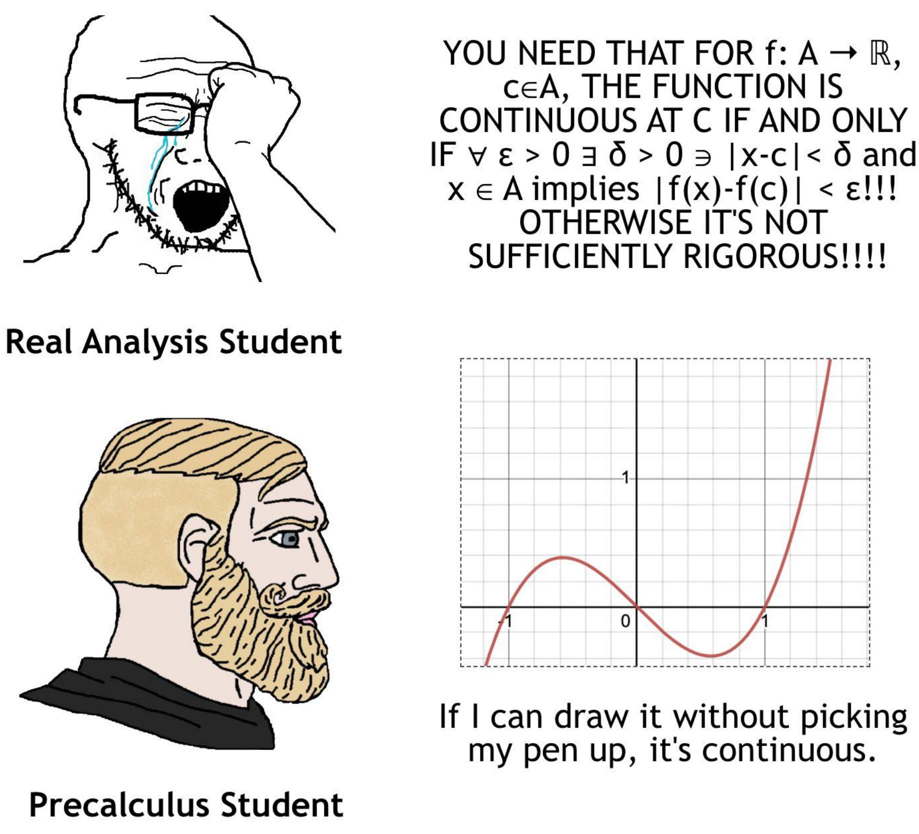
Fantuan's Academia

Fantuan's Math Notes series

Notes on Mathematical Analysis

Author: Jingxuan Xu



June 19, 2024

Contents

All the Sections with * are hard sections and can be skipped without losing coherence.

This note is referenced on **Understanding Analysis** by Stephen Abbott [1], **Principles of Mathematical Analysis** by Walter Rudin [2], **Analysis I** by Terence Tao [3], and MTH 117,118 notes of XJTLU.

Part I

Part I: The Real Line

Chapter 1

Real Numbers

1.1 Why Analysis?

Analysis, simply saying, is a course about ‘rigorous calculus’. Somebody may ask then: “why we need another course about calculus?”. Indeed, basic calculus concepts and various computing skills are introduced in Year I Calculus course. However, regarding calculus as a pure math object, it should maintain its full rigor. If we apply calculus in the real world problems without knowing where they came from and what is their constraints to be correctly applied, some pathological things will happen, as listed below.

Example 1.1.1. *Infinite Series*
Consider the divergent infinite series

$$S = 1 + 2 + 4 + 8 + 16 + \cdots \tag{1.1}$$

If we multiply it by 2,

$$2S = 2 + 4 + 8 + 16 + \cdots \tag{1.2}$$

Subtract (1.1) from (1.2), we will have the ridiculous result

$$S = -1$$

Example 1.1.2. *Interchanging Integrals*
We always change the order of double integral to make calculation easier. But, can we always do that in any cases? Consider

$$\int_0^\infty \int_0^1 \left(e^{-xy} - xye^{-xy}\right) \mathrm{d}y \, \mathrm{d}x$$

If we directly compute this, we can get,

$$\int_0^\infty \int_0^1 \left(e^{-xy} - xye^{-xy}\right) \mathrm{d}y \, \mathrm{d}x = \int_0^\infty \left[ye^{-xy}\right]_{y=0}^1 \mathrm{d}x = \int_0^\infty e^{-x} \, \mathrm{d}x = \left[-e^{-x}\right]_0^\infty = 1$$

However, if we change the order of integral

$$\int_0^1 \int_0^\infty \left(e^{-xy} - xye^{-xy}\right) \,\mathrm{d}x\,\mathrm{d}y = \int_0^1 \left[xe^{-xy}\right]_{x=0}^{\infty} \mathrm{d}y = \int_0^1 (0-0)\,\mathrm{d}y = 0$$

We arrive different answers!

Example 1.1.3. *Reordering Infinite Series*
Consider the alternating harmonic series

$$S = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots$$

We know that this infinite series converges at some point. Therefore, nothing similar as Example 1.1.1 could happen here. However, if we do the following computation:

$$\begin{aligned} \frac{1}{2}S &= \quad\ \frac{1}{2} \qquad\quad - \frac{1}{4} \qquad\quad + \frac{1}{6} \qquad\quad - \frac{1}{8} \qquad\quad + \frac{1}{10} \qquad\quad - \frac{1}{12} \qquad\quad + \frac{1}{14} \qquad\quad - \frac{1}{16} + \cdots \\ S &= 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} - \frac{1}{8} + \frac{1}{9} - \frac{1}{10} + \frac{1}{11} - \frac{1}{12} + \frac{1}{13} - \frac{1}{14} + \frac{1}{15} - \frac{1}{16} + \cdots \end{aligned}$$

$$\frac{3}{2}S = \left(1 + \frac{1}{3}\right) - \frac{1}{2} + \left(\frac{1}{5} + \frac{1}{7}\right) - \frac{1}{4} + \left(\frac{1}{9} + \frac{1}{11}\right) - \frac{1}{6} + \left(\frac{1}{13} + \frac{1}{15}\right) - \frac{1}{8} + \cdots$$

We see that $\frac{3}{2}S$ is just a reordering of our initial infinite series (with two positive terms following one negative term)! Therefore, we just change the convergent point by simply reordering the infinite series.
This doesn't make sense! Since by intuition, reordering the terms in an algorithm will not change its result. However, you see here, the situation changes in the infinite case.

As showed above, we indeed need this course to make analysis as a much more rigorous math topic than Year I Calculus. To get started, we will first talk about real numbers.

1.2 From Rational to Irrational Numbers

The simplest number system we can call to our mind is the **Natural Numbers**

$$\mathbb{N} = \{0, 1, 2, 3, 4, \cdots\}$$

Obviously, this number system is based on counting. It is enough for the simple use of counting things. However, this number system is not closed under subtraction (i.e., one natural number subtracts another natural number may not result in a natural number). Therefore, we introduce the **Integer Numbers**

$$\mathbb{Z} = \{\cdots, -3, -2, -1, 0, 1, 2, 3, \cdots\}$$

This number system is still not complete since it is not closed under division. For example, $3 \div 5$ is not in the list. We further introduce the **Rational Numbers**

$$\mathbb{Q} = \left\{ \frac{p}{q} : q \neq 0, p, q \in \mathbb{Z} \right\}$$

Back to Pythagoras's era (500-400 BC), he only believes the existance of rational numbers, and so did his followers in Pythagoreanism, except for one: Hippasus of Metapontum. After Pythagoras announced his famous Pythagorean Theorem, Hippasus directly used this theorem to discover $\sqrt{2}$: an irrational number!

Consider a right-angled triangle with two right-angled edges of length 1. Then by Pythagorean Theorem, length of the hypotenuse z should satisfy

$$z^2 = 1^2 + 1^2 = 2$$

We denote this number as $\sqrt{2}$, for which the square of it is 2. It seems that we cannot write this number in the form of a rational number. And indeed, we can prove that it is not a rational number.

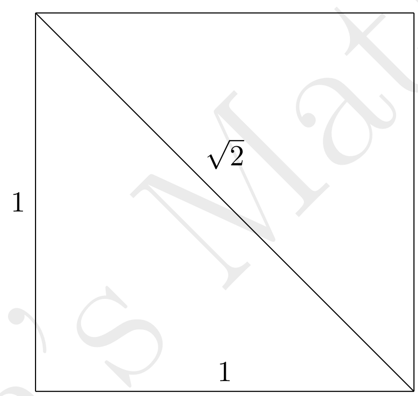


Figure 1.1: Pythagorean Theorem and $\sqrt{2}$

Proposition 1.2.1: Irrationality of $\sqrt{2}$

$\sqrt{2}$ is not a rational number.

Proof. We prove by contradiction. Suppose $\sqrt{2}$ is a rational number, then it can be written as

$$\sqrt{2} = \frac{p}{q}, \quad q \neq 0, p, q \in \mathbb{Z}, p, q \text{ are relatively prime}$$

Multiply by q on both sides and take square, we have

$$2q^2 = p^2$$

Because $2q^2$ is an even number (even number times a number must equal to an even number), p^2 is an even number. Therefore, p itself is an even number (if p is odd, then p^2 is odd, which is a contradiction). Hence, we can write p as

$$p = 2k, \quad k \in \mathbb{Z}$$

Substitute this in the previous equation, we have

$$2q^2 = 4k^2 \quad \Longrightarrow \quad q^2 = 2k^2$$

By the same arguement, we can also conclude that q is an even number. Then, p and q would have a common factor 2, which is a contradiction with our assumption that p, q are relatively prime. $\square$

Therefore, there is another kind of number except for rational numbers! This was a big shock for people during Pythagoras's time, and the discovery of $\sqrt{2}$ is called **'The First Mathematical Crisis'**. Since this discovery broke the belief of Pythagoreanism, Hippasus, who discovered this, was drowned at sea by Pythagoras's followers.

Fortunately, now we fully accept that there is 'irrational numbers'. Nobody would be sentenced to death for acknowledging the existence of irrational numbers. Rational and Irrational numbers together, are called **Real Numbers**, denoted by $\mathbb{R}$.

But, how we should construct real numbers from rational numbers? Could we construct a procedure just like what we did for extending integers to rational numbers? In section 1.8* we will introduce an elegant method, and another method would be introduced later in Chapter 2. Since these construction processes are hard, we should now temporarily just believe that there is indeed a set of numbers called real numbers. In next section we will state the axiom that real numbers should behave.

1.3 Axiom of Completeness I: Supremum Property

One of the most important property of real number is: **It is complete**. The rigorous definition of completeness would be introduced later. Heuristically, completeness of real numbers means that 'All points on the real line are described by real numbers".

Consider rational numbers, they are 'almost everywhere' on the real line, i.e., there is no such a rational number a that is 'closest' to the rational number b. Indeed, suppose there is a b that is 'closest' to a, then, the rational number $\frac{a+b}{2}$ is 'closer' to a, which is a contradiction. This property is called **dense**, and will be introduced later.

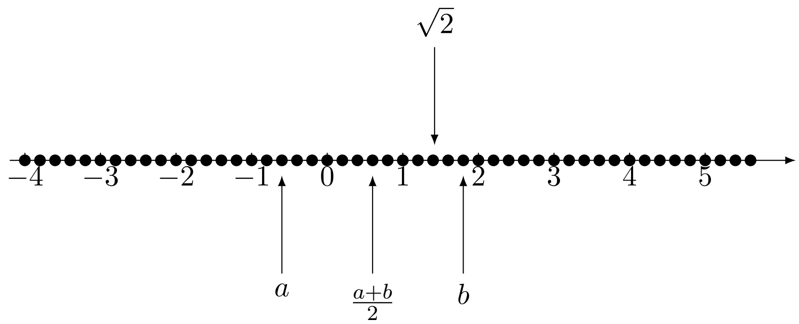


Figure 1.2: Rational Numbers $\mathbb{Q}$ is dense in Real Line $\mathbb{R}$

Even if $\mathbb{Q}$ is dense in $\mathbb{R}$, there are infinite many small 'holes' on the line that was not represented by any of the rational numbers. For example, the point at the distance of $\sqrt{2}$ from the origin, as showed in Figure 1.2. Completeness then means that these holes are exactly 'filled' by 'irrational numbers', so that each point on the line is represented by a unique real number.

To transform these discussions into mathematical language, we first introduce some simple definitions. In this whole note I will denote 'such that' by 's.t.' for simplicity.

Definition 1.3.1: Bounded Above/Bounded Below, Lower/Upper Bound

- A set $A \subseteq \mathbb{R}$ is **bounded above** if

$$\exists\, b \in \mathbb{R}, \text{ s.t. } \forall\, a \in A \implies a \leqslant b$$

The number b is called an **upper bound** for A.

- A set $A \subseteq \mathbb{R}$ is **bounded below** if

$$\exists\, l \in \mathbb{R}, \text{ s.t. } \forall\, a \in A \implies l \leqslant a$$

The number l is called an **lower bound** for A.

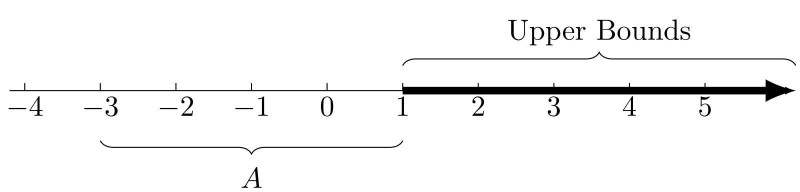


Figure 1.3: A set bounded above

Note that upper bound and lower bound of a set A may not be unique. In fact, if $A \subseteq \mathbb{R}$ is bounded above, upper bounds are always not unique. However, there would sometimes exist a 'least upper bound', which is the most important subject towards the construction of completeness axiom.

Definition 1.3.2: Supremum/Infimum

- A real number s is called the **supremum (least upper bound)** of a set $A \subseteq \mathbb{R}$ if

  1. s is an upper bound for A.
  2. For any upper bound b of A, we have $s \leqslant b$.

  This is denoted by $s = \sup A$. If $s \in A$, it is also the **maximum** of A.

- A real number l is called the **infimum (greatest lower bound)** of a set $A \subseteq \mathbb{R}$ if

  1. l is a lower bound for A.
  2. For any lower bound b of A, we have $b \leqslant l$.

  This is denoted by $l = \inf A$. If $l \in A$, it is also the **minimum** of A.

Note that although upper bound is not unique for sets, Supremum, if exists, is unique.

Proposition 1.3.3: Uniqueness of Supremum

A set $A \subseteq \mathbb{R}$ can have at most one supremum.

Proof. Suppose s_1, s_2 are suprema of a set A. Regard s_1 as an upper bound and s_2 as the supremum, we will arrive $s_2 \leqslant s_1$. Regard s_1 as the supremum and s_2 as an upper bound we will arrive $s_1 \leqslant s_2$. Therefore, $s_1 = s_2$. $\square$

Now we should have all tools for the construction of our completeness theorem. This theorem would be seen as an **axiom**, i.e., no need to be proved and it is raised by nature, so that it is an inherent property of the set of real numbers. (In Section 1.8* we will use an elegant method to prove this axiom)

Axiom 1.3.4: Supremum Property

Every nonempty set of real numbers that is bounded above has a supremum.

Why this axiom expresses the completeness of real numbers? We consider a counterexample. Suppose we only have rational number system. Then consider the set $A = (0, \sqrt{2}) \cap \mathbb{Q}$. If the supremum s is less than $\sqrt{2}$, say $s = \sqrt{2} - \epsilon \in \mathbb{Q}$. Then there would be a number $k = \sqrt{2} - \frac{\epsilon}{2} \in A$, such that $k > s$, which is a contradiction to the definition of supremum. Similarly, we can derive that the supremum also cannot be larger than $\sqrt{2}$. Since $\sqrt{2} \notin \mathbb{Q}$, we conclude that this set A, in rational number system, has no supremum.

Therefore, the rational number system $\mathbb{Q}$ does not have this supremum property. It's only for real number system! Actually, this is the first **Axiom of Completeness for real numbers** in this note. In later chapter there would be more, and we will later on examine the relashionships between these Axiom of Completeness.

Note: We state the axiom of completeness only regarding to supremum. There is no need to assert that infimum exists as part of the axiom. To see this, let A be nonempty and bounded below, define B as

$$B = \{b \in \mathbb{R} : \ b \text{ is a lower bound for } A\}$$

Then we will get $\sup B = \inf A$, by the definition of supremum and infimum. For set A, we can then state the axiom of completeness with respect to the set B, i.e., with respect to the supremum.

To conclude this section, a characterization of supremum would be stated below. This is an EXTREMELY USEFUL TOOL since sometimes it is very difficult to work on supremum directly using its definition.

Proposition 1.3.5: Characterization of Supremum

Let $s \in \mathbb{R}$ and set $A \subseteq \mathbb{R}$. $s = \sup A$ if and only if

- s is an upper bound of A.
- $\forall \epsilon > 0, \exists\, a \in A,$ s.t. $s - \epsilon < a$.

Proof.
($\Longrightarrow$) Suppose $s = \sup A$. Then, s is indeed an upper bound by definition. Also, $s - \epsilon$ is not an upper bound for any $\epsilon > 0$, since $s - \epsilon < s$. Therefore, by definition, there exists $a \in A$, such that $s - \epsilon < a$.
($\Longleftarrow$) Suppose s satisfy the conditions stated in the proposition. Then by the second condition, any number smaller than s is not an upper bound. Therefore, s is the least upper bound. $\square$

1.4 Properties of real numbers

There are many applications of the Axiom of Completeness. We will first introduce the important **Archimedean Property**, which states how $\mathbb{N}$ behaves inside $\mathbb{R}$.

Theorem 1.4.1: Archimedean Property

- For any $x \in \mathbb{R}$, there exists an $n \in \mathbb{N}$ satisfying that $n > x$.
- For any $y > 0$, there exists an $n \in \mathbb{N}$ satisfying that $\frac{1}{n} < y$

Proof.

1. To prove the first statement, we assume that there exists $x \in \mathbb{R}$ such that for all $n \in \mathbb{N}$ we have $n \leqslant x$. This is equivalent to say that, $\mathbb{N}$ is bounded above. By Supremum Property, supremum exists. Let $\alpha = \sup \mathbb{N}$. Then $\alpha + 1 \in \mathbb{N}$. This contradicts the definition of supremum since $\alpha + 1 > \alpha = \sup \mathbb{N}$. Thus, we arrive a contradiction.
2. The second statement follows from (1) by letting $x = 1/y$.

$\square$

Note: It seems that there is no need to prove the statement (1) in Archimedean Property. It just said that $\mathbb{N}$ is unbounded, and we know that as a common sense. However, it is worth noting that as a proper extension of $\mathbb{Q}$ (i.e., a set contains $\mathbb{Q}$ and not equal to $\mathbb{Q}$), the Archimedean Property is very unique for $\mathbb{R}$. Indeed, there **does exist** a proper extension of $\mathbb{Q}$ such that it is bounded (called the Extended-Real Numbers). Discussing this number system will go far out from the scope of this note. You should look for detailed explanation in my Real Analysis note. Now we see how $\mathbb{Q}$ and $\mathbb{R}\backslash\mathbb{Q}$ behaves inside $\mathbb{R}$.

Proposition 1.4.2: $\mathbb{Q}$ is dense in $\mathbb{R}$

For any $a, b \in \mathbb{R}$, $a < b$, there exists $r \in \mathbb{Q}$ such that $a < r < b$.

Proof. We need to produce $m, n \in \mathbb{Z}, n \neq 0$ such that $r = \frac{m}{n}$ and

$$a < \frac{m}{n} < b$$

The first thing we need to do is to choose sufficiently large n so that a 'step length' $\frac{1}{n}$ is less than the length of $b - a$, so that there would be some point of the form $\frac{m}{n}$ locating between the two points, as showed in the Figure 1.4.

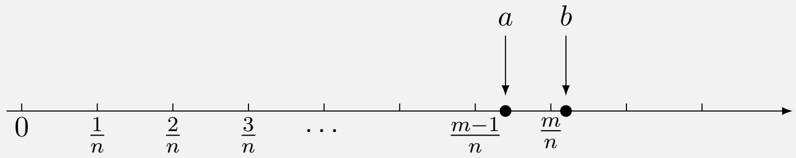


Figure 1.4: Choose sufficiently small step so that $\frac{m}{n}$ is between a and b

By Archimedean Property, we can actually take $n \in \mathbb{N}$ such that

$$\frac{1}{n} < b - a$$

Now, as showed in the picture, we need to choose $m \in \mathbb{Z}$ so that

$$\frac{m-1}{n} \leqslant a < \frac{m}{n} \quad \Longrightarrow \quad m - 1 \leqslant na < m$$

The only thing left is that to prove $\frac{m}{n} < b$. To do this, we see that

$$m \leqslant na + 1 < n\left(b - \frac{1}{n}\right) + 1 = nb$$

Therefore, the proposition is proved. □

We can use the same strategy to see that irrational number is also dense in $\mathbb{R}$. Before doing that, we need to prove some lemma about properties of operations in rational and irrational numbers.

Lemma 1.4.3: Operations in $\mathbb{Q}$ and $\mathbb{R}\backslash\mathbb{Q}$

1. If $a, b \in \mathbb{Q}$, then $a + b, ab \in \mathbb{Q}$.
2. If $a \in \mathbb{Q}, b \in \mathbb{R}\backslash\mathbb{Q}$, then $a + t, at(a \neq 0) \in \mathbb{R}\backslash\mathbb{Q}$

Proof. 1. Since $a, b \in \mathbb{Q}$, we have $a = \frac{m}{n}$, $b = \frac{r}{q}$ for $m, n, r, q \in \mathbb{Z}, n, q \neq 0$. Therefore,

$$a + b = \frac{mq + nr}{nq} \in \mathbb{Q}, \quad ab = \frac{mr}{nq} \in \mathbb{Q}$$

since $mq + nr, nq, mr \in \mathbb{Z}$.

2. Since $a \in \mathbb{Q}$, we have $a = \frac{m}{n}$ where $m, n \in \mathbb{Z}$, $n \neq 0$. Suppose $a + t \in \mathbb{Q}$, then we can write

$a + t = \frac{r}{q}, r, q \in \mathbb{Z}, q \neq 0$. Then,

$$t = (a+t) - a = \frac{r}{q} - \frac{m}{n} = \frac{rn - mq}{qn} \in \mathbb{Q}$$

which is a contradiction. Similarly, we can also see that $at \notin \mathbb{Q}$ (when $a \neq 0$). $\square$

Proposition 1.4.4: $\mathbb{R}\backslash\mathbb{Q}$ is dense in $\mathbb{R}$

For any $a, b \in \mathbb{R}, a < b$, there exists a $t \in \mathbb{R}\backslash\mathbb{Q}$ such that $a < t < b$.

Proof. As in Proposition 1.4.2, we first choose $n \in \mathbb{N}$ such that

$$\frac{1}{n} < b - a$$

Then, we choose $m \in \mathbb{Z}$ such that

$$m + \sqrt{2} - 1 \leqslant na < m + \sqrt{2}$$

Obviously, we have $a < \frac{m+\sqrt{2}}{n}$. Also,

$$m + \sqrt{2} \leqslant na + 1 < n\left(b - \frac{1}{n}\right) + 1 = nb$$

Thus, $a < \frac{m+\sqrt{2}}{n} < b$. Since $m, \frac{1}{n} \in \mathbb{Q}$, $\sqrt{2} \in \mathbb{R}\backslash\mathbb{Q}$, by Lemma 1.4.3, we have $\frac{m+\sqrt{2}}{n} \in \mathbb{R}\backslash\mathbb{Q}$. The statement is proved. $\square$

At this stage we have proved that $\sqrt{2}$ is not a rational number. But, we have not proved that it is a real number. Below we will prove this. Why we need to prove this obvious thing? Indeed, you will see from the prove below that the main theorem used is Supremum Property. Therefore, this inherent property of real numbers asserts that those irrational numbers we encounter often are all real numbers.

Proposition 1.4.5: $\sqrt{2}$ is a real number

There exists a real number $\alpha \in \mathbb{R}$ such that $\alpha^2 = 2$.

Proof. Consider the set

$$T = \{t \in \mathbb{R} : t^2 < 2\}$$

Set $\alpha = \sup T$.

- Suppose $\alpha^2 < 2$. Let $n \in \mathbb{N}$ be arbitrary, then

$$\left(\alpha + \frac{1}{n}\right)^2 = \alpha^2 + \frac{2\alpha}{n} + \frac{1}{n^2} < \alpha^2 + \frac{2\alpha}{n} + \frac{1}{n} = \alpha^2 + \frac{2\alpha+1}{n}$$

  If we choose $n \in \mathbb{N}$ such that

$$\frac{1}{n} < \frac{2-\alpha^2}{2\alpha+1}$$

  The existence of this n is promised by Archimedean Property. Note that $n > 0$ since we assume that $\alpha^2 < 2$. We would get

$$\left(\alpha + \frac{1}{n}\right)^2 < \alpha^2 + \frac{(2\alpha+1)(2-\alpha^2)}{2\alpha+1} = 2$$

  Therefore, $\alpha + \frac{1}{n} \in T$, which means that α is not an upper bound, which is a contradiction to that $\alpha = \sup T$.

- Suppose $\alpha^2 > 2$. Similarly, we can write

$$\left(\alpha - \frac{1}{n}\right)^2 = \alpha^2 - \frac{2\alpha}{n} + \frac{1}{n^2} > \alpha^2 - \frac{2\alpha}{n}$$

  If we choose $n \in \mathbb{N}$ such that

$$\frac{1}{n} < \frac{\alpha^2 - 2}{2\alpha}$$

  Again, the existance is promised by Archimedean Property. Then we would have

$$\left(\alpha - \frac{1}{n}\right)^2 > \alpha^2 - \alpha^2 + 2 = 2$$

  This shows that $\alpha - \frac{1}{n}$ is an upper bound of T, which means that α is not the least upper bound, i.e., not the supremum. This is a contradiction to the assumption that $\alpha = \sup T$.

By Supremum Property, the supremum of T exists. Then, it can only be $\sqrt{2}$. Since the supremum of a set is a real number (which is promised in Supremum Property), we finally arrive that $\sqrt{2}$ is actually a real number. $\square$

1.5 Axiom of Completeness II: Nested Interval Property

Another famous Axiom of Completeness of real numbers is called the **(Cantor) Nested Interval Property**. It says that for any nested closed interval sequence, the intersection of these intervals is not empty. Here is what all these words mean.

Axiom 1.5.1: Nested Interval Property

For each $n \in \mathbb{N}$, construct a closed interval $I_n = [a_n, b_n]$, where $a_n, b_n \in \mathbb{R}$. Assume $I_1 \supseteq I_2 \supseteq I_3 \supseteq \cdots$. Then, the intersection of these nested intervals

$$\bigcap_{n=1}^{\infty} I_n \neq \emptyset$$

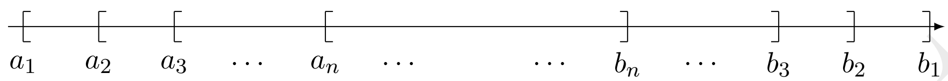


Figure 1.5: Nested Intervals

Actually, this axiom can be derived from Supremum Property, as showed below.

Proof. Consider the set

$$A = \{a_n : n \in \mathbb{N}\}$$

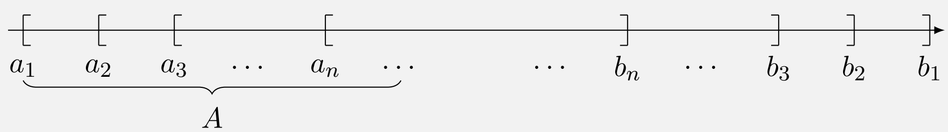


Figure 1.6: Nested Intervals with set A

We can see that each b_n is served as an upper bound for the set $[a_n, b_n]$, for every $n \in \mathbb{N}$. Set $x = \sup A$. Then $a_n \leqslant x$ for every n since x is the upper bound. Also, $b_n \geqslant x$ for every n since b_n are upper bounds and x is the supremum. Therefore we get,

$$\forall n \in \mathbb{N}, a_n \leqslant x \leqslant b_n$$

Hence, $x \in \bigcap_{n=1}^{\infty} I_n \neq \emptyset$. □

Can we, inversely, get Supremum Property from Nested Interval Property? The answer is no! Therefore, there is some 'strong' axioms and some 'weak' axioms. In this case, Supremum Property is 'strong' and Nested Interval Property is 'weak', since we can go from Supremum Property to Nested Interval, but not the reverse direction.

But, why we can't? In the 'proof' below I will show you a **circular reasoning**, indicating that we can only go from Nested Interval Property to Supremum Property if Archimedean Property exists.

Note: This is not a prove, but just a reasoning process.

Suppose Nested Interval Property is true. We want to prove Supremum Property. Let A be a nonempty set which is bounded above. Denote $\alpha = \sup A$. Choose $a \in A$ and b such that b is an upper bound of A. Then, consider the intervals $[a, \frac{a+b}{2}]$ and $[\frac{a+b}{2}, b]$. Then, α is at least in one of these two intervals. Choose an interval that contains it. Continue bisecting this interval as we did at the last step. Again, the supremum would contain in one of the intervals. Choose that interval. $\cdots\cdots$

After n steps, the length of the chosen interval would be $\frac{b-a}{2^n}$. The only thing we are left to do is to prove that $\frac{b-a}{2^n}$ converges to 0 (The rigorous definition of limit would be introduced in Chapter 2, here you can just recall what you have learnt in Year I Calculus). However, the proof of this statement needs Archimedean Property, since we want for every $\epsilon > 0$, we can find a $n \in \mathbb{N}$ such that $\frac{b-a}{2^n} < \epsilon$. This is equivalent to what is said in the second statement of Archimedean Property. Recall how Archimedean Property is derived. Yes, it is derived from Supremum Property! So we cannot directly use Archimedean Property if we assume we don't know Supremum Property and want to prove it. This is an example of **circular reasoning**, and in result, we have no way deriving Supremum Property from Nested Interval Property.

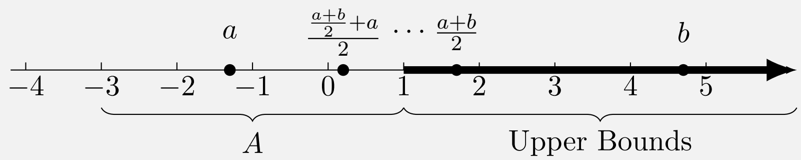


Figure 1.7: Bisecting intervals to approximate supremum

Therefore, we get the relationship between Supremum Property and Nested Interval Property. The relation is visually displayed below in Figure 1.8.

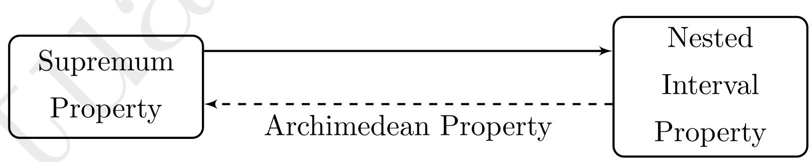


Figure 1.8: Relation between Axiom of Completeness

This graph would be further expanded when we encounter more and more Axiom of Completeness.

1.6 'Size of Infinity': Cardinality

Does infinity also have different 'sizes'? You may ask after seeing this section title. The answer is yes! We all know that there are infinitely many rational numbers and irrational numbers, and it seems that both of them are 'almost everywhere' on the real line. However, in this section we will introduce a surprising result: There are 'much more' irrational numbers than rational numbers!

Before we discuss this result, we need to first talk about the way of comparing two 'infinite sizes'. Let's start with finite one. To compare the number of elements in two finite sets, it is easy, just count them. For example, $A = \{1, 2, 3, 4\}$ and $B = \{5, 6, 7, 8\}$. They all have 4 elements, and naturally, have the same size. To extend this into infinite case, we need to use **bijective maps**.

Definition 1.6.1: Cardinality

Two sets A, B have the same **Cardinality** if there exists a bijective map $f : A \to B$ such that each element of A is mapped one-to-one and onto an element of B. Dented as $A \sim B$.

Then, the cardinality of a set just discribes the 'size' of that set. Let's see some examples first.

Example 1.6.2: $\mathbb{N}$ ***has the same cardinality as the set of even numbers***

This is weird at first glance, since intuitively the set of even numbers is a proper subset of $\mathbb{N}$, and they could not have the same size. Denote the set of even numbers as

$$E = \{2, 4, 6, 8, \cdots\}$$

Then we can construct a bijective map $f : \mathbb{N} \to E$ by $f(n) = 2n, \forall\, n \in \mathbb{N}$.

$$\begin{array}{rcccccccc} \mathbb{N}: & 1 & 2 & 3 & 4 & \cdots & n & \cdots \\ & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \cdots & \updownarrow & \cdots \\ E: & 2 & 4 & 6 & 8 & \cdots & 2n & \cdots \end{array}$$

Example 1.6.3: $\mathbb{N} \sim \mathbb{Z}$

We can construct a bijective map $f : \mathbb{N} \to \mathbb{Z}$ by

$$f(n) = \begin{cases} \dfrac{n-1}{2}, & \text{if } n \text{ is odd} \\ -\dfrac{n}{2}, & \text{if } n \text{ is even} \end{cases}$$

$$\begin{array}{rcccccccc} \mathbb{N}: & 1 & 2 & 3 & 4 & 5 & 6 & 7 & \cdots \\ & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \cdots \\ \mathbb{Z}: & 0 & -1 & 1 & -2 & 2 & -3 & 3 & \cdots \end{array}$$

In the two examples above, we see that we correspond elements of some set with natural numbers $1, 2, 3, 4, \cdots$, just as we are counting them in some order. This is such an important case that we gave it a name.

Definition 1.6.4: Countable/Uncountable Set

A set is A is called

- **Finite** if the number of elements in A is finite.
- **Countable** if $A \sim \mathbb{N}$.
- **Uncountable** if it is infinite and not countable.

We can see from Example 1.6.2 and 1.6.3 that E and $\mathbb{Z}$ are countable sets. What does uncountable set looks like? **The next theorem is central for this section.** It says that $\mathbb{R}$ has a somewhat 'bigger' size than $\mathbb{Q}$.

Theorem 1.6.5: Countability of $\mathbb{Q}$, Uncountability of $\mathbb{R}$

1. $\mathbb{Q}$ is a countable set.
2. $\mathbb{R}$ is an uncountable set.

Proof.

1. There are two popular ways of proving that $\mathbb{Q}$ is countable. I will show both of them.
   METHOD I: Arrage all the rational numbers in an infinite matrix such that mth row and nth column corresponds to the number $\frac{n}{m}$. Then, assign natural numbers to them 'meanderingly', as showed below. If there is a number that has the same value with some number that has been assigned, we delete it. For example, $\frac{1}{1}$ is assigned 1, $\frac{1}{2}$ is assigned 2, $\frac{2}{1}$ is assigned 3, $\frac{3}{1}$ is assigned 4, $\frac{2}{2}$ is deleted since it has the same value with $\frac{1}{1}$, and $\frac{1}{3}$ is assigned 5...... Continuing this fashion, we will have a bijective map from positive rational numbers to natural numbers.

With this, we can further map 0 to 0 and map negative rational numbers to negative integers. Then, this whole map is a bijective map from $\mathbb{Q}$ to $\mathbb{Z}$. Since there also exists biject map from $\mathbb{Z}$ to $\mathbb{N}$, we have $\mathbb{N} \sim \mathbb{Q}$.

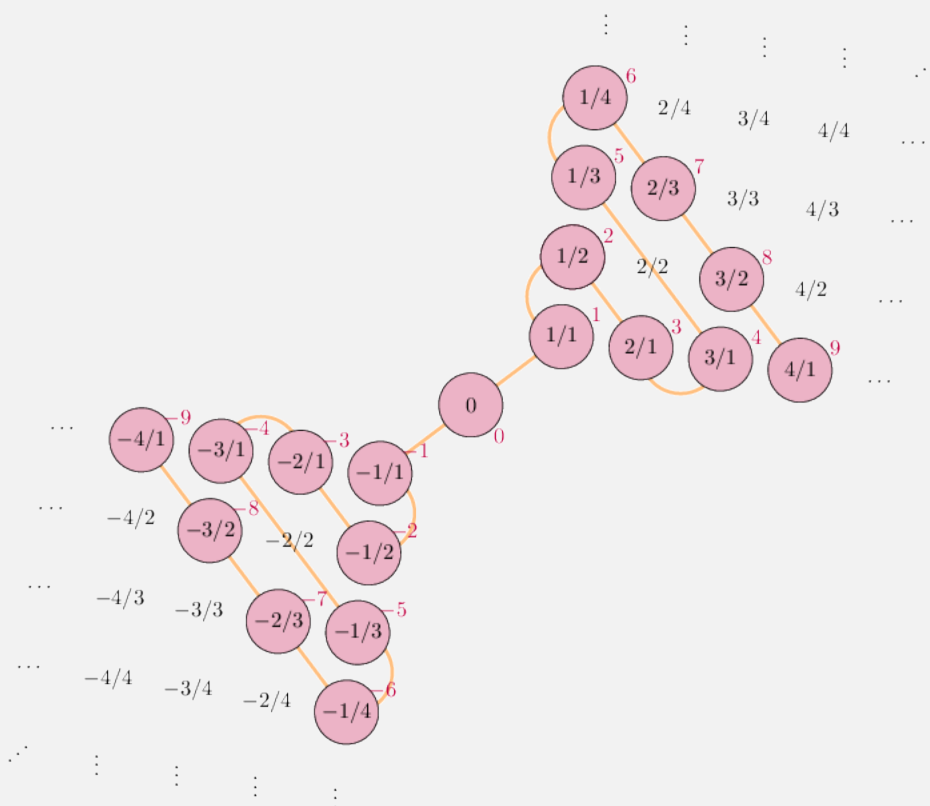
METHOD II: Set $A_1 = \{0\}$, and for all $n \geqslant 2$, set

$$A_n = \left\{ \pm\frac{p}{q} : p, q \in \mathbb{N} \text{ are relatively prime with } p + q = 0 \right\}$$

For example,

$$A_2 = \left\{ \frac{1}{1}, -\frac{1}{1} \right\}, \quad A_3 = \left\{ \frac{1}{2}, -\frac{1}{2}, \frac{2}{1}, -\frac{2}{1} \right\}, \quad A_4 = \left\{ \frac{1}{3}, -\frac{1}{3}, \frac{3}{1}, -\frac{3}{1} \right\}, \cdots$$

Each A_n is finite and every rational number appears in exactly one of these sets. Therefore, we can construct the bijective map by listing the elements in each A_n.



2. The main theorem used in this proof is the **Nested Interval Property**. We will prove by contradiction. Suppose there exists a bijective function $f : \mathbb{N} \to \mathbb{R}$. Then, each real number can be assigned to a natural number. Therefore, we can denote x_i as the real number being assigned

to the natural number i, and write $\mathbb{R}$ as

$$\mathbb{R} = \{x_1, x_2, x_3, x_4, \cdots\}$$

Now, consider the set $[0, 9]$. We can divide it into 3 parts: $[0, 3] \cup [3, 6] \cup [6, 9]$. Then, x_1 can at most belong to two of them. Choose the interval that x_1 does not belong to, denote it as I_1.

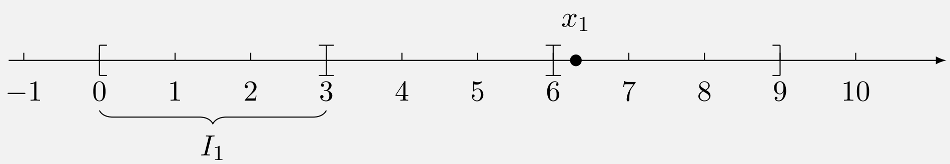


Figure 1.9: Construction Process of Nested Intervals, I

Then, we can divide I_1 into 3 equal parts just as the previous step. Again, x_2 can at most belong to one of these three intervals. Choose the one that x_2 does not belong to, and call it I_2. Continuing this fashion, for I_n, we divide it into 3 equal parts, and choose the interval that x_{n+1} does not belong to, call it I_{n+1} $\cdots$.

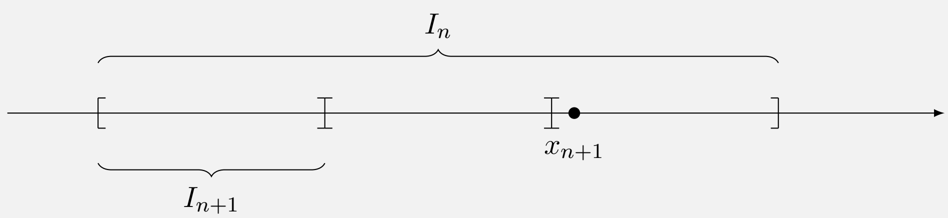


Figure 1.10: Construction Process of Nested Intervals, II

Using this procedure, we can produce nested intervals I_n such that

$$I_{n+1} \subseteq I_n, \forall\, n \in \mathbb{N}, \text{ and } \quad x_n \notin I_n$$

Therefore,

$$x_n \notin \bigcap_{n=1}^{\infty} I_n, \forall\, n \in \mathbb{N}$$

This shows that

$$\bigcap_{n=1}^{\infty} I_n = \emptyset$$

which is a contradiction to the Nested Interval Property. $\square$

Therefore, $\mathbb{R}$ is a 'bigger' set than $\mathbb{N}$! There does exist uncountable sets. In examples before, we have seen some other sets that is countable. In the next few examples, we will see what does uncountable sets look like.

Example 1.6.6: $(-1, 1) \sim \mathbb{R}$

Here we can construct the function $f : (-1, 1) \to \mathbb{R}$ by

$$f(x) = \frac{x}{x^2 - 1}$$

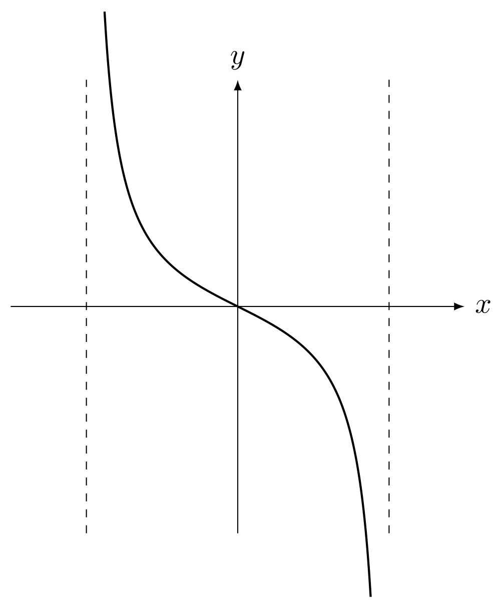


Figure 1.11: function $f(x) = \frac{x}{x^2-1}$

Example 1.6.7: $(a, b) \sim \mathbb{R}$

To extend the result from Example 1.6.6 to the case for every $a, b \in \mathbb{R}$, we can just do linear transformation on the function f in that example. Set

$$-1 < kx + c < 1, k > 0$$

We have

$$\frac{-1-c}{k} < x < \frac{1-c}{k}$$

Therefore, we can set

$$a = \frac{-1-c}{k}, \quad b = \frac{1-c}{k}$$

To get

$$k = \frac{2}{b-a}, \quad c = \frac{a+b}{a-b}$$

Thus

$$g(x) = f\left(\frac{2x}{b-a} + \frac{a+b}{a-b}\right)$$

will map (a, b) to the whole space $\mathbb{R}$.

Example 1.6.8: $(a, \infty) \sim \mathbb{R}$

We can construct the function $f : (a, \infty) \to \mathbb{R}$ by

$$f(x) = \log(x - a)$$

Example 1.6.9: $[0, 1) \sim (0, 1)$

This one is interesting. It seems very easy. However, if you try that, it is not. Let us consider one famous paradox in mathematics: **Hilbert's paradox of the Grand Hotel**. There is a hotel with countably infinitely many rooms in the hotel, and the hotel is full. Then, some travellers come and want a room. The reception just send a call to everybody in the room: For person who live in room n, just move to the room $n + 1$, and the room 1 would be available for this traveller. Weird! Since we seem to create a new empty room from nowhere.
This problem is just the same. We need to move some countable set accordingly to make a new room for this new comer 0. Therefore, we can construct the bijective function $f : [0, 1) \to (0, 1)$ like

$$f(x) = \begin{cases} 1/2, & \text{if } x = 0 \\ 1/4, & \text{if } x = \frac{1}{2} \\ 1/8, & \text{if } x = \frac{1}{4} \\ 1/16, & \text{if } x = \frac{1}{8} \\ \dots & \\ x, & \text{otherwise} \end{cases}$$

Now we should be familiar with countable sets and uncountable sets. In Chapter 1.7* we will further go into n-dimensions, to see that $\mathbb{R}$ are also uncountable sets, of the same cardinality with $\mathbb{R}$.

Finally, to end this section, we will see that whether the set of irrational numbers is countable or uncountable. Before doing that, let's prove some theorems about the properties of countable sets.

Theorem 1.6.10: Subsets and Unions of countable sets

1. If $A \subseteq B$, B is countable, then, A is either countable or finite.

2. If $A_1, A_2, ..., A_m$ are countable, then,
$$\bigcup_{n=1}^{m} A_n$$
is countable.

3. If A_n is countable for each $n \in \mathbb{N}$, then,
$$\bigcup_{n=1}^{\infty} A_n$$
is countable.

Proof.

