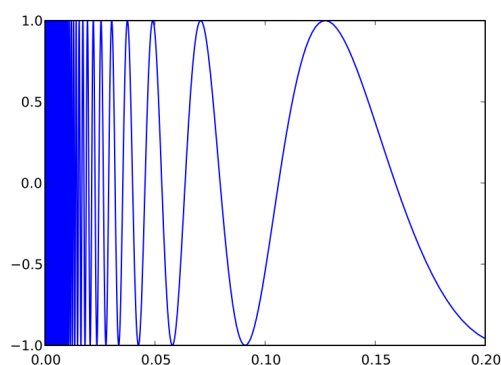


Figure 3.9: the topologist's sine curve

■ $\gamma : [0, 1] \rightarrow E$ from the unit interval $[0, 1]$ to E such that $\gamma(0) = x$ and $\gamma(1) = y$.

Theorem 3.6.4 Every non-empty path-connected set is connected.

Proof. Suppose for the sake of contradiction that there exist two non-empty disjoint open subsets $U, V \subset E$ such that $E = U \cup V$. Let $u \in U$ and $v \in V$. By the path-connectedness, there exists a continuous function $\gamma : [0, 1] \rightarrow E$ such that $\gamma(0) = u$ and $\gamma(1) = v$. Let $U' = \gamma^{-1}(U)$ and $V' = \gamma^{-1}(V)$ which are non-empty disjoint open subsets of $[0, 1]$ and $[0, 1] = U' \cup V'$. This contradicts the connectedness of $[0, 1]$. ■

■ **Example 3.68 — the topologist's sine curve.** Let $E = \{(0, y) \mid 0 \leq y \leq 1\} \cup \{(x, \sin(1/x)) \mid x > 0\}$. This is called *the topologist's sine curve*. E is connected but not path-connected.

The proof is a bit tricky, so it won't be detailed here. ■

Why do we want this notion of path-connectedness? Because we want an easy way to tell a set (e.g. $\mathbb{R}^n$ or a subset of $\mathbb{R}^n$) is connected.

■ **Example 3.69** The space $X = \mathbb{R}^2 \setminus \{(0, 0)\}$ is path-connected, hence connected. Given $x, y \in X$, if the straight line $\overline{xy}$ misses $(0, 0)$ then it gives a path in X connecting x and y . Otherwise, first follow a semicircle centered at $(0, 0)$ passing through x , then follow a straight path to y . Therefore any two points are connected by a path. Similarly, we can show that for any point $p \in \mathbb{R}^2$, the set $\mathbb{R}^2 \setminus \{p\}$ is path-connected.

By contrast, for any $x \in \mathbb{R}$, $\mathbb{R} \setminus \{x\} = (-\infty, x) \cup (x, \infty)$ is a disjoint union of nonempty open sets, so is disconnected. ■

Using this notion of path-connectedness, we can show $\mathbb{R}$ and $\mathbb{R}^2$ look “different”.

Proposition 3.6.5 The spaces $\mathbb{R}$ and $\mathbb{R}^2$ with the usual Euclidean metric are not homeomorphic. In other words, there is no continuous bijection $F : \mathbb{R} \rightarrow \mathbb{R}^2$ such that F^{-1} is also continuous.

Proof. Suppose for the sake of contradiction that there is a homeomorphism $F : \mathbb{R} \rightarrow \mathbb{R}^2$.

Let $x = F^{-1}((0, 0))$. Then since F^{-1} is continuous and $\mathbb{R}^2 \setminus \{(0, 0)\}$ is connected, by Theorem 3.6.2, we have $F^{-1}(\mathbb{R}^2 \setminus \{(0, 0)\})$ is also connected. However, $F^{-1}(\mathbb{R}^2 \setminus \{(0, 0)\}) = \mathbb{R} \setminus x$ is not connected, contradiction. ■

3.7 The contraction mapping theorem (Optional reading)

In this section, we are going to discuss the contraction mapping theorem, which is useful and powerful in the inverse function theorem, solutions to ordinary differential equations, etc.

