Lecture Notes for Math 104 (Follow Rudin's book)

Rui Wang

Contents

CHAPTER 1

The real number system

Main objects to study in analysis: Sequences, series, and functions. We are going to discuss their convergence, continuity, differentiation, and integration. All of these are based on accurate definition for numbers.

Outline for lecture 1, 2: We introduce number systems $\mathbb{Z} \to \mathbb{Q} \to \mathbb{R}(\to \mathbb{C})$.

1. Rational numbers $\mathbb{Q}$

As we are familiar with, integers include numbers $\cdots, -3, -2, -1, 0, 1, 2, 3, \cdots$. We use $\mathbb{Z}$ to denote the set of integers. Over $\mathbb{Z}$, we can do $+, -, \cdot$. However, $\mathbb{Z}$ is not closed under division, which means when you take two integers $a, b \in \mathbb{Z}$ to do division a/b, the outcome in general is not an integer any more. This motivates us to introduce rational numbers.

Rational numbers $\mathbb{Q}$ can be introduced following a general procedure called the construction of field of fractions. Every rational number can be written as

$$\frac{m}{n}, \quad m, n \in \mathbb{Z}, n \neq 0.$$

Moreover, every nonzero rational number can be uniquely written as

$$\frac{p}{q}, \quad p \in \mathbb{Z}^+, q \in \mathbb{Z}^*, \gcd(p, q) = 1.$$

E.g., $\frac{-18}{12} = \frac{3}{-2}$.

1.1. $\mathbb{Q}$ is a field. We explain the meaning of a field using $\mathbb{Q}$ as an example.

PROPOSITION 1.1. $(\mathbb{Q}, +, \cdot)$ *is a field, which means:*

- $(\mathbb{Q}, +)$ *is an abelian group with identity* 0*:*
  - $\mathbb{Q}$ *is closed under addition* $+$ *($+$ is a binary operation over* $\mathbb{Q}$*).*
  - *Addition* $+$ *is associative:* $(a + b) + c = a + (b + c)$.
  - *Addition* $+$ *has identity element, which is zero:* $a + 0 = 0 + a = a$.
  - *Any element has inverse element:* $a + (-a) = (-a) + a = 0$.
  - *Addition* $+$ *is commutative:* $a + b = b + a$.
- $(\mathbb{Q}, +, \cdot)$ *is a commutative, unital ring:*
  - $\mathbb{Q}$ *is closed under multiplication* $\cdot$ *($\cdot$ is a binary operation over* $\mathbb{Q}$*).*
  - *Multiplication* $\cdot$ *is associative:* $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.
  - *Multiplicaton* $\cdot$ *has unity element, which is* 1*:* $a \cdot 1 = 1 \cdot a = a$.
  - *Multiplication* $\cdot$ *is commutative:* $a \cdot b = b \cdot a$.
  - *Addition and multiplication satisfy distribution law:* $(a + b) \cdot c = a \cdot c + b \cdot c$.
- *Any nonzero element has multiplication inverse:* $a \cdot \frac{1}{a} = \frac{1}{a} \cdot a = 1$*, for any* $a \neq 0$.

EXAMPLE 1.2. $(\mathbb{Z}, +, \cdot)$ is a commutative, unital ring, but it is not a field.

PROOF. (1) Check $(\mathbb{Z}, +, \cdot)$ is a commutative, unital ring.

(2) The number $2 \in \mathbb{Z}$ (in fact, every nonzero number except ± 1) has no multiplication inverse in $\mathbb{Z}$.

□

1.2. $\mathbb{Q}$ is an ordered set.

DEFINITION 1.3. A set X is called an ordered set with an order $\prec$, if

(1) For any two elements $a, b \in X$, one and only one of the following three is true:
- $a \prec b$;
- $b \prec a$;
- $a = b$.

(2) If $a \prec b$ and $b \prec c$, then $a \prec c$.