1. Let B be a countable set. Then, there exists bijective map $f : \mathbb{N} \to B$. Let $A \subseteq B$ be an infinite subset of B. We must show that A is countable.
We now start to define a bijective map from $\mathbb{N}$ to A. Let
$$n_1 = \min\{n \in \mathbb{N} : f(n) \in A\},\ \text{Set } g(1) = f(n_1)$$
$$n_2 = \min\{n \in \mathbb{N}\backslash\{1, 2, \cdots, n_1\} : f(n) \in A\},\ \text{Set } g(2) = f(n_2)$$
$$n_3 = \min\{n \in \mathbb{N}\backslash\{1, 2, \cdots, n_2\} : f(n) \in A\},\ \text{Set } g(3) = f(n_3)$$
$$\vdots$$
Inductively, as we can easily verify, that this function is bijective from $\mathbb{N}$ to A.

2. We first prove that this is true for two countable sets, A_1, A_2. Let $B_2 = A_2\backslash A_1$. Then, A_1 and B_2 is disjoint, and $A_1 \cup A_2 = A_1 \cup B_2$. Therefore, we only need to prove that $A_1 \cup B_2$ is countable. By statement 1 above, we can see that $B_2 = A_2\backslash A_1 \subseteq A_2$ is countable or finite. First suppose that B_2 is countable. Then, we can write both sets in the enumerated form
$$A_1 = \{a_1, a_2, a_3, \cdots\}$$
$$B_2 = \{b_1, b_2, b_3, \cdots\}$$
Since they are disjoint, we can write their union as
$$A_1 \cup B_2 = \{a_1, b_1, a_2, b_2, a_3, b_3, \cdots\}$$

A bijective map then can be constructed as $f : \mathbb{N} \to A_1 \cup B_2$ such that

$$f(n) = \begin{cases} a_{\frac{n}{2}}, & \text{if } n \text{ is even} \\ b_{\frac{n+1}{2}}, & \text{if } n \text{ is odd} \end{cases}$$

Now suppose B_2 is finite. This time

$$A_1 = \{a_1, a_2, a_3, \cdots\}$$

$$B_2 = \{b_1, b_2, b_3, \cdots, b_m\}$$

Since they are disjoint, we can write their union as

$$A_1 \cup B_2 = \{b_1, b_2, b_3, \cdots, b_m, a_1, a_2, a_3, \cdots\}$$

A bijective map then can be constructed as $f : \mathbb{N} \to A_1 \cup B_2$ such that

$$f(n) = \begin{cases} b_n, & \text{if } n \leqslant m \\ a_{n-m}, & \text{if } n > m \end{cases}$$

We have proved that $A_1 \cup A_2$ is countable. Now if A_3 is countable, we can directly see that,

$$A_1 \cup A_2 \cup A_3 = (A_1 \cup A_2) \cup A_3$$

is countable. Inductively, we have

$$\bigcup_{n=1}^{m} A_n$$

is countable if $A_1, \cdots, A_m$ are countable sets.

3. Suppose A_n is countable for all $n \in \mathbb{N}$. Then each set can be written as

$$A_n = \{a_{n1}, a_{n2}, a_{n3}, \cdots\}$$

We can arrange the elements into a infinite matrix

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots \\ \vdots & \vdots & \vdots & \cdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Then, we can assign each element a natural number just as what we did in the proof of Theorem 1.6.5. This constructs a bijective map. □

Here is an interesting fact derived from the theorem: We can have a countable collection of disjoint open intervals, such as $(0,1),(1,2),(2,3),\cdots$ However, **we cannot have an uncountable collection of disjoint open intervals**. Suppose for contradiction, there is such a collection. Each open interval must contain a rational number (since $\mathbb{Q}$ is dense in $\mathbb{R}$). Then, for each interval, we can randomly select a rational number that is contained in it. This constructs a biject map from these intervals to a subset of $\mathbb{Q}$. By the first statement in the previous theorem, the number of intervals must be countable.

Let's go back to the topic. Naturally from the theorem proved above, we can have the result:

Corollary 1.6.11: $\mathbb{R}\backslash\mathbb{Q}$ is uncountable

The set of irrational numbers, $\mathbb{R}\backslash\mathbb{Q}$, is uncountable.

Proof. Suppose $\mathbb{R}\backslash\mathbb{Q}$ is countable. Then, by Theorem 1.6.10, the union $\mathbb{R} = \mathbb{R}\backslash\mathbb{Q} \cup \mathbb{Q}$ is countable since $\mathbb{Q}$ is countable, which contradicts the fact that $\mathbb{R}$ is uncountable. □

There are 'more' irrational numbers than rational numbers! Even if they are all dense in the real line. Indeed, this kind of pattern is what you will always encounter down the road of matheamtics study. The 'pathological' mathematical objects are far more than the well-behaved ones. For example, there are more discontinuous functions than continuous ones. There are more transcendental numbers than algebraic numbers......

1.7* Aleph Numbers and Continuum Hypothesis

This section is a hard section and can be skipped without losing coherence.

1.7.1 Cantor's Diagonalization Method

In 1891, Cantor offered another elegant proof of the fact that $\mathbb{R}$ is uncountable. This method is called **Cantor's Diagonalization Method**. It uses the **decimal representations** for real numbers.

Definition 1.7.1: Decimal Representation

A decimal representation of a non-negative real number r is its expression as a sequence of symbols consisting of decimal digits traditionally written with a single separator:

$$r = b_k b_{k-1} \cdots b_0 . a_1 a_2 a_3 \cdots , \quad b_i, a_j \in \{0, 1, 2, 3, \cdots, 9\}, i \in \{1, 2, 3, \cdots, k\}, j \in \mathbb{N}$$

and it represents the infinite sum

$$r = \sum_{i=0}^{k} b_i 10^i + \sum_{i=1}^{\infty} \frac{a_i}{10^i}$$

For the rigorous definition of Infinite sum, see Chapter 2. **Here is a problem, each real number has at least one decimal representation. Some real number has two decimal representation. The mapping from decimal representation to real numbers is not bijective**. To solve this problem, we observe that **a real number has two such representations if and only if one has a trailing infinite sequence of 0, and the other has a trailing infinite sequence of 9**. To make the mapping into a bijection, we will ban the use of decimal representations with a trailing infinite sequence of 9.
Now we state the Cantor's proof.

Proof. **CANTOR'S DIAGONALIZATION METHOD**
We have already seen that $(0,1) \sim \mathbb{R}$. If we can prove that $(0,1)$ is uncountable, we are done.
We prove by contradiction. Suppose there exists a bijective function $f : \mathbb{N} \to (0,1)$. Then, for each $m \in \mathbb{N}$, $f(m)$ is a real number that has decimal representation

$$f(m) = .a_{m1} a_{m2} a_{m3} \cdots$$

where $a_{mn} \in , \mathbb{1}, , \cdots , ,$. The bijective correspondence is summarized below as a table

$$\begin{array}{ccccccccc}
\mathbb{N} & & (0,1) & & & & & & \\
1 & \longleftrightarrow & f(1) & = & .\mathbf{a_{11}} & a_{12} & a_{13} & a_{14} & \cdots \\
2 & \longleftrightarrow & f(2) & = & .a_{21} & \mathbf{a_{22}} & a_{23} & a_{24} & \cdots \\
3 & \longleftrightarrow & f(3) & = & .a_{31} & a_{32} & \mathbf{a_{33}} & a_{34} & \cdots \\
4 & \longleftrightarrow & f(4) & = & .a_{41} & a_{42} & a_{43} & \mathbf{a_{44}} & \cdots \\
\vdots & & \vdots & & \vdots & \vdots & \vdots & \vdots & \ddots
\end{array}$$

Then, every decimal representation of numbers in $(0,1)$ would appear somewhere in the table. However, we can define a real number $x \in (0,1)$ with decimal representation of $x = .b_1 b_2 b_3 \cdots$ such that

$$b_n = \begin{cases} 1, & \text{if } a_{nn} \neq 1 \\ 2, & \text{if } a_{nn} = 1 \end{cases}$$

Then, $x \neq f(1)$ because $a_{11} \neq b_1$. $x \neq f(2)$ because $a_{22} \neq b_2$. $x \neq f(3)$ because $a_{33} \neq b_3 \cdots$ Therefore, we have for each $n \in \mathbb{N}$, $x \neq f(n)$, which is a contradiction. □

Note: This proof cannot be used on $\mathbb{Q}$ since, as we know from junior or high school, that rational numbers have infinite-repeating decimals, and the construction for x would lead to a real number.

This kind of method can be used not only on the proof of the fact that $\mathbb{R}$ is an uncountable set. Let's see another example.

Example 1.7.2: Set of infinite $0-1$ sequence is uncountable

Let

$$S = \{(a_1, a_2, a_3, \cdots) : a_n = 0 \text{ or } 1\}$$

S is uncountable.

Proof. Suppose it is countable. We can then write elements in S as $S = \{x_1, x_2, x_3, \cdots\}$, $x_n = (a_{n1}, a_{n2}, a_{n3}, \cdots)$, where $a_{nm} = 0$ or 1. Then, we can construct a similar table as in Cantor's Diagonalization Method. Now consider an infinite sequence

$$b_n = \begin{cases} 0, & \text{if } a_{nn} = 1 \\ 1, & \text{if } a_{nn} = 0 \end{cases}$$

It is not in the list, which is a contradiction. □

1.7.2* Schröder-Bernstein Theorem, Cardinality of Real Space $\mathbb{R}^n$

Sometimes it is very difficult to find a bijective function between two sets A and B with the same cardinality. However, it is almost always easy to find a injective function from A to B, and another injective function from B to A. Does this imply that there exists a bijective function? Yes! And it is called the **Schröder-Bernstein Theorem**.

Theorem 1.7.3: Schröder-Bernstein Theorem

Suppose there exists injective function $f : X \to Y$ and another injective function $f : Y \to X$. Then, there exists a bijective function from X to Y, hence $X \sim Y$.

Proof. There are many versions of proof of this theorem. The most famous one is presented by a 19-year old student, Bernstein, who was in Cantor's Seminar. Almost simultaneously, Schröder presents another proof.

To make this proof clear and well-structured, I will separate them into several STEPS. The basic idea is to partition X and Y into components

$$X = A \cup A' \text{ and } Y = B \cup B'$$

with $A \cap A' = \emptyset$ and $B \cap B' = \emptyset$, in such way that f maps A surjectively onto B, and g maps B' surjectively onto A'

STEP I: Set $A_1 = X \backslash g(Y)$. If $A_1 = \emptyset$ then $g(Y) = X$, g is bijective, then we are done. So, assume $A_1 \neq \emptyset$. Inductively define a sequence of sets by letting $A_{n+1} = g(f(A_n))$. We will show that $\{A_n : n \in \mathbb{N}\}$ is pairwise disjoint collection of subsets of X.

We first show that $A_1 \cap A_k = \emptyset$. This is obvious since $A_1 = X \backslash g(Y)$ and $A_k = g(f(k-1)) \in g(Y)$.

Now we turn to prove more general case that $A_j \cap A_k = \emptyset$. Define $h(x) = g(f(x))$. Because both f and g are injective, we have h is injective as well. Note that

$$h(A \cap B) = h(A) \cap h(B)$$

This can be proved by the following details:

($\Longrightarrow$) Suppose $x \in h(A \cap B)$, since h is injective, there exists unique $y \in A \cap B$ such that $h(y) = x$. Since $y \in A$ and $y \in B$, we have $x = h(y) \in h(A) \cap h(B)$. This shows that $h(A \cap B) \subseteq h(A) \cap h(B)$.

($\Longleftarrow$) Suppose $x \in h(A) \cap h(B)$, then $x \in h(A)$ and $x \in h(B)$. Since h is injective, there exists unique $y \in X$ such that $x = h(y)$. This means that $y \in A$ and $y \in B$. Thus, $y \in A \cap B, x = h(y)$. We conclude that $x \in h(A \cap B)$. Therefore, $h(A \cap B) \supseteq h(A) \cap h(B)$.

Denote $h^2 = h \circ h$, and inductively denote h^k as the kth composition of h. Note that h^k is injective. Therefore, we have

$$A_{j+1} \cap A_{k+1} = h^k(A_{j-k}) \cap h^k(A_1) = h^k(A_{j-k} \cap A_1) = h^k(\emptyset) = \emptyset, \quad j, k \in \mathbb{N}$$

STEP II: Now we prove that $\{f(A_n) : n \in \mathbb{N}\}$ is a pairwise disjoint collection of subsets of Y. This is easy. Since f is injective, we have $f(A_j) \cap f(A_k) = f(A_j \cap A_k) = f(\emptyset) = \emptyset$.

STEP III: Now we let $A = \bigcup_{n=1}^{\infty} A_n$ and $B = \bigcup_{n=1}^{\infty} f(A_n)$. Note that since A_n are pairwise disjoint,

$$f(A) = f\left(\bigcup_{n=1}^{\infty} A_n\right) = \bigcup_{n=1}^{\infty} f(A_n) = B$$

Therefore, f maps A surjectively onto B.

STEP IV: We finally show that for $A' = X\backslash A$ and $B' = X\backslash B$, we have g maps B' surjectively onto A'. To prove this, we need to show that $g(B') = A'$. We prove both directions by contradiction.
($\Longrightarrow$) To prove $g(B') \subseteq A'$, suppose for contradiction that there exists $b' \in B'$ such that $g(b') \in A$. Since $A_1 \cap g(Y) = \emptyset$, we must have $g(b') \notin A_1$, thus $g(b') \in A\backslash A_1$. Note that

$$g(B) = g\left(\bigcup_{n=1}^{\infty} f(A_n)\right) = \bigcup_{n=1}^{\infty} g(f(A_n)) = \bigcup_{n=1}^{\infty} A_{n+1} = A\backslash A_1$$

We have $g(b') \in g(B)$. This means that there exists $b \in B$ such that $b' \neq b$ (this is because $B \cap B' = \emptyset$), and $g(b') = g(b)$, which is a contradiction to the injectivity of g.
($\Longleftarrow$) To prove $g(B') \supseteq A'$, suppose for contradiction that there exists $a' \in A'$ such that $a' \notin g(B')$. Immediately, because $A' \in g(Y)$, we have $a' \in g(B)$ (since $a' \notin g(B')$). This will contradict the fact that $a' \in A'$ since $g(B) = A\backslash A_1 \subseteq A$.
STEP V: Now we know that $f : A \to B$ and $g : B' \to A'$ are bijective functions. Define

$$l(x) = \begin{cases} f(x), & \text{if } x \in A \\ g^{-1}(x), & \text{if } x \in A' \end{cases}$$

This is a bijective function from X to Y. $\square$

You see from this proof another fact in mathematics: Some theorems that seem to be easily structured can be very hard to prove.
Let's see some applications of this theorem. The very important one is that we can use Schröder-Bernstein Theorem to prove that $\mathbb{R}^n$ has the same cardinality with $\mathbb{R}$.

Example 1.7.4: $\mathbb{R}^n \sim \mathbb{R}$

Let us first consider the interval $(0, 1)$. Let

$$S = \{(x, y) : 0 < x, y < 1\}$$

The function $f : (0, 1) \to S$, where $f(x) = (x, x)$ maps $(0, 1)$ injectively into S, but not surjective. Now we want to find an injective function from S to $(0, 1)$. Recall the decimal representation of real numbers. We construct the map in the following way: Let $(x, y) \in S$, such that

$$x = .a_1a_2a_3\cdots, \quad y = .b_1b_2b_3\cdots$$

Let

$$z = .a_1b_1a_2b_2a_3b_3\cdots$$

We define $g : S \to (0,1)$ by $g(x,y) = z$. This is then an injective function, because the decimal representation is unique for each real number. By Schröder-Bernstein Theorem, there exists a bijective map from $(0,1)$ to S, thus $(0,1) \sim S$.

Now, if we define the function $h : S \to \mathbb{R}^2$ by

$$h(x,y) = \left(\frac{x}{x^2-1}, \frac{y}{y^2-1} \right)$$

We see from Example 1.6.6 that this function is bijective. Therefore, we have

$$\mathbb{R} \sim (0,1) \sim S \sim \mathbb{R}^2$$

Inductively, we can also have

$$\mathbb{R} \sim \mathbb{R}^n$$

This is actually a surprising result. We have $\mathbb{Q}$ is countable and $\mathbb{R}$ is uncountable, since we can think $\mathbb{Q}$ as a set of 'countable points', but $\mathbb{R}$ as a complete real 'line'. This jump from '1-dimensional' to '2-dimensional' space intuitively explained why $\mathbb{R}$ is a much bigger set. However, this intuition does not work for dimensions more than this. Hyper-Eulidean Spaces in all dimensions have the same cardinality! Then, a question would raise naturally: Would there be a set, that is even larger than $\mathbb{R}$? The answer is yes, and will be discussed in the next subsection.

1.7.3* Cantor's Theorem

In the same paper where Cantor published his Diagonalization Method, he also stated the proof of **Cantor's Theorem**, which says that **the power set of a set is strictly 'larger' than the original set**.

For those who don't know what is a power set, the power set of A is the collection of all subsets of A. For example, $A = \{1,2,3\}$, the power set is then $\mathcal{P}(A) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\}$. In finite case, it is easy to see that for a finite set with n elements, the power set of it has 2^n elements. Therefore, in finite case, it is obvious that there is no surjective function from original set to its power set. However, what is surprising, is that this is also true in infinite case.

Theorem 1.7.5: Cantor's Theorem

Given any set A, there does not exist a function $f : A \to \mathcal{P}(A)$ that is surjective.

Proof. Assume for contradiction, that $f : A \to \mathcal{P}(A)$ is surjective. Note that for each element $a \in A$, $f(a)$ is a *subset* of A. (This proof is super tortuous, please follow very carefully!)
Surjective means that for each $y \in \mathcal{P}(A)$, there exists $a \in A$, such that $f(a) = y$. To arrive a contradiction, we will produce a set $B \subseteq A$ such that it is not equal to $f(a)$ for any $a \in A$.

For each $a \in A$, consider $f(a) \subseteq A$. We will conclude all a in a set B such that $f(a)$ does not contain a. In precise,

$$B = \{a \in A : a \notin f(a)\}$$

Now, because f is surjective, there must be some $a' \in A$ such that $f(a') = B$.

- If $a' \in B$, then $a' \notin f(a')$ by the definition of B. Since $f(a') = B$, we have $a' \notin B$, which is a contradiction.
- If $a' \notin B$, then $a' \in f(a')$ by the definition of B. Since $f(a') = B$, we have $a' \in B$, which is a contradiction.

Therefore, we have a contradiction, f cannot be surjective. □

This theorem will directly mean that, the cardinality of $\mathcal{P}(\mathbb{R})$ is larger than the cardinality of $\mathbb{R}$. We find an even larger set than $\mathbb{R}$! Moreover, we will have a full spectrum of cardinalities, such that $\mathcal{P}(\mathcal{P}(\mathbb{R}))$ is larger than $\mathcal{P}(\mathbb{R})$, and $\mathcal{P}(\mathcal{P}(\mathcal{P}(\mathbb{R})))$ is larger than $\mathcal{P}(\mathcal{P}(\mathbb{R}))\cdots$ and there does not exist a 'largest set'. Thus, statement like 'Let U be a set of all possible things' would become a paradox, because we can immediately find a larger set by constructing the power set of U.

To see the most important application of this theorem, let's first prove that $\mathcal{P}(\mathbb{N}) \sim \mathbb{R}$.

Example 1.7.6: $\mathcal{P}(\mathbb{N}) \sim \mathbb{R}$

Define $f : \mathcal{P}(\mathbb{N}) \to S$, where

$$S = \{(a_1, a_2, a_3, \cdots) : a_n = 0 \text{ or } 1\}$$

by $f(A) = (a_1, a_2, a_3, \cdots)$ where $a_i = 0$ if $i \notin A$ and $a_i = 1$ if $i \in A$, for each $i \in \mathbb{N}$. This is a bijective map, thus $\mathcal{P}(\mathbb{N}) \sim S$. Recall Example 1.7.2, we have proved that $S \sim \mathbb{R}$. Therefore, $\mathcal{P}(\mathbb{N}) \sim \mathbb{R}$.

The cardinality of natural numbers and real numbers are so important, so they have a fancy name.

Definition 1.7.7: Aleph Number

1. The cardinality of natural number is called **Aleph 0**, and is denoted as $\aleph_0$.
2. The cardinality of real numbers is called **Aleph 1**, and is dentoed as $\aleph_1$, where $2^{\aleph_0} = \aleph_1$.

The more rigorous definition of aleph numbers would be stated in my Set Theory note. A question would naturally raise: Is there any other aleph numbers between $\aleph_0$ and $\aleph_1$? Cantor published a hypothesis that there is no such aleph number.

Conjecture 1.7.8: Continuum Hypothesis

There does not exist an aleph number c such that

$$\aleph_0 < c < \aleph_1$$

This connected back to the time of **'The Third Mathematical Crisis'**, when Bertrand Russell (1872-1970) stated his famous **'Russell's Paradox'**. It says that

$$\text{Let } R = \{x : x \notin x\}, \text{ then } R \in R \Longleftrightarrow R \notin R$$

A lively example of this is called **Barber Paradox**. The barber is the "one who shaves all those, and those only, who do not shave themselves". The question is, does the barber shave himself? If the barber shaves himself, then himself becomes the group of people who shaves themselves, so he cannot shave himself. If he doesn't, then he must shave himself because he shaves all those who do not shave themselves. A statement cannot be simultaneously right and wrong!

After few decades, there was a system of axiom constructed in set theory, called 'ZFC system', excluded the situation which Russell presented, and ended the Third Mathematical Crisis. Is that the end? No! **Kurt Gödel** (1906-1978) then stated his famous **Incompleteness Theorem**, which said that even in ZFC system, mathematics is not complete, i.e., there exists a theorem that we cannot prove it right or wrong.

What does this story relate to Continuum Hypothesis? Indeed, Continuum Hypothesis is one of the theorem that is 'undecidable', where it can be accepted or rejected without making any logical contradictions. Whether continuum Hypothesis is true or not, will be never known to us. To intuitively explain why this happens, you can go back to see the proof of Cantor's Theorem, which arrives the contradiction in the way like 'If $a \in B$, then $a \notin B$ and if $a \notin B$, then $a \in B$', which is a similar kind of paradox!

1.8* The Dedekind Cuts: Completion from $\mathbb{Q}$ to $\mathbb{R}$

This section is a hard section and can be skipped without losing coherence.

We refer the Supremum as an 'axiom', meaning that there is nothing to be proved. The real numbers were defined simply as an extension of the rational numbers in which bounded sets have supremum. No attempt was made to demonstrate that such extension is possible. Now, in this advanced section, we will actually prove that such extension exist.

1.8.1* Cut

We begin this chapter pretended that we don't know there exists a thing called 'real number', and we assume we know all the familiar addition, multiplication and order rule for rational numbers. The goal is to extend rational numbers into a larger set so that Supremum Property holds.

Definition 1.8.1: Cut

A subset A of the rational numbers is called a **cut** if

1. $A \neq \emptyset$ amd $A \neq \mathbb{Q}$.
2. If $r \in A$, then for all $q \in \mathbb{Q}$ such that $q < r$, we have $q \in A$.
3. A does not have a maximum. i.e., if $r \in A$, then there exists $s \in A$ with $r < s$.

This is the main tool used in constructing real numbers in this chapter. Let's see some examples of cut.

Example 1.8.2: Examples of Cuts

1. Fix $r \in \mathbb{Q}$. The set $C_r = \{t \in \mathbb{Q} : t < r\}$ is a cut.
2. $T = \{t \in \mathbb{Q} : t^2 < 2 \text{ or } t < 0\}$ is a cut.
3. $U = \{t \in \mathbb{Q} : t^2 \leqslant 2 \text{ or } t < 0\}$ is a cut. The proof of statement 3 in the definition of cut for U can follow the pattern in proof of Proposition 1.4.5.
4. **Counterexample:** $S = \{t \in \mathbb{Q} : t \leqslant 2\}$ is not a cut since it has maximum 2.

All the verification of 3 properties of cut above are easy, and thus are omitted here. Now we define our goal: The set of real numbers.

Definition 1.8.3: Real Number in Dedekind Cut Sense

The real numbers $\mathbb{R}$ is the set of all cuts in $\mathbb{Q}$.

You may ask: what? we define real numbers as sets! This looks very weird at first. The most intuitive (but heuristic) explanation to this weird definition is that, we can construct a bijection from each cut to each real number such that for a cut A, it is mapped to a real number on the 'cut point'. For example, for the cut T above, it is mapped to the real number $\sqrt{2}$. (**Note:** Since we assume at first we don't know real numbers, including $\sqrt{2}$, this explanation is just heuristic one.) This bijection is called a **isomorphism** if and only if the algebraic structure on $\mathbb{R}$ we defined above is the same as the set of real numbers in our common sense. Then the two sets are called **isomorphic**. If two sets are isomorphic, they are essentially the same, with just different notations.
Now we discuss exactly which algebraic structures $\mathbb{R}$ should obtain. Before that, we need to know what is a 'structure'.

1.8.2* Field and Ordering

Definition 1.8.4: (Binary) Operation

Given a set F, an **operation** on F is a function $f : F \times F \to F$.

For example, the 'addition' operation on rational numbers takes $(2, 3) \in \mathbb{Q} \times \mathbb{Q}$ to the element $5 \in \mathbb{Q}$. The 'multiplication' operation on rational numbers takes $(2, 3) \in \mathbb{Q} \times \mathbb{Q}$ to the element $6 \in \mathbb{Q}$. With this, we can define what is a field.

Definition 1.8.5: Field

A triple $(F, +, \times)$, where F is a set and $+, \times$ are two arbitrary operations, is a **field** if

- Commutativity: $x + y = y + x$ and $x \times y = y \times x$, $\forall\, x, y \in F$.
- Associativity: $(x + y) + z = x + (y + z)$ and $(x \times y) \times z = x \times (y \times z)$, $\forall\, x, y, z \in F$.
- Identity: There exists $0 \in F$ and $1 \in F$ such that $x + 0 = x$ and $x \times 1 = x$ for all $x \in F$.
- Inverse: Given $x \in F$, there exists $-x \in F$ such that $x + (-x) = 0$. If $x \neq 0$, there exists an element $x^{-1} \in F$ such that $x \times x^{-1} = 1$.
- Distributive Property: $x \times (y + z) = x \times y + x \times z$, $\forall\, x, y, z \in F$.

Note: If you are the first time encountering these definitions, note that the $+$ and $\times$ notation need not to represent addition and multiplication. As long as there are two operations on the set that satisfies these 5 properties correspondingly, it is a field. For example, sometimes in functional spaces, the function composition $\circ$ would take the position of $\times$.

Example 1.8.6: Examples of Fields

$\mathbb{Q}$ is a field, $\mathbb{N}$ and $\mathbb{Z}$ are not field, as you can verify.

Definition 1.8.7: (Binary) Relation

A **relation** on F is a subset of $F \times F$.

This definition is very abstract. However, the following definition would give a strict example of this.

Definition 1.8.8: Ordering/Ordered Field

An **Ordering** on a set F is a relation, represented by $\leqslant$, with properties

- At least one of the $x \leqslant y$ and $y \leqslant x$ is true, $\forall\, x, y \in F$.
- If $x \leqslant y$ and $y \leqslant x$, then $x = y$, $\forall\, x, y \in F$.
- If $x \leqslant y$ and $y \leqslant z$, then $x \leqslant z$, $\forall\, x, y, z \in F$.

A field F is called an **ordered field** if F is endowed with an ordering such that

- If $y \leqslant z$, then $x + y \leqslant x + z$, $\forall\, x, y, z \in F$.
- If $0 \leqslant x$, $0 \leqslant y$, then $0 \leqslant x \times y$, $\forall\, x, y \in F$.

Since $\mathbb{Q}$ is a field, and has an ordering on it, our goal now is to construct addition, multiplication, and ordering on $\mathbb{R}$, such that it is ordered, and it is a field.

1.8.3* Algebra on $\mathbb{R}$

We first define an ordering on $\mathbb{R}$.

Definition 1.8.9: Ordering on $\mathbb{R}$

Let $A, B \in \mathbb{R}$ be two cuts. Define $A \leqslant B$ to mean $A \subseteq B$.

Proof. Now we need to prove this definition satisfies the 3 properties of an ordering.

1. Suppose $A \nsubseteq B$. We need to prove that $B \subseteq A$. If $A \nsubseteq B$, there exists $a \in A$ such that $a \notin B$. This means that $\forall\, b \in B, b < a$. Then, by the second property in definition of cut, we have $\forall\, b \in B$, $b \in A$. This means that $B \subseteq A$.
2. If $A \subseteq B$ and $B \subseteq A$, then $A = B$ by definition of equality of set.
3. If $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$ by the property of set.

Therefore, this definition satisfy the properties of ordering. □

Further, we define an addition on $\mathbb{R}$.

Definition 1.8.10: Addition on $\mathbb{R}$

Let $A, B \in \mathbb{R}$ be cuts. Define

$$A + B = \{a + b : a \in A,\, b \in B\}$$

We first need to prove that $A + B \in \mathbb{R}$, i.e., $A + B$ is indeed a cut.

Proof. We need to go through the three properties of cuts.

- Since A, B are cuts, by property 1 of cut, $A, B \neq \emptyset$. Therefore, we can find $a \in A$ and $b \in B$, then $a + b \in A + B$. Thus $A + B \neq \emptyset$.
  Since A, B are cuts, by property 1 of cut, $A, B \neq \mathbb{Q}$. Therefore, we can find $a' \notin A$ and $b' \notin B$, then for all $a \in A$ and $b \in B$, we have $a < a'$, $b < b'$. Therefore, for all $a + b \in A + B$, we will have $a + b < a' + b'$. Thus $a' + b' \notin A + B$. We have $A + B \neq \mathbb{Q}$.

- To prove property 2 of cut, let $a + b \in A + B$ be arbitrary and let $s \in \mathbb{Q}$ satisfy $s < a + b$. Then, $s - b < a$, which implies that $s - b \in A$ because A is a cut. Then,

  $$s = (s - b) + b \in A + B$$

  Since s is arbitrary, we have for every $s \in \mathbb{Q}$ that $s < a + b$, we also have $s \in A + B$.

- To prove property 3, since A, B are cuts, for $a \in A, b \in B$, we can find $a' \in A, b' \in B$ such that $a < a', b < b'$. Then, for each $a + b \in A + B$, we can find $a' + b' \in A$ such that $a + b < a' + b'$.

Therefore, $A + B$ is indeed a cut. □

Then, we need to prove that this definition satisfies the related properties in the field, and also the ordering field property 1.

Proof.

- We first prove the **Commutativity**. Obviously,

  $$A + B = \{a + b : a \in A,\, b \in B\} = \{b + a : a \in A,\, b \in B\} = B + A$$

- Then, we prove the **Associativity**. Obviously,

$$(A+B)+C = \{(a+b)+c : a \in A,\ b \in B, c \in C\} = \{a+(b+c) : a \in A,\ b \in B, c \in C\} = A+(B+C)$$

- Next, we prove the existence of **Identity**, define

$$O = \{p \in \mathbb{Q} : p < 0\}$$

We want to prove that this is served as the identity.
($\Longrightarrow$) Let $a + o \in A + O$ where $a \in A$ and $o \in O$. Then $o < 0$. Therefore, $a + o < a$. By property of cut, $a + o \in A$. Thus $A + O \subseteq A$.
($\Longleftarrow$) Let $a \in A$. Then by property of cut, we can find $a < a' \in A$. Define $s = a - a' < 0$. We have $s \in O$. Then, $a = s + a' \in A + O$. Thus, $A \subseteq A + O$.
In conlcusion, we have $A = A + O$, proving that O is the identity.

- Now, we prove the existance of **Inverse**. This is a little bit more difficult than the identity, since the normal definiton of $-A$ would not be a cut. We alternatively, define

$$-A = \{r \in \mathbb{Q} : \text{ there exists } t \notin A \text{ with } t < -r\}$$

This is just like a 'reflection' of the 'cut point' with respect to the origin.

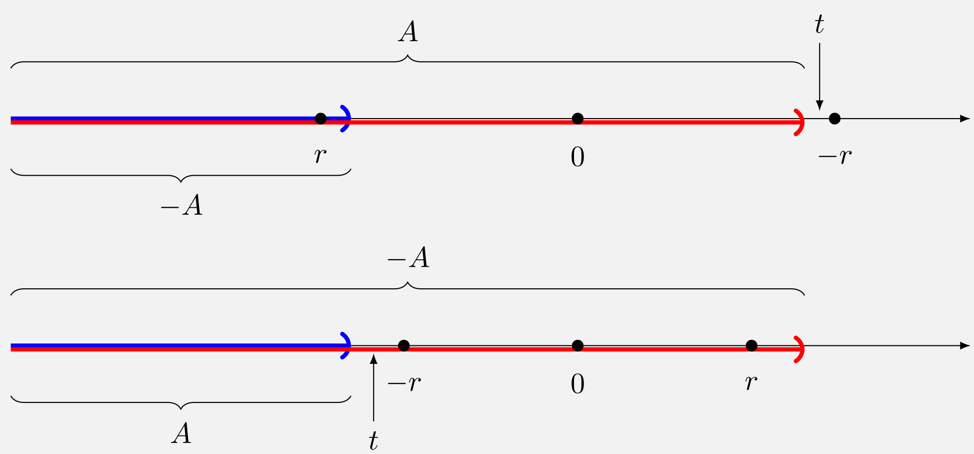


Figure 1.12: A and its inverse

We first need to prove that $-A$ is indeed a cut.

1. To prove the first property of cut, since A is a cut, we can find $t \notin A$, and since $\mathbb{Q}$ is unbounded, we can find some $-r \in \mathbb{Q}$ such that $t < -r$. Thus, $-A \neq \emptyset$.

Let $a \in A$. Then for all $t \notin A$, we have $t > a$. Thus $-a \notin -A$ since if it is, then $t < -(-a) = a$, which is a contradiction. Therefore, $-A \neq \mathbb{Q}$.

2. To prove the second property of cut, let $r \in -A$ and let $q < r$. Then, $t < -r < -q$ for the $t \notin A$ such that $t < -r$. Hence, by definition of $-A$, $q \in -A$.
3. To prove the third property of cut, let $r \in -A$ and let $t \notin A$ such that $t < -r$. Let $q = \frac{t-r}{2}$, we have $t < q < -r$. Thus, $-q \in -A$ and $-q > r$. There is no maximum.

After this, we need to prove that this indeed defines an inverse.
($\Longrightarrow$) If $a \in A$ and $r \in -A$, then there exists $t \notin A$ with $t < -r$. Since $t \notin A$, we have $t > a$. Then, $a + r < a - t < 0$, hence $a + r \in O$. This shows that $A + (-A) \subseteq O$.
($\Longleftarrow$) Now, let $o \in O$. We would like to find $a \in A$ and $r \in -A$ satisfying $o \leqslant a + r$. This would imply $O \subseteq A + (-A)$. Since $o \in \mathbb{Q}$, and $o < 0$, let $o = -p/q$ where $p, q \in \mathbb{N}, q \neq 0$. We first prove the lemma:
For any cut A and $n \in \mathbb{N}$, we can find $z \in A$ where

$$\frac{z}{n} \in A \quad \text{and} \quad \frac{z+1}{n} \notin A$$

To do this, start with $a \in A$ and $a' \notin A$. Find $N, M \in \mathbb{Z}$ such that

$$\frac{N}{n} < a \quad \text{and} \quad \frac{M}{n} > a'$$

Clearly $N/n \in A$ and $M/n \notin A$. Therefore, there must exist a transition point z such that the lemma holds.
Now if we let $a = \frac{n}{2q} \in A$ and $\frac{n+1}{2q} \notin A$, then $r = -\frac{n+2}{2q} \notin A$, and

$$a + r = -\frac{1}{q} \geqslant -\frac{p}{q} = o$$

Therefore, $O \subseteq A + (-A)$. In conclusion, $A + (-A) = O$.

- Finally, we prove the **first property of ordering field**.

Let $Y \subseteq Z$. Let $x \in X$ and $y \in Y$. Since $Y \subseteq Z$, we have $y \in Z$ either. This implies $x + y \in X + Z$. Thus, $X + Y \subseteq X + Z$. □

Note: We cannot simply define the inverse as

$$-A = \{r \in \mathbb{Q} : -r \notin A\}$$

which would be the first thought of most people. The counterexample is that for $A = \{t \in \mathbb{Q} : t < -2\}$,we will have $-A = \{t \in \mathbb{Q} : t \leqslant 2\}$, which is not a cut. Therefore, we need the additional 't' in the definiton of inverse to avoid this situation.

Now, we define a multiplication in $\mathbb{R}$. This is also difficult because negative set times negative set would be positive. Therefore, we first consider the case of $A \geqslant O$ and $B \geqslant O$.

Definition 1.8.11: Multiplication on $\mathbb{R}$, positive case

Given $A \geqslant O$ and $B \geqslant O$, define the product

$$AB = \{ab : a \in A, b \in B \text{ with } a, b \geqslant 0\} \cup O$$

Similarly, we need to first prove that the definition indeed results in a cut.

Proof.

- choose $-1 \in O$. Then, $-1 \in AB$, we have $AB \neq \emptyset$. Since A, B are cuts, choose $a' \notin A$,$b' \notin B$, then for all $a \in A, b \in B$, we have $a' > a, b' > b$. Also, $a', b' > 0$ since otherwise it will be in O. Thus, for all $r \in AB$, $a'b' > r$. This indicates that $AB \neq \mathbb{Q}$.
- Suppose $r \in AB$, let $q < r$. If $r < 0$ then $q < r < 0$, we have $q \in AB$. If $r > 0$, $r = ab$ for some $a \in A$ and $b \in B$ with $a, b \geqslant 0$. If $q < 0$, then obviously $q \in AB$. If $q > 0$, we have $\frac{q}{b} < \frac{r}{b} = a$, thus $\frac{q}{b} \in A$. This would indicate that $q = \frac{q}{b}b \in AB$.
- Let $r \in AB$. If $r < 0$ then obviously $\frac{r}{2} \in AB$ and $r < \frac{r}{2} \in AB$. If $r > 0$ then, $r = ab$ for some $a \in A, b \in B, a, b, \geqslant 0$. Sonce A, B are cuts, we could find $a' \in A$ and $b' \in B$ with $a' > a$ and $b' > b$. Then $a'b' \in AB$ and $a'b' > ab$. There is no maximum.

Therefore, AB is indeed a cut. $\square$

After defining this, we can accordingly define the negative cut cases such that

$$AB = \begin{cases} -[A(-B)], & \text{if } A \geqslant O \text{ and } B < O \\ -[(-A)B], & \text{if } A < O \text{ and } B \geqslant O \\ (-A)(-B), & \text{if } A < O \text{ and } B < O \end{cases}$$

We can accordingly, prove that these are cuts, and prove the commutativity, associativity, distributive property, and the second property of the ordering field. However, we will not do it here since it will be way tedious. The proving pattern is just like that for addition. Just note that the multiplicative identity is

$$I = \{t \in \mathbb{Q} : t < 1\}$$

and the multiplicative inverse for the positive cut A (defined in Definition 1.8.11) is defined as

$$A^{-1} = \left\{a \in \mathbb{Q} : \text{ there exists } t \notin A \text{ with } t < \frac{1}{a}\right\} \cup \{0\} \cup O$$

1.8.4* Rediscover Supremum Property

After proving that $\mathbb{R}$ satisfies all the **Ordered Field** Property, we finally, prove the **Supremum Property**, which we see as an axiom throughtout the first chapter.

Note that now, since we define real numbers as cuts, ‘a set of real numbers’ would be a collection of cuts (set of sets). For these collections we will denote them using the calligraphy font, such as $\mathcal{A}$. Obviously, for a set $\mathcal{A}$ which is nonempty and bounded above, the desired supremum would be the union of all cuts $A \in \mathcal{A}$

$$S = \bigcup_{A \in \mathcal{A}} A$$

Now we prove that S is indeed a cut.

Proof.

- Let $A \in \mathcal{A}$, and $a \in A$. Then $a \in S$. Thus $S \neq \emptyset$. Now let B be a bound on $\mathcal{A}$, let $b' \notin B$. Since for any $s \in S$, we will have $s \in A$ for some $A \in \mathcal{A}$, and $A \leqslant B$ for all A, we have $s \in B$. This will indicate that $b' > s$. Thus, $b' \notin S$, $S \neq \mathbb{Q}$.
- Consider arbitrary $s \in S$ with $s \in A \in \mathcal{A}$. Then, for any $q < s$ we have $q \in A$, thus $q \in S$.
- Continuing the proof of property 2, we can also find a $a \in A$ such that $s < a$. There is no maximum.

Therefore, S is indeed a cut. $\square$

Finally, we prove that S is the supremum for $\mathcal{A}$.

Proof. First, S is indeed an upper bound since for all $a \in A$, we have $a \in S$. Second, let B be an upper bound for $\mathcal{A}$. Now for any $s \in S$ with $s \in A \in \mathcal{A}$, we have $A \leqslant B$, so $s \in B$. Therefore, $S \leqslant B$, S is indeed a supremum! $\square$

We are not finished yet. Since this construction of $\mathbb{R}$ as a set of cuts, is a completely different notation from that of rational numbers. We are doing an extension, meaning that $\mathbb{Q}$ should be a subfield of $\mathbb{R}$.

However, note that if we write all rational numbers as 'rational cuts'

$$Q = \{t \in \mathbb{Q} : t < r, r \in \mathbb{Q}\}$$

This would be an **isomorphism** from rational numbers to rational cuts, and indeed, we can easily verify that the set of all rational cuts is an ordered field (just re-prove all the ordering field properties before, using the sets of all rational cuts).

In conclusion, our result is:

Theorem 1.8.12: Extension of $\mathbb{Q}$ to $\mathbb{R}$

There exists an ordered field in which every nonempty set that is bounded above has a supremum. In addition, this field contains $\mathbb{Q}$ as a subfield.

Chapter 2

Infinite Sequences and Series of Real Numbers

2.1 Limit of a Sequence

The first important mathematical object we need to analysis in this chapter is **sequences**.

Definition 2.1.1: Sequence

A **sequence** is a function whose domain is $\mathbb{N}$.

Considering the definition, we can reasonably write a sequence in the form of $(a_1, a_2, a_3, \cdots)$ where a_i is the element that $i \in \mathbb{N}$ maps to. We always denote it as $(a_n)_{n\in\mathbb{N}}$, or simply, (a_n).

Note: Sometimes sequences will start not with x_1, but with x_{n_0} where $n_0 > 1$, $n_0 \in \mathbb{N}$. This does not matter much, because we are only interested in how the sequence behaves at the infinite 'tail', i.e., the **limit**.

Definition 2.1.2: Convergence of a Sequence

A sequence (a_n) **converges** to a real number a if,

$$\forall\, \epsilon > 0, \exists\, N \in \mathbb{N}, \text{ s.t. } n > N \Longrightarrow |a_n - a| < \epsilon$$

We denote this as $\lim_{n\to\infty} a_n = a$.

This is **the** most important mathematical languange in analysis, called ϵ-δ language, or ϵ-N in this case. To fully understand the meaning of this definition, we first fix an ϵ, and by the definition, there exists a point, where all of the terms in the sequence after this point should be in the ϵ-range centered at a.

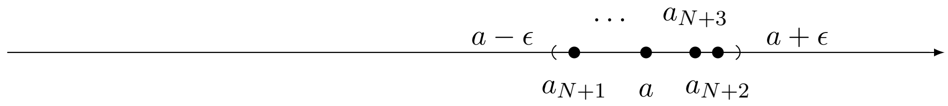


Figure 2.1: Definition of convergence

The critical point is that, we can choose all ϵ, regardless how small it is, we can always find N such that, all the points after are 'approaching' the limit a, and they can never jump out this ϵ range.

Definition 2.1.3: Divergence of a Sequence

A sequence that does not converge is said to **diverge**.

Note: We cannot identify divergence by the negation of the definition of convergence, i.e.,

$$\exists\,\epsilon > 0, \text{ s.t. } \forall\,N \in \mathbb{N}, \text{ s.t. } \exists\,n > N \Longrightarrow |a_n - a| \geqslant \epsilon$$

since by this statement, the sequence may converge, just not converges to the point a.

We can easily see that the limit of a sequence must be unique.

Proposition 2.1.4: Uniqueness of Limit

The limit of a sequence is unique.

Proof. Suppose a, a' are limits of a sequence (a_n). Then, by definition,

$$\text{Fix } \epsilon > 0, \exists\,N_1 \in \mathbb{N}, \text{ s.t. } n > N_1 \Longrightarrow |a_n - a| < \frac{\epsilon}{2}$$

$$\text{Fix } \epsilon > 0, \exists\,N_2 \in \mathbb{N}, \text{ s.t. } n > N_2 \Longrightarrow |a_n - a'| < \frac{\epsilon}{2}$$

Therefore, for current fixed ϵ and $n > \max\{N_1, N_2\}$, we have

$$|a - a'| = |a - a_n + a_n - a'| \leqslant |a - a_n| + |a_n - a'| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

This is true for arbitrary ϵ. Therefore, $a = a'$. $\square$

The proof of convergence using definition can generally follow these steps:

- 'Let $\epsilon > 0$ be arbitrary'.
- Demonstrate a choice for $N \in \mathbb{N}$. This step usually requires some work on draft paper, to see which N is suitable. Note that N may (and commonly will) depend on ϵ.