■ **Example 3.70** If a traveler places a map of the United States on the ground, the contraction mapping theorem tells us that there is precisely one point on the map, which corresponds directly to the point below it. ■

■ **Example 3.71** In reinforcement learning, a computer tries to learn the best long-term strategy by repeatedly updating its estimate of how good each decision is. The update rule is designed so that each new estimate is a little closer to the true answer than the previous one. Mathematically, this “getting closer each time” is exactly what a contraction mapping guarantees. Because of this property, we know the learning process will settle down to a unique correct solution instead of oscillating or diverging. ■

■ **Example 3.72** Similarly in economics, think of an economist trying to compute the best long-term savings strategy. They start with a guess, improve it using a formula, and repeat. If each improvement shrinks the error by a fixed percentage, then the guesses must settle down to a single best strategy. The contraction mapping theorem is what ensures that this convergence happens. ■

Definition 3.7.1 — contraction mapping. Let (X, d) be a metric space, and let $f : X \rightarrow X$ be a map. We say that f is a *contraction mapping* if there exists a constant $0 < c < 1$ such that $d(f(x), f(y)) \leq cd(x, y)$ for all $x, y \in X$; we call c the *contraction constant* of f .

Lemma 3.7.1 A contraction mapping is continuous.

Proof. This basically follows from the definition of continuity. ■

Definition 3.7.2 — fixed point. Let X be a metric space and f a map from X to X . A point $x_0 \in X$ is a *fixed point* of f if $f(x_0) = x_0$.

Exercise 3.24 1. Find a continuous function $f : \mathbb{R} \rightarrow \mathbb{R}$ that does not have a fixed point.
2. Find a continuous function $f : (0, 1) \rightarrow (0, 1)$ that does not have a fixed point.
3. Let $f : [0, 1] \rightarrow [0, 1]$ be continuous. Show that f has a fixed point.

Hint. For (3), you can use the intermediate value theorem. Notice that in general, the intermediate value theorem only tells you when there is a solution; it doesn't tell you when there isn't a solution. ■

Theorem 3.7.2 — Contraction mapping theorem. Let (X, d) be a nonempty complete metric space, and let $f : X \rightarrow X$ be a contraction mapping. Then f has a unique fixed point.

Proof. Suppose that the contraction constant of f is $c \in (0, 1)$.

Let x_1 be an element of X . Define $x_2 = f(x_1)$, $x_3 = f(x_2) = f(f(x_1)) = f^{\circ 2}(x_1)$, and in general, $x_{n+1} = f(x_n) = f^{\circ(n-1)}(x_1)$ for all $n \in \mathbb{N}^+$;

Then, if $n > m$, we have

$$\begin{aligned} d(x_n, x_m) &= d(f^{\circ(m-1)}(x_1), f^{\circ(n-1)}(x_1)) \\ &< c^{m-1} d(x_1, f^{\circ(n-m)}(x_1)) \\ &\leq c^{m-1} (d(x_1, x_2) + d(x_2, x_3) + \cdots + d(x_{n-m}, x_{n-m+1})) \\ &\leq c^{m-1} d(x_1, x_2) (1 + c + c^2 + \cdots) \\ &\leq \frac{c^{m-1}}{1-c} d(x_1, x_2). \end{aligned}$$

It follows that $(x_n)_{n=1}^\infty$ is a Cauchy sequence. Since X is complete, it converges to some $x_0 = \lim_{n \rightarrow \infty} x_n$. By the continuity of f , we have

$$f(x_0) = f(\lim_{n \rightarrow \infty} x_n) = \lim_{n \rightarrow \infty} f(x_n) = \lim_{n \rightarrow \infty} x_{n+1} = x_0.$$

If there exists another fixed point x'_0 of f , then we have

$$d(x_0, x'_0) = d(f(x_0), f(x'_0)) < cd(x_0, x'_0),$$

which implies that $d(x_0, x'_0) = 0$, i.e., $x_0 = x'_0$. ■

R Recall that we essentially did the key step of the proof in Practice Midterm Problem 1. So secretly you've already done most part of the proof.