EXAMPLE 1.4. Define for any $a, b \in \mathbb{Q}$, $a \prec b$ if and only if $a - b$ is a negative number. Then check that $(\mathbb{Q}, \prec)$ is an ordered set.

PROOF. Since a rational number must be one and only one of the following types: positive, negative or zero, Property (1) is satisfied. For (2), if $a - b$ and $b - c$ are both negative, then

$$a - c = (a - b) + (b - c)$$

is also negative. □

In fact, the order $\prec$ is just less than $<$ here.

Similarly, we can define $\prec$ on $\mathbb{Q}$ as "greater than" $>$.

EXERCISE 1.5. Check that $(\mathbb{Q}, \prec)$, with $\prec$ defined as $>$, is an ordered set.

LEMMA 1.6. *Any subset of an ordered set is also ordered.*

EXAMPLE 1.7. $(2^{\mathbb{Z}}, \subset)$ is not an ordered set but only a partial ordered set.

The next definition is about the compatibility of the field structure and the order.

DEFINITION 1.8. An field $(X, +, \cdot)$ is an ordered field, if over X is ordered by $\prec$ and the followings two conditions hold:

(1) If $x \prec y$, $z \in X$, then $x + z \prec y + z$;
(2) If $0 \prec x$ and $0 \prec y$, then $0 \prec x \cdot y$.

We use $(X, +, \cdot; \prec)$ to denote the ordered field.

Check that $(\mathbb{Q}, +, \cdot; <)$ is an ordered field, but $(\mathbb{Q}, +, \cdot; >)$ is not. From now on, when we mention $\mathbb{Q}$, we always means the set of rational number with the prescribed order field structure $(\mathbb{Q}, +, \cdot; <)$.

For an ordered set, we can talk about upper (lower) bounds and least upper (greatest lower) bounds for its subsets.

DEFINITION 1.9. Assume S is a subset of an order set $(X, <)$.

(1) If an element $x \in X$ satisfies that any $a \in S$, $a \leq x$, then call x an upper bound of subset S.
(2) If an element $x \in X$ satisfies that any $a \in S$, $x \leq a$, then call x a lower bound of subset S.

EXAMPLE 1.10. (1) Find an upper bound and a lower bound of $\{x \in \mathbb{Z} | x^2 \leq 2\} \subset \mathbb{Z}$.

(2) Find an upper bound and a lower bound of $\{x \in \mathbb{Q} | x^2 \leq 2\} \subset \mathbb{Q}$.

Are there a smallest upper bound or a biggest lower bound? For (1), the answers are yes, but for (2) are no.

DEFINITION 1.11. (1) Assume S is a subset of an order set $(X, <)$. If $x_0 \in X$ is an upper bound of S and any $x < x_0$ is not an upper bound of S, then we call x_0 the least upper bound of S. The least upper bound is also called the supremum of S and written as $\sup_X S$.

(2) Assume S is a subset of an order set $(X, <)$. If $x_0 \in X$ is a lower bound of S and any $x > x_0$ is not a lower bound of S, then we call x_0 the greatest lower bound of S. The greatest lower bound is also called the infimum of S and written as $\inf_X S$.

LEMMA 1.12. *The least upper bound or the greatest lower bound is unique if exists.*

PROOF. Assume both x_1 and x_2 are supremums of $S \subset X$. Then x_1 is an upper bound of S and hence must have $x_2 \leq x_1$ since x_2 is the least upper bound (This is because if $x_1 < x_2$, since x_2 is the least upper bound, x_1 can not be an upper bound which contradicts with the fact that x_1 is in fact an upper bound.). Similarly, $x_1 \leq x_2$, and this shows $x_1 = x_2$. □

The above example shows that not every subset of an ordered set has supremum or infimum, e.g., $\sup_{\mathbb{Q}}\{x \in \mathbb{Q} | x^2 \leq 2\}$ does not exist. We now give a rigorous proof of it.