- Assume $n > N$, show that $|a_n - a| < \epsilon$.

Example 2.1.5: Prove convergence using definition

Prove that
$$\lim \frac{2n^2}{n^3+3} = 0$$

Things that will appear on your draft paper:

$$\left|\frac{2n^2}{n^3+3} - 0\right| = \frac{2n^2}{n^3+3} < \frac{2n^2}{n^3} = \frac{2}{n} < \epsilon$$

it seems that $N = \frac{2}{\epsilon}$ would be a good choice.

Things that will appear on your answer sheet:

Proof. Let ϵ be arbitrary. Let $N = \frac{2}{\epsilon}$. Assume $n > N$. Then,

$$\left|\frac{2n^2}{n^3+3} - 0\right| = \frac{2n^2}{n^3+3} < \frac{2n^2}{n^3} = \frac{2}{n} < \frac{2}{2}\epsilon = \epsilon$$

Therefore, by definition, the sequence converges to 0. □

2.2 Properties of Limit

We first see that every convergent sequence are bounded. To rigorously state this, we need to define what is 'bounded'.

Definition 2.2.1: Bounded Sequence

A sequence (x_n) is bounded if there exists $M > 0$ such that $|x_n| \leqslant M$ for all $n \in \mathbb{N}$.

Proposition 2.2.2: Boundedness of Convergence Sequence

Every convergent sequence is bounded.

Proof. Suppose (x_n) converges to l. Then, fix $\epsilon = 1$, we have

$$\exists N \in \mathbb{N}, \text{ s.t. } n \geqslant N \implies |x_n - l| < 1$$

This means that for all $n \geqslant N$
$$|x_n| \leqslant |x_n - l| + |l| < |l| + 1$$

For the terms before N, since there are only finite terms, there must be a maximum. Let

$$M = \max\{|x_1|, |x_2|, \cdots, |x_{N-1}|, |l| + 1\}$$

We can conclude that $|x_n| \leqslant M$ for all $n \in \mathbb{N}$. □

Next, we state **Algebraic Limit Theorem**, which shows that limit of sequences behave very well under addition, multiplication and division.

Proposition 2.2.3: Algebraic Limit Theorem

Let $\lim a_n = a$ and $\lim b_n = b$. Then,

1. $\lim(ca_n) = ca, \forall c \in \mathbb{R}$
2. $\lim(a_n + b_n) = a + b$
3. $\lim(a_n b_n) = ab$
4. $\lim(a_n/b_n) = a/b$, provided $b \neq 0$

Proof.

1. When $c = 0$, it is trivial. So suppose $c \neq 0$. Since $\lim a_n = a$, we have

$$\forall \epsilon > 0, \exists N \in \mathbb{N}, \text{ s.t. } n > N \Longrightarrow |a_n - a| < \frac{\epsilon}{|c|}$$

Then we have for $n > N$,

$$|ca_n - ca| = |c||a_n - a| < |c|\frac{\epsilon}{|c|} = \epsilon$$

which shows our desired result.

2. Since $\lim a_n = a$ and $\lim b_n = b$, we have

$$\forall \epsilon > 0, \exists N_1 \in \mathbb{N}, \text{ s.t. } n > N_1 \Longrightarrow |a_n - a| < \frac{\epsilon}{2}$$

$$\forall \epsilon > 0, \exists N_2 \in \mathbb{N}, \text{ s.t. } n > N_2 \Longrightarrow |b_n - b| < \frac{\epsilon}{2}$$

Let $N = \max\{N_1, N_2\}$. Then, for $n > N$, we will have

$$|(a_n + b_n) - (a + b)| \leqslant |a_n - a| + |b_n - b| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

3. This is a little bit harder. The goal is to find an N such that for all $n > N$ we have $|a_n b_n - ab| < \epsilon$.

Note that

$$|a_nb_n - ab| = |a_nb_n - ab_n + ab_n - ab| \leqslant |a_nb_n - ab_n| + |ab_n - ab| = |b_n||a_n - a| + |a||b_n - b|$$

Since b_n converges, by Proposition 2.2.2, it is bounded. Therefore, $|b_n| \leqslant M$ for some M and for all $n \in \mathbb{N}$. Then if we choose N_1 and N_2 such that

$$\forall \epsilon > 0, \exists N_1 \in \mathbb{N}, \text{ s.t. } n > N_1 \Longrightarrow |a_n - a| < \frac{\epsilon}{2M}$$

$$\forall \epsilon > 0, \exists N_2 \in \mathbb{N}, \text{ s.t. } n > N_2 \Longrightarrow |b_n - b| < \frac{\epsilon}{2|a|}$$

and let $N = \max\{N_1, N_2\}$. Then for all $n > N$, we have

$$|a_nb_n - ab| \leqslant |b_n||a_n - a| + |a||b_n - b| < M\frac{\epsilon}{2M} + |a|\frac{\epsilon}{2|a|} = \epsilon$$

4. We can prove this statement if only if we can prove

$$(b_n) \longrightarrow b \quad \text{implies} \quad \left(\frac{1}{b_n}\right) \longrightarrow \frac{1}{b}$$

since we can then get the desired result from (3). The goal is to find N such that for all $n > N$ we have $\left|\frac{1}{b_n} - \frac{1}{b}\right| < \epsilon$. Note that

$$\left|\frac{1}{b_n} - \frac{1}{b}\right| = \frac{|b - b_n|}{|b||b_n|}$$

We know $|b - b_n|$, so we need to control the size of $\frac{1}{|b||b_n|}$. **This is a very important trick**. Now we are not concerning about the upper bound of b_n, we are concerning the lower bound. The trick is to use the convergence relation between (b_n) and b to construct desired inequality. Since $(b_n) \to b$, we can fix $\epsilon = \frac{|b|}{2}$, then

$$\forall \epsilon > 0, \exists N_1 \in \mathbb{N}, \text{ s.t. } n > N_1 \Longrightarrow |b_n - b| < \frac{|b|}{2}$$

Further simplify this relationship, we have for all $n > N_1$,

$$|b_n| = |(b_n - b) + b| > ||b_n - b| - |b|| = |b| - |b_n - b| > \frac{|b|}{2}$$

where the first inequality is by inverse triangular inequality. Therefore, if we choose N_2 such that

$$\forall \epsilon > 0, \exists N_2 \in \mathbb{N}, \text{ s.t. } n > N_2 \Longrightarrow |b_n - b| < \frac{\epsilon|b|^2}{2}$$

and let $N = \max\{N_1, N_2\}$. Then for all $n > N$ we have

$$\left|\frac{1}{b_n} - \frac{1}{b}\right| = \frac{|b - b_n|}{|b||b_n|} < \frac{\epsilon |b|^2}{2} \frac{1}{|b|\frac{|b|}{2}} = \epsilon$$

This shows the desired result. □

Next, we show **Order Limit Theorem**, which shows that the limit preserves the order of two related elements.

Proposition 2.2.4: Order Limit Theorem

Let $\lim a_n = a$ and $\lim b_n = b$. Then,

1. If $a_n \leqslant b_n$ for all $n \in \mathbb{N}$, then $a \leqslant b$
2. If $c \in \mathbb{R}$, and $a_n \leqslant c$ for all $n \in \mathbb{N}$, then $a \leqslant c$. Similarly, if $b_n \geqslant c$ for all $n \in \mathbb{N}$, then $b \geqslant c$

Proof.

1. For every $\epsilon > 0$ we can find $N \in \mathbb{N}$ such that for all $n > N$, we have

$$\forall \epsilon > 0, \exists N \in \mathbb{N}, \text{ s.t. } n > N \Longrightarrow a_n - a > -\epsilon$$

$$\forall \epsilon > 0, \exists N \in \mathbb{N}, \text{ s.t. } n > N \Longrightarrow b_n - b < \epsilon$$

Therefore, we can get

$$a - b \leqslant a - b + (b_n - a_n) = (b_n - b) - (a_n - a) < \epsilon - (-\epsilon) = 2\epsilon$$

where the first inequality hold because $a_n \leqslant b_n$. Therefore,

$$b > a - 2\epsilon$$

Since this holds for all $\epsilon > 0$, we have $b \geqslant a$.

2. Take $a_n = c$ or $b_n = c$, we can prove the second argument.

□

The **most useful** corollary of Proposition 2.2.4 is the famous **Squeeze Theorem**.

Corollary 2.2.5: Squeeze Theorem

If $x_n \leqslant y_n \leqslant z_n$ for all $n \in \mathbb{N}$, and $\lim x_n = \lim z_n = l$, then $\lim y_n = l$.

2.3 Completeness and Convergence

2.3.1 Axiom of Completeness III: The Monotone Convergence Theorem

Here we consider the third form of Axiom of Completeness of Real Number: Monotone Convergence Theorem. To state this, we first define what is a monotone sequence.

Definition 2.3.1: Monotone Sequence

A sequence (a_n) is **increasing** if $a_n \leqslant a_{n+1}$ for all $n \in \mathbb{N}$ and **decreasing** if $a_n \geqslant a_{n+1}$ for all $n \in \mathbb{N}$. A sequence is **monotone** if it is increasing or decreasing.

Now we state the theorem.

Theorem 2.3.2: Monotone Convergence Theorem (MCT)

If a sequence is monotone and bounded, then it converges. Specifically, if it is increasing, then it converges to the supremum of elements. If it is decreasing, then it converges to the infimum of elements.

Proof. Let (a_n) be monotone and bounded. Assume (a_n) is increasing, and the decreasing case can be handled similarly. We let

$$s = \sup\{a_n : n \in \mathbb{N}\}$$

We will then prove that $\lim a_n = s$. Let $\epsilon > 0$. Because s is the least upper bound, $s - \epsilon$ then, is not the upper bound. Then, there exists a point s_N in the sequence such that $s - \epsilon < a_N$. Since a_n is increasing, we have if $n \geqslant N$, we have $a_N \leqslant a_n$. Hence,

$$s - \epsilon < a_N < a_n \leqslant s < s + \epsilon$$

for all $n > N$, as desired. □

Actually, we could have used the MCT in place of Supremum Property as our starting axiom for building a proper theory of real numbers. Intuitively, for a nonempty set which is bounded above, there must exist a increasing bounded sequence in it, and it converges by MCT. The limit is then the supremum. One of the proof is stated below.

Proof. The idea is that, we prove Archimedean Property and Nested Interval Property using MCT without the use of Supremum Property, so that we can go from MCT to Nested Interval Property, finally to Supremum Property.
Since the sequence $(1/n)$ is monotone and bounded, by MCT, it converges. It can then only converge

to 0 by Order Limit Theorem. This shows that we can find n such that $|1/n - 0| \leqslant \epsilon$ for any ϵ, which is the Archimedean Property.

Consider an arbitrary collection of nested intervals $\{I_n\}_n$. For $I_n = [a_n, b_n]$, we have $\{a_n : n \in \mathbb{N}\}$ is bounded above by b_1, bounded below by a_1, and it is monotone. Similarly, $\{b_n : n \in \mathbb{N}\}$ is bounded below by a_1, bounded above by b_1, and it is monotone. Therefore, by MCT, both (a_n) and (b_n) converges, say $a_n \to a$ and $b_n \to b$. By order limit theorem, $a \leqslant b$ since for all n, $a_n \leqslant b_n$. Therefore, $a \in I_n$ for all n, and thus $a \in \bigcap_{i=1}^{\infty} I_n$ and thus $\bigcap_{i=1}^{\infty} I_n \neq \emptyset$. □

Below we extend the Relation graph 1.8.

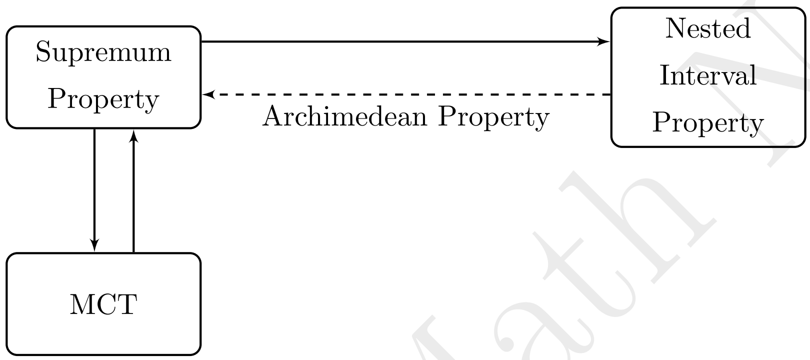


Figure 2.2: Relation between Axiom of Completeness

2.3.2 Axiom of Completeness IV: Bolzano-Weierstrass Theorem

A very important terminology in analysis is **subsequence**. A sequence can be divergent with some of its subsequence converges.

Definition 2.3.3: Subsequence

Let (a_n) be a sequence of real numbers, and let $n_1 < n_2 < n_3 < \cdots$ be an increasing sequence of natural numbers. Then the sequence

$$(a_{n_1}, a_{n_2}, a_{n_3}, \cdots)$$

is called a **subsequence** of (a_n), denoted by (a_{n_k}), where $k \in \mathbb{N}$.

Obviously, from intuition, if a sequence converges, then its subsequences also converge.

Proposition 2.3.4: Convergence of Subsequence

Subsequences of a convergent sequence converge to the same limit as the original sequence.

Proof. Assume $(a_n) \to a$, let (a_{n_k}) be a subsequence. Given $\epsilon > 0$, there exists N such that $|a_n - a| < \epsilon$ whenever $n \geqslant N$. Because $n_k \geqslant k$ for all k, the same N will suffice for the subsequence. □

Note that not all sequences contain a convergent subsequence. Consider $(a_n) = (1, 2, 3, 4, \cdots)$, there is no subsequence contained in it. However, for bounded sequence, the situation is changed.

Theorem 2.3.5: Bolzano-Weierstrass Theorem

Every bounded sequence contains a convergent subsequence.

Proof. Let (a_n) be a bounded sequence so that there exsits $M > 0$ such that $|a_n| \leqslant M$ for all $n \in \mathbb{N}$. Bisect the closed interval $[-M, M]$ into two parts $[-M, 0]$ and $[0, M]$. Now, it must be that at least one of these closed intervals contains an infinite number of the terms in the sequence (a_n). Select the half for which this is the case and label that interval as I_1. Then, let a_{n_1} be some term in the sequence (a_n) satisfying $a_{n_1} \in I_1$.

Now, bisect I_1 in the same way, choose I_2 as the interval with infinite terms in it. Since there are infinite terms, we can choose an a_{n_2} from the original sequence such that $n_2 > n_1$ and $a_{n_2} \in I_2 \cdots$. Continuing this fashion, we can form a nested interval $\{I_n\}$ and select $n_1 < n_2 < n_3 < \cdots$ so that each $a_{n_k} \in I_k$.

Now we argue that (a_{n_k}) is convergent. We first need to choose the candidate for the limit. By Nested Interval Property, there exsits at least one $x \in \bigcap_{k=1}^{\infty} I_k$. We choose this x as the candidate.

Let $\epsilon > 0$. By construction, the length of I_k is $M(1/2)^{k-1}$, which converges to 0, as you can verify. Choose N so that $k \geqslant N$ implies that the length of I_k is less than ϵ. Because a_{n_k} and x are both in I_k, it follows that $|a_{n_k} - x| < \epsilon$. □

As you can see, we again use Archimedean Property (to show that $M(1/2)^{k-1}$ converges to 0) and Supremum Property in the proof. This suggests that it can be seen as another Axiom of Completeness of real numbers. We do not show the proof of Bolzano-Weirestrass to Supremum Property here, since showing each connection within axioms would be extremely lengthy. Below we extend the relation graph again.

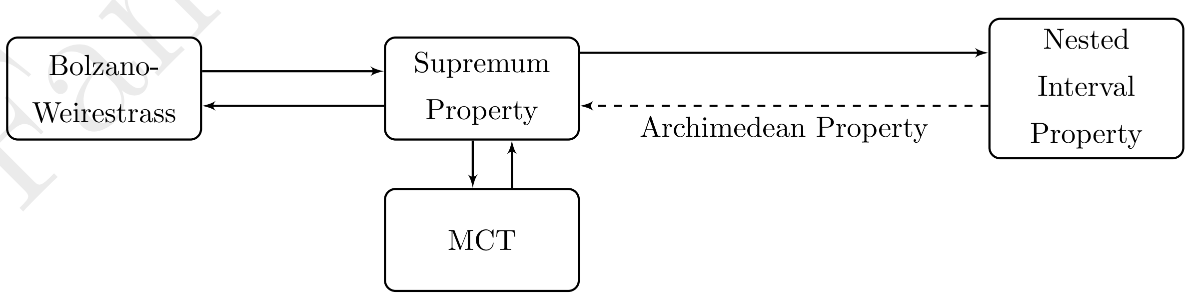


Figure 2.3: Relation between Axiom of Completeness

2.3.3 Axiom of Completeness V: Cauchy Criterion

When we consider the convergence of a sequence, we must first guess what is the limit first. However, there is a way to state the convergence without having any explicit knowledge of what the limit might be.

Definition 2.3.6: Cauchy Sequence

A sequence (a_n) is called a **Cauchy Sequence** if

$$\forall\, \epsilon > 0,\, \exists\, N \in \mathbb{N},\ \text{s.t. } m, n > N \quad \Longrightarrow \quad |a_n - a_m| < \epsilon$$

The definition just said that, when indices becomes larger, the distance between terms in sequence are getting closer and closer. It is clear that convergent sequences also have this property.

Proposition 2.3.7: Convergent Sequences are Cauchy

Every convergent sequence is a Cauchy sequence.

Proof. Suppose (a_n) is a convergent sequence. Since it is convergent, we have

$$\forall\, \epsilon > 0, \exists\, N \in \mathbb{N},\ \text{s.t. } m, n > N \Longrightarrow |a_n - a| < \frac{\epsilon}{2} \quad \text{and} \quad |a_m - a| < \frac{\epsilon}{2}$$

Then, by triangular inequality, we have

$$|a_n - a_m| = |a_n - a + a - a_m| \leqslant |a_n - a| + |a_m - a| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

This shows that it is a Cauchy sequence. □

Is the inverse also true? The answer is yes. Before proving that, we first need a lemma.

Lemma 2.3.8: Boundedness of Cauchy sequence

Cauchy sequences are bounded.

Proof. Let $\epsilon = 1$. there exists $N \in \mathbb{N}$ such that $|x_m - x_n| < 1$ for all $m, n > N$. Therefore, let $m = N + 1$, we must have $|x_n| = |x_n - x_{N+1} + x_{N+1}| < |x_{N+1}| + 1$ for all $n > N$, by triangular inequality. Take

$$M = \max\{|x_1|, |x_2|, |x_3|, \cdots, |x_N|, |x_{N+1}| + 1\}$$

(x_n) is then bounded by M. □

Now we are ready to prove the inverse direction. The idea is to use Bolzano-Weirestrass Theorem to find a convergent subsequence. Then, we can resort to the limit of this subsequence, and prove that the whole sequence converges to this limit.

Theorem 2.3.9: Cauchy Criterion

In $\mathbb{R}$, every Cauchy sequence converges.

Proof. Suppose (x_n) is Cauchy. By Lemma 2.3.8, it is bounded. By Bolzano-Weierstrass Theorem, it contains a convergent subsequence (x_{n_k}). Set

$$x = \lim x_{n_k}$$

Now we want to prove that the whole sequence converges to this limit. Let $\epsilon > 0$. Since (x_n) is Cauchy, there exists N_1 such that

$$|x_n - x_m| < \frac{\epsilon}{2}$$

for all $n, m > N_1$. We also know that $(x_{n_k}) \to x$, so choose $n_K > N_2$, we will have

$$|x_{n_K} - x| < \frac{\epsilon}{2}$$

Now if we choose $n > \max\{N_1, N_2\}$, by triangular inequality,

$$|x_n - x| = |x_n - x_{n_K} + x_{n_K} - x| \leqslant |x_n - x_{n_K}| + |x_{n_K} - x| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

which shows that it converges to x. □

Note that this Cauchy Criterion may fail if it is not on $\mathbb{R}$. Consider a sequence on the set of rational numbers, $\mathbb{Q}$, and the field we are interested in is only the set of rational numbers. Let (a_n) be the sequence such that the nth term is the nth decimal approximation of π, i.e.,

$$a_1 = 3.1, \quad a_2 = 3.14, \quad a_3 = 3.141 \quad , a_4 = 3.1415, \cdots$$

Then, all terms in this sequence are rational numbers, but it converges to π, a rational number, which is not in the rational field. We then cannot say this sequence 'converges' if we only consider $\mathbb{Q}$ as our field.

Therefore, Cauchy criterion implicitly reflects the completeness of real numbers, and it is indeed another Axiom of Completeness. **It is so important that it would be stated as a 'defining property' of (sequentially) compact sets.** However, this axiom, as Nested Interval Property, needs Archimedean Property to achive Supremum Property. Thus, it is a 'weak' axiom.

Here we attach Cauchy criterion to our axiom relation graph.

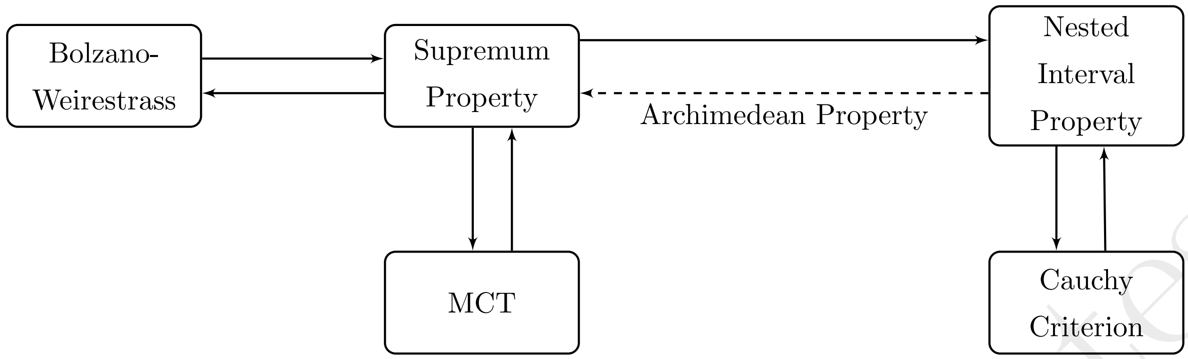


Figure 2.4: Relation between Axiom of Completeness

2.4 Convergence of Infinite Series

We are now transfer our sight from sequences to series. An **infinite series** is just the sum of all terms in a sequence.

Definition 2.4.1: Infinite Series

Let (b_n) be a sequence. An **infinite series** is an expression of the form

$$\sum_{n=1}^{\infty} b_n = b_1 + b_2 + b_3 + \cdots$$

Notice that the result of an infinite series can be infinity. For example, for sequence $(a_n) = (1, 1, 1, \cdots)$, its corresponding infinite series is then $\sum_{n=1}^{\infty} 1 = \infty \times 1 = \infty$. Therefore, a series can be convergent or divergent. Below we rigorously define the convergence of a series.

Definition 2.4.2: Convergence of Series

For a series $\sum_{n=1}^{\infty} b_n$, we define the corresponding sequence of **partial sum** (s_m) as

$$s_m = \sum_{n=1}^{m} b_n$$

and say that series converges to x if the sequence (s_m) converges to x. We write $\sum_{n=1}^{\infty} b_n = x$ in this case.

Now we see some classic examples of infinite series.

Example 2.4.3: Harmonic Series

The famous **harmonic series** is defined as

$$\sum_{n=1}^{\infty} \frac{1}{n}$$

It is a well-known divergent series. Consider the corresponding partial sum

$$s_m = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{m}$$

Notice the pattern that

$$s_4 = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} > 1 + \frac{1}{2} + \left(\frac{1}{4} + \frac{1}{4}\right) = 2$$

$$s_8 = 1 + \frac{1}{2} + \cdots + \frac{1}{8} > 1 + \frac{1}{2} + \left(\frac{1}{4} + \frac{1}{4}\right) + \left(\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}\right) = \frac{5}{2}$$

$$\vdots$$

$$\begin{aligned}
s_{2^k} &= 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) + \cdots + \left(\frac{1}{2^{k-1}+1} + \frac{1}{2^{k-1}+2} + \cdots + \frac{1}{2^k}\right) \\
&> 1 + \frac{1}{2} + \left(\frac{1}{4} + \frac{1}{4}\right) + \left(\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}\right) + \cdots + \left(\frac{1}{2^k} + \frac{1}{2^k} + \cdots + \frac{1}{2^k}\right) \\
&= 1 + \underbrace{\frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \cdots + \frac{1}{2}}_{k \text{ terms total}} = 1 + \frac{k}{2}
\end{aligned}$$

Therefore, s_m is not bounded, and the series does not converge.

Example 2.4.4: Basel problem

The Basel problem asks for the sum of the infinite series

$$\sum_{n=1}^{\infty} \frac{1}{n^2}$$

Though we can not solve the precise sum yet, we can give an approximation of the upper bound. Consider the partial sum

$$s_m = 1 + \frac{1}{4} + \frac{1}{9} + \cdots + \frac{1}{m^2}$$

Note that

$$s_m = 1 + \frac{1}{2} + \frac{1}{3 \times 3} + \frac{1}{4 \times 4} + \cdots + \frac{1}{m \times m}$$

$$
\begin{aligned}
&< 1 + \frac{1}{1\times 2} + \frac{1}{2\times 3} + \frac{1}{3\times 4} + \cdots + \frac{1}{m(m-1)} \\
&= 1 + \left(1 - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \cdots + \left(\frac{1}{m-1} - \frac{1}{m}\right) \\
&= 1 + 1 - \frac{1}{m} < 2
\end{aligned}
$$

Therefore, 2 is an upper bound for the partial sum. By Monotone Convergence Theorem, the partial sum converges to some limit less than 2. We will show later that this sum equals $\frac{\pi^2}{6}$, which is a surprising result.

Example 2.4.5: Geometric Series

A series is called **geometric** if it is of the form

$$\sum_{k=0}^{\infty} ar^k = a + ar + ar^2 + ar^3 + \cdots$$

Here we consider the case where $|r| < 1$. Since

$$(1-r)(1 + r + r^2 + r^3 + \cdots + r^{m-1}) = 1 - r^m$$

We can write the partial sum as

$$s_m = a + ar + ar^2 + \cdots + ar^{m-1} = \frac{a(1-r^m)}{1-r}$$

when $m \to \infty$, we can see that this partial sum converges to

$$\sum_{k=0}^{\infty} ar^k = \lim_{m\to\infty} s_m = \frac{a}{1-r}$$

Now we examine some properties of infinite series. We first note that since the convergence of infinite series is defined using convergence of sequences, we can immediately translate some result of sequence into statements about series.

Corollary 2.4.6: Algebraic Limit Theroem for Series

If $\sum_{k=1}^{\infty} a_k = A$ and $\sum_{k=1}^{\infty} b_k = B$, then

1. $\sum_{k=1}^{\infty} ca_k = cA, c \in \mathbb{R}$
2. $\sum_{k=1}^{\infty} (a_k + b_k) = A + B$

The reason that another two statements in Algebraic Limit Theorem is not transformed is, when dealing with $\lim s_n q_n$, where s_n, q_n are partial sums of a_n and b_n respectively, it does not equal to $\sum_{k=1}^{\infty} a_k b_k$, so it is of no practical use.

Moreover, since in $\mathbb{R}$, Cauchy sequences are convergent, we can use the definition of Cauchy sequence to restate the definition of convergence of series.

Proposition 2.4.7: Cauchy Criterion for Series

The series $\sum_{k=1}^{\infty} a_k$ converges if and only if,

$$\forall \epsilon > 0, \exists N \in \mathbb{N} \quad \text{s.t.} \quad n > m > N \quad \implies \quad |a_{m+1} + a_{m+2} + \cdots + a_n| < \epsilon$$

Proof. Observe that

$$|s_n - s_m| = |a_{m+1} + a_{m+2} + \cdots + a_n|$$

Thus, apply Cauchy criterion, we can get the result. □

For a series to be convergent, its corresponding sequence must have a extremely small 'tail', so that we are adding smaller and smaller terms to make the size controlled not to be large.

Proposition 2.4.8: Tail of convergent series

If the series $\sum_{k=1}^{\infty} a_k$ converges, then $(a_k) \to 0$.

Proof. Consider the special case $n = m + 1$ in the Cauchy Criterion for series, we have

$$|a_n| < \epsilon$$

for all $n > N$. Therefore, $(a_n) \to 0$. □

Note: The converse is not true. Consider the harmonic series, its corresponding sequence $(1/n)$ converges to 0, but the series itself diverges.

2.5 Convergence Tests for Series

2.5.1 Comparison Test

Sometimes it is very difficult to calculate the exact value of a infinite series. However, we have abundant tools to judge whether a series is convergent or not. One famous test is the **Comparison Test**, it uses another series as measuring stick to determine the convergency of another series.

Proposition 2.5.1: Comparison Test

Assume (a_k), (b_k) are sequences satisfying $0 \leqslant a_k \leqslant b_k$ for all $k \in \mathbb{N}$,

1. If $\sum_{k=1}^{\infty} b_k$ converges, then $\sum_{k=1}^{\infty} a_k$ converges.
2. If $\sum_{k=1}^{\infty} a_k$ diverges, then $\sum_{k=1}^{\infty} b_k$ diverges.

Proof.

1. If $\sum_{k=1}^{\infty} b_k$ converges, then by Cauchy criterion, for all $n > m > N$ where $N \in \mathbb{N}$ we have

$$|b_{m+1} + b_{m+2} + \cdots + b_n| < \epsilon$$

Notice that

$$|a_{m+1} + a_{m+2} + \cdots + a_n| \leqslant |b_{m+1} + b_{m+2} + \cdots + b_n| < \epsilon$$

We can conclude that $\sum_{k=1}^{\infty} a_n$ also converges.

2. This is just the contrapositive of (1). □

Note: The comparison test requires that the terms of series must be positive. These are called **positive infinite series**, and they are generally easier to manipulate than some arbitrary series.
Also, when we consider the limit of sequences and series, we are not very much interested in the first few terms, but only the tail of the sequence/series. Therefore, the condition in Comparison test can be relaxed to $0 \leqslant a_k \leqslant b_k$ for all $k > N$ with $N \in \mathbb{N}$.

A simple but important application of comparison tets is shown below.

Corollary 2.5.2: Application of Comparison Test

The series

$$\sum_{n=1}^{\infty} \frac{1}{n^p}$$

converges if and only if $p > 1$.

Proof. We consider the geometric series

$$\sum_{n=1}^{\infty} \frac{1}{p^n}$$

We have shown that it is convergent with $p > 1$. Choose large enough n, we will eventually have

$$\frac{1}{n^p} < \frac{1}{p^n}$$

Indeed, we can take n large enough so that $\log n/n < \log p/p$ since the sequence $\log n/n$ converges to 0, and the result follows. Then, by comparison test with relaxed condition, since geometric seires converges, we have that $\sum_{n=1}^{\infty} \frac{1}{n^p}$ also converges.
Now take $p \leqslant 1$. Since $1/n^p \geqslant 1/n$, by comparison test and the fact that harmonic series diverges, we have $\sum_{n=1}^{\infty} \frac{1}{n^p}$ diverges. □

2.5.2 Absolute and Conditional Convergence

To deal with more general series, we need to first define two terminologies.

Definition 2.5.3: Absolute and Conditional Convergence

- If $\sum_{n=1}^{\infty} |a_n|$ converges, then we say that the original series $\sum_{n=1}^{\infty} a_n$ **converges absolutely**.
- If the series $\sum_{n=1}^{\infty} a_n$ converges but $\sum_{n=1}^{\infty} |a_n|$ does not converge, we say that the original sequence $\sum_{n=1}^{\infty} a_n$ **converges conditionally**.

It may be easily noted that if a series converges absolutely, then it must converge.

Proposition 2.5.4: Absolute Convergence Test

If the series $\sum_{n=1}^{\infty} |a_n|$ converges, then $\sum_{n=1}^{\infty} a_n$ converges as well.

Proof. Since $\sum_{n=1}^{\infty} |a_n|$ converges, by Cauchy criterion, given $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$|a_{m+1}| + |a_{m+2}| + \cdots + |a_n| < \epsilon$$

for all $n > m > N$. By triangular inequality,

$$|a_{m+1} + a_{m+2} + \cdots + a_n| \leqslant |a_{m+1}| + |a_{m+2}| + \cdots + |a_n| < \epsilon$$

Therefore, the original series $\sum_{n=1}^{\infty} a_n$ also converges. □

Note: The converse is not true. Below we will show that the alternating **harmonic series**

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \cdots$$

indeed converges. However, taking absolute value each term, we have the divergent harmonic series.

How shouold we prove that the alternating harmonic series converges? The next proposition gives a more general result.

Proposition 2.5.5: Alternating Series Test

Let (a_n) be a sequence satisfying

1. $a_1 \geqslant a_2 \geqslant a_3 \geqslant \cdots \geqslant a_n \geqslant a_{n+1} \geqslant \cdots$
2. $(a_n) \to 0$

Then, the alternating series $\sum_{n=1}^{\infty}(-1)^{n+1}a_n$ converges.

Proof. Let $N \in \mathbb{N}$ be even and let $n > N$. Because the series is alternating and the terms is decreasing, we have

$$s_N \leqslant s_n \leqslant s_{N+1}$$

Since (a_n) converges to 0, we can make $|s_{N+1} - s_N| = |a_N|$ be arbitrarily small when we increase N. Set N large enough so that $|a_N| < \frac{\epsilon}{2}$, we have

$$|s_m - s_n| \leqslant |s_m - s_N| + |s_N - s_n| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

for all $n > m > N$. Therefore, (s_n) is Cauchy, thus it converges. $\square$

2.5.3 Root and Ratio Test

Before diving into the root test, we first introduce two extremely important concepts in analysis, the **limit superior** and the **limit inferior**.

Definition 2.5.6: Limit Superior/Inferior

- The **limit superior** of a sequence (x_n) is defined as

$$\limsup_{n\to\infty} x_n = \lim_{n\to\infty}\left(\sup_{m\geqslant n} x_m\right) = \inf_{n\geqslant 0}\left(\sup_{m\geqslant n} x_m\right)$$

- The **limit inferior** of a sequence (x_n) is defined as

$$\liminf_{n\to\infty} x_n = \lim_{n\to\infty}\left(\inf_{m\geqslant n} x_m\right) = \sup_{n\geqslant 0}\left(\inf_{m\geqslant n} x_m\right)$$

The reason for the two definitions above $\lim_{n\to\infty}\left(\sup_{m\geqslant n} x_m\right)$ and $\inf_{n\geqslant 0}\left(\sup_{m\geqslant n} x_m\right)$ are equal is, the sequence $(\sup_{m\geqslant n} x_m)_{n\in\mathbb{N}}$ is actually decreasing, as you should notice. Therefore, the inferior is equal to the limit.

Intuitively, limit superior and limit inferior is just the 'largest' and the 'smallest' number that some subsequence will converge to. For example, a sequence

$$a_n = \left(\frac{2}{3}, -\frac{2}{3}, 0, \frac{3}{4}, -\frac{3}{4}, 0, \frac{4}{5}, -\frac{4}{5}, 0, \frac{5}{6}, -\frac{5}{6}, 0, \cdots\right)$$

There are three obvious convergent subsequences,

$$a_{3k-2} = \left(\frac{k+1}{k+2}\right)$$

$$a_{3k-1} = \left(-\frac{k+1}{k+2}\right)$$

$$a_{3k} = (0)$$

they will converge to $1, -1$ and 0, respectively. Then $\limsup_{n\to\infty} a_n = 1$ and $\liminf_{n\to\infty} a_n = -1$. This can also be verified using the definition.

Limit superior and limit inferior has many greater properties that limit does not generally have.

Proposition 2.5.7: Properties of Limit Superior/Inferior

1. $\limsup$ and $\liminf$ will always exist for bounded sequence.

2. For every bounded sequence, $\liminf a_n \leqslant \limsup a_n$, the equality is attained if and only if $\lim a_n$ exists, and in this case $\liminf a_n = \limsup a_n = \lim a_n$.

Proof. 1. Since $\sup_{m\geqslant n} a_m$ and $\inf_{m\geqslant n} a_m$ are monotone sequences, by monotone convergence theorem, both of them will converge.

2. Since $\inf_{m\geqslant n} a_m < \sup_{m\geqslant n} a_m$ for all n, by order limit theorem, we have $\liminf a_n \leqslant \limsup a_n$. Moreover, since $\inf_{m\geqslant n} a_m \leqslant a_n \leqslant \sup_{m\geqslant n} a_m$, by squeeze theorem, if $\liminf a_n = \limsup a_n$, then $\lim a_n$ must exist. On the other hand, if $\lim a_n$ exists, by cauchy criterion,

$$\forall\, \epsilon > 0,\ \exists\, N \in \mathbb{N},\ \text{s.t. } m, n > N \quad \Longrightarrow \quad |a_n - a_m| < \epsilon$$

Then by the definition of superior and inferior, we should have

$$|\inf_{m\geqslant n} a_m - \sup_{m\geqslant n} a_m| \leqslant \epsilon$$

for all $n > N$. Since $\epsilon > 0$ is arbitrary, we have $\liminf a_n = \limsup a_n = \lim a_n$. □

With these definitions set up, we can state and prove the famous root and ratio tests for convergence of series.

Theorem 2.5.8: Root Test

Let $\sum_{n=1}^{\infty} a_n$ be a series of real numbers, let $\alpha = \limsup_{n\to\infty} |a_n|^{1/n}$.

(a) If $\alpha < 1$, then the series $\sum_{n=1}^{\infty} a_n$ is absolutely convergent.

(b) If $\alpha > 1$, then the series $\sum_{n=1}^{\infty} a_n$ is not convergent.

(c) If $\alpha = 1$, we cannot assert any conclusion.

Proof. (a) Suppose $\alpha < 1$. We must have $\alpha > 0$ since $|a_n|^{1/n} \geqslant 0$ for all n. Then, we can find $\epsilon > 0$ such that $0 < \alpha + \epsilon < 1$. Since

$$\alpha = \limsup_{n\to\infty} |a_n|^{\frac{1}{n}} = \inf_{n \geqslant 0} \left(\sup_{m \geqslant n} |a_n|^{\frac{1}{n}} \right)$$

By the definition of superior and inferior, and the fact that $\sup_{m \geqslant n} |a_n|^{\frac{1}{n}}$ is decreasing, there must exist $N \in \mathbb{N}$ such that

$$|a_n|^{\frac{1}{n}} \leqslant \alpha + \epsilon \quad \Longrightarrow \quad |a_n| \leqslant (\alpha + \epsilon)^n$$

for all $n \geqslant N$. But from the geometric series we have that

$$\sum_{n=N}^{\infty} (\alpha + \epsilon)^n$$

is absolutely convergent, since $0 < \alpha + \epsilon < 1$. Thus by the comparison test, we see that $\sum_{n=N}^{\infty} a_n$ is absolutely convergent, and thus $\sum_{n=1}^{\infty} a_n$ is absolutely convergent since first few terms will not influence the convergency of series.

(b) Now suppose that $\alpha > 1$. Then again, By the definition of superior and inferior, and the fact that $\sup_{m \geqslant n} |a_n|^{\frac{1}{n}}$ is decreasing, there exists $N \in \mathbb{N}$ such that

$$|a_n|^{\frac{1}{n}} > 1 \quad \Longrightarrow \quad |a_n| > 1$$

By the contrapositive of proposition 2.4.8, since (a_n) does not converge to 0, the corresponding series then does not converge. □

The root test is phrashed using limit superior, but of course if $|a_n|^{\frac{1}{n}}$ converges we can state it using just limit. Now we state and prove the **Ratio Test**.

Theorem 2.5.9: Ratio Test

Let $\sum_{n=1}^{\infty} a_n$ be a series of real numbers such that $a_n \neq 0$, let

$$r = \limsup_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right|$$

(a) If $r < 1$, then the series $\sum_{n=1}^{\infty} a_n$ is absolutely convergent.

(b) If $r > 1$, then the series $\sum_{n=1}^{\infty} a_n$ is not convergent.

(c) If $r = 1$, we cannot assert any conclusion.

Proof. (a) Suppose $r < 1$. Let r' satiesfies $r < r' < 1$. We have

$$r = \limsup_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| = \lim_{n\to\infty} \left(\sup_{m\geqslant n} \left|\frac{a_{m+1}}{a_m}\right|\right) < 1$$

By the definition of superior, and the fact that $\sup_{m\geqslant n} \left|\frac{a_{m+1}}{a_m}\right|$ is decreasing, we can find $N \in \mathbb{N}$ such that $n \geqslant N$ implies

$$\left|\frac{a_{n+1}}{a_n}\right| \leqslant r' \quad \Longrightarrow \quad |a_{n+1}| \leqslant |a_n| r'$$

Since the geometric series $|a_N| \sum_{n=N}^{\infty} (r')^n$ converges, by comparison test,

$$|a_{N+1}| \leqslant r'|a_N|$$

$$|a_{N+2}| \leqslant r'|a_{N+1}| \leqslant (r')^2|a_N|$$

$$|a_{N+3}| \leqslant (r')^3|a_N|$$

$$\vdots$$

$$|a_{N+n}| \leqslant (r')^n|a_N|$$

we have that $\sum_{n=N}^{\infty} |a_n|$ also converges. Therefore, $\sum_{n=1}^{\infty} |a_n|$ is convergent, the original sequence is absolutely convergent.

(b) If $r > 1$, By the definition of superior, and the fact that $\sup_{m\geqslant n} \left|\frac{a_{m+1}}{a_m}\right|$ is decreasing, we can find

$N \in \mathbb{N}$ such that $n \geqslant N$ implies

$$\left|\frac{a_{n+1}}{a_n}\right| \geqslant 1 \quad \Longrightarrow \quad |a_{n+1}| \geqslant |a_n|$$

for all $n > N$. Since $a_n \neq 0$ and $|a_n| > 0$ for all n, we conclude that (a_n) cannot converge to 0. Therefore, the corresponding series does not converge. $\square$

2.5.4* Cauchy Condensation Test, Abel's Test and Dirichlet's Test

To end this section, we introduce another three tests for convergence of series. Based on the proof process of divergency of harmonic series, we state the following general argument.

Theorem 2.5.10: Cauchy Condensation Test

Suppose (b_n) is decreasing and satisfies $(b_n) \geqslant 0$ for all $n \in \mathbb{N}$. Then, the series $\sum_{n=1}^{\infty} b_n$ converges if and only if the series

$$\sum_{n=0}^{\infty} 2^n b_{2^n} = b_1 + 2b_2 + 4b_4 + 8b_8 + 16b_{16} + \cdots$$

is convergent.

Proof.

- ($\Longrightarrow$) First assume that $\sum_{n=0}^{\infty} 2^n b_{2^n}$ converges. Then, by Proposition 2.2.2, the partial sum

  $$t_k = b_1 + 2b_2 + \cdots + 2^k b_{2^k}$$

  is bounded. That is, there exists $M > 0$ such that $t_k \leqslant M$ for all $k \in \mathbb{N}$. We want to prove that $\sum_{n=1}^{\infty} b_n$ converges. Since $b_n \geqslant 0$, we know that the partial sum s_m for this sequence is increasing. By Monotone Convergence Theorem, we only need to prove $s_m = b_1 + b_2 + \cdots + b_m$ is bounded.
  Fix m and let k be large enough to ensure $m \leqslant 2^{k+1} - 1$. Then, $s_m \leqslant s_{2^{k+1}-1}$ and

  $$\begin{aligned} s_{2^{k+1}-1} &= b_1 + (b_2 + b_3) + (b_4 + b_5 + b_6 + b_7) + \cdots + (b_{2^k} + \cdots + b_{2^{k+1}-1}) \\ &\leqslant b_1 + (b_2 + b_2) + (b_4 + b_4 + b_4 + b_4) + \cdots + \underbrace{(b_{2^k} + \cdots + b_{2^k})}_{2^k \text{ terms}} = t_k \end{aligned}$$

  Therefore, $s_m \leqslant t_k \leqslant M$. the partial sum is bounded. By monotone convergence theroem, we have that $\sum_{n=1}^{\infty} b_n$ is convergent.

- ($\Longleftarrow$) We now prove that if $\sum_{n=1}^{\infty} b_n$ converges, then $\sum_{n=0}^{\infty} 2^n b_{2^n}$ converges. Instead of proving this, we prove its contrapositive. Suppose $\sum_{n=0}^{\infty} 2^n b_{2^n}$ diverges. Since the partial sum t_k is increasing, by monotone converges theorem, it is unbounded (otherwise it is convergent). Therefore, for any $2M > 0$ we can find $k \in \mathbb{N}$ such that $t_k > 2M$. Therefore, with the fact that (b_n) is decreasing,

$$\begin{aligned} s_{2^k} &= b_1 + b_2 + \cdots + b_{2^k} \geqslant b_1 + b_2 + (b_4 + b_4) + (b_8 + b_8 + b_8 + b_8) + \cdots + \underbrace{(b_{2^k} + \cdots + b_{2^k})}_{2^{k-1} \text{ terms}} \\ &= b_1 + b_2 + 2b_4 + 4b_8 + \cdots + 2^{k-1} b_{2^k} = \frac{1}{2} t_k + \frac{1}{2} b_1 > M \end{aligned}$$

Since $M > 0$ is arbitrary, we have that s_{2^k} is unbounded. Therefore, the series $\sum_{n=1}^{\infty} b_n$ is not convergent. $\square$

There are two more tests that is very useful in judging whether a series of the form $\sum_{n=1}^{\infty} f(n)g(n)$ converges or not. For example, we may want to judge whether

$$\sum_{n=1}^{\infty} \frac{\cos(n)}{n}$$

converges or not. Then, the terms can be separated into $\frac{1}{n}$ and $\cos(n)$, and these theroems can be applied.