Exercise 3.25 We say that a function $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a *Lipschitz function* if there exists a constant C such that $\|f(x) - f(y)\| \leq C\|x - y\|$ for all $x, y \in \mathbb{R}^n$. Notice that from definition, f satisfies the Lipschitz condition with constant C for $0 < C < 1$ if and only if f is a contraction mapping.

1. Show that if f is Lipschitz, then f is uniformly continuous.
2. Find an example of a continuous function $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ that does not satisfy any Lipschitz condition.

In Bolter and Sally, they prove Picard's Theorem as an application, which is about the existence and uniqueness of solutions to first-order ordinary differential equations. Since we do not assume familiarity of differential equations in this course, we will not discuss it here.

3.8 The completion of a metric space (Optional reading)

Obviously, complete metric spaces play a special role among all metric spaces, and we have seen that not all metric spaces are complete. Throughout this section, we are going to show that by adding extra elements, we are able to make any metric space complete. This is similar to how we constructed $\mathbb{R}$ from $\mathbb{Q}$.

If you want to read this section, you might want to first read Section 3.4.4 to know what an isometry is.

Theorem 3.8.1 Let (X, d) be a metric space. Then there exists a complete metric space $(\tilde{X}, \tilde{d})$ and an injection $\phi : X \rightarrow \tilde{X}$, such that

1. $\phi : X \rightarrow \phi(X)$ is an isometry, and
2. $\phi(X)$ is dense in $\tilde{X}$.

Let X' be the set of all Cauchy sequences in X . We define an equivalence relation on X' by saying that $(x_n)_{n=1}^\infty$ is equivalent to $(y_n)_{n=1}^\infty$ if $\lim_{n \rightarrow \infty} d(x_n, y_n) = 0$.

Lemma 3.8.2 This is an equivalence relation.

Let $\tilde{X}$ be the set of equivalence classes, where the equivalence class of $(x_n)_{n=1}^\infty$ will be denoted by $[x_n]$. We can define a map $\tilde{d} : \tilde{X} \times \tilde{X} \rightarrow \mathbb{R}$ by

$$\tilde{d}([x_n], [y_n]) = \lim_{n \rightarrow \infty} d(x_n, y_n).$$

Lemma 3.8.3 $\tilde{d}$ is well-defined.

We define $\phi : X \rightarrow \tilde{X}$ by $\phi(x) = [x_n]$ where $\forall n \in \mathbb{N}^+, x_n = x$. It is clear that ϕ is an isometry from X to $\phi(X)$.

It suffices to show

1. $\phi(X)$ is dense in $\tilde{X}$.
2. $\tilde{X}$ is complete.

Let $\tilde{x} = [x_n] \in \tilde{X}$, and $(x_n)_{n=1}^\infty$ is a representative Cauchy sequence in X . For any fixed $\varepsilon > 0$. Since $(x_n)_{n=1}^\infty$ is Cauchy, there exists $N \in \mathbb{N}^+$, such that for all $n, m \geq N$, $d(x_n, x_m) < \varepsilon$. Hence, for all $n \geq N$,

$$\tilde{d}(\tilde{x}, x_n) = \lim_{m \rightarrow \infty} d(x_m, x_n) \leq \varepsilon.$$

This implies that $(\phi(x_n))_{n=1}^\infty$ converges to $\tilde{x}$ in the $\tilde{d}$ metric. Hence, $\phi(X)$ is dense in $\tilde{X}$.