PROOF. Assume $\sup_{\mathbb{Q}}\{x \in \mathbb{Q} | x^2 \leq 2\}$ exists and we denote it by x_0. Since $1 \in \{x \in \mathbb{Q} | x^2 \leq 2\}$, we know $x_0 > 0$.

First, we notice that $x_0^2 \neq 2$. This is because if $x_0^2 = 2$, we can write

$$x_0 = \frac{p}{q}, \quad p, q \in \mathbb{Z}^*, \gcd(p, q) = 1.$$

Then it follows

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

and then p must be a multiple of 2. Write $p = 2k$, $k \in \mathbb{Z}^*$, then we have

$$2k^2 = q^2,$$

and then q is also a multiple of q. This shows that the greatest common divisor of p and q is a multiple of 2, which contradicts with the assumption that $\gcd(p, q) = 1$.

Next we consider the number x designed as

$$x := x_0 - \frac{x_0^2 - 2}{x_0 + 2} = \frac{2x_0 + 2}{x_0 + 2} \in \mathbb{Q}^+.$$

We see from a short calculation that

$$x^2 - 2 = \frac{2(x_0^2 - 2)}{(x_0 + 2)^2}.$$

Now if $x_0^2 < 2$, then $x_0 < x$, and $x^2 < 2$, this contradicts with the assumption that x_0 is an upper bound since x lives in the set $\{x \in \mathbb{Q} | x^2 \leq 2\}$.

If $x_0^2 > 2$, then $x < x_0$ and $x^2 > 2$. It follows that x is also an upper bound of $\{x \in \mathbb{Q} | x^2 \leq 2\}$, since otherwise if there is some $a > x > 0$, then $a^2 > x^2 > 2$ will contradict with a is from $\{x \in \mathbb{Q} | x^2 \leq 2\}$. Then this contradicts with the assumption that x_0 is the least upper bound.

Above all, such x_0 doesn't exist. In another word, $\sup_{\mathbb{Q}}\{x \in \mathbb{Q} | x^2 \leq 2\}$ doesn't exist. □

EXERCISE 1.13. Show that see that $\{x \in \mathbb{Q} | x^2 \leq 2\} = \{x \in \mathbb{Q} | x^2 < 2\}$.

DEFINITION 1.14. We say an ordered set $(X, \prec)$ satisfies the least-upper-bound (l.u.b.) property if any nonempty subset of X with an upper bound has a least upper bound.

We just saw that $(\mathbb{Q}, <)$ doesn't satisfy the l.u.b. property.

QUESTION 1.15. *If you feel unfair that why we don't use the greatest-lower-bound (g.l.b.) property instead of the least-upper-bound property, then try to show the following statement which says that in fact these two properties are equivalent to each other.*

To be more specific, we define similarly: Say an ordered set $(X, \prec)$ satisfies the greatest-lower-bound (g.l.b.) property if any nonempty subset of X with a lower bound has the greatest lower bound. Then please show that $(X, \prec)$ satisfies the l.u.b. property if and only if $(X, \prec)$ satisfies the g.l.b. property.

The defect that $(\mathbb{Q}, <)$ doesn't satisfy the l.u.b. property is the main motivation that we want to extend $\mathbb{Q}$ to a larger number system. So we introduce real numbers next.

2. Real numbers $\mathbb{R}$

2.1. Dedekind cuts. The way of introducing real numbers via cuts was given by Dedekind in 1872. In the same year, Cantor used another way by Cauchy sequences to define real numbers. We now explain Dedekind's construction without detailed proofs (For the proof, you may refer Chapter 1 -Appendix in Rudin's book).

The goal is to find some bigger set R containing $\mathbb{Q}$, which preserves all nice properties of $\mathbb{Q}$ (as an ordered field), and enjoys the least-upper-bound property.

We now construct R from $\mathbb{Q}$ as a subset of $2^{\mathbb{Q}}$ (the power set of $\mathbb{Q}$).