Theorem 2.5.11: Abel's Test

Suppose $\sum_{n=1}^{\infty} a_n$ converges, and (b_n) is monotone and bounded. Then, the series

$$\sum_{n=1}^{\infty} a_n b_n$$

is convergent.

Proof. Since (b_n) is monotone and bounded, by monotone convergence theorem, it converges. Suppose $(b_n) \to b$. Let $a_0 = 0$, $s_m = \sum_{n=0}^{m} a_n$. Then, $a_m = s_m - s_{m-1}$ for all $m \geqslant 1$. We can first consider the finite sum, and decompose the target sequence as

$$\begin{aligned} \sum_{n=1}^{N} a_n b_n = \sum_{n=1}^{N} b_n (s_n - s_{n-1}) & \\ &= b_1 s_1 + b_2 s_2 - b_2 s_1 + b_3 s_3 - b_3 s_2 + \cdots + b_N s_N - b_N s_{N-1} \\ &= s_1(b_1 - b_2) + s_2(b_2 - b_3) + \cdots + s_{N-1}(b_{N-1} - b_N) + s_N b_N \end{aligned}$$

$$= \left(\sum_{n=1}^{N-1} s_n(b_n - b_{n+1})\right) + s_N b_N$$

Since (s_N) is convergent, and (b_N) is also convergent, by algebraic limit theorem, we have $(s_N b_N)$ must converge. Moreover, since s_N is convergent, it is bounded by some $M > 0$, we see that the sum is telescoping

$$\sum_{n=1}^{N-1} s_n(b_n - b_{n+1}) \leqslant M \sum_{n=1}^{N-1} (b_n - b_{n+1}) = M(b_1 - b_N)$$

Since b_N is convergent, by algebraic limit theorem, $\sum_{n=1}^{N-1} s_n(b_n - b_{n+1})$ also converges. These will finally show that

$$\lim_{n\to\infty} \sum_{n=1}^{N} a_n b_n = \sum_{n=1}^{\infty} a_n b_n = \left(\sum_{n=1}^{N-1} s_n(b_n - b_{n+1})\right) + s_N b_N$$

is convergent. □

Now we change our mind a little bit. In the previous theorem we get the convergence from the convergence of $\sum_{n=1}^{\infty} a_n$. Now we relax this condition mildly, to only require the partial sums to be bounded. As a compensation, we strengthen the another assumption so that (b_n) is not only monotone and bounded, but also converges to 0. **Note how the term $\mathbf{s_N b_N}$ is not necessarily convergent if we relax the assumption of a_n in Abel's test, or relax assumption of b_n in Dirichlet's Test.**

Theorem 2.5.12: Dirichlet's Test

Suppose that for two sequences (a_n) and (b_n), we have $\sum_{n=1}^{N} a_n$ is bounded for all N (but the series does not necessarily need to be convergent), i.e.,

$$\left|\sum_{n=1}^{N} a_n\right| \leqslant M \quad \text{for all } N \in \mathbb{N}$$

and (b_n) is monotone so that $\lim b_n = 0$. Then, the series

$$\sum_{n=1}^{\infty} a_n b_n$$

is convergent.

Proof. Similarly with the last proof, we can write the series as

$$\sum_{n=1}^{N} a_n b_n = \left(\sum_{n=1}^{N-1} s_n(b_n - b_{n+1})\right) + s_N b_N$$

Since s_N is bounded by M, and b_N converges to 0, we have

$$\lim_{N\to\infty} s_N b_N \leqslant M \lim_{N\to\infty} b_N = 0$$

is convergent. For the telescoping term, we still have

$$\sum_{n=1}^{N-1} s_n(b_n - b_{n+1}) \leqslant M \sum_{n=1}^{N-1}(b_n - b_{n+1}) = M(b_1 - b_N)$$

and it converges to $M(b_1 - 0) = Mb_1$. Therefore, the whole partial sum $\sum_{n=1}^{N} a_n b_n$ converges. □

2.6 Rearrangement of Series

We see in Example 1.1.3 of the opening section of this whole note that some series will change its value after reordering the terms. To refresh your memory, we will restate it here.

Example 2.6.1. *Reordering Infinite Series*
Consider the alternating harmonic series

$$S = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots$$

We know that this infinite series converges at some point. Therefore, nothing similar as Example 1.1.1 could happen here. However, if we do the following computation:

$$\begin{aligned}\frac{1}{2}S &= \frac{1}{2} - \frac{1}{4} + \frac{1}{6} - \frac{1}{8} + \frac{1}{10} - \frac{1}{12} + \frac{1}{14} - \frac{1}{16} + \cdots \\ S &= 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} - \frac{1}{8} + \frac{1}{9} - \frac{1}{10} + \frac{1}{11} - \frac{1}{12} + \frac{1}{13} - \frac{1}{14} + \frac{1}{15} - \frac{1}{16} + \cdots\end{aligned}$$

$$\frac{3}{2}S = \left(1 + \frac{1}{3}\right) - \frac{1}{2} + \left(\frac{1}{5} + \frac{1}{7}\right) - \frac{1}{4} + \left(\frac{1}{9} + \frac{1}{11}\right) - \frac{1}{6} + \left(\frac{1}{13} + \frac{1}{15}\right) - \frac{1}{8} + \cdots$$

We see that $\frac{3}{2}S$ is just a reordering of our initial infinite series (with two positive terms following one negative term)! Therefore, we just change the convergent point by simply reordering the infinite series.

We will rigorously define what is a rearrangement first.

Definition 2.6.1: Rearrangement

Let $\sum_{n=1}^{\infty} a_n$ be a series. A series $\sum_{n=1}^{\infty} b_n$ is a **rearrangement** of $\sum_{n=1}^{\infty} a_n$ if there exists a bijective function $f : \mathbb{N} \to \mathbb{N}$ such that

$$b_{f(k)} = a_k$$

for all $k \in \mathbb{N}$.

Now we are ready to see why this happens. It can be seen that for the alternating harmonic series, only taking the sum of positive terms or only take the sum of negative terms will both lead to a divergent series. This means that, we can reach 'any far' on the real line by choosing terms in this series and add them. **This only happens when the convergence is conditional.**

Theorem 2.6.2: Rearrangement Criterion

If a series converges absolutely, then any rearrangement of this series converges to the same limit.

Proof. Assume $\sum_{k=1}^{\infty} a_k$ converges absolutely to A, and let $\sum_{n=1}^{\infty} b_k$ be a rearrangement of $\sum_{n=1}^{\infty} a_k$. Denote s_n, t_m as the partial sum of (a_k) and (b_k), respectively. We want to show that $(t_m) \to A$. Let $\epsilon > 0$. Since $(s_n) \to A$, we can have $N_1 \in \mathbb{N}$ so that

$$|s_n - A| < \frac{\epsilon}{2}$$

for all $n \geqslant N_1$. Because the convergence is absolute, we can choose $N_2 \in \mathbb{N}$ so that

$$\sum_{k=m+1}^{n} |a_k| < \frac{\epsilon}{2}$$

for all $n > m \geqslant N_2$. Take $N = \max\{N_1, N_2\}$. We know that the finite set of terms $\{a_1, a_2, a_3, \cdots, a_N\}$ must all appear in the rearranged series since rearrangement is a bijection, and we want to move far enough out in the series $\sum_{n=1}^{\infty} b_n$ so that we have included all these terms. Thus, choose

$$M = \max\{f(k) : 1 \leqslant k \leqslant N\}$$

It should be now evident that if $m \geqslant M$, then $(t_m - s_N)$ consists of a finite set of terms, and the absolute values of which appear in the tail $\sum_{k=m+1}^{\infty} |a_k|$. Thus the choice of N_2 guarantees that $|t_m - s_N| < \frac{\epsilon}{2}$, and

$$|t_m - A| = |t_m - s_N + s_N - A| \leqslant |t_m - s_N| + |s_N - A| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

for all $m > M$, which shows that $(t_n) \to A$. $\square$

2.7* Cauchy Sequence Method: Completion from $\mathbb{Q}$ to $\mathbb{R}$

Note: This session is just an introduction to this topic. There is no full detail since it needs tedious proof procedure just like the Dedekind cut, and it nees many prerequisites from abstract algebra.
To end this whole session, we mention another different way from Dedekind Cut, to construct real numbers from rational numbers. The main tool we use here is the Cauchy Sequences.

Here is the intuitive idea. We know that all convergent sequences are Cauchy. However, the inverse would only be true in complete spaces such as $\mathbb{R}$. This is because some Cauchy sequence will converge to a point that is outside of the space (an example is given in the corresponding Cauchy sequence chapter). In our case, some Cauchy sequence in rational field will converge to a point in real number field, and we cannot say that is convergent in this sense. However, if we define these points using the 'limiting property' of Cauchy sequence, i.e., two points in the sequence is getting closer and closer, we can actually define those points using Cauchy sequences!

Before we formally start to prove that, we need to first introduce some terminologies that is maybe too early to appear. The definition of limit and convergence depends on the definition of 'distance' between to points. For example, on the real line, we measure the distance between two points x and y by calculating the absolute value $|x - y|$. In any other space, if we can define such a distance function on some general space X to be $d : X \times X \to [0, \infty)$, which takes elements in X and take values on nonnegative real numbers, we can also perform convergence analysis on it. To make this distance function 'regular' (i.e., not pathological) enough, we assume

- $d(x, x) = 0, x \in X$
- $d(x, y) = d(y, x), x, y \in X$
- $d(x, z) \leqslant d(x, y) + d(x, z), x, y, z \in X$

This is just the idea of **Metric Space**, which will formally appear in PART III of this note. A distance function is called a **metric**.

The second terminology is **isometry**. Intuitively, two spaces are isometry if they have exactly the same topological properties, but just with different notations. We define formally, an **isometry** $\phi : X \to \tilde{X}$, is a function from a space to another space such that $\tilde{d}(\phi(x), \phi(y)) = d(x, y)$, where d and $\tilde{d}$ are metrics on X and $\tilde{X}$, respectively. Two spaces are **isometric** if there exists such a isometry between them.

Finally, we will introduce the **equivalence relation**, which will often appear in the context of abstract algebra. A **binary relation** R on a set X is a subset of $X \times X$. Two elements $x, y \in X$ are equivalent if

$(x, y) \in R$, denoted by $x \sim y$. For example, we can define a relation '$\geqslant$' on the set $\mathbb{R}$ so that it is the set $\{(x, y) : y \geqslant x, x, y \in \mathbb{R}\}$. Then, an **equivalence relation** will have properties

- **Reflexive:** $x \sim x, \forall x \in X$
- **Symmetric:** $x \sim y \Leftrightarrow y \sim X, \forall x, y \in X$
- **Transitive:** $x \sim y, y \sim z \Rightarrow x \sim z, \forall x, y, z \in X$

An **equivalence class** $[x]$ is then a set of all equivalence element to some x, i.e., $[x] = \{y \in X : x \sim y\}$.

With all these three tools, we can state our result:

Theorem 2.7.1: Cauchy Sequence construction of $\mathbb{Q}$ to $\mathbb{R}$

There exists a complete space $\mathbb{R}$ with a properly defining metric such that it has a subspace W which is isometric with $\mathbb{Q}$, and is dense in $\mathbb{Q}$. The space $\mathbb{R}$ is unique up to isometry.

By 'unique up to isometry', we mean that if two spaces satisfy the conditions, and they are isometric, we see them as the same. Now we prove the theorem. It is actually way more simple and concise than Dedekind Method.

Define the equivalence relation: Let $(x_n), (x'_n)$ be rational Cauchy sequences. Define $(x_n) \sim (x'_n)$ if

$$\lim_{n \to \infty} |x_n - x'_n| = 0$$

Let $\mathbb{R}$ be defined as the set of all equivalence classes of Cauchy sequences. Then, as in the procedure of Dedekind Cut, we can prove all the properties of real numbers, including the supremum property.

Note: Here we construct $\mathbb{R}$ using the metric $|x - y|$. For other metrics, we can construct other complete spaces called the **p-adic numbers**.

Chapter 3

Topology on the Real Line

Here we move on to another interesting topic of mathematical analysis, the **point set topology**, which sees sets as geometric objects and analyze their properties such as closedness, compactness, separability and connectedness, etc. In this chapter, we mainly focus on the real line. We consider the real line as a geometric object, and consider those topological properties on it.

3.1 Open and Closed Sets

3.1.1 Open Sets

To define what is an open set, we first introduce the interior point.

Definition 3.1.1: Interior Point

A point a is an **interior point** of a set A if

$$\exists\, \epsilon > 0, \text{ s.t. } (a-\epsilon, a+\epsilon) \subseteq A$$

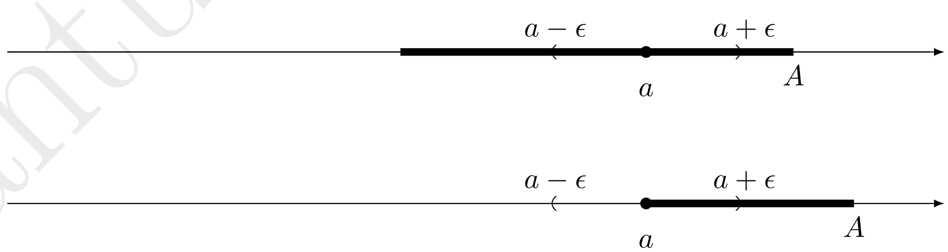


Figure 3.1: Up: Interior Point Down: Not Interior Point

As drawn in the figure, if the point it in the 'interior' of the set, there would always be a small ϵ region that is completely contained in the set. However, if the point is at the boundary, then half of the ϵ region would always be out of the set. This explains why it is called interior point.

Definition 3.1.2: Open Set

A set G is **open** if all points $a \in G$ are interior points.

Example 3.1.1.

- $\mathbb{R}$ is an open set. For any point $a \in \mathbb{R}$, we are free to choose ϵ and always $(a-\epsilon, a+\epsilon) \subseteq \mathbb{R}$.
- Open intervals are open sets. Consider open interval (a,b). For any point $x \in (c,d)$, we can take $\epsilon = \min\{x-a, b-x\}$, i.e., we choose the smallest distance from x to the boundary of this open interval. Then $(x-\epsilon, x+\epsilon) \subseteq (a,b)$.
- Closed intervals are not open set. Consider closed interval $[a,b]$. Choose $a \in [a,b]$, for all $\epsilon > 0$, the set $(a-\epsilon, a+\epsilon)$ will always fall out the interval.

An important result is that union of open intervals is open. The next theorem states this fact. It is very important so that it is later introduced as the defining property of a general topological space.

Theorem 3.1.3: Topological property of open set

1. The union of **arbitrary collection** of open sets is open.
2. The intersection of **a finite collection** of open sets is open.

Proof.

1. Let $\{G_\lambda : \lambda \in I\}$ be a collection of open sets and let $G = \bigcup_{\lambda \in I} G_\lambda$. Let $a \in G$. Then, we have $a \in G_{\lambda'}$ for some specific $\lambda' \in I$. Since $G_{\lambda'}$ is open, there exists $\epsilon > 0$ such that
$$(a-\epsilon, a+\epsilon) \subseteq G_{\lambda'} \subseteq G$$
Therefore, G is open.
2. Let $\{G_1, G_2, G_3, \cdots G_n\}$ be finite collection of open sets. If $a \in \bigcap_{k=1}^n G_k$, then $a \in G_k$ for all k. Since all G_k are open sets, we can find a ϵ_k for each set G_k such that $(a-\epsilon_k, a+\epsilon_k) \subseteq G_k$. Take $\epsilon = \min\{\epsilon_1, \epsilon_2, \cdots, \epsilon_n\}$. If follows that $(a-\epsilon, a+\epsilon) \subseteq (a-\epsilon_k, a+\epsilon_k) \subseteq G_k$ for all k and thus $(a-\epsilon, a+\epsilon) \subseteq \bigcap_{k=1}^n G_k$.

This follows that $\bigcap_{n=1}^k G_k$ is open. $\square$

Note that the intersection must be finite, since in the case such as

$$G = \bigcap_{n=1}^{\infty} \left(-\frac{1}{n}, \frac{1}{n} \right) = \{0\}$$

is not open. This can also be shown in the proving process that we are allowed to choose the minimum of all the ϵ_k, which does not always exist in infinite case.
To end this subsection, we introduce a terminology regard with open sets.

Definition 3.1.4: Interior

The **interior** of E, denoted by E°, is defined as the set of all interior points of E.

It is trivial to see that all interiors are open.

3.1.2 Closed Sets

For closed sets, there is also a kind of point that is related to.

Definition 3.1.5: Limit Point

A point x is a **limit point (accumulation point)** of a set A if

$$\forall \epsilon > 0, (a - \epsilon, a + \epsilon) \cap (A \backslash \{a\}) \neq \emptyset$$

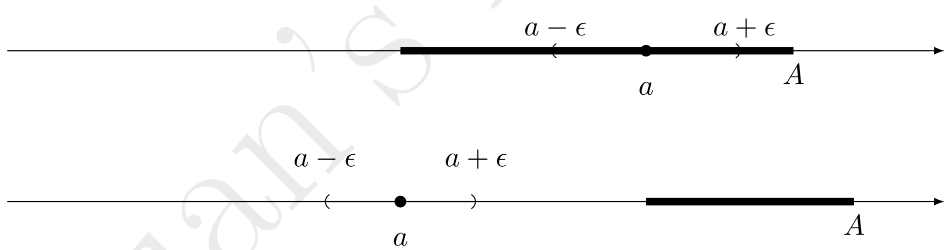


Figure 3.2: Up: Limit Point Down: Not Limit Point

As showed in the figure, if the point has other points 'accompanied closely' to the point a, then it is a cumulative point. If it is completely isolated, it is not.

The reason that it is called limit point, is that this point can be a limit of a sequence in this set.

Proposition 3.1.6: Limit Point Characterization

A point x is a limit point of A if and only if $x = \lim a_n$ for some $(a_n) \in A$ satisfying $a_n \neq x$ for all $n \in \mathbb{N}$.

Proof.
($\Longrightarrow$) Suppose x is a limit point. Then,

$$\forall \epsilon > 0, (x-\epsilon, x+\epsilon) \cap (A \backslash \{x\}) \neq \emptyset$$

Therefore, we can choose $\epsilon = 1/n$ such that there exists $a_n \in A$ with

$$a_n \in (x-\epsilon, x+\epsilon) \cap (A \backslash \{x\})$$

This indicates that for any $\epsilon > 0$, choose N such that $1/N < \epsilon$, we have $|a_n - x| < \epsilon$ for all $n \geqslant N$. Thus, $(a_n) \to x$.
($\Longleftarrow$) Assume $x = \lim a_n$ for some $(a_n) \in A$ satisfying $a_n \neq x$ for all $n \in \mathbb{N}$. This directly indicates that for any $\epsilon > 0$, there will always exist $a_n \in A$ such that $a_n \in (x-\epsilon, x+\epsilon) \cap (A \backslash \{x\}) \neq \emptyset$. $\square$

Therefore, Proposition 3.1.6 can be stated as the definition of limit point. There are two ways to define a limit point, one through topology, and one through analysis.

Definition 3.1.7: Isolated Point

A point $a \in A$ is an **isolated point** of A if it is not a limit point.

Note that an isolated point is always an element of the set. However, a limit point can be not in the set. Consider open interval (a, b), then, it is easy to verify that a and b are limit points, but they do not belong to the interval.

Definition 3.1.8: Closed Set

A set $F \subseteq \mathbb{R}$ is **closed** if it contains all its limit points.

Example 3.1.2.

- A closed interval $[a, b]$ is a closed set. For each $x \in [a, b]$, we have $(x-\epsilon, x+\epsilon) \cap ([a, b]] \backslash \{x\}) \neq \emptyset$ for all $\epsilon > 0$ by the density of rational numbers. We also need to verify that all points in $[a, b]^c$ are not limit points. This is trivial, and we omit it here.

- The set
$$A = \left\{ \frac{1}{n} : n \in \mathbb{N} \right\} \cup \{0\}$$
is closed. Given $1/n \in A$, choose $\epsilon = 1/n - 1/(n+1)$, then $(1/n - \epsilon, 1/n + \epsilon) \cap A = \{1/n\}$. Therefore, all points with the form $1/n$ are not limit points. Similarly, all points that not belong to A are not

limit point. Finally, 0 is the only limit point since for all $\epsilon > 0$, we can always find some $n \in \mathbb{N}$ such that $1/n \in (-\epsilon, \epsilon)$.

- $\mathbb{Q}$ is not a closed set. For all points $x \in \mathbb{R}$, and for all $\epsilon > 0$, there exists $a \in \mathbb{Q}$ such that $a \in (x-\epsilon, x+\epsilon)$, by the density of rational number. Therefore, the set of limit points of $\mathbb{Q}$ is all of $\mathbb{R}$.

We can construct a closed set from a non-closed set by including all limit points of it. This is called the **closure** of the set.

Definition 3.1.9: Closure

Given a set $A \subseteq \mathbb{R}$. Let B be the set of all limit points of A. Then, $\bar{A} = A \cup B$ is called the **closure** of A.

There is still possibility that, after including these limit points, there are potentially new limit points produced. The next proposition told us this can not happen.

Proposition 3.1.10: Closedness of Closure

For any $A \subseteq \mathbb{R}$, the closure $\bar{A}$ is the smallest closed set containing A.

Proof.

- Let B be the set of all limit points of A. We first show that B is a closed set. To do this, suppose x is a limit point of B, we want to show that it belongs to B, i.e., it is a limit point of A. Since x is a limit point of B, there exists $(a_n) \in B$ with $a_n \neq x$ such that for $n > N_1$ where $N_1 \in \mathbb{N}$, $|a_n - x| < \frac{\epsilon}{2}$. Since $(a_n) \in B$, each a_n is a limit point of A. Therefore, for each $n \in \mathbb{N}$, there exists $(b_n^k)_{k=1}^{\infty} \in A$ such that for $k > N_2$ where $N_2 \in \mathbb{N}$, $|b_n^k - a_n| < \frac{\epsilon}{2}$. Therefore, consider the sequence $(b_n^k)_{n,k}$, when $n > N_1$ and $k > N_2$, we have

$$|b_n^k - x| \leqslant |b_n^k - a_n| + |a_n - x| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

  Therefore, $(b_n^k)_{n,k} \to x$, x is a limit point of A.

- Now consider the set $\bar{A} = A \cup B$. If x is a limit point of $A \cup B$, then $x = \lim x_n$, where $(x_n) \in A \cup B$, and $x_n \neq x$. Since (x_n) is infinite, there must exist a subsequence $(x_{n_k}) \to x$ such that all $x_{n_k} \in A$ or all $x_{n_k} \in B$. If $x_{n_k} \in A$, then x is a limit point of A. If $x_{n_k} \in B$, then it belongs to B, and it is also a limit point of A. Therefore, If x is a limit point of $A \cup B$, then it must be a limit point of A. This shows that $A \cup B$ does not produce any new limit points. Therefore, $A \cup B$ contains all its limit points, it is closed.

- Finally, any closed set containing A must contain B as well, since at least it needs to conclude all limit points of A. Therefore, $\bar{A}$ is the smallest closed set containing A. □

In Chapter 1 we see that rational and irrational numbers are 'dense' in $\mathbb{R}$. Now we can use the concept of closure to rigorously define what does it mean by 'dense'.

Definition 3.1.11: Dense

A subset $A \subseteq \mathbb{R}$ is **dense** in $\mathbb{R}$ if $\bar{A} = \mathbb{R}$. Equivalently, A is dense in $\mathbb{R}$ if for all $c \in \mathbb{R}$, there exists a sequence $(x_n) \subseteq A$ such that $(x_n) \to c$.

There is also an equivalent definition of closed set. It is trivial to see and it is defined based on open sets.

Theorem 3.1.12: Complement of Open or Closed Sets

A set A is open if and only if A^c is closed. A set B is closed if and only if B^c is open.

Proof.

- Let A be open. Suppose x is a limit point of A^c. Then, for all $\epsilon > 0$, $(x-\epsilon, x+\epsilon) \cap (A^c \backslash \{x\}) \neq \emptyset$. If $x \in A$, then there exists $\epsilon > 0$ such that $(x-\epsilon, x+\epsilon) \in A$, which is impossible since then $(x-\epsilon, x+\epsilon) \cap (A^c \backslash \{x\}) = \emptyset$. Therefore, $x \in A^c$. This shows that A^c is closed.
- To prove the second argument, we prove its contrapositive. Suppose B^c is closed. Then, all points $x \in B$ are not limit point of B^c. By the definition of limit point, this implies that there exists $\epsilon > 0$ such that $(x-\epsilon, x+\epsilon) \cap (B^c \backslash \{x\}) = \emptyset$, which means that $(x-\epsilon, x+\epsilon) \in B$. Therefore, B is open.

□

Therefore, by the preceding theorem, and consider theorem 3.1.3, using **De Morgan's Law**,

$$\left(\bigcup_{i\in I} E_i\right)^c = \bigcap_{i\in I} E_i^c, \quad \text{and} \quad \left(\bigcap_{i\in I} E_i\right)^c = \bigcup_{i\in I} E_i^c$$

we can easily get the following corollary.

Corollary 3.1.13: Topological Property of closed set

1. The union of **a finite collection** of closed sets is closed.
2. The intersection of **arbitrary collection** of closed sets is closed.

3.2 Compactness on the real line

3.2.1 Compact Sets

Definition 3.2.1: Compact Set

A set $K \subseteq \mathbb{R}$ is **compact** if every sequence in K has a subsequence that converges to a limit that is also in K.

This is the usual definition of compact set, and it can be generalized into spaces other than $\mathbb{R}$. However, sometimes it is difficult to manipulate. The following theorem gives a good property of compact sets in $\mathbb{R}$.

Theorem 3.2.2: Characterization of Compact Sets in $\mathbb{R}$

A set $K \subseteq \mathbb{R}$ is compact if and only if it is closed and bounded.

Proof.
($\Longrightarrow$) Suppose K is compact. We will first prove that K is bounded. Suppose for contradiction, K is unbounded. Then, we can find $x_n \in K$ such that $|x_n| > n$ for all $n \in \mathbb{N}$. The sequence (x_n) does not have convergent subsequence, which is a contradiction.
Now we show that K is closed. Let x be a limit point of K. Then there exists $(x_n) \in K$ with $x_n \neq x$ such that $x = \lim x_n$. Then, since K is compact, (x_n) has a convergent subsequence (x_{n_k}). Since (x_n) is convergent, (x_{n_k}) converges to the same point x. By the definition of compactness, $x \in K$. Therefore, K is closed.
($\Longleftarrow$) Suppose K is closed and bounded. Let $(x_n) \in K$. By Bolzano-Weierstrass Theorem, it must contain a convergent subsequence (x_{n_k}). Suppose $(x_{n_k}) \to x$. Since K is closed, we must have $x \in K$. Therefore, K is compact. □

3.2.2 Axiom of Completeness IV: Heine-Borel Theorem

Now, excitingly, we are prepared to state our **last** Axiom of completeness. There is some terminologies needed to be stated first.

Definition 3.2.3: Open Cover/Finite Subcover

Let $A \subseteq \mathbb{R}$.

- An **open cover** of A is a collection of open sets $\{G_i : i \in I\}$ such that $A \subseteq \bigcup_{i\in I} G_i$.
- Given an open cover of A, a **finite subcover** is a finite subcollection of open sets from the original open cover $\{G_{i_k} : i_k \in I\} \subseteq \{G_i : i \in I\}$ such that $A \subseteq \bigcup_{i_k \in I} G_{i_k}$.

Theorem 3.2.4: Heine-Borel Theorem

Let $K \subseteq \mathbb{R}$. K is compact if and only if every open cover for K has a finite subcover.

Proof.
($\Longleftarrow$) Suppose every open cover for K has a finite subcover. We are going to show that K is closed and bounded. To show that K is bounded, we construct an open cover $\{G_x : x \in K\}$ by letting $G_x = (x-1, x+1)$ for all $x \in K$. This must have a finite subcover $\{G_{x_1}, G_{x_2}, G_{x_3}, \cdots, G_{x_n}\}$. Since each G_{x_k} is bounded, $\bigcup_k G_{x_k}$ is bounded. Thus, $K \subseteq \bigcup_k G_{x_k}$ must be bounded.
To prove that K is closed, we prove by contradiction. Let $(y_n) \in K$ with $y = \lim y_n$. To show that K is closed, we must argue that $y \in K$. Therefore, suppose for contradiction, that $y \notin K$. This indicates $|x-y| > 0$. Therefore, we can construct an open cover $\{G_x : x \in K\}$ by taking

$$G_x = \left(x - \frac{|x-y|}{2}, x + \frac{|x-y|}{2}\right)$$

This must have a finite subcover $\{G_{x_1}, G_{x_2}, G_{x_3}, \cdots, G_{x_n}\}$. If we set

$$\epsilon_0 = \min\left\{\frac{|x_i - y|}{2} : 1 \leqslant i \leqslant n\right\}$$

Because $(y_n) \to y$, we can find $N \in \mathbb{N}$ such that $|y_N - y| < \epsilon_0$. But such a y_N must necessarily be excluded from each G_{x_i}, thus,

$$y_N \notin \bigcup_{i=1}^{n} G_{x_i}$$

Thus, our supposed subcover does not actually cover all of K, which is a contradiction.
($\Longrightarrow$) Now suppose K is compact. Also, we prove by contradiction. Let $\{G_i : i \in I\}$ be an open cover of K. Suppose for contradiction, that no finite subcover exists.
Since K is compact, it is closed and bounded. Then, there exists a closed interval $I_0 = [a, b]$ such that $K \subseteq I_0$. Bisect I_0 such that it is divided in $[a, (a+b)/2]$ and $[(a+b)/2, b]$. Since there is no finite subcover of $\{G_i\}$, one of the set

$$K \cap [a, (a+b)/2] \quad \text{or} \quad K \cap [(a+b)/2, b]$$

cannot be covered by finitely many sets in $\{G_i\}$ (Otherwise, the union of two finite subcover would be a finite subcover of K). Let I_1 be the bisection of I_0 such that $K \cap I_1$ does not have a finite subcover of $\{G_i\}$. Continuing this fashion, we can construct a nested sequence of closed intervals $I_0 \supseteq I_1 \supseteq I_2 \supseteq \cdots$

such that for each n, $I_n \cap K$ cannot be finitely covered, and

$$\lim |I_n| = \lim \frac{1}{2^n} |I_0| = 0$$

where $|I_n|$ denotes the length of the set I_n. Note that $\{I_n \cap K\}_n$ is a sequence of nested compact sets (since finite intersection of closed sets are closed, and all these sets are bounded). We want to show that $\bigcap_n (I_n \cap K) \neq \emptyset$. Let $K_n = I_n \cap K$. For each $n \in \mathbb{N}$, pick $x_n \in K_n$. Then $(x_n) \in K_1$ since the sets K_n are nested. Therefore, (x_n) must contain a convergent subsequence (x_{n_k}) with limit $x = \lim x_{n_k} \in K_1$. In fact, $x \in K_n$ for every K_n for essentially the same reason (just consider the set $(x_k)_{k \geqslant n}$). Therefore, $x \in \bigcap_{n=1}^{\infty} K_n$. Therefore, there exists $x \in K$, such that $x \in I_n$ for each n. Because $x \in K$, there must exist an open set G_{i_0} from the open cover that contains x as an element. Since $\lim |I_n| = 0$, there must exist $N \in \mathbb{N}$ such that for all $n > N$, $I_n \subseteq G_{i_0}$, which is a contradiction with that each $I_n \cap K$ cannot be finitely covered by the subcover of $\{G_i\}$. $\square$

Now we have a complete sight of the **six Axiom of Completeness** of real numbers.

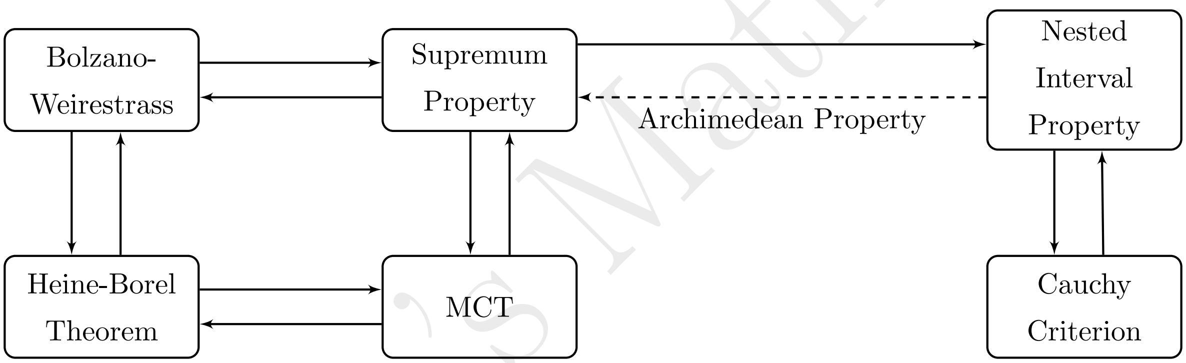


Figure 3.3: Relation between Axiom of Completeness

The two weak axioms are:

- Nested Interval Property
- Cauchy Criterion

and they does not imply Archimedean Property. The four strong axioms are:

- Bolzano-Weierstrass Theorem
- Supremum Property
- Monotone Convergence Theorem
- Heine-Borel Theorem

and from them, Archimedean Property can be derived.

3.3 The Cantor Set

One of the most important example in real line topology is **the Cantor Set**. It is a pathological set endowed with many surprising properties.

Let C_0 be the closed interval $[0, 1]$. Let C_1 be the set that the open middle third of C_0 is removed, i.e.,

$$C_1 = C_0 \backslash \left(\frac{1}{3}, \frac{2}{3} \right) = \left[0, \frac{1}{3} \right] \cup \left[\frac{2}{3}, 1 \right]$$

Construct C_2 to be the set that the open middle third of two parts of C_1 is removed, i.e.,

$$C_2 = \left[0, \frac{1}{9} \right] \cup \left[\frac{2}{9}, \frac{1}{3} \right] \cup \left[\frac{2}{3}, \frac{7}{9} \right] \cup \left[\frac{8}{9}, 1 \right]$$

Continuing this fashion, we can get a sequence of set $\{C_n\}$ for all $n \in \mathbb{N}$. C_n consists of 2^n closed intervals with each having length $1/3^n$.

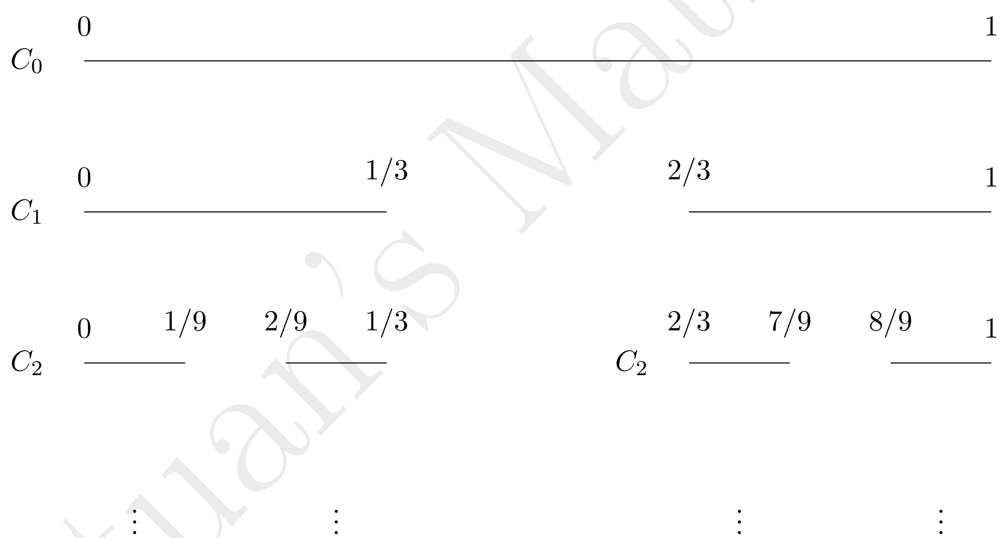


Figure 3.4: The Cantor Set

Definition 3.3.1: Cantor Set

The **Cantor Set** C is the intersection

$$C = \bigcap_{n=0}^{\infty} C_n$$

Before we formally state its properties, we first introduce another terminology about the topology of sets.

3.3.1 Perfect Sets

Definition 3.3.2: Perfect Set

A set $P \subseteq \mathbb{R}$ is **perfect** if it is closed and contains no isolated points.

Clearly the closed intervals $[a, b]$ is an example of perfect set. The most important observation about perfect sets is that, they are uncountable.

Theorem 3.3.3: Perfect Sets are uncountable

A nonempty perfect set is uncountable.

Proof. Notice that if P is perfect and nonempty, then it must be infinite because otherwise it would consist only of isolated points. Therefore, suppose for contradiction, P is countable. Then, we can write

$$P = \{x_1, x_2, x_3, \cdots\}$$

Let I_1 be a closed interval that contains x_1 in its interior (i.e., $x_1 \in I_1^\circ$). Since x_1 is not isolated, there exists some other point $y_2 \in P$ such that $y_2 \in I_1^\circ$. Construct a closed interval centered on y_2, such that $I_2 \subseteq I_1$ but $x_1 \notin I_2$. This can be done by taking

$$\epsilon = \min\{y_2 - a, b - y_2, |x_1 - y_2|\}$$

and let $I_2 = [y_2 - \epsilon/2, y_2 + \epsilon/2]$.

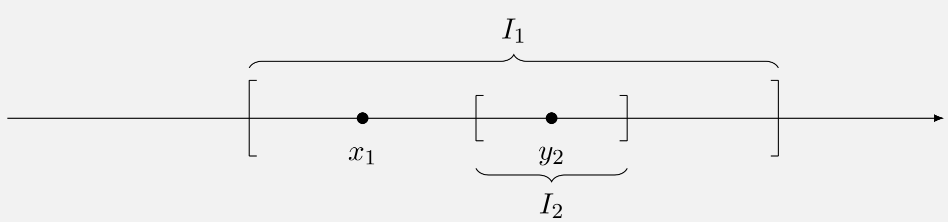


Figure 3.5: Construction of I_1 and I_2

Continuing this fashion, since $y_2 \in P$ is not isolated, there must exist another point $y_3 \in P$ in I_2°. We can choose $y_3 \neq x_2$, and construct I_3 centered on y_3 such that $x_2 \notin I_3$ and $I_3 \subseteq I_2 \cdots$
If we do this iteratively, the result sequence is a sequence of closed intervals I_n satisfying

- $I_{n+1} \subseteq I_n$
- $x_n \notin I_{n+1}$
- $I_n \cap P \neq \emptyset$ since at least $y_n \in I_n \cap P$.

Let $K_n = I_n \cap P$. Then, K_n is compact for all n. As shown in the proof of Heine-Borel Theorem, the intersection of nested compact sets is not empty, i.e.,

$$\bigcap_{i=1}^{\infty} K_n \neq \emptyset$$

However, K_n is a subset of P, and $x_n \notin I_{n+1}$ leads to the result that $\bigcap_{i=1}^{\infty} K_n = \emptyset$, which is a contradiction. □

3.3.2 Properties of the Cantor Set

Now we are ready to see some fancy properties of the Cantor Set. First of all, we can intuitively see that the Cantor Set is a very 'small' set, in the sense of its length. For each C_n, it has 2^n closed intervals with each having length $1/3^n$. Therefore, $|C_n| = (2/3)^n$. Therefore, length of the Cantor set is then

$$|C| = \lim_{n\to\infty} |C_n| = 0$$

However, in the sense of cardinality, it is a 'large set'. We are going to show that C is a perfect set, thus it is an uncountable set.

Proposition 3.3.4: Cantor Set is Perfect

Cantor Set is a perfect set. Therefore, it is uncountable.

Proof. Since each C_n is closed, by Corollary 3.1.13, $C = \bigcap_n C_n$ is closed.
Let $x \in C$ be arbitrary. Then, $x \in C_n$ for all $n \in \mathbb{N}$. Because $x \in C_1$, it must be contained in one of the closed interval in C_1, i.e., $[0, 1/3]$ or $[2/3, 1]$. Suppose without losing generality, that $x \in [0, 1/3]$. There are at least two points in $[0, 1/3] \cap C$, i.e., the two endpoints 0 and $1/3$. Therefore, there must exist another point $x_1 \in C \cap C_1$ with $x_1 \neq x$ such that $|x - x_1| \leqslant 1/3$.
Similarly, because $x \in C_2$, it must be contained in one of the four closed interval in C_2, each interval has length $1/3^2$. For the interval which x is contained in, at least the two end points are contained in C. Therefore, there exists $x_2 \in C \cap C_2$ with $x_2 \neq x$ such that $|x - x_2| \leqslant 1/3^2$.
Continuing this fashion, for each $n \in \mathbb{N}$, there exists $x_n \in C \cap C_n = C$ different from x, satisfying $|x - x_n| \leqslant 1/3^n$. This shows that x is a limit point, it is not isolated. □

There is another elegant way to prove that C is uncountable. This method recalls us back to the topic of cardinality.

Proof. A real number $x \in [0,1]$ in **Ternary numeral system** is

$$x = \sum_{n=1}^{\infty} \frac{c_n}{3^n}, \quad c_n = \{0,1,2\}$$

Just as in decimal expression, if we ban the use of infinitely many $2's$ at the tail of the series expressing x (this kind of number has another ternary expression, with infinitely many $0's$ at the tail), then each x has a unique ternary expression. This means that $x \in [0,1]$ has a bijection to the sequence $\{c_n\}$.
Note that in C_1, the middle third is excluded, this means that $c_1 = \{0,2\}$. If it belongs to the left interval, $c_1 = 0$. If right, $c_1 = 2$. Similarly, in C_2, the middle third of the two closed intervals are all excluded, this means that $c_2 = \{0,2\}$, $\cdots$ Inductively, the points in Cantor Set has a bijection to the sequence $\{c_n\}$ where $c_n = 0$ or 2 for all n.
Here we use the **Cantor's Diagonalization Method**. Suppose for contradiction, that C is countable. Then, C can be expressed as

$$C = \{x_1, x_2, x_3, \cdots\}$$

By the definiton of cardinality, there exists a bijection from $\mathbb{N}$ to the set C. Suppose for x_n, its corresponding ternary expression is denoted as $\{c_n k\}_{k=1}^{\infty}$. We can construct a bijection like below.

$$\begin{array}{cccccccccc}
\mathbb{N} & & C & & & & & & & \\
1 & \longleftrightarrow & x_1 & = & .\mathbf{c_{11}} & c_{12} & c_{13} & c_{14} & \cdots \\
2 & \longleftrightarrow & x_2 & = & .c_{21} & \mathbf{c_{22}} & c_{23} & c_{24} & \cdots \\
3 & \longleftrightarrow & x_3 & = & .c_{31} & c_{32} & \mathbf{c_{33}} & c_{34} & \cdots \\
4 & \longleftrightarrow & x_4 & = & .c_{41} & c_{42} & c_{43} & \mathbf{c_{44}} & \cdots \\
\vdots & & \vdots & & \vdots & \vdots & \vdots & \vdots & \ddots
\end{array}$$

Now, we can set an element $b \in C$ such that its ternary expression is $\{b_n\}$.

$$b_n = \begin{cases} 0, & \text{if } a_{nn} = 2 \\ 2, & \text{if } a_{nn} = 0 \end{cases}$$

Clearly, this b_n does not appear in this bijection, which is a contradiction. Therefore, C is uncountable. $\square$

Therefore, the Cantor set has a pathological property, where it is 'small' in some sense, but very 'large' in another sense.

Chapter 4

Continuity of Functions on $\mathbb{R}$

4.1 Limit of Functions

We have examined the limit of a sequence. You can see the limit of a sequence as a 'discrete' process, where each step is length 1. The limit of a function is a 'continuous' process.

Definition 4.1.1: Limit of Function

Let $f : A \to \mathbb{R}$. Let c be a limit point of A. We say $\lim_{x\to c} f(x) = L$ if

$$\forall\, \epsilon > 0, \exists\, \delta > 0, \text{ s.t. } 0 < |x - c| < \delta \quad \Longrightarrow \quad |f(x) - L| < \epsilon$$

This is another classical example of $\epsilon - \delta$ language used in analysis. To use this language, we need to first fix an $\epsilon > 0$, then find corresponding $\delta > 0$ such that the requirement is meet.