To see that $\tilde{X}$ is complete, take a Cauchy sequence $(\tilde{y}_n)_{n=1}^\infty$ in $\tilde{X}$. For each $n \in \mathbb{N}$, by density, choose $\tilde{z}_n = \phi(z_n) \in \phi(X)$ such that $\tilde{d}(\tilde{y}_n, \tilde{z}_n) < \frac{1}{n}$. Then

$$\tilde{d}(\tilde{z}_n, \tilde{z}_m) \leq \tilde{d}(\tilde{z}_n, \tilde{y}_n) + \tilde{d}(\tilde{y}_n, \tilde{y}_m) + \tilde{d}(\tilde{y}_m, \tilde{z}_m) \leq \frac{1}{n} + \tilde{d}(\tilde{y}_n, \tilde{y}_m) + \frac{1}{m},$$

and thus, $(\tilde{z}_n)_{n=1}^\infty$ is a Cauchy sequence in $\tilde{X}$. Since ϕ is an isometry, $(z_n)_{n=1}^\infty$ is Cauchy in X , so we can let $\tilde{z} = [z_n] \in \tilde{X}$. Then,

$$\tilde{d}(\tilde{y}_n, \tilde{z}) \leq \tilde{d}(\tilde{y}_n, \tilde{z}_n) + \tilde{d}(\tilde{z}_n, \tilde{z}) < \frac{1}{n} + \tilde{d}(\tilde{z}_n, \tilde{z}).$$

Observe that $\lim_{n \rightarrow \infty} \tilde{d}(\tilde{z}_n, \tilde{z}) = 0$, so $\lim_{n \rightarrow \infty} \tilde{y}_n = \tilde{z} \in \tilde{X}$.

Definition 3.8.1 — completion. The metric space $\tilde{X}$ in the above theorem is called the *completion* of X .

Lemma 3.8.4 If (X', d') and $(\tilde{X}, \tilde{d})$ are completions of X , then (X', d') and $(\tilde{X}, \tilde{d})$ are isometric. Therefore, the completion is unique up to isometry.

3.9 Review

1. Convergence and limit in $\mathbb{R}$

You should understand the situation in real numbers, because this is the basis for what we do in general metric spaces. For example, if you don't understand the supremum, then it's also hard for you to understand the sup metric.

(1) construction and property of real numbers

- Here the goal is not to test you on any specific construction. Rather the goal is to understand the key concept of Cauchy sequence and how to pass properties between a Cauchy sequence of rational numbers and the real number it represents (i.e. its limit), and know the basic properties of $\mathbb{R}$ (e.g. inf/sup).

(2) sequences in $\mathbb{R}$

- Here the goal is to understand the definition and the relationship between three important concepts: convergent, Cauchy, and bounded (e.g. Cauchy criterion, bounded monotone implies convergent, etc.)
- Also you should be able to do some limit computations using the theorems, limit laws and squeeze theorem, etc.

(3) topology of $\mathbb{R}$

- Here you should understand the key topological concepts (accumulation points, open, closed, compact, sequentially compact), know how to use their definitions,

and how they relate to each other (e.g. arbitrary union of open sets are open, characterization of closed sets using accumulation points, Heine-Borel theorem, etc.).

2. Metric spaces

(1) definition of metric space and examples

- definition of metric space: know how to show some space is a metric space: either verify the three properties directly, or use a norm on a vector space to get a metric space. Should also know how to relate the properties between metric spaces.
- examples: $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{Z}$, d_p metric on $\mathbb{R}^n$, discrete metric, subspace

(2) topology

- open sets, closed sets, interior, exterior, boundary, accumulation points, isolated points, adherent points, closure, and how these concepts are related
- compact vs sequentially compact vs closed and bounded
- definition of a connected set

(3) sequences

- the relationship between convergent, Cauchy, and bounded: the general situation vs in a complete metric space vs in $(\mathbb{R}^n, d_p)$ vs in $(\mathbb{R}, d_2)$

(4) functions

- the metric space $\mathcal{B}(X, \mathbb{R})$, $\mathcal{BC}(X, \mathbb{R})$ with sup-metric
- for a sequence of functions: pointwise convergence vs uniform convergence
- limiting values of functions $\lim_{x \rightarrow x_0} f(x)$; limit computations using limit laws; interchange double limits
- continuity, uniform continuity; how continuity interacts with openness/closedness; how continuity is preserved under various operations
- how to check if a function on $\mathbb{R}$ or $\mathbb{R}^2$ is continuous
- how continuity interacts with compactness: continuity implies uniform continuity on compact sets, etc.
- how continuity interacts with connectedness: the intermediate value theorem