(1) We can regard $\mathbb{Q}$ as certain subsets of $\mathbb{Q}$ of the forms

$$\{x \in \mathbb{Q} | x < q\}$$

for $q \in \mathbb{Q}$. Given a $q \in \mathbb{Q}$, we can write a set as this way. Conversely, $q = \sup_{\mathbb{Q}}\{x \in \mathbb{Q} | x < q\}$.

More precisely, that is to say, we consider the map

$$F : \mathbb{Q} \to 2^{\mathbb{Q}}, \quad q \mapsto \{x \in \mathbb{Q} | x < q\}.$$

It is an injective map, so we can identify its image $F(\mathbb{Q}) \subset \mathbb{Q}$ with $\mathbb{Q}$. There is an inverse map, the sup, which maps $F(\mathbb{Q})$ back to $\mathbb{Q}$. This map F induces all structures on $\mathbb{Q}$ to $F(\mathbb{Q})$. For example, we define addition on $F(\mathbb{Q})$ as

$$F(a) + F(b) := F(a + b).$$

We have seen that $\mathbb{Q}$ contains subsets which have no supremum, so this map is not surjective.

EXERCISE 2.1. Show that $\{x \in \mathbb{Q} | x < 0 \text{ or } x^2 \leq 2\}$ is a cut, but is not coming from $\mathbb{Q}$.

(2) Call a subset C of $\mathbb{Q}$ a cut, if
   (a) $C \neq \emptyset$, $X \neq \mathbb{Q}$;
   (b) If $c \in C$, then any $x < c$ also in C;
   (c) If $c \in C$, then there exists some $x > c$ that lives in C.

Clearly, the subset $\{x \in \mathbb{Q} | x < q\}$, with $q \in \mathbb{Q}$, i.e., any element in $F(\mathbb{Q})$, is a cut. However, a cut may not be element in $F(\mathbb{Q})$. For example, the set

$$S = \{a \in \mathbb{Q} | a < 0 \text{ or } a^2 \leq 2\}$$

is a cut, but it is not in $F(\mathbb{Q})$ from Exercise 2.1.

We define R as the set of cuts in $\mathbb{Q}$. Then one can check:

(1) R is ordered by $\subset$.
(2) R has addition defined as

$$C_1 + C_2 := \{x_1 + x_2 | x_1 \in C_1, x_2 \in C_2\}.$$

(3) R has multiplication is kind of bothersome but can be defined.
(4) $(R, +, \cdot)$ is a field.
(5) $(R, +, \cdot; \subset)$ is an ordered field.
(6) R satisfies l.u.b. property.

We use $\mathbb{R}$ to denote the ordered field R constructed above and this is the field of real numbers.

EXERCISE 2.2. Consider the set $S = \{a \in \mathbb{Q} | a^2 \leq 2\}$. What is $\sup_{\mathbb{R}} S$? (Recall that we have seen $\sup_{\mathbb{Q}} S$ doesn't exist.)

From now on, when we consider the supremum or infimum in $\mathbb{R}$ for a subset $S \subset \mathbb{R}$, we omit the subindex $\mathbb{R}$ from $\sup_{\mathbb{R}}$ and just write as $\sup S$. For a subset $S \subset \mathbb{R}$, if $\sup S$ exists (This is equivalent to say S has an upper bound.) and is in S, then we say this supremum is the maximum of S, which we denote as $\max S$. Similarly, if $\inf S$ exists (This is equivalent to say S has a lower bound.) and is in S, then we say this infimum is the minimum of S, which we denote as $\min S$. For example,

$$\sup(1, 2] = \max(1, 2],$$

but since $\inf(1, 2] = 1 \notin (1, 2]$, $\min(1, 2]$ does not exist.

Denote by

$$\mathbb{R}^+ := \{x \in \mathbb{R} | x > 0\}, \quad \mathbb{R}^* := \{x \in \mathbb{R} | x \neq 0\}.$$

In the next sections, we show two important properties of $\mathbb{R}$. Both are essentially based on the fact that $\mathbb{R}$ has the l.u.b. property.