Example 4.1.2: Example of calculating limit using definition

Prove for $g : \mathbb{R} \to \mathbb{R}$, $g(x) = x^2$,

$$\lim_{x\to c} g(x) = c^2$$

Proof. Let $\epsilon > 0$ be arbitrary. We need to find $\delta > 0$ such that whenever $0 < |x - c| < \delta$, we have

$$|g(x) - c^2| = |x - c||x + c| < \epsilon$$

We can easily control the size of $|x - c|$. The difficult point is $|x + c|$. The crucial point is that, δ should be very small so that it can fix arbitrarily small ϵ. Therefore, if $\delta < 1$, then $|x - c| < 1$, then

$$|x + c| = |x - c + 2c| \leqslant |x - c| + 2|c| < 1 + 2|c|$$

Therefore, if we choose

$$\delta = \min\left\{\frac{\epsilon}{2|c|+1}, 1\right\}$$

Then, we will have, if $0 < |x-c| < \delta$,

$$|g(x) - c^2| = |x-c||x+c| < \frac{\epsilon}{2|c|+1} \times (2|c|+1) = \epsilon$$

which is our desired result. □

There is a strong relationship between sequence limit and function limit. This relationship has a important and useful application on detecting a point where the function limit does not exist.

Theorem 4.1.3: Sequential Criterion for Function Limit

Let $f : A \to \mathbb{R}$. Let c be a limit point of A. the following are equivalent:

1. $\lim_{x\to c} f(x) = L$
2. For all sequences $(x_n) \subseteq A$ satisfying $x_n \neq c$ for all n and $(x_n) \to c$, it follows that $f(x_n) \to L$.

Proof.
($\Longrightarrow$) Suppose $\lim_{x\to c} f(x) = L$. Consider an arbitrary sequence (x_n) which converges to c and satisfies $x_n \neq c$ for all n. Let $\epsilon > 0$. Then there exists $\delta > 0$ such that whenever $0 < |x-c| < \delta$, we have $|f(x) - L| < \epsilon$. Since $(x_n) \to c$, there does exists $N \in \mathbb{N}$ such that for all $n \geqslant N$, $0 < |x_n - c| < \delta$. If follows that for all $n \geqslant N$, $|f(x_n) - L| < \epsilon$.
($\Longleftarrow$) To prove the inverse direction, we prove by contradiction. Suppose (2) is true, and suppose for contradiction, (1) is wrong. To say that

$$\lim_{x\to c} f(x) \neq L$$

means that there exists at least one particular $\epsilon_0 > 0$ such that for all $\delta > 0$ with $0 < |x-c| < \delta$, we have $|f(x) - L| > \epsilon_0$. Now consider $\delta_n = 1/n$. From the preceding discussion, it follows that for each $n \in \mathbb{N}$ we may pick $0 < |x_n - c| < \delta_n$ and $|f(x_n) - L| > \epsilon_0$. However, this (x_n) is a sequence that does not converge to L, which is a contradiction to condition (2). □

Using this relation between functional and sequential limit, we have the following application of establishing that certain limits do not exist.

Corollary 4.1.4: Divergence Criterion of Function Limit

Let $f : A \to \mathbb{R}$. Let c be a limit point of A. If there exists two sequences (x_n) and (y_n) in A with $x_n \neq c$ and $y_n \neq c$ for all n, and

$$\lim x_n = \lim y_n = c, \quad \text{but} \quad \lim f(x_n) \neq \lim f(y_n)$$

Then the function limit $\lim_{x\to c} f(x)$ does not exist.

As a classic example, consider the limit $\lim_{x\to 0} \sin(1/x)$. Let

$$x_n = \frac{1}{2n\pi}, \quad y_n = \frac{1}{2n\pi + \frac{\pi}{2}}$$

Then both $\lim(x_n) = \lim(y_n) = 0$. However, $\sin(1/x_n) = 0$ for all $n \in \mathbb{N}$ and $\sin(1/y_n) = 1$ for all $n \in \mathbb{N}$. Therefore,

$$\lim \sin(1/x_n) \neq \lim \sin(1/y_n)$$

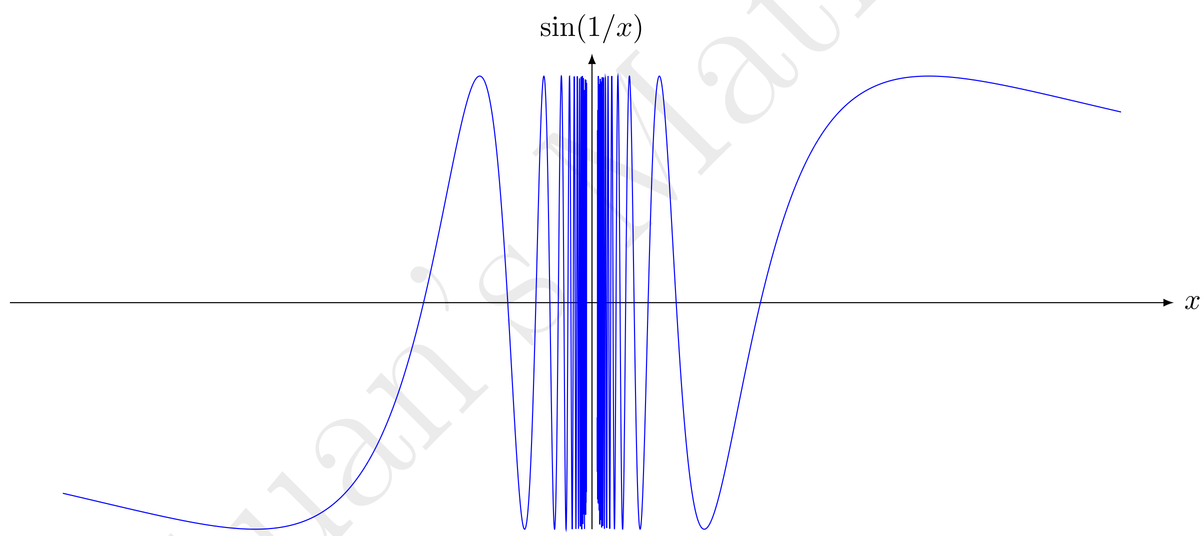


Figure 4.1: Graph of $\sin(1/x)$

There is another simple and straight result that can be proved easily using the relation 4.1.3. The proof idea is the same with the proof of Algebraic Limit Theorem for sequences, just use 4.1.3 to transform the function limit to sequence limit.

Corollary 4.1.5: Algebraic Limit Theorem for Function Limit

Let $f, g : A \to \mathbb{R}$ on the same domain $A \subseteq \mathbb{R}$. Let $\lim_{x\to c} f(x) = L$ and $\lim_{x\to c} g(x) = M$ for limit point $c \in A$. Then,

1. $\lim_{x\to c} kf(x) = kL$ for all $k \in \mathbb{R}$.
2. $\lim_{x\to c}(f(x) + g(x)) = L + M$.
3. $\lim_{x\to c}(f(x)g(x)) = LM$.
4. $\lim_{x\to c} f(x)/g(x) = L/M$ provided $M \neq 0$.

4.2 Continuity of Functions

Definition 4.2.1: Continuity of Function

$f : A \to \mathbb{R}$ is **continuous** at $c \in A$ if

$$\forall \epsilon > 0, \exists \delta > 0 \text{ s.t. } |x - c| < \delta \implies |f(x) - f(c)| < \epsilon$$

Note that the only difference of the definition of continuity compared with the definition of function limit, is that L is replaced by $f(c)$, so that it is ensured the function is convergent to the point it exactly 'should be'.

Naturally, we can define that f is continuous on A if it is continuous at every $c \in A$. Moreover, there is another way of defining continuity.

Corollary 4.2.2: Characterization of Continuity

$f : A \to \mathbb{R}$ is continuous at $c \in A$ if

$$\lim_{x\to c} f(x) = f(c)$$

Similarly, the continuity of functions behaves good algebraic properties, and there is a way using sequence limit to define discontinuity.

Corollary 4.2.3: Algebraic Continuity Theorem

Let $f, g : A \to \mathbb{R}$ are continuous at $c \in A$. Then, for all $k \in \mathbb{R}$, $kf(x)$, $f(x) + g(x)$ and $f(x)g(x)$ are all continuous at c. Provided that $g(x) \neq 0$, $f(x)/g(x)$ is also continuous at c.

Corollary 4.2.4: Discontinuity Criterion

Let $f : A \to \mathbb{R}$. Let $c \in A$ be a limit point of A. If there exists a sequence $(x_n) \subseteq A$ such that $(x_n) \to c$ but $f(x_n)$ does not converge to $f(c)$, then f is not continuous at c.

As another classic example, consider the limit $\lim_{x\to 0} g(x)$ where

$$g(x) = \begin{cases} x\sin(1/x), & \text{if } x \neq 0 \\ 0, & \text{if } x = 0 \end{cases}$$

Since

$$|g(x) - g(0)| = |x\sin(1/x)| \leqslant |x|$$

Given $\epsilon > 0$, set $\delta = \epsilon$, then whenever $|x - 0| = |x| < \delta$, we have $|g(x) - g(0)| < \epsilon$.

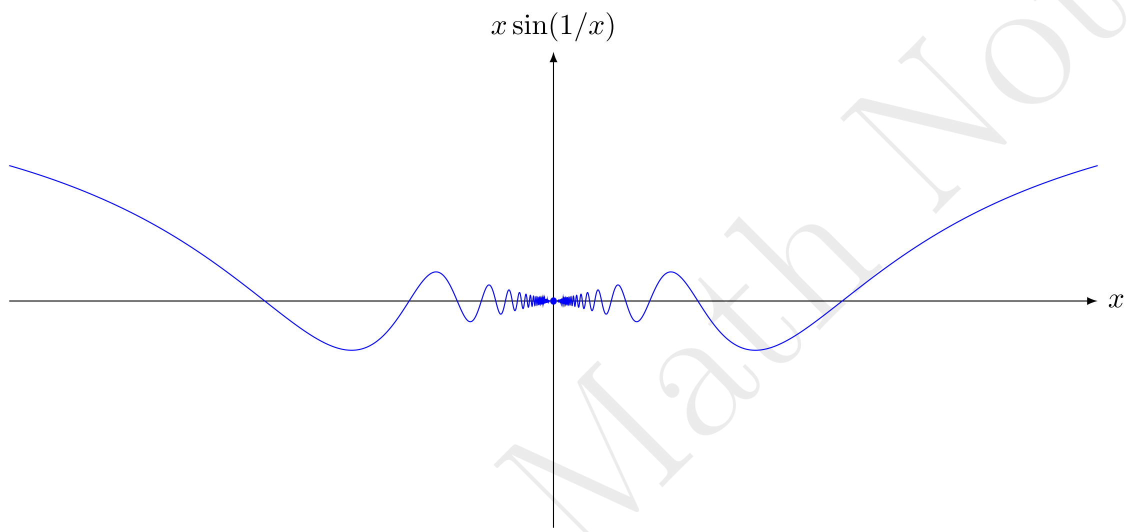


Figure 4.2: Graph of $x\sin(1/x)$

There is another fancy topological way to define continuity of functions. This definition would be extremely important in general topology, but here, we just make an introduction.

Theorem 4.2.5: Topological Definition of Continuity

Let $f : \mathbb{R} \to \mathbb{R}$. Define the **preimage (inverse image)** of f on $B \subseteq \mathbb{R}$ by

$$f^{-1}(B) = \{x \in \mathbb{R} : f(x) \in B\}$$

Then, f is continuous if and only if $f^{-1}(G)$ is open for all open set $G \subseteq \mathbb{R}$.

Proof.
Before proving all these, we prove a simple lemma about map and its preimage: $f(A) \subseteq B$ if and only

if $A \subseteq f^{-1}(B)$, this is true since

$$f(A) \subseteq B \implies A \subseteq f^{-1}(f(A)) \subseteq f^{-1}(B) \text{ and } A \subseteq f^{-1}(B) \implies f(A) \subseteq B$$

($\Longrightarrow$) Suppose f is continuous. Then,

$$\forall \epsilon > 0, \exists \delta > 0, \text{ .s.t } |x - c| < \delta \implies |f(x) - f(c)| < \epsilon$$

Let $G \subseteq \mathbb{R}$ be an arbitrary open set. Let $c \in f^{-1}(G)$. Then $f(c) \in G$. Since G is open, there exists $\epsilon > 0$ such that

$$(f(c) - \epsilon, f(c) + \epsilon) \subseteq G$$

By the assumption of continuity, there exists $\delta > 0$ such that $c - \delta < x < c + \delta$, and $f((c - \delta, c + \delta)) \subseteq (f(c) - \epsilon, f(c) + \epsilon) \subseteq G$. Therefore, we have

$$(c - \delta, c + \delta) \subseteq f^{-1}(G)$$

which shows that $f^{-1}(G)$ is open since c is arbitrary.

($\Longleftarrow$) Suppose $f^{-1}(G)$ is open for all open set $G \subseteq \mathbb{R}$. Fix $c \in \mathbb{R}$, we know that $f^{-1}((f(c) - \epsilon, f(c) + \epsilon))$ is open. Since $f(c) \in f^{-1}((f(c) - \epsilon, f(c) + \epsilon))$ meaning that there exists $\delta > 0$ such that

$$(c - \delta, c + \delta) \subseteq f^{-1}((f(c) - \epsilon, f(c) + \epsilon))$$

which implies

$$f((c - \delta, c + \delta)) \subseteq (f(c) - \epsilon, f(c) + \epsilon)$$

and this shows that f is continuous. □

To end this section, we examine the composition of two continuous functions.

Proposition 4.2.6: Composition of Continuous Functions

Let $f, g : A \to \mathbb{R}$. Assume the range $f(A) = \{f(x) : x \in A\} \subseteq B$ so that $g \circ f(x) = g(f(x))$ is defined on A. If f is continuous at $c \in A$, g is continuous at $f(c) \in B$, then $g \circ f$ is continuous at c.

Proof. Let $\epsilon > 0$ be arbitrary. Our goal is to show that whenever $|x - c| < \delta$, we have $|g(f(x)) -$

$g(f(c))| < \epsilon$. Since g is continuous at $f(c)$, we have

$$\exists\, \delta_1 > 0,\ \text{s.t. } |y - f(c)| < \delta_1 \quad \Longrightarrow \quad |g(y) - g(f(c))| < \epsilon$$

Since f is continuous at c, we have

$$\exists\, \delta_2 > 0,\ \text{s.t. } |x - c| < \delta_2 \quad \Longrightarrow \quad |f(x) - f(c)| < \epsilon$$

Let $y = f(x)$ in the first condition, and combining these two, we have

$$\exists\, \delta_2 > 0,\ \text{s.t. } |x - c| < \delta_2 \quad \Longrightarrow \quad |f(x) - f(c)| < \delta_1 \quad \Longrightarrow \quad |g(f(x)) - g(f(c))| < \epsilon$$

Therefore, $g \circ f$ is continuous at c. □

4.3 Uniform Continuity

There is another terminology of continuity defined as below.

Definition 4.3.1: Uniform Continuity

$f : A \to \mathbb{R}$ is **uniformly continuous** on A if

$$\forall\, \epsilon > 0, \exists\, \delta > 0,\ \text{s.t. } \forall\, x, y \in A, |x - y| < \delta \quad \Longrightarrow \quad |f(x) - f(y)| < \epsilon$$

This is a **strictly stronger** property than continuity. If you compare this definition with that one of continuity, you can see that the only difference is that the definition of uniform continuity does not set a specific point $c \in A$. If we fix $y = c$ in the definition of uniform continuity, it becomes the definition of continuity.

Then what does uniform continuity say? We can see from previous example that when we fix ϵ, and try to find δ that satisfies the definition, δ may be dependent on the point c. **However, for uniform continuity, δ does not dependent on any fixed point, and we can choose the same δ for the continuity of any point. This is why it is called 'uniform'**. This can be better understood by examples.

Example 4.3.1. Consider the function

$$f(x) = 2x + 3$$

It is obviously continuous everywhere. Here we show that by definition. Let $c \in \mathbb{R}$ be arbitrary. Fix $\epsilon > 0$. Now, if we choose $\delta = \epsilon/2$, and let $|x - c| < \delta$, we have

$$|f(x) - f(c)| = |2x + 3 - 2c - 3| = 2|x - c| < \epsilon$$

which shows that it is continuous. Moreover, note that $\delta = \epsilon/2$ does not depend on the point c, which shows that it is uniformly continuous. If we use the definition of uniform continuity, and choose $\delta = \epsilon/2$, with $|x-y| < \delta$, we indeed have

$$|f(x) - f(y)| = |2x + 3 - 2y - 3| = 2|x - y| < \epsilon$$

which shows that it is uniformly continuous.

Example 4.3.2. Consider the function

$$g(x) = x^2$$

It is obviously continuous everywhere. Recall Example 4.1.2, if we choose

$$\delta = \min\left\{\frac{\epsilon}{2|c|+1}, 1\right\}$$

we can prove the continuity of $g(x)$. However, note that in this case, δ does depend on c, for larger $|c|$, we need smaller δ to control the deviation of x from the limit point. This is because $g(x) = x^2$ increases more and more rapidly at two tails, and it needs more narrower constraint of variable to control the difference between function value and the function limit.

To disprove the uniform continuity of $g(x) = x^2$, suppose there exists a $\delta > 0$ that satisfies the expected condition. Then, $\epsilon/2 < |x - y| < \delta$ would be enough. If we let $x, y > \epsilon/\delta$, then

$$|g(x) - g(y)| = |x - y||x + y| > \frac{\epsilon}{2}\frac{2\epsilon}{\delta} = \epsilon$$

which is a contradiction. Besides this, there is also a sequential way to disprove the uniform continuity.

Corollary 4.3.2: Sequential Criterion for disproving uniform continuity

Let $f : A \to \mathbb{R}$. If there exists $\epsilon_0 > 0$ and two sequences (x_n) and (y_n) in A such that

$$|x_n - y_n| \to 0 \quad \text{but} \quad |f(x_n) - f(y_n)| \geqslant \epsilon_0$$

Then f is not uniformly continuous.

Example 4.3.3. Consider the function

$$h(x) = \sin\left(\frac{1}{x}\right), x \in (0, 1)$$

We have seen that it does not have limit at 0. It is also not uniformly continuous on this interval, and the problem occurs again around point 0. It is oscillated faster and faster around the origin. Consider two sequences

$$x_n = \frac{1}{2n\pi} \quad \text{and } y_n = \frac{1}{2n\pi + \frac{\pi}{2}}$$

Each sequence tends to 0, so we would have $|x_n - y_n| \to 0$. However,

$$|h(x_n) - h(y_n)| = 1 \text{ for all } n \in \mathbb{N}$$

Therefore it is not uniformly continuous ($\epsilon_0 = 1/2$ would be enough).

Example 4.3.4. Although $g(x) = x^2$ is not uniformly continuous on $\mathbb{R}$, it is uniformly continuous on a bounded set, for example, on $[-1, 1]$. In this case, $|x + y| \leqslant 2$ for all x and y. Therefore, we can take $\delta = \epsilon/2$, and let $|x - y| < \delta$, then

$$|g(x) - g(y)| = |x - y||x + y| < \left(\frac{\epsilon}{2}\right) 2 = \epsilon$$

which shows that it is uniformly continuous.

From previous examples, there may now be a though in your mind that, a function with finite 'slope' would be uniformly continuous (though we have not rigorously define the derivative of a function). The next definition and proposition would be related to this thought.

Definition 4.3.3: Lipschitz Function

A function $f : A \to \mathbb{R}$ is called **Lipschitz** if there exists $M > 0$ such that

$$\left|\frac{f(x) - f(y)}{x - y}\right| \leqslant M$$

for all $x \neq y \in A$.

Proposition 4.3.4: Lipschitz and Uniform Continuity

If $f : A \to \mathbb{R}$ is Lipschitz, then it is uniformly continuous on A. The converse statement is not true.

Proof.

- Fix $\epsilon > 0$. If f is Lipschitz, then we can choose $\delta = \epsilon/M$, so that whenever $|x - y| < \delta$, we have

$$|f(x) - f(y)| \leqslant M\delta = \epsilon$$

for all $x, y \in A$, which shows that it is uniformly continuous.

- The converse is not true, for example, consider the function

$$f(x) = \sqrt{x}, x \in [0, 1]$$

We can see that $\sqrt{x} + \sqrt{y} \leqslant 2$ in this interval. Therefore, fix $\epsilon > 0$, let $\delta = \epsilon^2$, then whenever $|x - y| < \delta$, we have

$$|\sqrt{x} - \sqrt{y}|^2 \leqslant |\sqrt{x} - \sqrt{y}||\sqrt{x} + \sqrt{y}| = |x - y| < \epsilon^2$$

Therefore, $|\sqrt{x} - \sqrt{y}| < \epsilon$, which shows that it is uniformly continuous. However, it is not Lipschitz, since

$$\left|\frac{f(x) - f(y)}{x - y}\right| = \left|\frac{\sqrt{x} - \sqrt{y}}{x - y}\right| = \left|\frac{1}{\sqrt{x} + \sqrt{y}}\right|$$

If we fix $y = 0$, we can see that for all $M \in \mathbb{N}$, we can find $x < 1/M^2$ such that

$$\left|\frac{f(x) - f(y)}{x - y}\right| = \left|\frac{1}{\sqrt{x} + \sqrt{y}}\right| > M$$

Therefore, the slope is unbounded.

□

To end up this section, we examine some algebraic properties of uniformly continuous functions. Since it is a stronger property than continuity, the algebraic properties are not quite well-behaved compared to those continuous function. Uniform continuity may lose after doing some algebraic operations.

Proposition 4.3.5: Algebraic Properties of Uniform Continuity

Let $f, g : A \to \mathbb{R}$ be uniformly continuous on A. Then,

1. $f(x) + g(x)$ and $f(g(x))$ are uniformly continuous.
2. Provided that they are defined, $f(x)g(x)$ and $f(x)/g(x)$ may not be uniformly continuous.

Proof.

1. Fix $\epsilon > 0$. For f, g both uniformly continuous, there exists $\delta > 0$ such that when $|x - y| < \delta$, we have $|f(x) - f(y)| < \epsilon/2$ and $|g(x) - g(y)| < \epsilon/2$. Therefore,

$$|f(x) + g(x) - f(y) - g(y)| \leqslant |f(x) - f(y)| + |g(x) - g(y)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} > \epsilon$$

which shows that $f + g$ is uniformly continuous.
Fix $\epsilon > 0$. Since f is uniformly continuous,

$$\exists\, \delta_1 > 0, \text{ s.t. } |g(x) - g(y)| < \delta_1 \quad \Longrightarrow \quad |f(g(x)) - f(g(y))| < \epsilon$$

Since g is uniformly continuous,

$$\exists\, \delta_2 > 0, \text{ s.t. } |x - y| < \delta_2 \implies |g(x) - g(y)| < \delta_1$$

Combining these two, we can see that $f \circ g$ is uniformly continuous.

2. For $f(x) = g(x) = x$, they are both uniformly continuous, but $f(x)g(x) = x^2$ is not. For $f(x) = 1$, $g(x) = x$, they are both uniformly continuous, but $f(x)/g(x)$ is not.

□

4.4 Properties Derived from Continuity

4.4.1 Continuous Functions on Compact Sets

Continuous functions on compact sets have more good properties than continuous functions on some general sets. In this section we will prove some properties of this kind of functions.

We first note that, for a continuous function f, the image

$$f(A) = \{f(x) : x \in A\}$$

on an open interval A may not be an open interval again. For example, $f(x) = x^2$ on open interval $(0, 1)$ has image $[0, 1)$, which is neither open nor closed.

For images on closed intervals, this is also the case. Consider the function

$$g(x) = \frac{1}{x^2 + 1}$$

and the closed set $B = [0, \infty)$. The image on B is $(0, 1]$, which is neither open nor closed.

However, a continuous function will preserve the compactness of a set after mapping into its image.

Theorem 4.4.1: Preservation of Compactness on Continuous Map

Let $f : A \to \mathbb{R}$ be continuous on A. Let $K \subseteq A$ be compact. Then, $f(K)$ is also compact.

Proof. Suppose $(y_n) \subseteq f(K)$. The goal is to prove that there exists a subsequence (y_{n_k}) such that the limit $y = \lim y_{n_k}$ also belongs to $f(K)$.

Since $y_n \in f(K)$, for each n, there exists a $x_n \in K$ such that $f(x_n) = y_n$. This yields a sequence $(x_n) \subseteq K$. Since K is compact, there exists a subsequence (x_{n_k}) such that $(x_{n_k}) \to x \in K$. Since f is continuous, by Theorem 4.1.3, we have that $f(x_{n_k}) = y_{n_k} \to f(x)$. Let $y = f(x)$. Since $x \in K$, we have $f(x) \in f(K)$. Therefore, there exists a subsequence y_{n_k} such that it converges to a point in $f(K)$, which shows that $f(K)$ is compact. ☐

An important and famous corollary of this proerty is the extreme value theorem.

Corollary 4.4.2: Extreme Value Theorem

If $f : K \to \mathbb{R}$ is continuous on a compact set $K \subseteq \mathbb{R}$, then it attains its maximum and minimum value. i.e.,

$$\exists\, x_0, x_1 \in K, \text{ s.t. } \forall x \in K \implies f(x_0) \leqslant f(x) \leqslant f(x_1)$$

Proof. We first show that, if K is compact and nonempty, then $\sup K$ and $\inf K$ both exist and belong to K.

Let $s = \sup K$. By the characterization of supremum, for all $\epsilon > 0$ there exists $x \in K$ such that $s - \epsilon < x$. If we choose $\epsilon_n = 1/n$ and choose $x_n \in K$ such that $s - \epsilon_n < x_n$, taking limit on both sides, by order limit theorem,

$$\lim_{n\to\infty} (s - \epsilon_n) = s \leqslant \lim_{n\to\infty} x_n$$

Moreover, since s is supremum, for all $x_n \in K$, we must have $x_n < s$. Taking limit on both sides, by order limit theorem,

$$\lim_{n\to\infty} x_n \leqslant s$$

Therefore, by squeeze theorem, we have $(x_n) \to s$. Then each subsequence of (x_n) converges to s. Since K is compact, we have that $s \in K$. The same argument follows for $\inf K$.

Now we are ready to prove the corollary. By Theorem **??**, $f(K)$ is compact. Therefore, it attains its supremum and infimum, say s and l. Then, since $s, l \in f(K)$, there exists $x_0, x_1 \in K$ such that $l = f(x_0)$ and $s = f(x_1)$. ☐

The most important theorem of this subsection is stated as below. It says that a continuous function on a compact set is inherently uniformly continuous. This theorem is so important that it will be the main tool of proving that all continuous functions on closed bounded intervals are Riemann Integrable.

Theorem 4.4.3: Uniform Continuity of continuous Function on Compact Set

Let $f : K \to \mathbb{R}$ be continuous on compact set $K \subseteq \mathbb{R}$. Then, f is uniformly continuous on K.

Proof. We prove by contradiction. Suppose for contradiction, that f is not uniformly continuous. Then, by Corollary 4.3.2, there exists $\epsilon_0 > 0$ and two sequences (x_n) and (y_n) in K such that

$$|x_n - y_n| \to 0 \quad \text{but} \quad |f(x_n) - f(y_n)| \geqslant \epsilon_0$$

Since K is compact, there exists subsequences $(x_{n_k}), (y_{n_p})$ that it converges to $x \in K$ and $y \in K$, respectively. By Algebraic Limit Theorem,

$$\lim_{n\to\infty} y_{n_p} = \lim_{n\to\infty} (y_{n_p} - x_{n_p}) + x_{n_p} = 0 + x$$

This means that $y = x$. Since f is continuous, we have $f(x_{n_k}) = f(y_{n_p}) = x$, which shows that,

$$\lim_{n\to\infty} (f(x_{n_k}) - f(y_{n_p})) = 0$$

which is a contradiction with the assumption that $|f(x_n) - f(y_n)| \geqslant \epsilon_0$ for all $n \in \mathbb{N}$. Therefore, f is uniformly continuous. $\square$

4.4.2 Intermediate Value Theorem

Recall the starting meme of this note. Precalculus students always explain continuity as 'graph can be drawn without lifting up the pen'. This is an intuitive way to say that a continuous function goes across every point that goes between.

Theorem 4.4.4: Intermediate Value Theorem (IVT)

Let $f : [a, b] \to \mathbb{R}$ be continuous. If L satisfies $f(a) < L < f(b)$ or $f(b) < L < f(a)$, then there exists a point $c \in (a, b)$ such that $f(c) = L$.

Proof. Without losing generality, we will assume $f(a) < L < f(b)$. To make it simple, we can let $g(x) = f(x) - L$ so that our assumption becomes $g(a) < 0 < g(b)$. Our goal is to show that there exists $c \in (a, b)$ such that $g(c) = 0$. We will do this using Supremum Property.
Let

$$B = \{x \in [a, b] : g(x) \leqslant 0\}$$

Note that B is bounded above by b, and $a \in B$, so it is nonempty. By Supremum Property, its supremum $c = \sup B$ exists.

- If $g(c) > 0$. Since c is the supremum, it is the least upper bound, which means that for all $\delta > 0$, there exists $x \in B$ such that $c - \delta < x$. i.e., $c - x < \delta$ with $g(x) \leqslant 0$ (Therefore $g(c) - g(x) \geqslant g(c)$).

However, since g is continuous (by Algebraic Continuity Theorem), we have

$$\forall \epsilon > 0, \exists \delta > 0, \text{ s.t. } |x - c| < \delta \implies |g(x) - g(c)| < \epsilon$$

This is a contradiction since we cannot find $\delta > 0$ such that $|g(x) - g(c)| < g(c)$, which is a contradiction.

- If $g(c) < 0$. Since c is the supremum, it is an upper bound. That is, for all $x \in B$, we have $x \leqslant c$. However, since g is continuous, for $\epsilon = |g(c)|/2$, we can find $k = c + \delta/2$ for some $\delta > 0$ such that

$$|g(k) - g(c)| < \frac{|g(c)|}{2}$$

Since $g(c) < 0$, this shows that $g(k) < g(c) - \frac{g(c)}{2} = \frac{g(c)}{2} < 0$, which means that $k \in B$. This is again a contradiction since $k > c$.

Therefore, the only possibility is that $g(c) = 0$. To prove that $c \in (a, b)$, we only need to notice that $g(a) < 0$ and $g(b) > 0$. □

The inverse is not true. To correctly state the inverse of IVT, we need to first define the Intermediate Value Property.

Definition 4.4.5: Intermediate Value Property

A function $f : [a, b] \to \mathbb{R}$ has **intermediate value property** if for all $x, y \in [a, b]$ with $x < y$, and all L between $f(x)$ and $f(y)$, there exists $c \in (x, y)$ such that $f(c) = L$.

The inverse statement of IVT is then: All functions with intermediate value property is continuous. The obvious counterexample is

$$f(x) = \begin{cases} \sin(1/x), & \text{if } x \neq 0 \\ 0, & \text{if } x = 0 \end{cases}$$

since it is not continuous at point 0. However, if we add some more strict constraints on the function, we can arrive this beautiful result.

Definition 4.4.6: Monotone Function

A function $f : A \to \mathbb{R}$ is **increasing** on A if for all $x, y \in A$ with $x < y$, we have $f(x) \leqslant f(y)$. It is **decreasing** on A if for all $x, y \in A$ with $x < y$, we have $f(x) \geqslant f(y)$. We say that it is **monotone** if it is either increasing or decreasing.

The next proposition said that if the function is monotone, the inverse statement of IVT is true.

Proposition 4.4.7: Inverse of IVT for Monotone Functions

If $f : [a, b] \to \mathbb{R}$ is monotone, and it has intermediate value property, then it is continuous.

Proof. Without losing generality, we assume that f is increasing. The situation of decreasing is of the same idea.
Let $c \in (a, b)$. Let $\epsilon > 0$. If $f(c) - \epsilon/2 < f(a)$, take $x_1 = a$. If $f(a) \leqslant f(c) - \epsilon/2$, then by intermediate value property, there exists $x_1 < c$ where $f(x_1) = f(c) - \epsilon/2$. In either case, since f is increasing, $x \in (x_1, c]$ implies

$$f(c) - \frac{\epsilon}{2} \leqslant f(x_1) \leqslant f(x) \leqslant f(c)$$

Similarly, we can also deduce that there exists $x_2 > c$ such that $x \in [c, x_2)$ implies

$$f(c) \leqslant f(x) \leqslant f(x_2) \leqslant f(c) + \frac{\epsilon}{2}$$

If we set $\delta = \min\{c - x_1, x_2 - c\}$, then for $|x - c| < \delta$, we have

$$f(c) - \frac{\epsilon}{2} \leqslant f(x) \leqslant f(c) + \frac{\epsilon}{2}$$

which means that $|f(x) - f(c)| < \epsilon$. This shows that f is continuous. $\square$

4.5 Sets of Discontinuity

4.5.1 Dirichlet's Function and Thomae's Function

In this section we transfer our sight from continuous functions to discontinuous functions. To begin with, we first introduce two extremely pathological, but interesting functions.

The first one to introduce is the Dirichlet's Function.

Definition 4.5.1: Dirichlet's Function

The **Dirichlet's Function** is defined as

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

It is hard to actually draw the graph of Dirichlet's function, since both rational numbers and irrational numbers are 'dense' in $\mathbb{R}$. Virtually it would be two parallel horizontal lines with $y = 0$ and $y = 1$. However, we know that there exists infinitely many 'holes' in these lines.

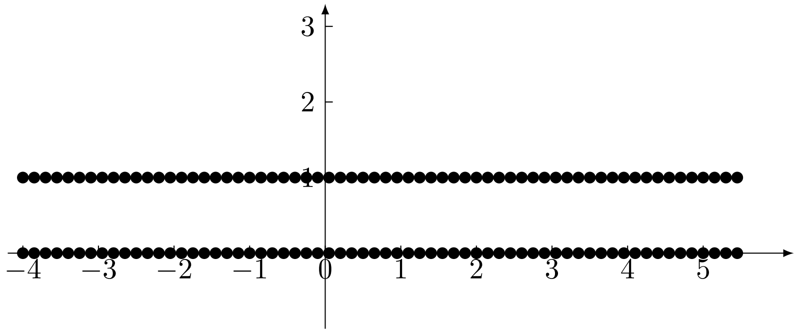


Figure 4.3: Dirichlet's Function

Property 4.5.2: Set of Discontinuity of Dirichlet's Function

Dirichlet's Function is everywhere discontinuous. i.e., the set of discontinuity of Dirichlet's function is $\mathbb{R}$.

Proof. Consider an arbitrary point $c \in \mathbb{R}$. Let $(x_n) \subseteq \mathbb{Q}$ be a rational sequence and $(y_n) \subseteq \mathbb{R}\backslash\mathbb{Q}$ be a irrational sequence such that both $(x_n), (y_n) \to c$. These two sequences can be found since rational and irrational numbers are dense in $\mathbb{R}$. Then, we have

$$\lim_{n\to\infty} f(x_n) = 1 \quad \text{but} \quad \lim_{n\to\infty} f(y_n) = 0$$

By Corollary 4.3.2, it is not continuous on c. Since c is arbitrary, it is not continuous on all points. $\square$

If we slightly modify the Dirichlet's function, then we can have a different set of discontinuity.

Property 4.5.3: Modified Dirichlet's Function

The function

$$f(x) = \begin{cases} x, & \text{if } x \in \mathbb{Q} \\ 0, & \text{if } x \notin \mathbb{Q} \end{cases}$$

is discontinuous at every point $c \neq 0$, and it is continuous at 0. i.e., the set of discontinuity is $\mathbb{R}\backslash\{0\}$.

Proof. We can use the similar way with what is done in the proof of Dirichlet's function to prove that it is discontinuous at every point except 0. The critical part is to prove that this function is continuous at 0.

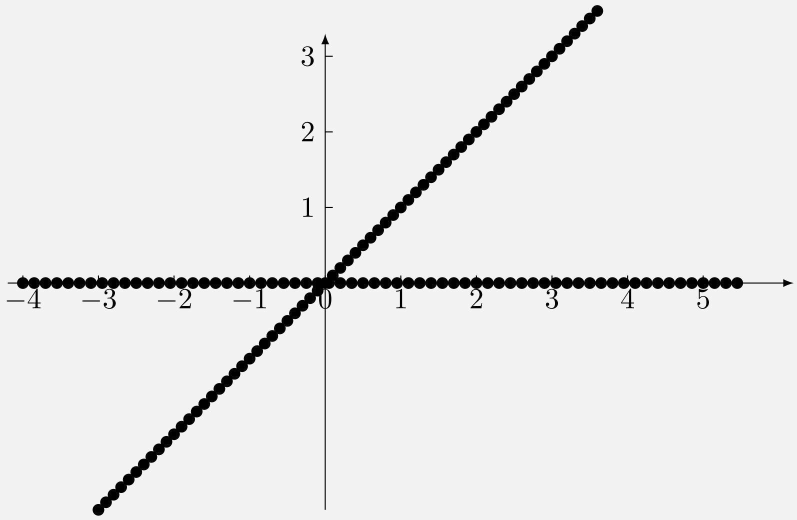


Figure 4.4: Modified Dirichlet's Function

Fix $\epsilon > 0$. Let $\delta = \epsilon$. Then, if we have $|x - 0| = |x| < \delta$, we have that

$$|f(x) - f(0)| = |f(x)| = \begin{cases} |x| < \epsilon, & \text{if } x \in \mathbb{Q} \\ 0 < \epsilon, & \text{if } x \notin \mathbb{Q} \end{cases}$$

Therefore, it is continuous at 0. □

The second one, which is much more fascinating, is called Thomae's Function. It is discovered in 1875 by K.J.Thomae.

Definition 4.5.4: Thomae's Function

The **Thomae's Function** is defined by

$$f(x) = \begin{cases} 1, & \text{if } x = 0 \\ 1/n, & \text{if } x = m/n \text{ with } \gcd(m, n) = 1 \text{ and } n > 0, m \neq 0 \\ 0, & \text{if } x \notin \mathbb{Q} \end{cases}$$

In case that some readers do not know, 'gcd' in the definition means **greatest common divisor**. If $\gcd(m, n) = 1$, it means that m and n are **relatively prime**, and the fraction is actually in its lowest term. For example, $2/4$ is not in its lowest term since $\gcd(4, 2) = 2$. $1/2$ is in its lowerst term.

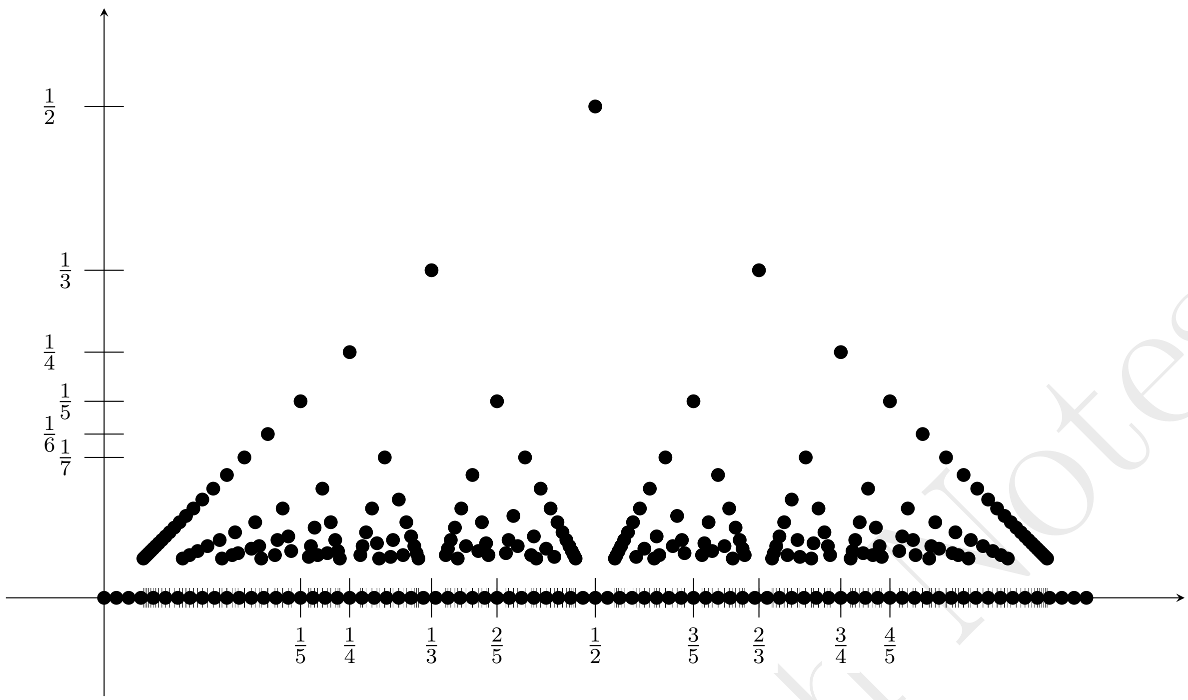


Figure 4.5: Thomae's Function (on interval $(0,1)$)

Property 4.5.5: Set of Discontinuity of Thomae's Function

Thomae's function is continuous at each irrational point, but discontinuous at every rational point. i.e., the set of discontinuity of Thomae's function is $\mathbb{Q}$.

Proof. We first see that it is discontinuous at each rational point. Since irrational numbers are dense in $\mathbb{R}$, we can find a irrational sequence $(x_n) \subseteq \mathbb{R}$ such that $(x_n) \to q$ for each $q \in \mathbb{Q}$. Therefore,

$$\lim_{n \to q} f(x_n) = 0 \neq f(q)$$

which shows that it is not continuous at q.

Now we prove that it is continuous at each irrational point. Let c be an arbitrary irrational point. Fix $\epsilon > 0$. By Archimedean Property, there exists $n_0 \in \mathbb{N}$ such that $1/n_0 < \epsilon$. For some interval $(c-k, c+k)$, there are only finitely many number of rationals with denominator less than n_0 in this interval. This is because the distance between two rationals with denominator n_0 has at least distance of $1/n_0$, and there are only finitely many natural numbers that is less than n_0. Therefore, we can find $\delta > 0$ such that there is no rational numbers that has denominator less than n_0 in the interval

$(c-\delta, c+\delta)$. Then, we would have, for all $|x-c|<\delta$,

$$|f(x)-f(c)| = |f(x)-0| = \begin{cases} < \dfrac{1}{n_0} < \epsilon, & \text{if } x \in \mathbb{Q} \\ 0, & \text{if } x \notin \mathbb{Q} \end{cases}$$

which shows that the function is continuous at c. □

4.5.2 Categorization of Discontinuity

Before we formally examine the topological properties of set of discontinuity, we first note that there are different kinds of discontinuity. At first glance, the discontinuity problem is always produced by the 'jump' of function values. To describe this situation, we introduce the terminology of 'one-sided limits'.

Definition 4.5.6: One-side limit

- Given function $f : A \to \mathbb{R}$ and a limit point c of A, the **limit from the right**

$$\lim_{x \to c^+} f(x) = L$$

means that:

$$\forall \epsilon > 0, \exists \delta > 0 \text{ s.t. } 0 < x - c < \delta \implies |f(x) - L| < \epsilon$$

- Given function $f : A \to \mathbb{R}$ and a limit point c of A, the **limit from the left**

$$\lim_{x \to c^-} f(x) = L$$

means that:

$$\forall \epsilon > 0, \exists \delta > 0 \text{ s.t. } 0 < c - x < \delta \implies |f(x) - L| < \epsilon$$

These definitions are crucial to the existence of limit of function, therefore also crucial to the continuity of a function.

Theorem 4.5.7: One-side limit and existence of limit

Given function $f : A \to \mathbb{R}$ and a limit point c of A, $\lim_{x \to c} f(x) = L$ if and only if both one-sided limits exist and equal to L.

Proof.

($\Longrightarrow$) Suppose $\lim_{x\to c} f(x) = L$. Then,

$$\forall\, \epsilon > 0, \exists\, \delta > 0 \text{ s.t. } |x - c| < \delta \quad \Longrightarrow \quad |f(x) - L| < \epsilon$$

which means that for both $0 < x - c < \delta$ and $0 < c - x < \delta$, the limit exists.
($\Longleftarrow$) Suppose both one-sided limits exist and equal to L. This means that

$$\forall\, \epsilon > 0, \exists\, \delta > 0 \text{ s.t. } 0 < x - c < \delta \text{ or } 0 < c - x < \delta \quad \Longrightarrow \quad |f(x) - L| < \epsilon$$

which is equivalent to the $\epsilon - \delta$ statement of function limit. □

Now we formally state these types of discontinuities. They are:

1. **Removable Discontinuity:** $\lim_{x\to c} f(x)$ exists but $\lim_{x\to c} f(x) \neq f(c)$.
2. **Jump Discontinuity:** $\lim_{x\to c^+} f(x) \neq \lim_{x\to c^-} f(x)$. This can be further categorized as:
   - **Finite Jump:** Both one-side limits exist but $\lim_{x\to c^+} f(x) \neq \lim_{x\to c^-} f(x)$.
   - **Infinite Jump:** One of the one-side limit does not exist.
3. **Essential Discontinuity:** $\lim_{x\to c} f(x)$ does not exist for other reasons.

4.5.3 Sets of Discontinuity of Monotone Functions

We have seen that monotone functions obtain some good properties related to intermediate value property. It also obtains good properties related to discontinuities. Specifically, a monotone function can only have one kind of discontinuity.

Theorem 4.5.8: Jump Discontinuity of Monotone Functions

The only type of discontinuity a monotone function can have is a finite jump discontinuity.

Proof. Without losing generality, suppose $f : A \to \mathbb{R}$ is increasing. The goal is to show that for any $c \in A$, $\lim_{x\to c^+} f(x)$ and $\lim_{x\to c^-} f(x)$ exist.
Let $\epsilon > 0$, set $L = \sup\{f(x) : x < c\}$. Then $L - \epsilon$ would not be an upper bound. Hence, there exists $\delta_1 > 0$ such that $f(c - \delta_1) > L - \epsilon$. Since f is increasing, $0 < c - x < \delta_1$ implies $|f(x) - L| < \epsilon$. Likewise, we can also show that we can find $\delta_2 > 0$ such that $0 < x - c < \delta_2$ implies $|f(x) - L| < \epsilon$. Therefore, both one-sided limits exist, and only finite jump discontinuity is possible. □

The type of discontinuity is uniquely determined. We can imagine that finite jump discontinuities must be separated. Then, the number of jump discontinuities of monotone function is also determined.

Corollary 4.5.9: Cardinality of Set of Discontinuity of Monotone Function

The set of discontinuity of a monotone function f must be either finite or countable.

Proof. Again, we suppose that f is increasing. We know that for arbitrary point $c \in A$, $\lim_{x\to c^+} f(x)$ and $\lim_{x\to c^-} f(x)$ exist. Let $\{c_\lambda\}_{\lambda\in I}$ be the set of discontinuity. Let $\lim_{x\to c_\lambda^+} f(x) = L_\lambda$ and $\lim_{x\to c\lambda^-} f(x) = M_\lambda$. Since f is increasing, for each λ, $M_\lambda < L_\lambda$. Then, choose a rational point q_λ from each (M_λ, L_λ). This can be done since rational numbers are dense in $\mathbb{R}$. In this manner, we can construct a bijection between the set of finite jump discontinuities and a subset of $\mathbb{Q}$, by mapping each jump discontinuity to the chosen q_λ. Since the cardinality of subset of $\mathbb{Q}$ must be finite or countable, we conclude that the cardinality of the set of discontinuity must be finite or countable. $\square$

4.5.4* General Topology of Sets of Discontinuity

Now we eliminate the restriction of monotone function, and consider the set of discontinuity of an arbitrary function. Can the set of discontinuities of a particular function be arbitrary? Or it need to follow some topological properties?

Before diving into this topic, we need to first introduce some advanced terminologies. We have seen the topological properties of open and closed sets, and we know that arbitrary union of closed sets may not be closed, and arbitrary intersection of open sets, also, may not be open. However, these two kinds of set do have their own names.

Definition 4.5.10: F_σ Sets and G_δ Sets

- A set H is F_σ **set** if it is a countable union of closed sets, i.e., $H = \cup_k F_k$, where F_k are closed.
- A set H is G_δ **set** if it is a countable intersection of open sets, i.e., $H = \cap_k G_k$, where G_k are open.

The notation σ comes from French *somme* and means 'summation'. δ comes from German *Durchschnitt* and means 'multiplication'. The notation F and G comes from French *fermé* and German *Gebiet*, respectively. Now we see some examples.

Example 4.5.1.

- Half open intervals are both F_σ and G_δ sets, since it can be written as

$$[a,b) = \bigcap_{k=1}^{\infty}\left(a-\frac{1}{k}, b\right) = \bigcup_{k=1}^{\infty}\left[a, b-\frac{1}{k}\right]$$

- $\mathbb{Q}$ is F_σ set since it is a countable union of singletons $\{x\}$, each of which is closed.

- $\mathbb{R}\backslash\mathbb{Q}$ is G_δ set since it can be written as the countable intersection of U_k, where

$$U_k = (-\infty, r_k) \cup (r_k, \infty), \quad r_k \text{ are rational numbers}$$

We focus on the analysis of F_σ set in this section. We can imagine that the collection of all F_σ set is extremely huge. However, does that collection contain all subsets of $\mathbb{R}$? We do not know yet. To examine this, we first need some lemma.

Lemma 4.5.11: Intersection of Countable Dense Sets

If $\{G_1, G_2, G_3, \cdots\}$ is a countble collection of dense, open sets, then the intersection $\bigcap_{n=1}^{\infty} G_n$ is not empty.

Proof. Because G_1 is open, there exists an open interval $(a_1, b_1) \subseteq G_1$. Let $I_1 = [c_1, d_1] \subseteq (a_1, b_1) \subseteq G_1$. Because G_2 is open, $(c_1, d_1) \cap G_2$ is also open. Because G_2 is dense, $(c_1, d_1) \cap G_2$ is nonempty, thus containing an open interval (a_2, b_2). Let $I_2 = [c_2, d_2] \subseteq (a_2, b_2) \subseteq (c_1, d_1) \cap G_2$.
Continuing this fashion, we can construct a colletion of nested closed interval $I_1 \supseteq I_2 \supseteq I_3 \supseteq \cdots$. By nested interval property, $\bigcap_{n=1}^{\infty} I_n \neq \emptyset$. Therefore, $\bigcap_{n=1}^{\infty} G_n \neq \emptyset$. □

Lemma 4.5.12

It is impossible to write

$$\mathbb{R} = \bigcup_{n=1}^{\infty} F_n$$

where F_n is a closed set containing no nonempty open intervals for each $n \in \mathbb{N}$.

Proof. This is just the complement of previous lemma. Set $G_n = F_n^c$ for each n. Each G_n is open since it is a complement of a closed set. Moreover, each G_n is dense. To see this, let $a, b \in \mathbb{R}$ such that $a < b$. Since $(a, b) \nsubseteq F_n$, there exists $c \in (a, b)$ such that $c \in F_n^c = G_n$.
Therefore, $\mathbb{R} = \bigcup_{n=1}^{\infty} F_n$ will imply $\emptyset = \bigcap_{n=1}^{\infty} G_n$, which is a contradiction with Lemma 4.5.11. □

With these two lemmas, we can prove that the set of all irrational numbers cannot be an F_σ set, and set of all rational numbers cannot be an G_δ set, thus answering our question that F_σ set does not contain all subsets of $\mathbb{R}$.

Corollary 4.5.13

$\mathbb{R}\backslash\mathbb{Q}$ is not a F_σ set, and $\mathbb{Q}$ is not a G_δ set.

Proof.

- Recall that $\mathbb{Q}$ is an F_σ set. Suppose for contradiction, $\mathbb{R}\backslash\mathbb{Q}$ is also an F_σ set. Then we could write
$$\mathbb{Q} = \bigcup_{n=1}^{\infty} F_n, \quad \mathbb{R}\backslash\mathbb{Q} = \bigcup_{n=1}^{\infty} F'_n$$
Each F_n/F'_n must contain no nonempty open intervals since otherwise it will conclude irrationals/rationals. Therefore, $\mathbb{R} = \mathbb{Q} \cup \mathbb{R}\backslash\mathbb{Q}$ is a countable union of closed sets containing no nonempty open intervals, which is a contradiction to Lemma 4.5.12.

- To see that $\mathbb{Q}$ is not a G_δ set, we only need to show that, a set is G_δ if and only if its complement is F_σ. This may be useful afterwards in your study, so I also state it as a lemma.

Lemma 4.5.14

A set $A \subseteq \mathbb{R}$ is a G_δ set if and only if its complement is an F_σ set.

This can be easily seen by De Morgan's Law and the definition of these two kinds of set. Therefore, if $\mathbb{R}\backslash\mathbb{Q}$ is not a F_σ set, then $\mathbb{Q}$ cannot be a G_δ set.

□

We can easily see that, the set of discontinuities $\mathbb{R}$ of Dirichlet's function, the set of discontinuities $\mathbb{R}\backslash\{0\}$ of modified Dirichlet's function, the set of discontinuities $\mathbb{Q}$ of Thomae's function, are all F_σ sets. Does there exist any function that has a set of discontinuity that is not a F_σ set? The answer is no. Acutally, all sets of discontinuities are F_σ set. Before proving this, we need some lemma.

Definition 4.5.15: α-continuous

Let $f : \mathbb{R} \to \mathbb{R}$. f is α-continous at $c \in \mathbb{R}$ if

$$\exists\, \delta > 0, \text{ s.t. } \forall\, y, z \in (c-\delta, c+\delta) \quad \Longrightarrow \quad |f(y) - f(z)| < \alpha$$

For the convenience of notation, we write

- $D_f = \{x \in \mathbb{R} : f \text{ is not continuous at } x\}$ the set of discontinuities.
- $D_f^\alpha = \{x \in \mathbb{R} : f \text{ is not } \alpha\text{-continuous at } x\}$ the set of α-discontinuities.

There is a good topological property of D_f^α that becomes our basis of proving the topological property of D_f.

Lemma 4.5.16: D_f^α is closed

For a fix $\alpha > 0$, the set D_f^α is closed.

Proof. We prove this by showing that the complement is open. Let $x \in (D_f^\alpha)^c$. Then, f is α-continuous at x. i.e.,

$$\exists\, \delta > 0, \text{ s.t. } \forall\, y, z \in (x - \delta, x + \delta) \implies |f(y) - f(z)| < \alpha$$

Then we can see that for all $x' \in (x - \delta/2, x + \delta/2)$, let $\delta' = \delta/2$, we will have

$$\forall\, y, z \in (x' - \delta', x' + \delta') \subseteq (x - \delta, x + \delta) \implies |f(y) - f(z)| < \alpha$$

Therefore, $x' \in (D_f^\alpha)^c$. This means that there exists an open interval $(x - \delta', x + \delta')$ for each $x \in (D_f^\alpha)^c$ such that $(x - \delta', x + \delta') \subseteq (D_f^\alpha)^c$, which shows that $(D_f^\alpha)^c$ is open. □

It is also clear that, for smaller α, the constraint of α-continuity is stricter. This indicates the inclusion relationship of different set of α-discontinuity.

Lemma 4.5.17: Relation between different α-continuity

If $\alpha < \alpha'$, then $D_f^{\alpha'} \subseteq D_f^\alpha$.

Proof. If $|f(y) - f(z)| < \alpha$ then it must be that $|f(y) - f(z)| < \alpha'$. This shows that if a function is α-continuous at x, it must be α'-continuous at x. Thus, $(D_f^\alpha)^c \subseteq (D_f^{\alpha'})^c$. □

The next lemma says that all continuous functions are α-continuous at corresponding points.

Lemma 4.5.18: Continuous functions are α-continuous

Let $\alpha > 0$. If f is continuous at c, then it is α-continuous at c. This shows that $D_f^\alpha \subseteq D_f$.

Proof. Since f is continuous at c,

$$\forall\, \epsilon > 0, \exists\, \delta > 0 \text{ s.t. } |x - c| < \delta \implies |f(x) - f(c)| < \epsilon$$

Let $\epsilon = \alpha/2$. Then, if $y, z \in (c - \delta, c + \delta)$, by triangular inequality, we have

$$|f(y) - f(z)| \leqslant |f(y) - f(c)| + |f(z) - f(c)| < \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha$$

which shows that it is α-continuous at c. □

Now we are ready to prove that all sets of discontinuity are F_σ set. (**The converse is also true, that for all F_σ set, we can find a function that is discontinuous exactly on this set. However, the proof is intricate, and it is not introduced here.**)

Theorem 4.5.19: Topological Property of Sets of Discontinuity

Let $f : \mathbb{R} \to \mathbb{R}$. Then, D_f is an F_σ set.

Proof. If f is not continuous at some point c, then

$$\exists\, \epsilon_0 > 0, \text{ s.t. } \forall\, \delta > 0, 0 < |x - c| < \delta \implies |f(x) - f(c)| \geqslant \epsilon_0$$

Let $\alpha_n = 1/n$. Once $\alpha_n < \epsilon_0$, we will have f is not α_n-continuous at c. This guarantees that

$$D_f = \bigcup_{n=1}^{\infty} D_f^{\alpha_n}$$

since for n going to infinity, eventually $\alpha_n < \epsilon_0$, and all the discontinuous points will be α_n-discontinuous, thus included at the right hand side. Since each $D_f^{\alpha_n}$ is closed by Lemma 4.5.16, we have that D_f is indeed an F_σ set. $\square$

Finally, since $\mathbb{R}\backslash\mathbb{Q}$ is not an F_σ set, we have our milestone corollary.

Corollary 4.5.20

There does not exist a function such that it is continuous at every rational point and discontinuous at every irrational point.

Part II

PART II: Calculus on the Real Line

Chapter 5

Differentiation

5.1 Derivatives and Differentiability

5.1.1 Definition and Examples

After introducing rigorously the definition of function limit, we can rigorously state the definition of derivative.

Definition 5.1.1: Derivative

Let $f : A \to \mathbb{R}$ be a function. Given $c \in A$, the **derivative** of f at c is defined by

$$f'(c) = \lim_{x \to c} \frac{f(x) - f(c)}{x - c}$$

provided that this limit exists. If it does exist, we say f is **differentiable** at c. If the derivative exists at all points $c \in A$, we say that g is **differentiable** on A.

Normally we will consider A as an open interval (a, b) to make the definition intuitive. Geometrically, $\frac{f(x)-f(c)}{x-c}$ represents the slope of the line through two points $(x, f(x))$ and $(c, f(c))$. By taking the limit as x approaches c, we arrive at well-defined mathematical meaning for the slope of the tangent line at $x = c$.

Let us see some examples of differentiation.

Example 5.1.1. The function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = x^n, \quad n \in \mathbb{N}$$

is differentiable everywhere. To prove this using definition, let $c \in \mathbb{R}$ be arbitrary. We can calculate that

$$f'(c) = \lim_{x \to c} \frac{x^n - c^n}{x - c} = \lim_{x \to c} \frac{(x - c)(x^{n-1} + cx^{n-2} + c^2x^{n-3} + \cdots + c^{n-1})}{x - c} = \underbrace{c^{n-1} + c^{n-1} + \cdots + c^{n-1}}_{n \text{ terms}} = nc^{n-1}$$

Example 5.1.2. The function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = |x|$$

is not differentiable at 0. To see this, we note that the one-sided limits

$$\lim_{x \to 0^+} \frac{|x|}{x} = 1 \quad \text{and} \quad \lim_{x \to 0^-} \frac{|x|}{x} = -1$$

are not equal, thus the limit does not exist.

Moreover, the definition of derivative can be written in a few other different forms. Indeed, there are many, and we just introduce two simplest here, other forms can be derived using the same logic.

Proposition 5.1.2: Alternative Definition of Derivative

Let f be defined on A. Let $c \in A$. Suppose f is differentiable at c.

- $f'(c)$ can be alternatively defined by

$$f'(c) = \lim_{h \to 0} \frac{f(c+h) - f(c)}{h}$$

- If A is open, $f'(c)$ can be alternatively defined by

$$f'(c) = \lim_{h \to 0} \frac{f(c+h) - f(c-h)}{2h}$$

Proof. The first one is just a change of variable. Set $h = x - c$, and we have

$$f'(c) = \lim_{x \to c} \frac{f(x) - f(c)}{x - c} = \lim_{h \to 0} \frac{f(h+c) - f(c)}{h + c - c} = \lim_{h \to 0} \frac{f(h+c) - f(c)}{h}$$

The second one can be derived from simple algebra. Since f is differentiable at c, we can separate this formula as

$$\lim_{h \to 0} \frac{f(c+h) - f(c-h)}{2h} = \frac{1}{2} \left(\lim_{h \to 0} \frac{f(c+h) - f(c)}{h} + \lim_{h \to 0} \frac{f(c) - f(c-h)}{h} \right)$$

where the first term at right hand side is just $f'(c)$ and the second term is also $f'(c)$ with substitution of $h = c - x$. $\square$

These alternative forms of definition is actually very useful in approximation theory and numerical analysis of partial differential equations.

5.1.2 Relation with Continuous functions

We can see from the second example that not all continuous functions are differentiable everywhere. However, the converse statement is true, and it is a already well-known knowledge of Chinese high school students.

Theorem 5.1.3: Differentiability implies Continuity

If f is differentiable at a point c, then f is continuous at c as well.

Proof. We are assuming that

$$f'(c) = \lim_{x \to c} \frac{f(x) - f(c)}{x - c}$$

exists, and we want to prove that $\lim_{x \to c} f(x) = f(c)$. Note that by Algebraic Limit Theorem for function limit, we indeed have

$$\lim_{x \to c}(f(x) - f(c)) = \lim_{x \to c} \left(\frac{f(x) - f(c)}{x - c} \right)(x - c) = f'(c) \times 0 = 0$$

It follows that $\lim_{x \to c} f(x) = f(c)$. □

5.1.3 Algebraic Properties

Now we prove some well-known algebraic properties of differentiable functions.

Theorem 5.1.4: Algebraic Differentiability Theorem

Let $f, g : A \to \mathbb{R}$ be differentiable at $c \in A$. Then,

1. $(f + g)'(c) = f'(c) + g'(c)$.
2. $(kf)'(c) = kf'(c)$ for all $k \in \mathbb{R}$.
3. $(fg)'(c) = f'(c)g(c) + f(c)g'(c)$.
4. $(f/g)'(c) = \frac{f'(c)g(c) - f(c)g'(c)}{g(c)^2}$, provided that $g(c) \neq 0$.

Proof.

1. Since f, g are differentiable at c, we can write it as

$$(f + g)'(c) = \lim_{x \to c} \frac{f(x) + g(x) - f(c) - g(c)}{x - c} = \lim_{x \to c} \frac{f(x) - f(c)}{x - c} + \lim_{x \to c} \frac{g(x) - g(c)}{x - c} = f'(c) + g'(c)$$

by algebraic limit theorem of function limit.

2. Since f is differentiable at c, we can write it as

$$(kf)'(c) = \lim_{x\to c} \frac{kf(x) - kf(c)}{x - c} = k \lim_{x\to c} \frac{f(x) - f(c)}{x - c} = kf'(c)$$

by algebraic limit theorem of function limit.

3. We can write the equation as

$$\begin{aligned}\frac{(fg)(x) - (fg)(c)}{x - c} &= \frac{f(x)g(x) - f(x)g(c) + f(x)g(c) - f(c)g(c)}{x - c} \\ &= f(x)\left(\frac{g(x) - g(c)}{x - c}\right) + g(c)\left(\frac{f(x) - f(c)}{x - c}\right)\end{aligned}$$

By Algebraic limit theorem of function limit, since f, g are differentiable at c, f is also continuous at c. Therefore, we have

$$(fg)'(c) = \lim_{x\to c} f(x) \lim_{x\to c} \frac{g(x) - g(c)}{x - c} + g(c) \lim_{x\to c} \frac{f(x) - f(c)}{x - c} = f(c)g'(c) + g(c)f'(c)$$

4. We can write the equation as

$$\frac{(f/g)(x) - (f/g)(c)}{x - c} = \frac{f(x)g(c) - f(c)g(x)}{g(x)g(c)(x - c)}$$

By Algebraic limit theorem of function limit, since f, g are differentiable at c, they are also continuous at c. Therefore, we have

$$\begin{aligned}(f/g)'(c) &= \frac{1}{g(c)} \lim_{x\to c} \frac{1}{g(x)} \lim_{x\to c} \frac{f(x)g(c) - f(c)g(c) + f(c)g(c) - f(c)g(x)}{(x - c)} \\ &= \frac{1}{g(c)^2}\left(g(c) \lim_{x\to c} \frac{f(x) - f(c)}{(x - c)} - f(c) \lim_{x\to c} \frac{g(x) - g(c)}{x - c}\right) \\ &= \frac{g(c)f'(c) - f(c)g'(c)}{g(c)^2}\end{aligned}$$

which completes the proof. □

Now we consider the composition of two differentiable functions. Fortunately, the composition results in another differentiable function, which is called the **Chain Rule**. Indeed, we can actually write the formula of derivative as

$$(g \circ f)'(c) = \lim_{x\to c} \frac{g(f(x)) - g(f(c))}{x - c} = \lim_{x\to c} \frac{g(f(x)) - g(f(c))}{f(x) - f(c)} \cdot \frac{f(x) - f(c)}{x - c} = g'(f(c))f'(c)$$

However, there would be possibility that $f(x) = f(c)$ for all x value in an arbitrarily small neighborhood of

c, which makes the previous formula undefined. To solve this problem, we need a more delicate proof.

Theorem 5.1.5: Chain Rule

Let $f : A \to \mathbb{R}$ and $g : B \to \mathbb{R}$ with $f(A) \subseteq B$ so that $g \circ f$ is defined. If f is differentiable at $c \in A$ and g is differentiable at $f(c) \in B$, then $g \circ f$ is differentiable at c with

$$(g \circ f)'(c) = g'(f(c))f'(c)$$

Proof. Since g is differentiable at $f(c)$, we can define

$$d(y) = \begin{cases} \dfrac{g(y) - g(f(c))}{y - f(c)}, & \text{if } y \neq f(c) \\ g'(f(c)), & \text{if } y = f(c) \end{cases}$$

and observe that $\lim_{y \to c} d(y) = g'(f(c))$. Therefore, $d(y)$ is continuous at $f(c)$. Note that for $y \neq f(c)$, we have

$$g(y) - g(f(c)) = d(y)(y - f(c)) \tag{5.1}$$

and this equation holds also for $y = f(c)$. Therefore, we are free to substitute $y = f(t)$ for any arbitrary $t \in A$. If $t \neq c$, we can divide equation 5.1 by $(t - c)$ to get

$$\frac{g(f(t)) - g(f(c))}{t - c} = d(f(t))\frac{f(t) - f(c)}{t - c}$$

for all $t \neq c$. Finally, take $t \to c$ and apply algebraic limit theorem of function limit, we have

$$\lim_{t \to c} \frac{g(f(t)) - g(f(c))}{t - c} = (g \circ f)'(c) = \lim_{t \to c} d(f(t)) \lim_{t \to c} \frac{f(t) - f(c)}{t - c} = g'(f(c))f'(c)$$

which completes the proof. □

To end this section, we introduce the algebra of derivative of inverse function.

Definition 5.1.6: Inverse Function

If $f : A \to \mathbb{R}$ is one-to-one (injective), we can define the inverse function $f^{-1} : f(A) \to \mathbb{R}$ by

$$f^{-1}(y) = x \quad \text{where} \quad y = f(x)$$

Note that the above definition only works for injective function. Otherwise there would be many x correspond to one y, which makes f^{-1} not a well-defined function.

Theorem 5.1.7: Derivative of Inverse Function

Let $f : [a, b] \to \mathbb{R}$ be injective and f^{-1} its inverse function. Then,

1. If f is continuous on $[a, b]$, then f^{-1} is also continuous on its domain.
2. If f is differentiable on (a, b) with $f'(x) \neq 0$ for all $x \in (a, b)$, then f^{-1} is also differentiable on its domain and
$$(f^{-1})'(y) = \frac{1}{f'(x)} \quad \text{where} \quad y = f(x)$$

Proof.

1. We first show that **a continuous injective function must be strictly monotone**. i.e., $f(x) < f(y)$ (or $f(x) > f(y)$) for all $x, y \in [a, b]$ with $x < y$.

   We prove this by contradiction. Suppose that, f is not strictly monotone. Then, there must exist point $x, y, z \in [a, b]$ such that $f(x) \leqslant f(y)$ and $f(y) \geqslant f(z)$. If $f(x) = f(y)$ or $f(y) = f(z)$, then the function is not injective, which is a contradiction. So, suppose $f(x) < f(y)$ and $f(y) > f(z)$. Then by Intermediate Value Theorem, for some L with $\max\{f(x), f(z)\} < L < f(y)$ and $c_1 \in (x, y), c_2 \in (y, z)$ such that $f(c_1) = f(c_2) = L$, which is also a contradiction.

   Knowing that f is continuous and strictly monotone, we note that by definition, f^{-1} is also strictly monotone. Therefore, if we prove the Intermediate Value Property, we can use Proposition 4.4.7 to conclude that f^{-1} is continuous.

   Without losing generality, we assume that f is strictly increasing. The decreasing case is similar. Then for $x < y$, we have $f(x) < f(y)$, thus $f^{-1}(f(x)) < f^{-1}(f(y))$, which shows that f^{-1} is also increasing. Now consider $f(x_1) = y_1 < f(x_2) = y_2$, and L between $f^{-1}(y_1)$ and $f^{-1}(y_2)$. Clearly, their exists $y_3 = f(L) \in (y_1, y_2)$ such that $f^{-1}(y_3) = L$. Therefore, it attains intermediate value property, thus f^{-1} is continuous.

2. Let $x_0 \in (a, b)$ and let $y_0 = f(x_0)$. Since f is differentiable at x_0, we have
$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = f'(x_0)$$
   Hence, by algebraic differentiability theorem, we have
$$\lim_{x \to x_0} \frac{x - x_0}{f(x) - f(x_0)} = \frac{1}{f'(x_0)}$$

This can be done since $f(x) \neq f(x_0)$ by strict monotonicity. This indicates that

$$d(x) = \begin{cases} \dfrac{x - x_0}{f(x) - f(x_0)}, & \text{if } x \neq x_0 \\ \dfrac{1}{f'(x_0)}, & \text{if } x = x_0 \end{cases}$$

is continuous at x_0. Then, the composition $d \circ f^{-1}$ is continuous at $y_0 = f(x_0)$. Hence,

$$\frac{1}{f'(f^{-1}(y_0))} = \frac{1}{f'(x_0)} = d(f^{-1}(y_0)) = \lim_{y \to y_0} d(f^{-1}(y)) = \lim_{y \to y_0} \frac{f^{-1}(y) - f^{-1}(y_0)}{y - y_0}$$

So f^{-1} is differentiable at y_0. □

5.2 Intermediate Value Property for Derivatives

The derivatives have a surprising property that although they are not in general continuous, but they do possess the intermiediate value property! This observation would be a direct corollary to what is called **Interior Extremum Theorem**.

5.2.1 Interior Extremum Theorem

The interior extremum theorem says that, differentiable functions attain maximums and minimums on an open interval only at points where the derivative is equal to zero.

Theorem 5.2.1: Interior Extremum Theorem

Let f be differentiable on (a, b). If f attains maximum/minimum at $c \in (a, b)$, then $f'(c) = 0$.

Proof. Without losing generality, we assume $f(c)$ is the maximum. The minimum case is similar. Because c is in the open interval (a, b), we can construct two sequences (x_n) and (y_n), which converges to c and satisfies $x_n < c < y_n$ for all $n \in \mathbb{N}$. Since $f(c)$ is the maximum, we have

$$\frac{f(y_n) - f(c)}{y_n - c} \leqslant 0$$

since numerator is negative and denominator is positive. By Order Limit Theorem, we have

$$f'(c) = \lim_{n \to \infty} \frac{f(y_n) - f(c)}{y_n - c} \leqslant 0$$

In a similar way, we can have

$$\frac{f(x_n)-f(c)}{x_n-c} \geqslant 0 \quad \text{and thus} \quad f'(c)=\lim_{n\to\infty}\frac{f(x_n)-f(c)}{x_n-c} \geqslant 0$$

And therefore $f'(c)=0$, as desired. □

5.2.2 Darboux's Theorem

Now we state the intermediate value property for derivatives. This statement has its unique name, called **Darboux's Theorem**.

Theorem 5.2.2: Darboux's Theorem

If f is differentiable on $[a,b]$, and if α satisfies $f'(a)<\alpha<f'(b)$ or $f'(b)<\alpha<f'(a)$, then there exists $c\in(a,b)$ such that $f'(c)=\alpha$.

Proof. To simplify the matters, we define $g(x)=f(x)-\alpha x$. Then, g is differentiable on $[a,b]$ and $g'(x)=f'(x)-\alpha$. Our hypothesis then becomes $g'(a)<0<g'(b)$ or $g'(b)<0<g'(a)$, and we need to show that there exists $c\in(a,b)$ such that $g'(c)=0$.

Without losing generality, we will assume $g'(a)<0<g'(b)$. Since g is differentiable at a, we have for all $\epsilon>0$, there exists $\delta>0$ such that whenever $0<x-a<\delta$, we have

$$g'(a)-\epsilon<\frac{g(x)-g(a)}{x-a}<g'(a)+\epsilon$$

rewriting this inequality, since we choose $x-a>0$, we have

$$g(x)<g(a)+(x-a)(g'(a)+\epsilon)$$

Since $g'(a)<0$, we can choose ϵ small enough, so that $g'(a)+\epsilon<0$, and it leads to that $g(x)<g(a)+(x-a)(g'(a)+\epsilon)<g(a)$. Therefore, there exists $x\in(a,b)$ such that $g(x)<g(a)$.
Similarly, since g is differentiable at b, we have for all $\epsilon>0$, there exists $\delta>0$ such that whenever $0<b-y<\delta$, we have

$$g'(b)-\epsilon<\frac{g(y)-g(b)}{y-b}<g'(b)+\epsilon$$

rewriting this inequality, since we choose $y-b<0$, we have

$$g(y)<g(b)+(y-b)(g'(b)-\epsilon)$$

Since $g'(b) > 0$, we can choose ϵ small enough, so that $g'(b) - \epsilon > 0$, and it leads to that $g(y) < g(b) + (y - b)(g'(b) - \epsilon) < g(b)$. Therefore, there exists $y \in (a, b)$ such that $g(y) < g(b)$.

Now using Extremum Value Theorem 4.4.2, g attains its maximum and minimum on $[a, b]$. Note that we have shown that there exists $x, y \in (a, b)$ such that $g(x) < g(a)$ and $g(y) < g(b)$, thus the minimum does not occur at a, or b. Therefore, the minimum is attained at (a, b). By Interior Extremum Theorem 5.2.1, we have that for the minimum point $c \in (a, b)$, we have $g'(c) = 0$, which completes our proof. □

5.3 The Mean Value Theorems (MVT)

Now we are going to introduce **the most important** theorems in this chapter, which is the different forms of mean value theorem.

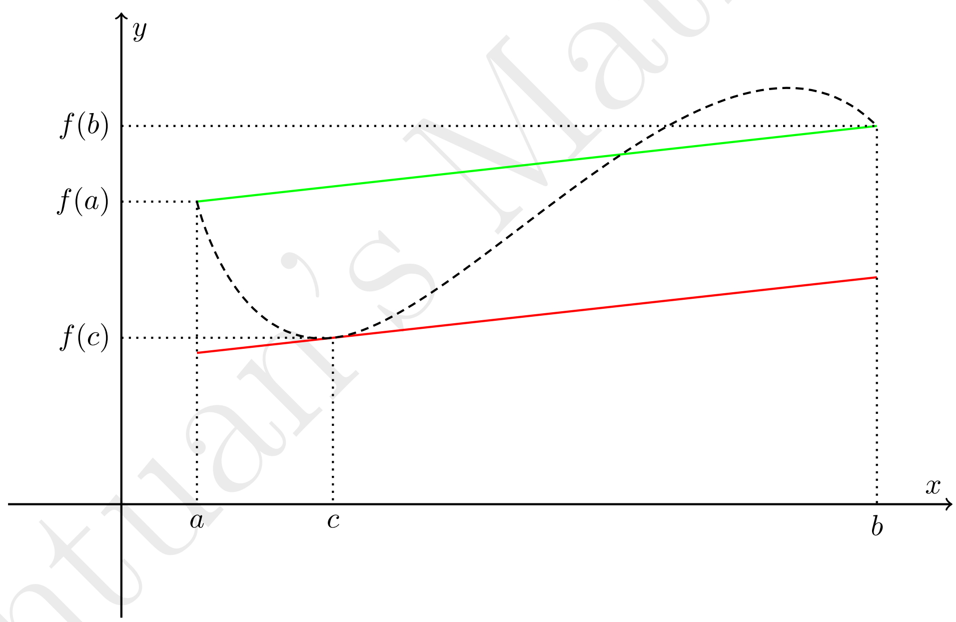


Figure 5.1: The Mean Value Theorems

The theorem says that a differentiable function f on an interval $[a, b]$ will, at some point c, attain a slope equal to the slope of the line through the endpoints $(a, f(a))$ and $(b, f(b))$, as shown in the figure above. There are many different forms of Mean Value Theorem, each with different generality and application background. The mean value theorems are important also because they will be the main tool for us to prove the famous L'Hôpital's Rule in the next section.

5.3.1 Rolle's MVT

Rolle's Mean Value Theorem is the simplest special case of these mean value statements. It deals with the case that the function value at two endpoints are equal. It is just a corollary combining Extremum Value Theorem and Interior Extremum Theorem.

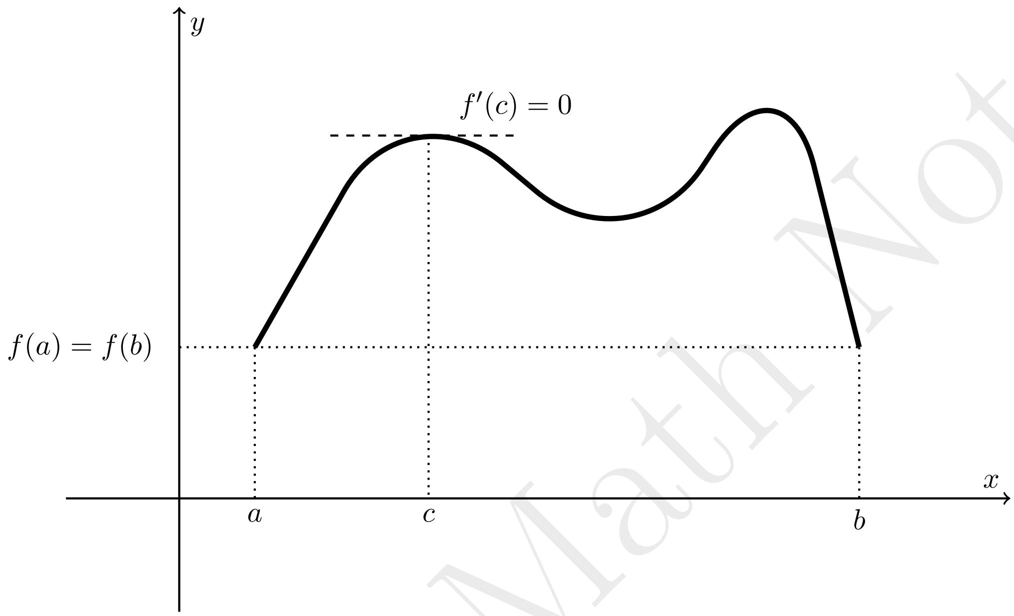


Figure 5.2: Rolle's Mean Value Theorems

Theorem 5.3.1: Rolle's Mean Value Theorem

Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b). If $f(a) = f(b)$, then there exists $c \in (a, b)$ such that $f'(c) = 0$.

Proof. Because f is continuous on a compact set $[a, b]$, by Extremum Value Theorem 4.4.2, it attains its minimum and maximum value.

- If both maximum and minimum occurs at the two endpoints, then f must be a constant function and $f'(c) = 0$ for all $c \in (a, b)$.
- If otherwise, either maximum or minimum occurs at some point $c \in (a, b)$, then by Interior Extremum Theorem 5.2.1, we have $f'(c) = 0$.

In conclusion, for both two situations, we can always find $c \in (a, b)$ such that $f'(c) = 0$, as desired. □

5.3.2 Lagrange's MVT

Lagrange's Mean Value Theorem is just the statement that we refer to when we say 'Mean Value Theorem'. It is the most useful form and will be appeared over and over again in later sections.

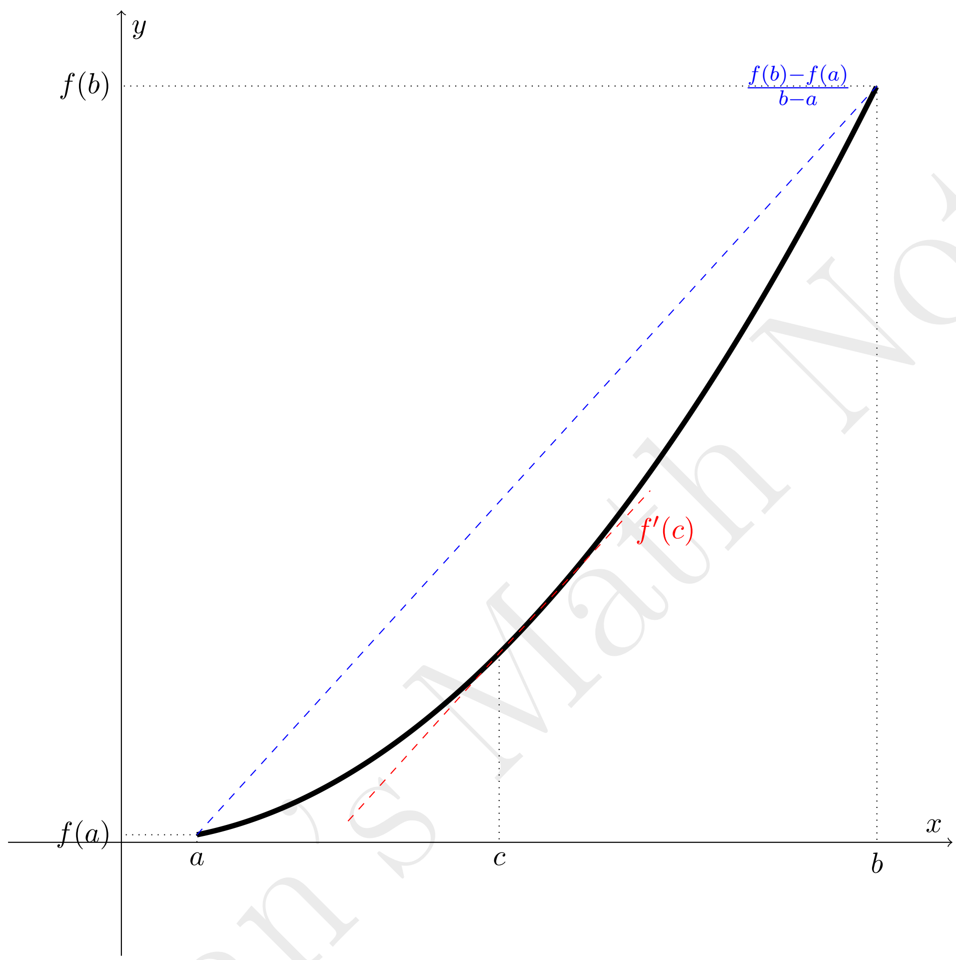


Figure 5.3: Lagrange's Mean Value Theorems

Theorem 5.3.2: (Lagrange's) Mean Value Theorem

If $f : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$ and differentiable on (a, b), then there exists $c \in (a, b)$ such that

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

Proof. Note that our theorem reduces to Rolle's MVT when $f(a) = f(b)$. The strategy of the proof is to reduce the more general statement to this special case.

The equation of the line through $(a, f(a))$ and $(b, f(b))$ is

$$y = \left(\frac{f(b) - f(a)}{b - a}\right)(x - a) + f(a)$$

Now we consider the distance between this line and the function $f(x)$, which is

$$d(x) = f(x) - \left[\left(\frac{f(b) - f(a)}{b - a}\right)(x - a) + f(a)\right]$$

By Algebraic Continuity Theorem 4.2.3 and Algebraic Differentiability Theorem 5.1.4, we see that $d(x)$ is continuous on $[a, b]$ and differentiable on (a, b), with $d(a) = d(b) = 0$. Thus, by Rolle's Mean Value Theore, there exists $c \in (a, b)$ such that $d'(c) = 0$. Because

$$d'(x) = f'(x) - \frac{f(b) - f(a)}{b - a}$$

we get

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

which completes the proof.

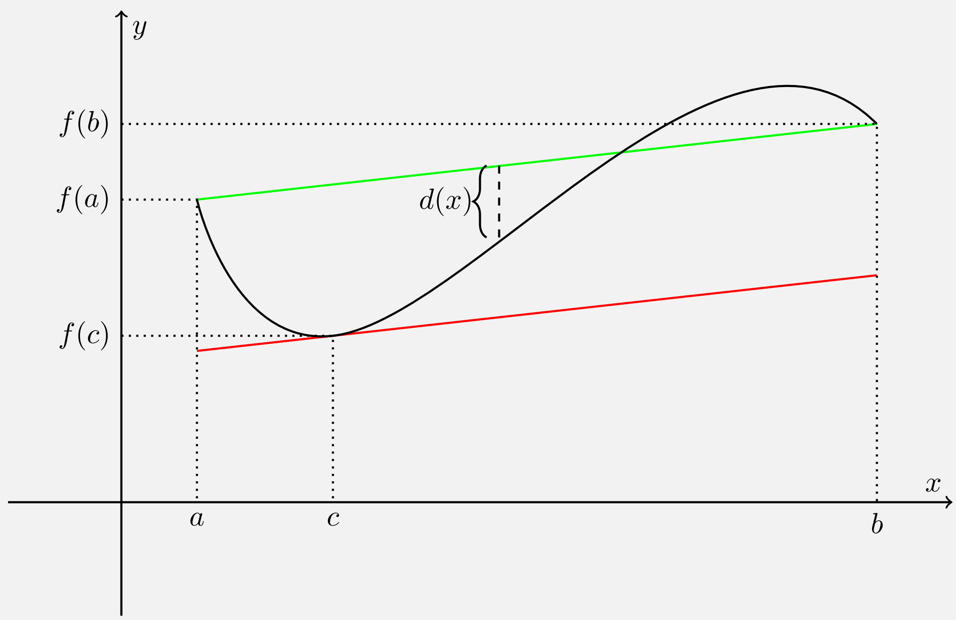


Figure 5.4: The newly defined $d(x)$

□

Lagrange's Mean Value Theorem has many simple applications. The next corollary states a high-school level simple result.

Corollary 5.3.3: Derivative and Monotonicity

Suppose $f : (a, b) \to \mathbb{R}$ is differentiable.

1. If $f'(x) \geqslant 0$ for all $x \in (a, b)$, then f is monotonically increasing.
2. If $f'(x) \leqslant 0$ for all $x \in (a, b)$, then f is monotonically decreasing.
3. If $f'(x) = 0$ for all $x \in (a, b)$, then $f(x) = k$ is constant, $k \in \mathbb{R}$.

Proof.

1. Choose arbitrary $c, d \in (a, b)$ with $c < d$. Consider the smaller interval (c, d). By MVT, there exists $c \in (c, d)$ such that
$$f'(c) = \frac{f(d) - f(c)}{d - c}$$
The assumption is $f'(c) \geqslant 0$. Since $d - c > 0$, we have $f(d) \geqslant f(c)$. Since c, d are arbitrary, the function is increasing.
2. This proof is similar to the first one.
3. We prove by contradiction. Suppose f is not constant. Then, there exists $x_1, x_2 \in (a, b)$ with $x_1 \neq x_2$ such that $f(x_1) \neq f(x_2)$. Without losing generality, suppose $x_1 < x_2$, then by MVT, there exists $c \in (x_1, x_2)$ such that
$$f'(c) = \frac{f(x_2) - f(x_1)}{x_2 - x_1} \neq 0$$

which is a contradiction. □

Corollary 5.3.4: Functions with same derivative are parallel

If f, g are differentiable on (a, b) and $f'(x) = g'(x)$ for all $x \in (a, b)$. Then, $f(x) = g(x) + k$ for some $k \in \mathbb{R}$.

Proof. Let $h(x) = f(x) - g(x)$. Then

$$h'(x) = f'(x) - g'(x) = 0 \text{ for all } x \in (a, b)$$

By Corollary 5.3.3, we have $h'(x) = k$ for some $k \in \mathbb{R}$, which completes the proof. □

5.3.3 Generalized MVT

Cauchy has a more general form of mean value theorem. It is also very important for further proof and anlaysis.