2.2. $\mathbb{R}$ is archimedean.

THEOREM 2.3. *For any $x \in \mathbb{R}^+$ and $y \in \mathbb{R}$, there exists some $n \in \mathbb{Z}^+$ so that*

$$n \cdot x > y.$$

In particular, if we take $x = 1$ from this theorem, we immediately get the following statement.

PROPOSITION 2.4. *For any $y \in \mathbb{R}$, there exists some positive integer n so that $n > y$.*

We now give a proof of Proposition 2.4 directly without using Theorem 2.3, and then we prove Theorem 2.3 from Proposition 2.4. This shows that these two statements are in fact equivalent, though Proposition 2.4 looks much simpler.

PROOF. Assume such $n \in \mathbb{Z}^+$ doesn't exist. That is to say that the set of positive integers $\mathbb{Z}^+$ has an upper bound y. Then using the l.u.b. property of $\mathbb{R}$, the least upper bound $\sup \mathbb{Z}^+$ exists and we use $x_0 \in \mathbb{R}$ to denote it.

Now we look at $x_0 - 1$. This is not an upper bound since we assume x_0 is the least upper bound, which means there exists some $N \in \mathbb{Z}^+$ so that

$$x_0 - 1 < N.$$

Then it follows $x_0 < N + 1$. Notice that $N + 1 \in \mathbb{Z}^+$, this contradicts with the assumption that x_0 is an upper bound.

Hence, our original assumption can not be true and there exists $n \in \mathbb{Z}^+$ with $n > y$. □

PROOF OF THEOREM 2.3. For any $x \in \mathbb{R}^+$ and $y \in \mathbb{R}$, consider $y \cdot x^{-1} \in \mathbb{R}$. From Proposition 2.4, there exits some $n \in \mathbb{Z}^+$ so that

$$n > y \cdot x^{-1}.$$

Then this is equivalent to $n - yx^{-1} > 0$. Since $x > 0$, and $\mathbb{R}$ is an ordered field, we have

$$(n - y \cdot x^{-1}) \cdot x > 0.$$

This is equivalent to $n \cdot x > y$.

□

REMARK 2.5. The archimedean property guarantees that we can use decimals to represent real numbers. (See Rudin's 1.22 for a discussion.)

2.3. $\mathbb{Q}$ is dense in $\mathbb{R}$.

THEOREM 2.6. *For any $a, b \in \mathbb{R}$ with $a < b$, there exists some $x \in \mathbb{Q}$ so that $a < x < b$.*

PROOF. This is equal to say that one can find some $m \in \mathbb{Z}$ and $n \in \mathbb{Z}^+$ so that

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

which is further equivalent to find $m \in \mathbb{Z}$ and $n \in \mathbb{Z}^+$ so that

$$an < m < bn.$$

Notice that $b - a > 0$, so by the archimedean property, there exits $n \in \mathbb{Z}^+$ so that

$$bn - an = (b - a)n > 1.$$

Let's argue that there exists some integer between two real numbers, whenever their difference is bigger than 1.

LEMMA 2.7. *For any $\alpha, \beta \in \mathbb{R}$ with $\beta - \alpha > 1$, there exists some integer m so that $\alpha < m < \beta$.*

PROOF OF LEMMA 2.7. We prove this lemma by finding such m. First, using archimedean property of $\mathbb{R}$, we can find some integer $N > 0$ so that

$$-N < \alpha < \beta < N.$$

Then consider the integers which are smaller than N and greater than α, i.e., the set

$$A := \{k \in \mathbb{Z} | \alpha < k \leq N\}.$$

It is not empty since $N \in A$. Since this is a subset of $\{-N + 1, -N + 2, \cdots, N - 2, N - 1, N\}$ which is finite set, it contains only finite elements. We can pick the smallest one from it and denote it by m, i.e., $m := \min A$. We claim this m is just the one we are looking for.