Theorem 5.3.5: Generalized Mean Value Theorem

If f, g are continuous on $[a, b]$ and differentiable on (a, b), then there exists $c \in (a, b)$ such that

$$(f(b) - f(a))g'(c) = (g(b) - g(a))f'(c)$$

If g' is never zero on (a, b) and $g(a) \neq g(b)$, it can be stated more delicate as

$$\frac{f'(c)}{g'(c)} = \frac{f(b) - f(a)}{g(b) - g(a)}$$

Proof. Let

$$h(x) = (f(b) - f(a))g(x) - (g(b) - g(a))f(x)$$

By algebraic differentiability theorem and algebraic continuity theorem, h is continuous on $[a, b]$ and differentiable on (a, b). We also notice that

$$h(a) = f(b)g(a) - f(a)g(a) - g(b)f(a) + g(a)f(a) = f(b)g(a) - g(b)f(a)$$

and

$$h(b) = f(b)g(b) - f(a)g(b) - g(b)f(b) + g(a)f(b) = f(b)g(a) - g(b)f(a)$$

Thus $h(a) = h(b)$. By Rolle's MVT, there exists $c \in (a, b)$ such that

$$h'(c) = (f(b) - f(a))g'(c) - (g(b) - g(a))f'(c) = 0$$

which completes the proof. □

The geometric interpretation of this theorem is more obscure than the other two. Indeed, it's the mean value theorem for parametric curves. If we denote $x = f(t)$, $y = g(t)$ as a parametric curve, we can plot this curve and find points $(f(a), g(a))$ and $(f(b), g(b))$. Consider the line l passing through these two points, the generalized mean value theorem then said, there exists some point $(f(c), g(c))$ such that the tangent line at that point has the same slope with l. i.e.,

$$\frac{x'(c)}{y'(c)} = \frac{dx}{dy} = \frac{x(b) - x(a)}{y(b) - y(a)}$$

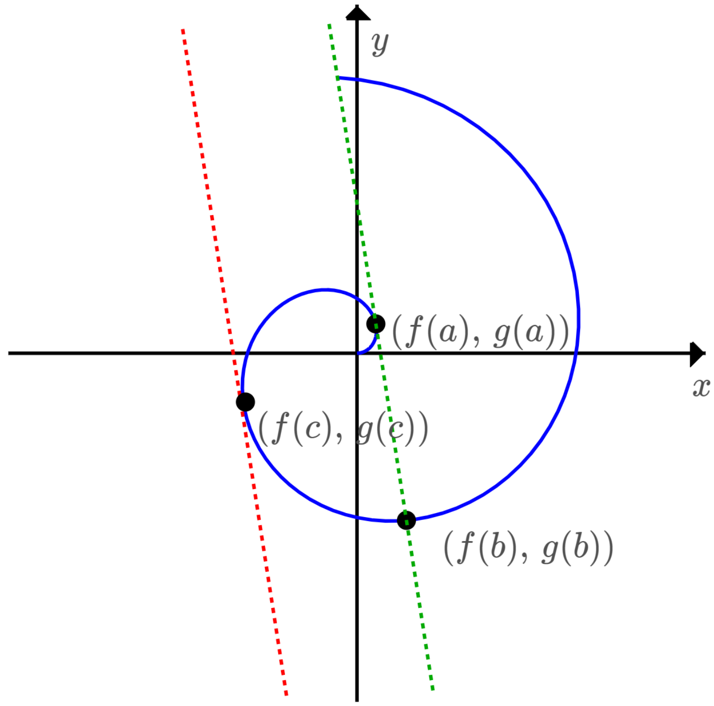


Figure 5.5: Graphical interpretation of Generalized MVT

5.4 L'Hôpital's Rule

Sometimes we meet the difficulty that when we calculate

$$f'(c) = \lim_{x \to c} \frac{f(x) - f(c)}{x - c}$$

Both numerator and denominator converge to 0, and we have no clear idea of what is the derivative, so we cannot use the definition. In this case the algebraic differentiality theorem does not apply, so we need a new tool to effectively deal with this difficulty.

5.4.1 $0/0$ case

Theorem 5.4.1: L'Hôpital's Rule: $0/0$ Case

Let $f, g : (a, b) \to \mathbb{R}$ be continuous on (a, b), and differentiable on (a, b) except with $c \in (a, b)$. If $f(c) = g(c) = 0$ and $g'(x) \neq 0$ for all $x \neq c$, then

$$\lim_{x \to c} \frac{f'(x)}{g'(x)} = L \quad \text{implies} \quad \lim_{x \to c} \frac{f(x)}{g(x)} = L$$

Proof. By Generalized MVT, since $f(c) = g(c) = 0$,

$$\frac{f(x)}{g(x)} = \frac{f(x) - f(c)}{g(x) - g(c)} = \frac{f'(y)}{g'(y)}$$

for some y between x and c. Therefore,

$$\lim_{x \to c} \frac{f(x)}{g(x)} = \lim_{x \to c} \frac{f'(y)}{g'(y)} = \lim_{x \to c} \frac{f'(x)}{g'(x)}$$

since $y \to c$ when $x \to c$ (this is because y is between x and c). □

5.4.2 ∞/∞ case

There is yet another case of difficulty. Sometimes the limit of numerator and denominator does not exist, but the limit of their quotient does exist. In this case, it leads to another form of L'Hôpital's Rule.

Definition 5.4.2: Infinitiy Limit

- Let $f : A \to \mathbb{R}$ and $c \in A$ be a limit point. We say $\lim_{x \to c} f(x) = \infty$ if

$$\forall\, M > 0, \exists\, \delta > 0, \text{ s.t. } 0 < |x - c| < \delta \quad \implies \quad f(x) > M$$

- Let $f : A \to \mathbb{R}$ and $c \in A$ be a limit point. We say $\lim_{x \to c} f(x) = -\infty$ if

$$\forall\, L < 0, \exists\, \delta > 0, \text{ s.t. } 0 < |x - c| < \delta \quad \implies \quad f(x) < L$$

Theorem 5.4.3: L'Hôpital's Rule: ∞/∞ Case

Let $f, g : (a, b) \to \mathbb{R}$ be differentiable on (a, b) and $g'(x) \neq 0$ for all $x \in (a, b)$. If

$$\lim_{x \to a} g(x) = \infty \text{ or } -\infty$$

Then,

$$\lim_{x \to a} \frac{f'(x)}{g'(x)} = L \quad \text{implies} \quad \lim_{x \to a} \frac{f(x)}{g(x)} = L$$

Proof. Let $\epsilon > 0$. Because $\lim_{x \to a} \frac{f'(x)}{g'(x)} = L$, there exists $\delta_1 > 0$ such that

$$\left| \frac{f'(x)}{g'(x)} - L \right| < \frac{\epsilon}{2}$$

for all $a < x < a + \delta_1$. For convenience of notation, let $t = a + \delta_1$ and note that t is fixed later on in the proof.

Our functions are not defined at a, but for any $x \in (a, t)$, we can apply Generalized Mean Value Theorem on the interval $[x, t]$ so that

$$\frac{f(x) - f(t)}{g(x) - g(t)} = \frac{f'(c)}{g'(c)}$$

for some $c \in (x, t)$. Our choice of t then implies

$$L - \frac{\epsilon}{2} < \frac{f(x) - f(t)}{g(x) - g(t)} < L + \frac{\epsilon}{2} \tag{5.2}$$

for all $x \in (a, t)$.

To isolate the target form $f(x)/g(x)$, we need to multiply 5.2 by $(g(x) - g(t))/g(x)$. However, we need to ensure that this quantity is positive. Since t is fixed and $\lim_{x\to a} g(x) = \infty$, we can choose $\delta_2 > 0$ so that $g(x) > g(t)$ for all $a < x < a + \delta_2$. Then, we will have

$$\frac{g(x) - g(t)}{g(x)} = 1 - \frac{g(t)}{g(x)} > 0$$

Then, multiply this to 5.2, we have

$$\left(L - \frac{\epsilon}{2}\right)\left(1 - \frac{g(t)}{g(x)}\right) < \frac{f(x) - f(t)}{g(x)} < \left(L + \frac{\epsilon}{2}\right)\left(1 - \frac{g(t)}{g(x)}\right)$$

adding $f(t)/g(x)$ on both sides, we have

$$\left(L - \frac{\epsilon}{2}\right)\left(1 - \frac{g(t)}{g(x)}\right) + \frac{f(t)}{g(x)} < \frac{f(x)}{g(x)} < \left(L + \frac{\epsilon}{2}\right)\left(1 - \frac{g(t)}{g(x)}\right) + \frac{f(t)}{g(x)}$$

which after some algebraic manipulateions, becomes

$$L - \frac{\epsilon}{2} + \frac{-Lg(t) + \frac{\epsilon}{2}g(t) + f(t)}{g(x)} < \frac{f(x)}{g(x)} < L + \frac{\epsilon}{2} + \frac{-Lg(t) - \frac{\epsilon}{2}g(t) + f(t)}{g(x)}$$

Since t is fixed and $\lim_{x\to a} g(x) = \infty$, we can choose $\delta_3 > 0$ such that whenever $a < x < a + \delta_3$ we have $g(x)$ is large enough to ensure that both

$$\frac{-Lg(t) + \frac{\epsilon}{2}g(t) + f(t)}{g(x)} \quad \text{and} \quad \frac{-Lg(t) - \frac{\epsilon}{2}g(t) + f(t)}{g(x)}$$

are less than $\epsilon/2$ in absolute value. Putting all these together, choose $\delta = \min\{\delta_1, \delta_2, \delta_3\}$, we have

$$\left|\frac{f(x)}{g(x)} - L\right| < \epsilon$$

for all $a < x < a + \delta$. $\square$

Chapter 6

Infinite Sequences and Series of Functions

In this chapter we consider the behavior of infinite sequences and series of functions. Before we have discussed about infinite sequences and series of real numbers, and we can see sequences and series of functions as a real number sequence/series at each point $x \in A$ in its domain, so that each point converge to the final function value. This though leads to the definition of 'pointwise convergence'.

6.1 Pointwise Convergence

Definition 6.1.1: Pointwise Convergence

Let $\{f_n\}_{n \in \mathbb{N}}$ be a collection of functions defined on a set $A \subseteq \mathbb{R}$. The sequence (f_n) of functions **converges pointwise** on A to a function f if,

$$\forall\, x \in A, f_n(x) \to f(x)$$

We write this as $\lim_{n\to\infty} f_n(x) = f(x)$, or simply $f_n \to f$.

Note that for a specific x, $f_n(x)$ is just a sequence of real numbers, so the statement $f_n(x) \to f(x)$ is defined in Chapter 2.

Example 6.1.1. Consider

$$f_n(x) = \frac{x^2 + nx}{n}$$

defined on $\mathbb{R}$. We can compute

$$\lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} \frac{x^2 + nx}{n} = \lim_{n\to\infty} \frac{x^2}{n} + x = x$$

Thus, (f_n) converges to $f(x) = x$ on $\mathbb{R}$.

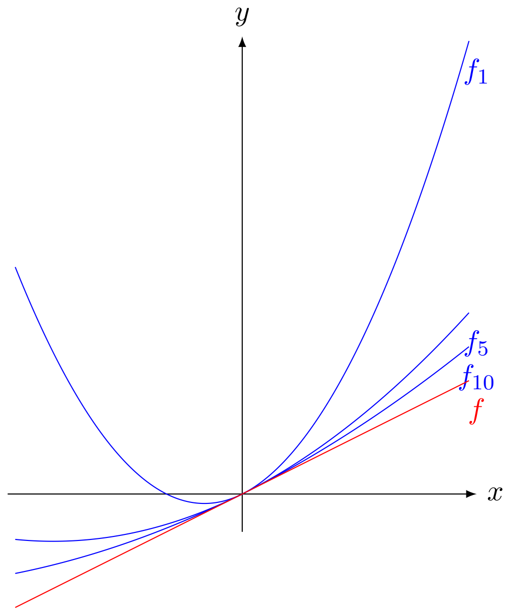


Figure 6.1: Graph of f_1, f_5, f_{10} and f

Example 6.1.2. Consider

$$g(x) = x^n, \quad x \in [0, 1]$$

If $0 \leqslant x < 1$, we have $x^n \to 0$. However, if $x = 1$, then $x^n \to 1$. Therefore, g_n converges pointwise to the function

$$g(x) = \begin{cases} 0, & \text{if } 0 \leqslant x < 1 \\ 1, & \text{if } x = 1 \end{cases}$$

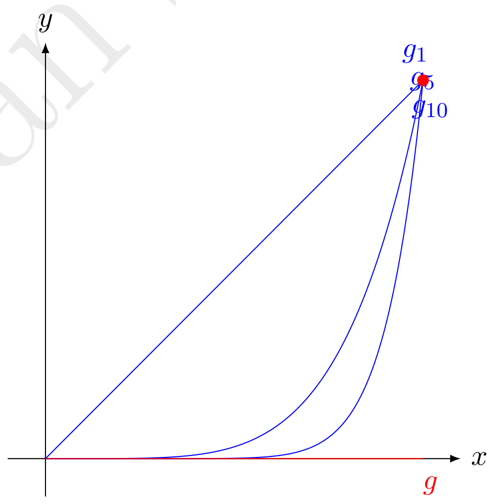


Figure 6.2: Graph of g_1, g_5, g_{10} and g

Note that in this case, the limit is not continuous.

Example 6.1.3. Consider

$$h_n(x) = x^{1+\frac{1}{2n-1}}, \quad x \in [-1, 1]$$

It can be computed that

$$\lim_{n\to\infty} h_n(x) = x \lim_{n\to\infty} x^{\frac{1}{2n-1}} = |x| = h(x)$$

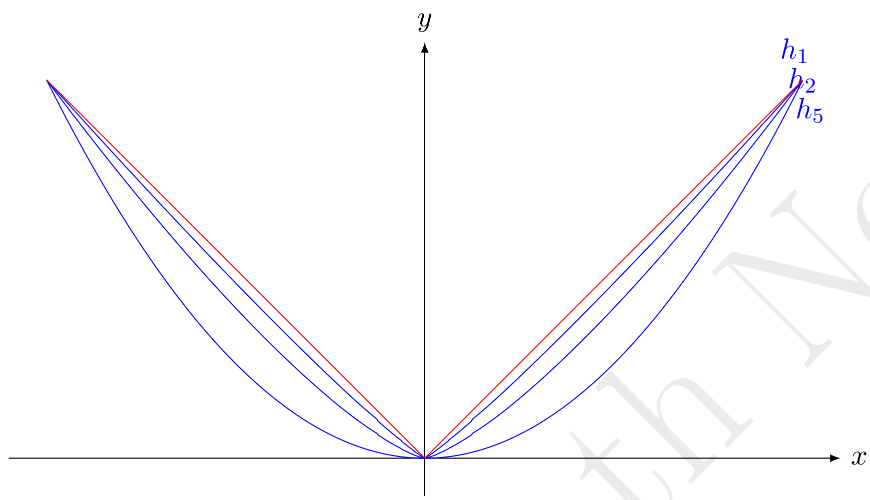


Figure 6.3: Graph of h_1, h_2, h_5 and h

Note that in this case the limit function is not differentiable at 0.

It seems that, pointwise convergence does not inherit continuity and differentiability to the limit. This is not an expected property of a good definition of limit of sequence of function. To resolve this problem, we need a refined, stronger definition of convergence.

6.2 Uniform Convergence

6.2.1 Definition and Examples

Definition 6.2.1: Uniform Convergence

Let $f_n : A \to \mathbb{R}$ for $n \in \mathbb{N}$. (f_n) **converges uniformly** on A to f if

$$\forall \epsilon > 0, \exists N \in \mathbb{N}, \text{ s.t. } n \geqslant N, x \in A \implies |f_n(x) - f(x)| < \epsilon$$

Note the difference between this definition and the definition of pointwise convergence. For pointwise convergence, if we expand the definition using ϵ-δ language, we have

$$\forall x \in A, \forall \epsilon > 0, \exists N \in \mathbb{N}, \text{ s.t. } n \geqslant N \implies |f_n(x) - f(x)| < \epsilon$$

You can see that the only difference is the position of the assertion '$\forall\, x \in A$'. For pointwise convergence, the assertion $\forall\, x \in A$ is at the start, meaning that the chosen N may depend on x, with different x, it needs a different scale of N to meet the ϵ requirement. However, for the uniform convergence, N does not need to depend on x. This is why it is called 'uniform'.

Geometrically speaking, the uniform convergence of f_n to a limit f on a set A can be visualized by constructing a band of radius $\pm\epsilon$ around the limit function f. If $f_n \to f$ uniformly, then there exists a point in the sequence after which each f_n is completely contained in this ϵ-strip.

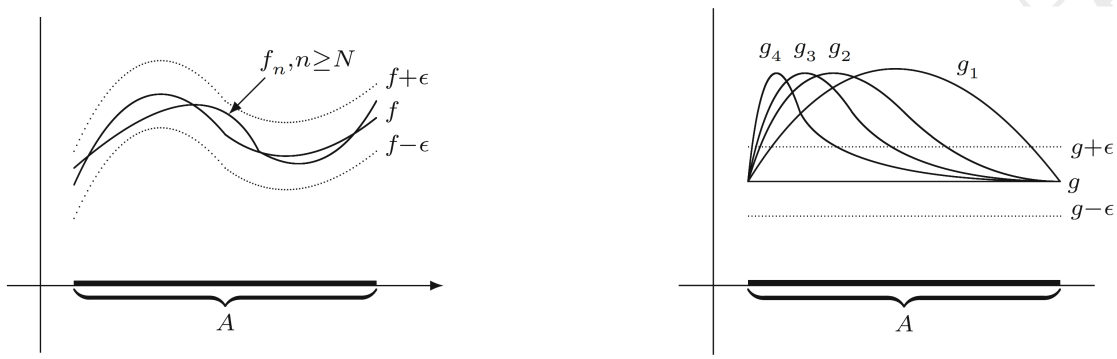


Figure 6.4: $f_n \to f$ uniformly, but $g_n \to g$ pointwise, not uniformly

Example 6.2.1. Consider

$$g_n(x) = \frac{1}{n(1+x^2)}, \quad x \in \mathbb{R}$$

We can see that for all $x \in \mathbb{R}$, $\lim g_n(x) = 0$, so the pointwise limit of g_n is $g(x) = 0$. Moreover, we see that

$$|g_n(x) - g(x)| = \left| \frac{1}{n(1+x^2)} - 0 \right| \leqslant \frac{1}{n}$$

Therefore, given $\epsilon > 0$, we can choose $N > 1/\epsilon$ so that $n \geqslant N$ implies

$$|g_n(x) - g(x)| \leqslant \frac{1}{n} \leqslant \frac{1}{N} < \epsilon$$

Hence, $g_n \to g$ uniformly on $\mathbb{R}$.

Example 6.2.2. Consider

$$f_n(x) = \frac{x^2 + nx}{n}, \quad x \in \mathbb{R}$$

we have shown that it converges pointwise to $f(x) = x$. However, the convergence is not uniform, since

$$|f_n(x) - f(x)| = \left| \frac{x^2 + nx}{n} - x \right| = \frac{x^2}{n}$$

and we need $N > x^2/\epsilon$ to make $|f_n(x) - f(x)| < \epsilon$. There is no way to choose a single N that works with all value of $x \in \mathbb{R}$.

Example 6.2.3. Consider

$$f_n(x) = \frac{x^2 + nx}{n}$$

again, but this time let $x \in [-b, b]$. This time we have

$$|f_n(x) - f(x)| = \left| \frac{x^2 + nx}{n} - x \right| = \frac{x^2}{n} \leqslant \frac{b^2}{n}$$

Therefore, we can choose $N > b^2/\epsilon$, which is not dependent on x, so that $|f_n(x) - f(x)| < \epsilon$. Therefore, it is uniformly convergent on $[-b, b]$.

Example 6.2.4. Consider

$$g(x) = x^n, \quad x \in [-1, 1]$$

We have shown that it converges pointwise to

$$g(x) = \begin{cases} 0, & \text{if } 0 \leqslant x < 1 \\ 1, & \text{if } x = 1 \end{cases}$$

However, note that to make

$$|g_n(x) - g(x)| = \begin{cases} x^n, & \text{if } 0 \leqslant x < 1 \\ 0, & \text{if } x = 1 \end{cases} < \epsilon$$

we need to choose $n > \log \epsilon / \log x$, which is dependent on x. Since $\log \epsilon / \log x$ can be arbitrarily large for $x \in [-1, 1]$, there is no way to choose one N for all x in this range.

As sequence of real numbers, there is also a Cauchy criterion of convergence for sequence of functions.

Proposition 6.2.2: Cauchy Criterion for Uniform Convergence

A sequence of functions (f_n) on $A \subseteq \mathbb{R}$ converges uniformly on A if and only if

$$\forall \epsilon > 0, \exists N \in \mathbb{N}, \text{ s.t. } m, n \geqslant N, x \in A \implies |f_n(x) - f_m(x)| < \epsilon$$

Proof.
($\Longrightarrow$) Suppose (f_n) converges uniformly on A to f. Then,

$$\forall \epsilon > 0, \exists N_1 \in \mathbb{N}, \text{ s.t. } n \geqslant N_1, x \in A \implies |f_n(x) - f(x)| < \frac{\epsilon}{2}$$

and

$$\forall \epsilon > 0, \exists N_2 \in \mathbb{N}, \text{ s.t. } m \geqslant N_2, x \in A \implies |f_m(x) - f(x)| < \frac{\epsilon}{2}$$

Therefore, if we choose $N = \max\{N_1, N_2\}$, we have

$$|f_n(x) - f_m(x)| \leqslant |f_n(x) - f(x)| + |f_m(x) - f(x)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

($\Longleftarrow$) Suppose

$$\forall\, \epsilon > 0, \exists\, N \in \mathbb{N}, \text{ s.t. } m, n \geqslant N, x \in A \quad \Longrightarrow \quad |f_n(x) - f_m(x)| < \frac{\epsilon}{2}$$

Fix $x \in A$, and by Cauchy Criterion, we know that the sequence of real number (f_n) converges to a limit L. Then doing this for all $x \in A$ will give us the pointwise limit of (f_n), denote it as f. We will prove that this pointwise limit is actually the uniform limit. Indeed, if we take m to infinity with $n \geqslant N$ and $x \in A$, by the pointwise convergence and order limit theorem, we have

$$\lim_{m\to\infty} |f_n(x) - f_m(x)| = |f_n(x) - f(x)| \geqslant \frac{\epsilon}{2} < \epsilon$$

which completes the proof. □

6.2.2 Presevation of Continuity

Now with this stronger definition, limit, as expected, will preserve the continuity of sequence of functions.

Theorem 6.2.3: Continuous Limit Theorem

Let (f_n) be a sequence of functions on A that converges uniformly on A to f. If each f_n is continuous at $c \in A$, then f is also continuous at c.

Proof. Fix $c \in A$, and let $\epsilon > 0$. Choose $N \in \mathbb{N}$ so that

$$|f_N(x) - f(x)| < \frac{\epsilon}{3}$$

for all $x \in A$. Because f_N is continuous, there exists $\delta > 0$ for which

$$|f_N(x) - f_N(c)| < \frac{\epsilon}{3}$$

whenever $|x - c| < \delta$. These two implies,

$$|f(x) - f(c)| \leqslant |f(x) - f_N(x)| + |f_N(x) - f_N(c)| + |f_N(c) - f(c)| < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon$$

Thus, f is continuous at $c \in A$. □

6.2.3 Presevation of Differentiability

From Example 6.1.3, we notice that even if (f_n) converges to f, the derivative (f_n') may not converge to f', even not exist. The next theorem indicates sufficient condition ensure the convergence of derivatives.

Theorem 6.2.4: Differentiable Limit Theorem

Let $f_n \to f$ pointwise on the closed interval $[a,b]$, and assume that each f_n is differentiable. If (f_n') converges uniformly on $[a,b]$ to a function g, then the function f is differentiable and $f' = g$.

Proof. Fix $c \in [a,b]$ and let $\epsilon > 0$. We want to argue that $f'(c)$ exists and equals to $g(c)$. Because f' is defined by the limit

$$f'(c) = \lim_{x \to c} \frac{f(x) - f(c)}{x - c}$$

Our task is to produce a $\delta > 0$ so that

$$\left| \frac{f(x) - f(c)}{x - c} - g(c) \right| < \epsilon$$

whenever $0 < |x - c| < \delta$.

Since (f_n') converges at c to g, we can choose $N_1 \in \mathbb{N}$ such that

$$|f_m'(c) - g(c)| < \frac{\epsilon}{3} \tag{6.1}$$

for all $m \geqslant N_1$. Moreover, by Proposition 6.2.2, since (f_n') converges uniformly to g, we can choose $N_2 \in \mathbb{N}$ such that for all $m, n \geqslant N_2$, we have

$$|f_m'(x) - f_n'(x)| < \frac{\epsilon}{3}$$

for all $x \in [a,b]$. Let $N = \max\{N_1, N_2\}$.

The function f_N is differentiable at c, so there exists a $\delta > 0$ such that

$$\left| \frac{f_N(x) - f_N(c)}{x - c} - f_N'(c) \right| < \frac{\epsilon}{3} \tag{6.2}$$

whenever $0 < |x - c| < \delta$. Fix x satisfying $0 < |x - c| < \delta$, and let $m \geqslant N$. Apply Mean Value Theorem to the function $f_m - f_N$ on the interval $[c,x]$ (if $x < c$, then on the interval $[x,c]$), there exists $\alpha \in (c,x)$ such that

$$f_m'(\alpha) - f_N'(\alpha) = \frac{(f_m(x) - f_N(x)) - (f_m(c) - f_N(c))}{x - c}$$

Recall that our choice of N implies

$$|f_m'(\alpha) - f_N'(\alpha)| < \frac{\epsilon}{3}$$

and so it follows that

$$\left| \frac{f_m(x) - f_m(c)}{x - c} - \frac{f_N(x) - f_N(c)}{x - c} \right| < \frac{\epsilon}{3}$$

Because $f_m \to f$ pointwise, we can take $m \to \infty$ and by order limit theorem, we have

$$\left| \frac{f(x) - f(c)}{x - c} - \frac{f_N(x) - f_N(c)}{x - c} \right| \leqslant \frac{\epsilon}{3} \tag{6.3}$$

Now, combining Equations 6.1, 6.2 and 6.3, we have if $0 < |x - c| < \delta$, we have

$$\begin{aligned} \left| \frac{f(x) - f(c)}{x - c} - g(c) \right| &\leqslant \left| \frac{f(x) - f(c)}{x - c} - \frac{f_N(x) - f_N(c)}{x - c} \right| + \left| \frac{f_N(x) - f_N(c)}{x - c} - f_N'(c) \right| + |f_N'(c) - g(c)| \\ &< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon \end{aligned}$$

which completes our goal. □

We see that the assumption in this theorem is relatively strong. We assume that the derivatives converge uniformly, which seems like assuming what we need to prove. Indeed, there is a version of Differentiable Limit Theorem that relaxes the assumption, and lead to a stronger statement. Instead of pointwise convergence on the whole domain, we need only to ensure one point convergence so that the expected function is not different by a constant.

Lemma 6.2.5: Differentiability Condition of Uniform Convergence

Let (f_n) be a sequence of differentiable functions defined on $[a, b]$, and assume (f_n') converges uniformly on $[a, b]$. If there exists a point $x_0 \in [a, b]$ where $f_n(x_0)$ is convergent, then (f_n) converges uniformly on $[a, b]$.

Proof. Let $\epsilon > 0$. Our goal is to show that there exists $N \in \mathbb{N}$ so that

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

whenever $n, m \geqslant N_1$ and $x \in [a, b]$. Then by Cauchy Criterion for uniform convergence, we can arrive our desired property.

Since (f_n') converges uniformly on $[a,b]$, we can choose $N \in \mathbb{N}$ such that

$$|f_n'(x) - f_m'(x)| < \frac{\epsilon}{2(b-a)}$$

for all $m, n \geqslant N$ and $x \in [a,b]$. Apply Mean Value Theorem on the function $h_{m,n} = f_n(x) - f_m(x)$ within the interval $[x, x_0]$ for some $x < x_0$ (the process is similar for $x \geqslant x_0$), we have

$$\frac{(f_n(x) - f_m(x)) - (f_n(x_0) - f_m(x_0))}{x - x_0} = f_n'(x_1) - f_m'(x_1)$$

for some $x_1 \in [a,b]$. Since our choice N makes $|f_n'(x) - f_m'(x)| < \frac{\epsilon}{2(b-a)}$, we can make

$$|(f_n(x) - f_m(x)) - (f_n(x_0) - f_m(x_0))| = |x - x_0||f_n'(x_1) - f_m'(x_1)| < (b-a) \times \frac{\epsilon}{2(b-a)} = \frac{\epsilon}{2}$$

for all $n, m \geqslant N$ and $x \in [a,b]$.

Moreover, sine f_n converges at point x_0, we can take $N_2 \in \mathbb{N}$ such that for all $n, m \geqslant N_2$, we have

$$|f_n(x_0) - f_m(x_0)| < \frac{\epsilon}{2}$$

Therefore, if we choose $N = \max\{N_1, N_2\}$, then for all $n, m \geqslant N$ and $x \in A$, we have

$$|f_n(x) - f_m(x)| \leqslant |(f_n(x) - f_m(x)) - (f_n(x_0) - f_m(x_0))| + |f_n(x_0) - f_m(x_0)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

which completes the proof. □

Combining Theorem 6.2.4 and Lemma 6.2.5, we have our desired corollary as the stronger version of Differentiable Limit Theorem.

Corollary 6.2.6: Stronger Version of Differentiable Limit Theorem

Let (f_n) be a sequence of differentiable functions defined on $[a,b]$, and assume (f_n') converges uniformly on $[a,b]$ to a function g. If there exists a point $x_0 \in [a,b]$ where $f_n(x_0)$ is convergent, then (f_n) converges uniformly on $[a,b]$ to a function f. Moreover, f is differentiable and satisfies $f' = g$.

6.3 Series of Functions

In this simple section, we naturally extend our language from sequence of functions to series of functions. Also, both preservation of continuity and differentiability of uniform convergence will have corresponding result in the language of series. Similarly as series of real numbers, we define convergence of series of functions by partial sum.

Definition 6.3.1: Pointwise/Uniform convergence of Series of Functions

For each $n \in \mathbb{N}$, let f_n and f be defined on $A \subseteq \mathbb{R}$. Consider the infinite series

$$\sum_{n=1}^{\infty} f_n(x) = f_1(x) + f_2(x) + \cdots$$

and its corresponding partial sum

$$s_k(x) = \sum_{n=1}^{k} f_n(x) = f_1(x) + f_2(x) + \cdots + f_k(x)$$

- $\sum_{n=1}^{\infty} f_n(x)$ **converges pointwise** on A to $f(x)$ if the sequence of partial sums $s_k(x)$ converges pointwise on A to $f(x)$.
- $\sum_{n=1}^{\infty} f_n$ **converges uniformly** on A to f if the sequence of partial sums s_k converges uniformly on A to f.

Now we state the corresponding two theorems of Continuous/Differentiable Limit Theorem.

Theorem 6.3.2: Term-by-Term Continuity Theorem

Let f_n be continuous functions on $A \subseteq \mathbb{R}$. Assume $\sum_{n=1}^{\infty} f_n(x)$ converges uniformly on A to f. Then, f is continuous on A.

Proof. By Algebraic Continuity Theorem 4.2.3, the partial sums s_k are continuous functions on A. Therefore, by Continuous Limit Theorem 6.2.3, f is continuous on A. □

Theorem 6.3.3: Term-by-Term Differentiability Theorem

Let f_n be differentiable on $[a, b]$. Assume $\sum_{n=1}^{\infty} f_n'(x)$ converges uniformly to $g(x)$ on $[a, b]$. If there exists $x_0 \in [a, b]$ such that $\sum_{n=1}^{\infty} f_n(x_0)$ converges, then the series $\sum_{n=1}^{\infty} f_n(x)$ converges uniformly on $[a, b]$ to a function $f(x)$ satisfying $f'(x) = g(x)$. i.e,

$$f(x) = \sum_{n=1}^{\infty} f_n(x) \quad \text{and} \quad f'(x) = \sum_{n=1}^{\infty} f_n'(x)$$

Proof. By Algebraic Differentiability Theorem 5.1.4, the partial sums are differentiable and $s_k' = f_1' + f_2' + \cdots + f_k'$. Then this theorem follows from Stronger Version of Differentiable Limit Theorem 6.2.6. □

Correspondingly, we have Cauchy Criterion for uniform convergence here:

Corollary 6.3.4: Cauchy Criterion for Uniform Convergence of Series of Functions

A series $\sum_{n=1}^{\infty} f_n(x)$ converges uniformly on $A \subseteq \mathbb{R}$ if and only if

$$\forall \epsilon > 0, \exists N \in \mathbb{N} \text{ s.t. } n > m \geqslant N, x \in A \quad \Longrightarrow \quad |f_{m+1}(x) + f_{m+2}(x) + \cdots + f_n(x)| < \epsilon$$

There is a convenience way of judging uniform convergence of series that will be useful later.

Theorem 6.3.5: Weierstrass M-test

Let f_n be defined on $A \subseteq \mathbb{R}$. Let $M_n > 0$ be real numbers such that

$$\forall x \in A, |f_n(x)| \leqslant M_n$$

If $\sum_{n=1}^{\infty} M_n$ converges, then $\sum_{n=1}^{\infty} f_n$ converges uniformly on A.

Proof. Since $\sum_{n=1}^{\infty} M_n$ converges, by Cauchy Criterion for series of real numbers, there exists $n > m \geqslant N$ such that $|M_{m+1} + M_{m+2} + \cdots + M_n| = M_{m+1} + M_{m+2} + \cdots + M_n < \epsilon$. Then,

$$|f_{m+1}(x)+f_{m+2}(x)+\cdots+f_n(x)| \leqslant |f_{m+1}(x)|+|f_{m+2}(x)|+\cdots+|f_n(x)| \leqslant M_{m+1}+M_{m+2}+\cdots+M_n < \epsilon$$

By Cauchy Criterion for Series of Functions 6.3.4, we have $\sum_{n=1}^{\infty} f_n(x)$ converges uniformly. □

6.4 Power Series

In this section we are going to discuss about a specific kind of series of functions: Power Series.

Definition 6.4.1: Power Series

Series of the form

$$f(x) = \sum_{n=0}^{\infty} a_n x^n = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots$$

are called **power series**.

6.4.1 Radius of Convergence

It is intuitive that for smaller x, the series is easier to converge. Heuristically, if the power series converges at a large x, then for smaller x, it would also converge. The next theorem proves this thought.

Theorem 6.4.2: Convergence Interval of Power Series

If $\sum_{n=0}^{\infty} a_n x^n$ converges at $x_0 \in \mathbb{R}$, then it converges absolutely for any x with $|x| < |x_0|$.

Proof. If $\sum_{n=0}^{\infty} a_n x_0^n$ converges, then the sequence of terms $(a_n x_0^n)$ is bounded. Suppose $|a_n x_0^n| \geqslant M$ for some $M > 0$. If $x \in \mathbb{R}$ satisfies $|x| < |x_0|$, then

$$|a_n x^n| = |a_n x_0^n| \left|\frac{x}{x_0}\right|^n \leqslant M \left|\frac{x}{x_0}\right|^n$$

Notice that

$$\sum_{n=0}^{\infty} M \left|\frac{x}{x_0}\right|^n$$

is a geometric series with ratio $|x/x_0| < 1$, thus converges. By Comparison Test, $\sum_{n=0}^{\infty} a_n x^n$ converges absolutely. $\square$

Considering the previous theorem, the set of convergence for power series must be the form of $\{0\}$, $\mathbb{R}$, or a bounded interval centered at $x = 0$. These intervals can be the form of $(-R, R)$, $[-R, R)$, $(-R, R]$ or $[-R, R]$.

Definition 6.4.3: Radius of Convergence

- The value R in the set of convergence is called the **radius of convergence** of a power series.
- If the set of convergence is $\{0\}$ or $\mathbb{R}$, then the radius of convergence is defined as 0 and ∞, respectively.

To evaluate the radius of convergence of a power series, we can just use a modified version of the Root Test 2.5.8 and the Ratio Test 2.5.9 in Chapter 2.

Corollary 6.4.4: Root Test for Power Series

Let $\sum_{n=1}^{\infty} a_n x^n$ be a power series, let $\alpha = \limsup_{n\to\infty} |a_n|^{1/n}$.

(a) If $\alpha = \infty$, then the series converges only at $x = 0$.

(b) If $0 < \alpha < \infty$, then the series converges absolutely in $(-1/\alpha, 1/\alpha)$. The convergency on the two endpoints are not decided.

(c) If $\alpha = 0$, the series converges for all $x \in \mathbb{R}$.

Proof. By Theorem 2.5.8, the series converges absolutely when

$$\limsup_{n\to\infty} |a_n|^{1/n}|x^n|^{1/n} = |x| \limsup_{n\to\infty} |a_n|^{1/n} = |x|\alpha < 1$$

Therefore, it converges when $|x| < 1/\alpha$. If $\alpha = \infty$, then the only x that satisfies is 0. If $\alpha = 0$, then all x will satisfy. □

Corollary 6.4.5: Ratio Test for Power Series

Let $\sum_{n=1}^{\infty} a_n x^n$ be a power series such that $a_n \neq 0$, let

$$r = \limsup_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right|$$

(a) If $r = \infty$, then the series converges only at $x = 0$.

(b) If $0 < r < \infty$, then the series converges absolutely in $(-1/r, 1/r)$. The convergency on the two endpoints are not decided.

(c) If $r = 0$, the series converges for all $x \in \mathbb{R}$.

Proof. By Theorem 2.5.9, the series converges absolutely when

$$\limsup_{n\to\infty} \left|\frac{a_{n+1}x^{n+1}}{a_n x^n}\right| = |x| \limsup_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| = |x|r < 1$$

Therefore, it converges when $|x| < 1/r$. If $r = \infty$, then the only x that satisfies is 0. If $r = 0$, then all x will satisfy. □

6.4.2 Uniform Convergence of Power Series on Intervals

One of the most satisfactory result for power series is that, absolute convergence at a point implies the uniform convergence on the corresponding closed interval.

Theorem 6.4.6: Absolute Convergence implies Uniform Convergence

If power series $\sum_{n=0}^{\infty} a_n x^n$ converges absolutely at x_0, then it converges uniformly on $[-|x_0|, |x_0|]$.

Proof. Since it converges aboslutely at x_0, $\sum_{n=0}^{\infty} |a_n x_0^n|$ converges. By Weirestrass M-test, since for all $x \in [-|x_0|, |x_0|]$,

$$|a_n x^n| \leqslant |a_n x_0^n|$$

we have $\sum_{n=0}^{\infty} a_n x^n$ converges uniformly on $[-|x_0|, |x_0|]$. $\square$

A direct application of this theorem is the continuity of power series on an open interval.

Proposition 6.4.7: Continuity of Power Series on an Open Interval

If power series $\sum_{n=0}^{\infty} a_n x^n$ converges on $(-R, R)$, then $f(x) = \sum_{n=0}^{\infty} a_n x^n$ is continuous on this interval.

Proof. First by Theorem 6.4.2, If $\sum_{n=0}^{\infty} a_n x^n$ converges on $(-R, R)$, then it converges absolutely on $(-R, R)$.

Now, since any $x \in (-R, R)$ in contained in the interior of some $[-c, c] \subseteq (-R, R)$, by Theorem 6.4.6, it is continuous at any $x \in (-R, R)$. $\square$

6.4.3 Uniform Convergence of Power Series on Endpoints: Abel's Theorem

Note that on the endpoints of set of convergence, we have no idea yet. If at $x = R$ it is absolutely converge, then we can still apply Theorem 6.4.6. But how about conditional convergence? First note that it is possible that it converges conditionally at $x = R$, and does not converge at $x = -R$. For example, the series

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

when $R = 1$, it converges conditionally. However, at $R = -1$, it does not converge. This subsection is dedicated to solve this problem.

Lemma 6.4.8: Abel's Lemma

Let b_n satisfy $b_1 \geqslant b_2 \geqslant b_3 \geqslant \cdots \geqslant 0$, and let $\sum_{n=1}^{\infty} a_n$ be a series with partial sums are bounded, i.e., there exists $A > 0$ such that for all $n \in \mathbb{N}$,

$$|a_1 + a_2 + \cdots + a_n| \leqslant A$$

Then, for all $n \in \mathbb{N}$, we have

$$|a_1 b_1 + a_2 b_2 + a_3 b_3 + \cdots + a_n b_n| \leqslant A b_1$$

Proof. Let $s_n = a_1 + a_2 + a_3 + \cdots + a_n$. We can rewrite

$$\begin{aligned}\sum_{k=1}^{n} a_k b_k &= \sum_{k=1}^{n} b_k(s_k - s_{k-1}) \\ &= b_1 s_1 + b_2 s_2 - b_2 s_1 + b_3 s_3 - b_3 s_2 + \cdots + b_n s_n - b_n s_{n-1} \\ &= s_1(b_1 - b_2) + s_2(b_2 - b_3) + \cdots + s_{n-1}(b_{n-1} - b_n) + s_n b_n \\ &= \left(\sum_{k=1}^{n-1} s_k(b_k - b_{k+1})\right) + s_n b_n\end{aligned}$$

Therefore,

$$\begin{aligned}\left|\sum_{k=1}^{n} a_k b_k\right| &= \left|\left(\sum_{k=1}^{n-1} s_k(b_k - b_{k+1})\right) + s_n b_n\right| \\ &\leqslant A b_n + A\sum_{k=1}^{n-1}(b_k - b_{k+1}) \\ &= A b_n + A(b_1 - b_n) = A b_1\end{aligned}$$

which completes the proof. □

Proposition 6.4.9: Abel's Theorem

Let power series $\sum_{n=0}^{\infty} a_n x^n$ converge at $x = R > 0$. Then the series converges uniformly on $[0, R]$. Similar results hold for $x = -R$.

Proof. Write

$$g(x) = \sum_{n=0}^{\infty} a_n x^n = \sum_{n=0}^{\infty} (a_n R^n) \left(\frac{x}{R}\right)^n$$

Let $\epsilon > 0$. Since we assume $\sum_{n=0}^{\infty} a_n R^n$ converges, by Cauchy Criterion for convergent series of real numbers, there exists $N \in \mathbb{N}$ such that

$$\left|a_{m+1}R^{m+1} + a_{m+2}R^{m+2} + \cdots + a_n R^n\right| < \frac{\epsilon}{2}$$

whenever $n > m \geqslant N$. If we see $(a_n R^n)$ as a_n, and $(x/R)^n$ as b_n, which is monotone decreasing. By Abel's Lemma 6.4.8,

$$\left|(a_{m+1}R^{m+1})\left(\frac{x}{R}^{m+1}\right) + (a_{m+2}R^{m+2})\left(\frac{x}{R}^{m+2}\right) + \cdots + (a_n R^n)\left(\frac{x}{R}^{n}\right)\right| \leqslant \frac{\epsilon}{2}\left(\frac{x}{R}\right)^{m+1} < \epsilon$$

By Cauchy Criterion for Uniform Convergence of Series 6.3.4, it is indeed uniformly convergent. □

Combining all these we have done throughout the two subsections on uniform convergence, we have the concluding corollary below.

Corollary 6.4.10: Pointwise Convergence implies Uniform Convergence on Compact Set

If a power series converges pointwise on $A \subseteq \mathbb{R}$, then it converges uniformly on any compact set $K \subseteq \mathbb{R}$.

Proof. A compact set contains its maximum and minimum $x_1, x_0 \in A$. Abel's Theorem implies that the series converges uniformly on $[x_0, x_1] \supseteq K$. □

Corollary 6.4.11: Continuity of Convergent Power Series

A power series is continuous at every point at which it converges.

6.4.4 Derivative and Anti-differentiation of Power Series

The more interesting thing about power series is that, it is infinitely differentiable without changing its radius of convergence.

Lemma 6.4.12: Presevation of Radius of Convergence after Differentiation

If $\sum_{n=0}^{\infty} a_n x^n$ converges on $(-R, R)$, then the term-by-term differentiation $\sum_{n=0}^{\infty} n a_n x^{n-1}$ converges also on $(-R, R)$.

Proof. We know $|x| \limsup_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| < 1$ for all $x \in (-R, R)$. By ratio test again,

$$\limsup_{n\to\infty} \left|\frac{(n+1)a_{n+1}x^n}{n a_n x^{n-1}}\right| = |x| \limsup_{n\to\infty} \left|\frac{n+1}{n}\right| \limsup_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| < 1$$

which leads to the same result. □

With this, we can state our conclusion that, power series can be differentiated term-by-term without changing the convergence on interior of set of convergence.

Theorem 6.4.13: Term-by-term Differentiation of Power Series

Assume

$$f(x) = \sum_{n=0}^{\infty} a_n x^n$$

converges on an interval $A \subseteq \mathbb{R}$. Then $f(x)$ is differentiable on any $(-R, R) \subseteq A$. The derivative is

$$f'(x) = \sum_{n=1}^{\infty} n a_n x^{n-1}$$

Moreover, it is infinitely differentiable on $(-R, R)$, the successive derivatives can be obtained by continuing this term-by-term process.

Proof. By Lemma 6.4.12, along with Abel's Theorem, the power series and its derivative satisfy all assumptions of Term-by-Term differentiability theorem 6.3.3. This process is inductive, so it is infinitely differentiable. □

Accordingly, the 'inverse algebra' of differentiation could also be applied on power series. This is called anti-derivatives.

Definition 6.4.14: Anti-differentiation

A function F is an **anti-derivative** of f if $F' = f$. The anti-derivative is not unique up to an addition of constant c.

Theorem 6.4.15: Term-by-Term Anti-Differentiation of Power Series

Assume

$$f(x) = \sum_{n=0}^{\infty} a_n x^n$$

converges on an interval $A \subseteq \mathbb{R}$. Then, its anti-derivative

$$F(x) = \sum_{n=0}^{\infty} \frac{a_n}{n+1} x^{n+1}$$

is convergent on $(-R, R)$, with $F'(x) = f(x)$.