First since $m \in A$, $m > \alpha$. Then we only need to check $m < \beta$. If this is not true, i.e., $m \geq \beta$, then we consider $m - 1$. It follows

$$m - 1 \geq \beta - 1 \geq \alpha.$$

This contradicts with the fact that m is the smallest integer which is greater than α.

Above all, we are done with the lemma.

$\square$

At last, apply the lemma to $\alpha = an$ and $\beta = bn$, we are done.

$\square$

2.4. $\mathbb{R}$ is closed under taking roots. Using the Dedkind's cuts construction, it is easy to see the following result, but here we outline a proof (try to fill details or refer Rudin 1.21 Theorem for details) which is based on the l.u.b. property of $\mathbb{R}$.

THEOREM 2.8. *For every $y \in \mathbb{R}^+$ and every $n \in \mathbb{Z}^+$, there exists a unique $x \in \mathbb{R}^+$ so that $x^n = y$.*

PROOF. We first claim that such $x \in \mathbb{R}^+$, if exists, must be unique. Otherwise, assume that both $x_1, x_2 \in \mathbb{R}^+$ are solutions of the equation

$$x^n = y, \quad y \in \mathbb{R}^+, n \in \mathbb{Z}^+.$$

Assume now $x_1 < x_2$, then from that fact that $\mathbb{R}$ is an ordered field, we have $x_1^n < x_2^n$ (why?) and that is a contradiction. Similarly, $x_1 > x_2$ also leads to contradiction, and so $x_1 = x_2$.

Now we look for a solution for the equation. Consider a subset of $\mathbb{R}$ as

$$S := \{a \in \mathbb{R}^+ | a^n < y\}.$$

Try to check that

(1) $S \neq \emptyset$;
(2) S has upper bound.

Then using the fact that $\mathbb{R}$ has the l.u.b. property, $\sup S$ exists. Define it as x, clearly, $x \in \mathbb{R}^+$. We show that x solves the equation. (The idea of the proof is similar to the proof of $\sup_{\mathbb{Q}}\{x \in \mathbb{Q} | x^2 \leq 2\}$ does not exist.)

First, we show that if $x^n < y$, then we can construct some $x_0 \in S$ which is greater than x, which says x is not an upper bound of S. So $x^n \geq y$.

Second, we show that if $x^n > y$, then we can find an upper bound of S which is smaller than x, which says x is not the least upper bound. So $x^n \leq y$.

Above all, we must have $x^n = y$. $\square$

From now on, we use $y^{\frac{1}{n}}$ to denote the unique solution for the equation

$$x^n = y, \quad y \in \mathbb{R}^+, n \in \mathbb{Z}^+,$$

and call it the n-th real root of y. The property

$$(ab)^{\frac{1}{n}} = a^{\frac{1}{n}} b^{\frac{1}{n}}$$

immediately follows from the uniqueness of n-th real root.

2.5. The extended real number system. We can add $\pm\infty$ to $\mathbb{R}$, and call the union $\mathbb{R} \cup \{\pm\infty\}$ the extended real number system. The advantage for the extended real number system is that now any nonempty set in $\mathbb{R}$ has the least upper bound and the greatest lower bound, since we can simply define

$$\sup S = +\infty, \quad \text{if } S \text{ has no upper bound in } \mathbb{R};$$

and

$$\inf S = -\infty, \quad \text{if } S \text{ has no lower bound in } \mathbb{R}.$$

Formally, we can define operations of finite numbers with $\pm\infty$, but $\mathbb{R} \cup \{\pm\infty\}$ is not a field (Why?).

3. The Euclidean plane $\mathbb{R}^2$

We consider the Cartesian product of $\mathbb{R}$ with $\mathbb{R}$, i.e.,

$$\mathbb{R}^2 := \mathbb{R} \times \mathbb{R} := \{(x_1, x_2) | x_1, x_2 \in \mathbb{R}\}.$$

Over $\mathbb{R}^2$, we can define operations

(1) Addition $+$: $(x_1, x_2) + (y_1, y_2) = (x_1 + y_1, x_2 + y_2)$;
(2) Scaler multiplication $\mathbb{R} \times \mathbb{R}^2 \to \mathbb{R}^2 : c \cdot (x_1, x_2) = (c \cdot x_1, c \cdot x_2)$.

This two operations make $\mathbb{R}^2$ a 2-dimensional vector space (linear space) over the real field $\mathbb{R}$. We also say $\mathbb{R}^2$ is a $\mathbb{R}$-linear space of real dimension 2. For example, $\{(1, 0), (0, 1)\}$ forms a basis of $\mathbb{R}^2$.

Moreover, over the linear space $\mathbb{R}^2$, one can define an inner product as

$$\langle (x_1, x_2), (y_1, y_2) \rangle = x_1 y_1 + x_2 y_2 = \Sigma_{j=1,2} x_j y_j.$$

The inner product induces a norm

$$|(x_1, x_2)| = \sqrt{\langle (x_1, x_2), (x_1, x_2) \rangle} = \sqrt{x_1^2 + x_2^2}.$$

From now on, we use $\vec{x}$ to denote (x_1, x_2).

PROPOSITION 3.1. (1) $|\vec{x}| \geq 0$ *and it is* 0 *if and only if* $\vec{x} = \vec{0}$.
(2) $|c \cdot \vec{x}| = |c||\vec{x}|$.
(3) $|\vec{x} + \vec{y}| \leq |\vec{x}| + |\vec{y}|$.
(4) $|\langle \vec{x}, \vec{y} \rangle| \leq |\vec{x}||\vec{y}|$.

All constructions here can be easily generalized to any $\mathbb{R}^n$ with $n \in \mathbb{Z}^+$.

4. The complex numbers $\mathbb{C}$

Over $\mathbb{R}^2$, we can define a multiplication $\cdot$ as

$$(a, b) \cdot (c, d) = (ac - bd, ad + bc).$$

If we identity $\mathbb{R}^2$ with

$$\mathbb{C} := \{x + yi | x, y \in \mathbb{R}\}$$

via $(x, y) \mapsto x + yi$, then all structures defined above are induced to $\mathbb{C}$. In particular, the multiplication is induced to $\mathbb{C}$ via requiring $i^2 = -1$. A nontrivial fact is that $(\mathbb{C}, +, \cdot)$ is a field. A element in $\mathbb{C}$ is called a complex number. Usually, people prefer to use $z = x + yi$, $x, y \in \mathbb{R}$, to denote a complex number. Here x is called the real part of z and y is called the imaginary part of z. We use $|z|$ to denote its norm.

CHAPTER 2

Some basic topology

We are going to understand the system of real numbers $\mathbb{R}$ from the viewpoint of topology.

1. Countable sets and uncountable sets

So far, we have met several subsets of $\mathbb{C}$, which are

$$\{1, 2, \cdots, N\} \subset \mathbb{Z}^+ \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C},$$

where N is some positive integer. Our first goal for this section is to put them into three different classes of sets, which are:

(1) Finite sets: $\{1, 2, \cdots, N\}$, for some $N \in \mathbb{Z}^+$;
(2) Countable sets: $\mathbb{Z}^+$, $\mathbb{Z}$, $\mathbb{Q}$;
(3) Uncountable sets: $\mathbb{R}$, $\mathbb{C}$.

Two sets X, Y are called 1-1 correspondent, if there exists a bijective map from X to Y. We use $X \sim Y$ to denote.

DEFINITION 1.1. A relation $\sim$ for a collection of objects C is called an equivalence relation, if

(1) for any $x \in C$, $x \sim x$;
(2) for any $x, y \in C$, if $x \sim y$, then $y \sim x$;
(3) for any $x, y, z \in C$, if $x \sim y$, $y \sim z$, then $x \sim z$.