Example 6.4.1. We have known that the geometric series

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots, \quad |x| < 1$$

By differentiation, we have

$$\frac{1}{(1-x)^2} = 1 + 2x + 3x^2 + 4x^3 + \cdots, \quad |x| < 1$$

If we use the simple substitution, changing x to $-x^2$, we will have

$$\frac{1}{1+x^2} = 1 - x^2 + x^4 - x^6 + x^8 - \cdots, \quad |x| < 1$$

Then anti-differentiation results in

$$\arctan(x) = x - \frac{1}{3}x^3 + \frac{1}{5}x^5 - \frac{1}{7}x^7 + \cdots, \quad |x| < 1$$

Indeed, this formula also works for $x = \pm 1$. Indeed, we can prove that this power series is convergent at $x = 1$. By Abel's Theorem, it converges uniformly at $x = \pm 1$, and thus it is continuous. Taking limit, we will have, at $x = 1$,

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots$$

Example 6.4.2. From previous example, we have

$$\frac{1}{(1-x)^2} = 1 + 2x + 3x^2 + 4x^3 + \cdots, \quad |x| < 1$$

If we choose $x = 1/2$, we will have the result

$$\sum_{n=1}^{\infty} \frac{n}{2^n} = \frac{x}{(1-x)^2} = 2$$

Differentiate again,

$$\left(\sum_{n=1}^{\infty} nx^n\right)' = \sum_{n=1}^{\infty} n^2 x^{n-1} = \frac{1}{(1-x)^2} + \frac{2x}{(1-x)^3}$$

Therefore, with $x = 1/2$, we have

$$\sum_{n=1}^{\infty} \frac{n^2}{2^n} = \frac{x}{(1-x)^2} + \frac{2x^2}{(1-x)^3} = 6$$

6.5 Taylor Series

All of the works in the previous section is for this moment, the generation of Taylor Series, one of the most important tools in many fields of mathematics such as differential equations, numerical analysis and mathematical physics.

Consider the power series

$$f(x) = \sum_{n=0}^{\infty} a_n x^n = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots, \quad x \in A$$

where A is an interval. Clearly $a_0 = f(0)$. If we differentiate,

$$f'(x) = \sum_{n=0}^{\infty} n a_n x^{n-1} = a_1 + 2a_2 x + 3a_3 x^2 + \cdots, \quad x \in (-R, R) \subseteq A$$

Again, let $x = 0$, we have $a_1 = f'(0)$. Inductively, we can have $a_2 = f''(0)/2$, $a_3 = f^{(3)}(0)/6$, $a_4 = f^{(4)}(0)/24, \cdots$ Similar derivation can be done with different value of center x such that f is defined as

$$f(x) = \sum_{n=0}^{\infty} a_n (x-a)^n = a_0 + a_1(x-a) + a_2(x-a)^2 + a_3(x-a)^3 + \cdots, \quad x \in A$$

Definition 6.5.1: Taylor Series/Maclaurin Series

Suppose $f(x)$ can be expressed as a power series

$$f(x) = \sum_{n=0}^{\infty} a_n (x-a)^n = a_0 + a_1(x-a) + a_2(x-a)^2 + a_3(x-a)^3 + \cdots$$

on some nontrivial interval centered at $a \in \mathbb{R}$. Then the **Taylor Series** of this function takes

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

If $a = 0$, the **Maclaurin Series** of this function takes

$$a_n = \frac{f^{(n)}(0)}{n!}$$

6.5.1 A Counterexample that Taylor Series does not Converge to Original Function

Does all Taylor Series converges to the function that it represents? The answer is no, and we show a classic counterexample here.

Consider the function

$$f(x) = \begin{cases} e^{-1/x^2}, & \text{if } x \neq 0 \\ 0, & \text{if } x = 0 \end{cases}$$

Clearly that $f(0) = 0$. If we take derivative,

$$a_1 = f'(0) = \lim_{x \to 0} \frac{f(x) - f(0)}{x - 0} = \lim_{x \to 0} \frac{e^{-1/x^2}}{x}$$

By L'Hôpital's Rule,

$$a_1 = \lim_{x\to 0} \frac{e^{-1/x^2}}{x} = \lim_{x\to 0} \frac{1/x}{e^{1/x^2}} = \lim_{x\to 0} \frac{-1/x^2}{-2e^{1/x^2}x^{-3}} = \lim_{x\to 0} \frac{x}{2e^{1/x^2}} = 0$$

Continuing this fashion, we can see that $f^{(n)}(0) = 0$ for all $n \in \mathbb{N}$. Therefore, the Taylor Series of $f(x)$ at $x = 0$ is actually $f(x) = 0$, which does not converge to itself.

The unmistakable conclusion is that not every infinitely differentiable function can be represented by its Taylor series.

6.5.2 Lagrange's Remainder

Though not all Taylor Series converges to the function it represents, we can analyze the magnitude of error term

$$E_N(x) = f(x) - S_N(x)$$

where $S_N(x)$ is the partial sum to Nth term.

Theorem 6.5.2: Lagrange's Remainder Theorem

Let f be $(n+1)$-times differentiable on (a, b). Then, for any $c, x \in (a, b)$, there exists y between c and x such that

$$f(x) = f(c) + f'(c)(x-c) + \frac{f''(c)}{2!}(x-c)^2 + \cdots + \frac{f^{(n)}(c)}{n!}(x-c)^n + \frac{f^{(n+1)}(y)}{(n+1)!}(x-c)^{n+1}$$

Proof. Denote

$$S_n(x) = f(c) + f'(c)(x-c) + \frac{f''(c)}{2!}(x-c)^2 + \cdots + \frac{f^{(n)}(c)}{n!}(x-c)^n$$

the partial sum of n terms, and let $E_n(x) = f(x) - S_n(x)$. Note that since f and S_n are both $(n+1)$-times differentiable, so does E_n. To simplify notation, let's assume $x > c$ and apply Generalized Mean Value Theorem to the function $E_n(x)$ and $(x-c)^{n+1}$ on interval $[c, x]$. There exists $x_1 \in (c, x)$ such that

$$\frac{E_n(x)}{(x-c)^{n+1}} = \frac{E_n'(x)}{(n+1)(x-c)^n}$$

Now apply the Generalized Mean Value Theorem to the functions $E_n'(x)$ and $(n+1)(x-c)^n$ on the interval $[c, x_1]$, there exists $x_2 \in (c, x_1)$ such that

$$\frac{E_n(x)}{(x-c)^{n+1}} = \frac{E_n'(x_1)}{(n+1)(x_1-c)^n} = \frac{E_n''(x_2)}{(n+1)n(x_2-c)^{n-1}}$$

Continuing this fashion, we will find a $x_{n+1} \in (c, x)$ such that

$$\frac{E_n(x)}{(x-c)^{n+1}} = \frac{E_n^{(n+1)}(x_{n+1})}{(n+1)!}$$

Set $y = x_{n+1}$. Because $S_n^{n+1}(x) = 0$, we have $E_n^{n+1}(x) = f^{n+1}(x)$. Therefore,

$$\frac{E_n(x)}{(x-c)^{n+1}} = \frac{f^{(n+1)}(y)}{(n+1)!} \quad \Longrightarrow \quad E_n(x) = \frac{f^{(n+1)}(y)}{(n+1)!}(x-c)^{n+1}$$

which completes the proof. □

We see that in the remainder, on denominator $(n+1)!$ helps to make E_n small as n tends to infinity. On numerator, $(x-c)^{n+1}$ potentially grows depending on the size of x. Thus, we should expect that a Taylor series is less likely to converge the farther x is chosen from the origin c. $f^{(n+1)}(y)$ is then a mystery. However, for functions with explicit derivatives, we can always deal with this term using an upper bound.

Example 6.5.1. In this example, we will prove that

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \cdots = \ln 2$$

Proof. Recall that the geometric series

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots, \quad |x| < 1$$

If we substitute x by $-x$, we have

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + \cdots, \quad |x| < 1$$

Do anti-differentiation on this formula, we will have

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \frac{x^5}{5} - \frac{x^6}{6} + \cdots, \quad |x| < 1$$

We know that this series also converges at $x = 1$. To see whether this taylor series converges to the expected function, let $f(x) = \ln(1+x)$ and consider the Lagrange's Remainder

$$E_n(x) = \frac{f^{(n+1)}(y)}{(n+1)!}x^{n+1}, \quad y \in (0, x)$$

We see that there is an upper bound

$$\left|f^{(n+1)}(y)\right| = n!(1+c)^{-n-1}$$

Therefore, for $c < x = 1$ we have that the remainder

$$|E_n(x)| \leqslant \frac{n!(1+c)^{-n-1}}{(n+1)!}x^{n+1} \leqslant \frac{1}{n+1} \to \infty$$

as n goes to infinity. Therefore, the Taylor series converges to $\ln(1+x)$ on $(0,1)$. Since $\ln(1+x)$ is continuous at $x=1$, and the Taylor series is continuous at $x=1$ by Abel's Theorem, the equality can also be extended to $x=1$ by taking limit. $\square$

6.5.3 Peano's Remainder

If we are not concerning about the convergence of Taylor Series, we can write the remainder term in a more compact way.

Definition 6.5.3: Little o Notation

We write $f(x) = o(g(x))$ as $x \to c$ if
$$\lim_{x\to c}\frac{f(x)}{g(x)} = 0$$

Heuristically, this indicates that the convergence rate of $f(x)$ respect to point 0 is much faster than $g(x)$. Therefore, if both $f(x)$ and $g(x)$ converges to 0, then $f(x)$ would be a 'negligible' term in a formula containing both $f(x)$ and $g(x)$ when approximating x to c.

Proposition 6.5.4: Peano's Remainder Theorem

Let f be n-times differentiable on (a,b). Then, for $c \in (a,b)$ there exists a little o form such that

$$f(x) = f(c) + f'(c)(x-c) + \frac{f''(c)}{2!}(x-c)^2 + \cdots + \frac{f^{(n)}(c)}{n!}(x-c)^n + o\left((x-c)^n\right)$$

Since we have $f^{(k)}(c) = S_n^{(k)}(c)$ for all $0 \leqslant k \leqslant n$, where S_n is defined as

$$S_n(x) = f(c) + f'(c)(x-c) + \frac{f''(c)}{2!}(x-c)^2 + \cdots + \frac{f^{(n)}(c)}{n!}(x-c)^n$$

Then, $E_n(x) = f(x) - S_n(x)$ will have the property that

$$E_n^{(k)}(c) = 0 \text{ for all } 0 \leqslant k \leqslant n$$

Therefore, by iterative L'Hôpital's Rule,

$$\lim_{x\to c}\frac{E_n(x)}{(x-c)^n}=\lim_{x\to c}\frac{E_n'(x)}{n(x-c)^{n-1}}=\lim_{x\to c}\frac{E_n''(x)}{n(n-1)(x-c)^{n-2}}$$
$$=\cdots=\lim_{x\to c}\frac{E_n^{(n)}(x)}{n!}=0$$

6.5.4* Cauchy's Remainder

There is another form of remainder, which is quite similar with the Lagrange's version.

Theorem 6.5.5: Cauchy's Remiander Theorem

Let f be $(n+1)$-times differentiable on (a,b). Then, for any $c,x\in(a,b)$, there exists y between c and x such that

$$f(x)=f(c)+f'(c)(x-c)+\frac{f''(c)}{2!}(x-c)^2+\cdots+\frac{f^{(n)}(c)}{n!}(x-c)^n+\frac{f^{(n+1)}(y)}{n!}(x-y)^n(x-c)$$

Proof. Let $S_n(x,c)$ be the partial sum of taylor series of f centered at c,

$$S_n(x,c)=f(c)+f'(c)(x-c)+\frac{f''(c)}{2!}(x-c)^2+\cdots+\frac{f^{(n)}(c)}{n!}(x-c)^n$$

Let $E_n(x,c)=f(x)-S_n(x,c)$. Now if we fix $x\neq 0$ and consider $E_n(x,c)$ as a function of c. Then,

$$E_n(x,x)=f(x)-S_n(x,x)=f(x)-f(x)=0$$

Moreover, $E_n(x,c)$ is differentiable with respect to a, since

$$E_n'(x,c)=(f(x)-S_n(x,c))'=-S_n'(x,c)=-\left(\sum_{k=0}^{n}\frac{f^{(k)}(c)}{k!}(x-c)^k\right)'$$
$$=-\sum_{k=0}^{n}\frac{f^{(k+1)}(c)}{k!}(x-c)^k+\sum_{k=1}^{n}\frac{f^{(k)}(c)}{(k-1)!}(x-c)^{k-1}$$
$$=-\sum_{k=0}^{n}\frac{f^{(k+1)}(c)}{k!}(x-c)^k+\sum_{k=0}^{n-1}\frac{f^{(k+1)}(c)}{k!}(x-c)^k$$
$$=-\frac{f^{(n+1)}(c)}{n!}(x-c)^n$$

Therefore, by Mean Value Theorem, there exists y between c and x such that

$$\frac{E_n(x,x) - E_n(x,c)}{x-c} = E_n'(x,y)$$

Sine $E_n(x,x) = 0$, we finally arrive

$$E_n(x,c) = -E_n'(x,y)(x-c) = \frac{f^{(n+1)}(y)}{n!}(x-y)^n(x-c)$$

which completes the proof. $\square$

6.6 A Continuous Nowhere-Differentiable Function

To end this chapter, we introduce a extremely pathological function as an important counterexample.

We have proved in Section 5.1.2 that a differentiable function must be continuous. How about the inverse? The answer is obviously false. The simplest counterexample is $h(x) = |x|$. But how pathological could a continuous function could be? In this section we will show that is it possible to construct a function that is continuous on all of $\mathbb{R}$ but fails to be differentiable at every point.

We start from the periodic function

$$h(x) = |x|, \quad x \in [-1,1], \qquad h(x+2) = h(x)$$

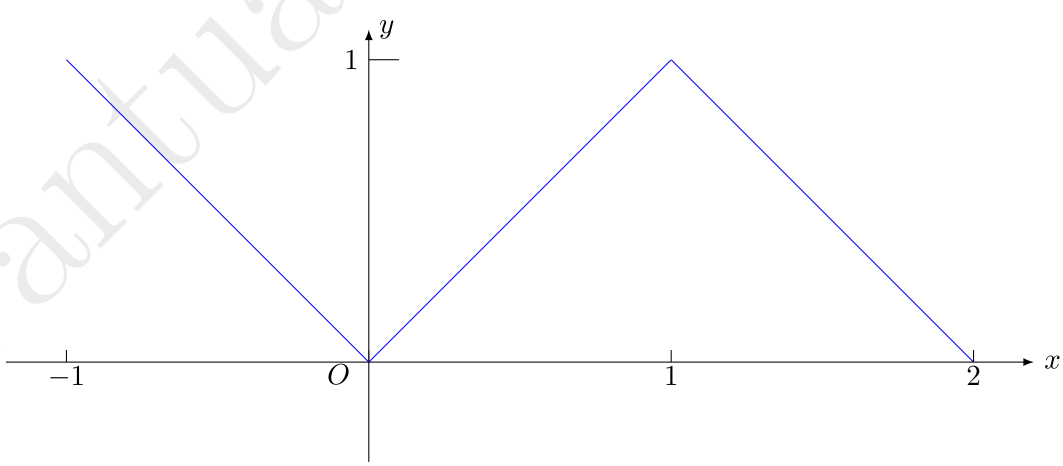


Figure 6.5: Function $h(x)$ with periodic extension

Now we define

$$h_n(x) = \left(\frac{3}{4}\right)^n h(4^n x)$$

These functions are just the n times 1/4-shrinkage of the original function $f(x)$ on x scale and 3/4-shrinkage on y scale.

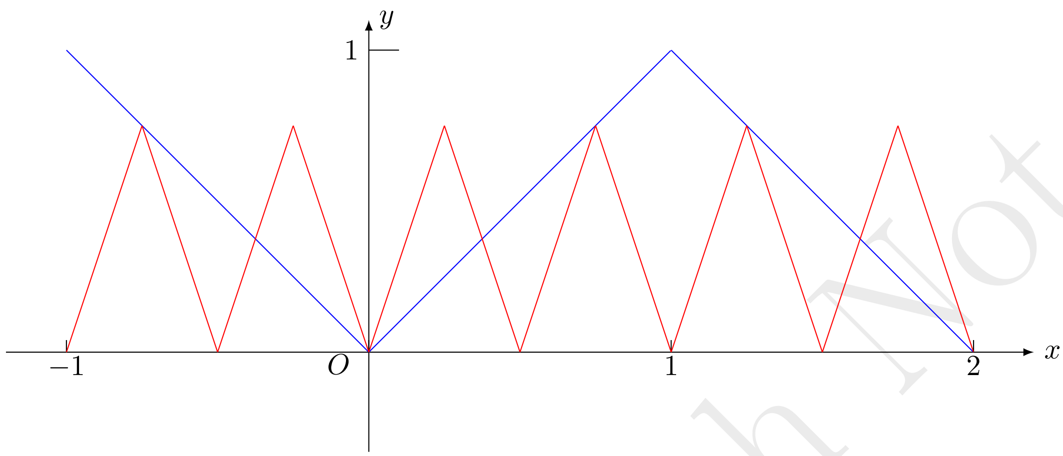


Figure 6.6: Blue: $h_0(x)$, Red: $h_1(x)$

If we define

$$g(x) = \sum_{n=0}^{\infty} h_n(x) = \sum_{n=0}^{\infty} \left(\frac{3}{4}\right)^n h(4^n x)$$

The claim is that $g(x)$ is continuous on all of $\mathbb{R}$ but fails to be differentiable at any point.

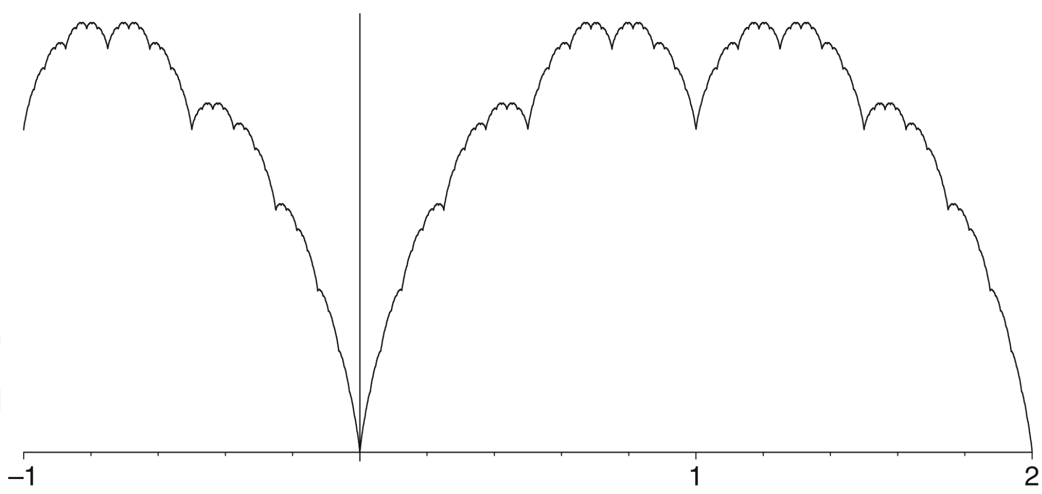


Figure 6.7: Graph of $g(x)$

Before doing any analysis, we need to show that the series converges on $\mathbb{R}$, and thus the function g is well-defined. Fix $x \in \mathbb{R}$, we know that

$$h_n(x) = \left(\frac{3}{4}\right)^n h(4^n x) \leqslant \left(\frac{3}{4}\right)^n$$

We know that $\sum_{n=1}^{\infty} 3^n/4^n$ converges. By comparison test, g does converge on x.

Now we show that g is continuous everywhere. We only need to show that this convergence is a uniform convergence. Indeed, by Weierstrass M-test, since $|h_n(x)| \leqslant |(3/4)^n|$, and $\sum_{n=1}^{\infty}(3/4)^n$ converges, it is uniformly continuous. Therefore, by Continuous limit theorem 6.2.3, g is continuous on $\mathbb{R}$.

The most difficult part is about the differentiation. We fix an arbitrary real number x. Choose $\delta_m = \pm 4^{-m}/2$, where the sign is chosen so that no integers lies between $4^m x$ and $4^m(x + \delta_m)$. This can be done since the length between $4^m x$ and $4^m(x + \delta_m)$ is $1/2$. Define

$$\gamma_{n,m} = \frac{h(4^n(x + \delta_m)) - h(4^n x)}{\delta_m}$$

When $n > m$, $4^n \delta_m$ is an even integer, so that $\gamma_{n,m} = 0$ since h is periodic with a period 2, which means $h(4^n(x + \delta_m)) = h(4^n x)$ in this case. By definition of h, it follows that

$$|h(x) - h(y)| \leqslant |x - y| \text{ for any } x, y \in \mathbb{R} \text{ with equality if there is no integer between } x \text{ and } y$$

This implies that for all $0 \leqslant n \leqslant m$,

$$|\gamma_{n,m}| = \left|\frac{h(4^n(x + \delta_m)) - h(4^n x)}{\delta_m}\right| \leqslant \left|\frac{4^n(x + \delta_m) - 4^n x}{\delta_m}\right| = 4^n$$

Since there is no integer between $4^m x$ and $4^m(x + \delta_m)$ for $n = m$, we have the equality

$$|\gamma_{m,m}| = \left|\frac{h(4^m(x + \delta_m)) - h(4^m x)}{\delta_m}\right| = \left|\frac{4^m(x + \delta_m) - 4^m x}{\delta_m}\right| = 4^m$$

With these, we deduce that

$$\begin{aligned}
\left|\frac{g(x + \delta_m) - g(x)}{\delta_m}\right| &= \left|\frac{\sum_{n=0}^{\infty}\left(\frac{3}{4}\right)^n h(4^n(x + \delta_m)) - \sum_{n=0}^{\infty}\left(\frac{3}{4}\right)^n h(4^n x)}{\delta_m}\right| \\
&= \left|\sum_{n=0}^{\infty}\left(\frac{3}{4}\right)^n \gamma_{n,m}\right| = \left|\sum_{n=0}^{m}\left(\frac{3}{4}\right)^n \gamma_{n,m}\right| \\
&\geqslant \left|\left(\frac{3}{4}\right)^m \gamma_{m,m}\right| - \left|\sum_{n=0}^{m-1}\left(\frac{3}{4}\right)^n \gamma_{n,m}\right|
\end{aligned}$$

$$\geqslant 3^m - \sum_{n=0}^{m-1} 3^n = \frac{3^m + 1}{2}$$

with $m \to \infty$, we have $\delta_m \to 0$, and

$$\left| \frac{g(x+\delta_m) - g(x)}{\delta_m} \right| \to \infty$$

Therefore, it is not differentiable at x. Since x is arbitrary, we have that g is not differentiable everywhere.

Chapter 7

Riemann Integration

7.1 Definition of Riemann Integral

Throughout this section, it is assumed that we are working with a bounded function f on a closed interval $[a, b]$.

7.1.1 Upper and Lower Sums

Definition 7.1.1: Partition

A **partition** $P = \{x_0, x_1, \cdots, x_n\}$ of $[a, b]$ is a finite set of points from $[a, b]$ that includes both a and b with

$$a = x_0 < x_1 < x_2 < \cdots < x_n = b$$

For each interval $[x_{k-1}, x_k]$ of P, let

$$m_k = \inf\{f(x) : x \in [x_{k-1}, x_k]\}, \quad \text{and} \quad M_k = \sup\{f(x) : x \in [x_{k-1}, x_k]\}$$

Definition 7.1.2: Lower/Upper Riemann Sum

The **lower sum** of f with respect to P is

$$L(f, P) = \sum_{k=1}^{n} m_k(x_k - x_{k-1})$$

The **upper sum** of f with respect to P is

$$U(f, P) = \sum_{k=1}^{n} M_k(x_k - x_{k-1})$$

Obviously, for a particular partition P, $U(f, P) \geqslant L(f, P)$. Moreover, we have more inequalities holds with respect to different partitions.

Definition 7.1.3: Refinement

A partition Q is a **refinement** of P if Q contains all points of P, denoted as $P \subseteq Q$.

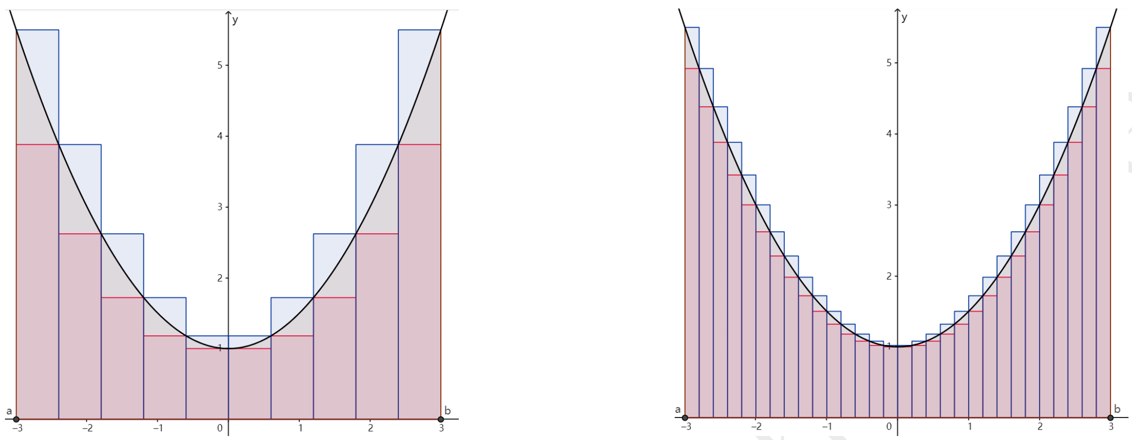


Figure 7.1: Blue: Upper Sum, Red: Lower Sum. Right panel is a refinement of left one

Lemma 7.1.4: Inequalities with Riemann Sums and Partition Refinement

If $P \subseteq Q$, then

$$L(f,P) \leqslant L(f,Q) \leqslant U(f,Q) \leqslant U(f,P)$$

Proof. Obviously $L(f,Q) \leqslant U(f,Q)$. We need to prove the other two.

Suppose $P = \{x_0, x_1, \cdots, x_n\}$ and $Q = \{x'_0, x'_1, \cdots, x'_N\}$ is a refinement. Then, for each $j = 1, 2, \cdots, n$, there exists $k \in \{0, 1, \cdots, N-1\}$ and a positive integer m such that

$$x_{j-1} = x'_k < x'_{k+1} < \cdots < x'_{k+m} = x_j$$

Then, if we denote

$$m'_{k+i} = \inf\{f(x) : x \in [x_{k+i-1}, x_{k+i}]\}$$

we have for each $j = 1, 2, \cdots, n$,

$$m_j(x_j - x_{j-1}) = \sum_{i=1}^{m} m_k(x'_{k+i} - x'_{k+i-1}) \leqslant \sum_{i=1}^{m} m'_{k+i}(x'_{k+i} - x'_{k+i-1})$$

Sum them up, we have $L(f,P) \leqslant L(f,Q)$. Similarly, we can derive another one for upper sum. $\square$

Lemma 7.1.5: Inequality with Riemann Sums of different partitions

If P_1 and P_2 are any two partitions of $[a, b]$, then

$$L(f, P_1) \leqslant U(f, P_2)$$

Proof. Let $Q = P_1 \cup P_2$. By Lemma 7.1.4, we have

$$L(f, P_1) \leqslant L(f, Q) \leqslant U(f, Q) \leqslant U(f, P_2)$$

which easily completes the proof. $\square$

7.1.2 Riemann Integrability

We take supremum/infimum of Riemann Lower/Upper sum as our basis for definition of Riemann integrability.

Definition 7.1.6: Lower/Upper Riemann Integral

Let $\mathcal{P}$ be the collection of all possible partitions of $[a, b]$.

- The **upper integral** of f is
$$U(f) = \inf\{U(f, P), P \in \mathcal{P}\}$$
- The **lower integral** of f is
$$L(f) = \sup\{L(f, P) : P \in \mathcal{P}\}$$

A direct corollary can be derived from previous lemmas by order preservation property of supremum and infimum.

Corollary 7.1.7: Inequality of Lower/Upper Integral

For any bounded function f, $U(f) \geqslant L(f)$.

Now we can define what is called 'integrable'.

Definition 7.1.8: Riemann Integrable

A bounded function f on $[a, b]$ is **Riemann Integrable** if $U(f) = L(f)$. In this case, we define

$$\int_a^b f = U(f) = L(f)$$

The criterion of integrability can be states as ϵ-language.

Proposition 7.1.9: Integrability Criterion

A bounded function f on $[a, b]$ is integrable if and only if

$$\forall\, \epsilon > 0, \exists\, \text{partition } P_\epsilon \in [a, b] \text{ s.t. } U(f, P_\epsilon) - L(f, P_\epsilon) < \epsilon$$

Proof.
($\Longleftarrow$) Let $\epsilon > 0$. Suppose such P_ϵ exists, then by definition

$$U(f) - L(f) \leqslant U(f, P_\epsilon) - L(f, P_\epsilon) < \epsilon$$

Since ϵ is arbitrary, it must be $U(f) = L(f)$.
($\Longrightarrow$) Suppose f is integrable. Then by characterization of supremum, there exists P_1 such that

$$U(f, P_1) < U(f) + \frac{\epsilon}{2}$$

Similarly, there exists P_2 such that

$$L(f, P_2) > L(f) - \frac{\epsilon}{2}$$

Let $P_\epsilon = P_1 \cup P_2$, then

$$U(f, P_\epsilon) - L(f, P_\epsilon) \leqslant U(f, P_1) - U(f, P_2) < U(f) + \frac{\epsilon}{2} - L(f) + \frac{\epsilon}{2} = \epsilon$$

which completes the proof. □

7.1.3 Relation with continuous functions

Theorem 7.1.10: Continuity implies Integrability

If f is continuous on $[a, b]$, then it is integrable.

Proof. Because f is continuous on a compact set, it must be bounded and uniformly continuous by Theorem 4.4.1 and Theorem 4.4.3. Thus, given $\epsilon > 0$, there exists $\delta > 0$ such that whenever $|x - y| < \delta$, we have

$$|f(x) - f(y)| < \frac{\epsilon}{b - a}$$

Let P be a partition on $[a, b]$ such that $\Delta x_k = (x_k - x_{k-1})$ is less than δ for each k. By Extremum Value Theorem 4.4.2, there exists $z_k, y_k \in [x_{k-1}, x_k]$ such that $M_k = f(z_k)$ and $m_k = f(y_k)$ for all k.

Then, this means that $|z_k - y_k| < \delta$. Therefore,

$$M_k - m_k = f(z_k) - f(y_k) < \frac{\epsilon}{b-a}$$

Finally,

$$U(f,P) - L(f,P) = \sum_{k=1}^{n} (M_k - m_k)\Delta x_k < \frac{\epsilon}{b-a} \sum_{k=1}^{n} \Delta x_k = \frac{\epsilon}{b-a}(b-a) = \epsilon$$

which complete the proof by Proposition 7.1.9. □

We note an important non-integrable function here. It is the **Dirichlet's Function**

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

For any subinterval $[x_{k-1}, x_k] \subseteq [a,b]$ with $x_{k-1} < x_k$, we see that there would always exist a rational and an irrational point because the density of these two kinds of numbers in $\mathbb{R}$. Therefore, $M_k = 1$ and $m_k = 0$ whatever the partition chosen.

However, if a function is different from a continuous function on only finitely many points, it is still integrable.

Proposition 7.1.11: Continuous Except for Finitely Many Points implies Integrability

If f is continuous on $[a,b]$ except finitely many points, then f is integrable.

Proof. Suppose $y_1, y_2, \cdots, y_m$ are discontinuous points and let $M = \sup_{x\in[a,b]} |f(x)|$. Let $\epsilon > 0$. For each i, choose the interval

$$\left[y_i - \frac{\epsilon}{8mM}, y_i + \frac{\epsilon}{8mM}\right]$$

Without losing generality, suppose these intervals does not overlap (If it overlaps, then just choose the even smaller interval that does not overlap and contains y_i, which does not influence the progression of the proof). Outside these intervals, f is continuous. Hence we can apply the previous theorem on each of the connected components (intervals) of the set

$$A = [a,b] \backslash \bigcup_{i=1}^{m} \left(y_i - \frac{\epsilon}{8mM}, y_i + \frac{\epsilon}{8mM}\right)$$

so that for each component $[a', b'] \subseteq A$, we can obtain a partition P' such that

$$|U(P) - L(P)| \leqslant \frac{\epsilon}{2(m+1)}$$

Note that A has at most $(m+1)$ connected components, and let P be the union of the partitions of all components of A and endpoints a, b. Then P is a partition of $[a, b]$ with

$$\begin{aligned} U(f,P) - L(f,P) &\leqslant (m+1) \times \frac{\epsilon}{2(m+1)} + \sum_{i=1}^{m} (M_i - m_i) \frac{\epsilon}{4mM} \\ &\leqslant \frac{\epsilon}{2} + M \sum_{i=1}^{m} \frac{\epsilon}{4mM} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \end{aligned}$$

which completes the proof. □

7.2 Properties of Riemann Integral

7.2.1 Algebraic Properties

Proposition 7.2.1: Interval Additivity

Assume $f : [a, b] \to \mathbb{R}$ is bounded, let $c \in (a, b)$. Then,

- f is integrable on $[a, b]$ if and only if f is integrable on $[a, c]$ and $[c, b]$.
- In this case, we have
$$\int_a^b f = \int_a^c f + \int_c^b f$$

Proof. ($\Longrightarrow$) Suppose f is integrable on $[a, b]$. Then, for $\epsilon > 0$ there exists P such that $U(f, P) - L(f, P) < \epsilon$. Then, let $P_1 = P \cap [a, c]$ and $P_2 = P \cap [c, b]$. Because refining a partition can only potentially bring the upper and lower sums closer together, in this case we refine by a single point c, it follows that

$$U(f, P_1) - L(f, P_1) < \epsilon \quad \text{and} \quad U(f, P_2) - L(f, P_2) < \epsilon$$

implying that f is integrable on $[a, c]$ and $[c, b]$.
($\Longleftarrow$) Conversely, if f is integrable on $[a, c]$ and $[c, b]$, then given $\epsilon > 0$ we can produce partition P_1 of $[a, c]$ and P_2 of $[c, b]$, such that

$$U(f, P_1) - L(f, P_1) < \frac{\epsilon}{2} \quad \text{and} \quad U(f, P_2) - L(f, P_2) < \frac{\epsilon}{2}$$

Let $P = P_1 \cup P_2$ produce a partition of $[a, b]$, and thus

$$U(f, P) - L(f, P) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

which indicates that it is integrable on $[a, b]$.

Continuing to let $P = P_1 \cup P_2$ as earlier, we have

$$\int_a^b f \leqslant U(f, P) < L(f, P) + \epsilon = L(f, P_1) + L(f, P_2) + \epsilon \leqslant \int_a^c f + \int_c^b f + \epsilon$$

Since ϵ is arbitrary, it must be that $\int_a^b f \leqslant \int_a^c f + \int_c^b f$. Conversely,

$$\int_a^c f + \int_c^b f \leqslant U(f, P_1) + U(f, P_2) < L(f, P_1) + L(f, P_2) + \epsilon = L(f, P) + \epsilon \leqslant \int_a^b f + \epsilon$$

Since ϵ is arbitrary, it must be that $\int_a^b f \geqslant \int_a^c f + \int_c^b f$. Combining the two, we have

$$\int_a^b f = \int_a^c f + \int_c^b f$$

as desired. □

Some other already familiar properties of Riemann Integration are listed below.

Proposition 7.2.2: Algebraic Properties of Riemann Integration

Let $f, g : [a, b] \to \mathbb{R}$ be integrable. Then,

1. For any $s, t \in \mathbb{R}$, $sf + tg$ is integrable with

$$\int_a^b (sf + tg) = s\int_a^b f + t\int_a^b g$$

2. If for all $x \in [a, b]$, we have $m \leqslant f(x) \leqslant M$, then

$$m(b - a) \leqslant \int_a^b f \leqslant M(b - a)$$

3. If for all $x \in [a, b]$, we have $f(x) \leqslant g(x)$, then

$$\int_a^b f \leqslant \int_a^b g$$

The proof is easy and omitted here.

We next consider the integrability of composition of functions.

Theorem 7.2.3: Integrability of Composition of Functions

Let $f : [a, b] \to \mathbb{R}$ be integrable with $m \leqslant f(x) \leqslant M$ for all $x \in [a, b]$. Let $\phi : [m, M] \to \mathbb{R}$ be continuous. Then the composition $\phi \circ f : [a, b] \to \mathbb{R}$ is integrable.

Proof. Note that ϕ is uniformly continuous. Hence,

$$\forall\, \epsilon > 0, \exists\, \delta < \epsilon \text{ s.t. } |x - y| < \delta, \quad \Longrightarrow \quad |\phi(x) - \phi(y)| < \epsilon$$

Since f is integrable, there exists partition P such that

$$U(f, P) - U(L, P) = \sum_{k=1}^{n} (M_k - m_k)(x_k - x_{k-1}) < \delta^2$$

where m_k, M_k are the minimal/maximal value of f over $[x_{k-1}, x_k]$, and we denote the minimal and maximal values of $\phi \circ f$ over $[x_{k-1}, x_k]$ by m_k^*, M_k^*. Let

$$A = \{k : M_k - m_k < \delta\} \quad \text{and} \quad B = \{k : M_k - m_k \geqslant \delta\}$$

Note that f ranges from m_k to M_k on $[x_{k-1}, x_k]$, thus $\phi \circ f$ ranges from m_k^* to M_k^*. By uniform continuity of ϕ, if $k \in A$, we will have $|M_k^* - m_K^*| < \epsilon$. For $k \in B$, we have the bound $|M_k^* - m_k^*| < 2M^*$, where $M^* = \sup |\phi(x)|$. However, for $k \in B$, the condition $M_k - m_k \geqslant \delta$ implies that

$$\sum_{k \in B} \delta(x_k - x_{k-1}) \leqslant \sum_{k \in B} (M_k - m_k)(x_k - x_{k-1}) < \delta^2$$

which implies that $\sum_{k \in B} (x_k - x_{k-1}) < \delta$. Therefore,

$$\begin{aligned} U(\phi \circ f, P) - L(\phi \circ f, P) &= \sum_{k \in A} (M_k^* - m_k^*)(x_k - x_{k-1}) + \sum_{k \in B} (M_k^* - m_k^*)(x_k - x_{k-1}) \\ &\leqslant \sum_{k \in A} \epsilon(x_k - x_{k-1}) + \sum_{k \in B} 2M^*(x_k - x_{k-1}) \\ &\leqslant \epsilon(b - a) + 2M^* \delta < \epsilon(b - a + 2M^*) \end{aligned}$$

Since ϵ is arbitrary, we have $U(\phi \circ f, P) = L(\phi \circ f, P)$, thus it is integrable. $\square$

Notice that the condition of ϕ be continuous is necessary. A composition of two integrable functions may be not integrable. For example, recall that Thomae's function

$$f(x) = \begin{cases} 1, & \text{if } x = 0 \\ 1/n, & \text{if } x = m/n \text{ with } \gcd(m,n) = 1 \text{ and } n > 0, m \neq 0 \\ 0, & \text{if } x \notin \mathbb{Q} \end{cases}$$

We will show that **Thomae's Function is integrable on** $[0,1]$. First of all, since irrational number is dense in $\mathbb{R}$, for each parition P, irrational numbers exist in each subintervals, therefore $L(f,P) = 0$ for all P. Now we consider the size of the upper sum. Let $\epsilon > 0$, and consider the set of points

$$D_{\epsilon/2} = \{x \in [0,1] : f(x) \geqslant \epsilon/2\}$$

This is the size of the set of rational numbers which, when expressed in lowest terms as m/n, have $n \leqslant \epsilon/2$. Therefore, this set $D_{\epsilon/2}$ must be finite. Denote the set as

$$D_{\epsilon/2} = \{r_1, r_2, \cdots, r_n\}$$

Then, we can choose intervals for each i,

$$\left[r_i - \frac{\epsilon}{4n}, r_i - \frac{\epsilon}{4n}\right]$$

Without losing generality, suppose these intervals does not overlap (If it overlaps, then just choose the even smaller interval that does not overlap and contains r_i, which does not influence the progression of the proof). In this case, we form a partition P_ϵ such that the endpoints of these intervals are set as partition points, i.e.,

$$P_\epsilon = \left\{r_i - \frac{\epsilon}{4n}, r_i - \frac{\epsilon}{4n}, 0, 1 : i \in 1, 2, \cdots, n\right\}$$

Using this partition, we have,

$$U(f, P_\epsilon) < \frac{\epsilon}{2} + \frac{\epsilon}{2n}n = \epsilon$$

Since ϵ is arbitrary, we have that

$$U(f) = 0 = L(f)$$

and Thomae's function is integrable. However, **the composition of Thomae's function with function** $g(x)$**, where**

$$g(x) = \begin{cases} 1, & \text{if } x = 1/n, n \in \mathbb{N} \\ 0, & \text{otherwise} \end{cases}$$

make $g \circ f$, is **Dirichlet's function**, which is not integrable.

Next we find an important inequality about Riemann Integration.

Proposition 7.2.4

Let $f : [a, b] \to \mathbb{R}$ be integrable. Then, $|f|$ is integrable, and

$$\left|\int_a^b f\right| \leqslant \int_a^b |f|$$

Proof. Choose $\phi(x) = |x|$. Since ϕ is continuous, by Theroem 7.2.3, the composition $g \circ f = |f|$ is integrable. If we choose α such that $\alpha \int_a^b f \geqslant 0$, we have

$$\left|\int_a^b f\right| = \alpha \int_a^b f = \int_a^b \alpha f \leqslant \int_a^b |f|$$

which completes the proof. □

7.2.2 Uniform Convergence Preserves Integrability

Sometimes pointwise convergence may not preserve the value of integral, even not preserve integrability for example, let

$$f(x) = \begin{cases} n, & \text{if } 0 < x < 1/n \\ 0, & \text{if } x = 0 \text{ or } x \geqslant 1/n \end{cases}$$

It is integrable with $\int_a^b f_n = 1$, and for each $x \in [0, 1]$, we have $f_n(x) \to 0$. Thus it converges pointwise to $f = 0$, however,

$$0 \neq \lim_{n\to\infty} \int_0^1 f_n$$

This means that we cannot change the order of the limit and integration. Similar as previous dicussion in Chapter 6, we can resolve this using uniform convergence.

Theorem 7.2.5: Integrable Limit Theorem

Let $f_n \to f$ uniformly on $[a, b]$, with each f_n integrable. Then, f is integrable and

$$\lim_{n\to\infty} \int_a^b f_n = \int_a^b \lim_{n\to\infty} f_n = \int_a^b f$$

Proof. We first prove that f is integrable. Fix $\epsilon > 0$. Since f_n is uniformly convergent, there exists

$N \in \mathbb{N}$ such that for all $n \geqslant N$ and $x \in [a, b]$ we have

$$|f_n(x) - f(x)| < \frac{\epsilon}{4(b-a)}$$

This means that when $n \geqslant N$,

$$f_n - \frac{\epsilon}{4(b-a)} < f < f_n + \frac{\epsilon}{4(b-a)}$$

This shows that,

$$L\left(f_n - \frac{\epsilon}{4(b-a)}\right) \leqslant L(f) \leqslant U(f) \leqslant U\left(f_n + \frac{\epsilon}{4(b-a)}\right)$$

since by Proposition 7.2.2, the functions on both ends of the inequality is integrable, therefore

$$\int_a^b \left(f_n - \frac{\epsilon}{4(b-a)}\right) \leqslant L(f) \leqslant U(f) \leqslant \int_a^b \left(f_n + \frac{\epsilon}{4(b-a)}\right)$$

With Proposition 7.2.2 again, we have

$$\left(\int_a^b f_n\right) - \frac{\epsilon}{4} \leqslant L(f) \leqslant U(f) \leqslant \left(\int_a^b f_n\right) + \frac{\epsilon}{4}$$

Therefore, we have

$$|U(f) - L(f)| \leqslant \left|U(f) - \int_a^b f_n\right| + \leqslant \left|L(f) - \int_a^b f_n\right| \leqslant \frac{\epsilon}{4} + \frac{\epsilon}{4} < \epsilon$$

which shows that f is integrable.

Now we prove the formula of limit-integral order change. This is easily derive from Proposition 7.2.4 such that for $n \geqslant N$,

$$\left|\int_a^b f_n - \int_a^b f\right| = \left|\int_a^b (f_n - f)\right| \leqslant \int_a^b |f_n - f| \leqslant \int_a^b \frac{\epsilon}{4(b-a)} < \epsilon$$

which completes the proof. □

7.3 The Fundamental Theorem of Calculus (FTC)

The derivative and the integral have been independently defined, each in its own rigorous mathematical terms. The definition of the derivative is motivated by the problem of finding slopes of tangent lines and is given in terms of function limits of difference quotients. The definition of the integral grows out of the desire to calculate areas under nonconstant functions and is given in terms of supremums and infimums of finite sums. The Fundamental Theorem of Calculus reveals the remarkable inverse relationship between the two processes.

Theorem 7.3.1: Fundamental Theorem of Calculus

1. If $f : [a, b] \to \mathbb{R}$ is integrable, and $F : [a, b] \to \mathbb{R}$ satisfies $F'(x) = f(x)$ for all $x \in [a, b]$, then

$$\int_a^b f = F(b) - F(a)$$

2. Let $g : [a, b] \to \mathbb{R}$ be integrable, and for $x \in [a, b]$, define

$$G(x) = \int_a^x g$$

Then G is continuous on $[a, b]$. If g is continuous at some point $c \in [a, b]$, then G is differentiable at c and

$$G'(c) = g(c)$$

7.4 Improper Integrals

Chapter 8

Introduction to Fourier Series

Part III

PART III: Metric and Normed Space

Part IV

PART IV: Calculus on the Real Space

Bibliography

[1] Abbott, S. (2015). *Understanding analysis*. Springer.

[2] Rudin, W. (1953). *Principles of mathematical analysis*.

[3] Tao, T. (2006). *Analysis i*, volume 1. Springer.