PROPOSITION 1.2. *1-1 correspondence is an equivalence relation for sets.*

PROOF. (1) For any set X, the identity map $id_X : X \to X$ is bijective. So $X \sim X$.

(2) For any two sets X, Y, if $X \sim Y$, then there is a bijective map

$$f : X \to Y.$$

The inverse map $f^{-1} : Y \to X$ is defined and bijective, and hence $Y \sim X$.

(3) For any three sets X, Y, Z, if $X \sim Y$ and $Y \sim Z$, then there are two bijective maps

$$f : X \to Y, \quad g : Y \to Z.$$

Consider their composition $g \circ f : X \to Z$. It is then bijective and hence $X \sim Z$.

□

A set X is called finite, if X contains finitely many elements. Equivalently, there exists a positive integer N so that $\{1, 2, \cdots, N\} \sim X$. If X is a finite set of N elements, we write $|X| := N$. Obviously from definition, if $X \sim Y$ and with one of them finite, then the other is also finite with $|X| = |Y|$. The bijective map between $\{1, 2, \cdots, N\}$ and X is in fact the counting map which counts the elements of X.

A set X is called infinite, if it is not a finite set. We use the following notations for finite or infinite sets: If X is finite, we write $|X| < \infty$; If X is infinite, we write $|X| = \infty$.

DEFINITION 1.3. (1) An infinite set X is called countable, if $\mathbb{Z}^+ \sim X$.

(2) An infinite set X is called uncountable, if X is not countable.

A set is called at most countable, if it is either finite or countable.

We check the definition for the following examples.

EXAMPLE 1.4. $\mathbb{Z}$ is countable.

PROOF. Let's show $\mathbb{Z}^+ \sim \mathbb{Z}$. Construct a map

$$f : \mathbb{Z} \to \mathbb{Z}^+$$

as

$$\begin{aligned} f(k) &= 2k, \quad \text{if } k \in \mathbb{Z}^+; \\ f(k) &= 2(-k)+1, \quad \text{if } k \in \mathbb{Z}^-; \\ f(0) &= 1. \end{aligned}$$

Then check that f is both injective and surjective. □

EXAMPLE 1.5. $\mathbb{Q}$ is countable.

PROOF. Since any rational number $x \in \mathbb{Q}$ can be written as a fraction $\frac{m}{n}$, with some $m, n \in \mathbb{Z}$, this defines a map from $\mathbb{Q}$ to $\mathbb{Z} \times \mathbb{Z}$. (Though there are many different ways to represent the same rational number, we just pick one for each rational number to define this map. In fact, we can pick-up one is guaranteed by the Choice Axiom.) This map is injective, but not surjective. Then $\mathbb{Q}$ is countable follows from the following two general propositions: Proposition 1.6 and Proposition 1.7. Their statements and proofs are as follows. □

PROPOSITION 1.6. *If both X, Y are countable, then the cartesian product $X \times Y$ is also countable.*

PROOF. From the assumption X and Y are both countable, it follows $X \times Y \sim \mathbb{Z}^+ \times \mathbb{Z}^+$. Hence, it is enough to show $\mathbb{Z}^+ \times \mathbb{Z}^+$ is countable.

Elements in $\mathbb{Z}^+ \times \mathbb{Z}^+$ are of the form (m, n), $m, n \in \mathbb{Z}^+$. We can assign a counting by regrouping them into disjoint union of diagonals. Define

$$S_i := \{(m, n) \in \mathbb{Z}^+ \times \mathbb{Z}^+ | m + n = i + 1\}, \quad i = 1, 2, \cdots.$$

Equally, S_i contains i elements as

$$S_i = \{(1, i), (2, i-1), (3, i-2), \cdots, (i, 1)\}.$$

It is not hard to check that $S_i \cap S_j = \emptyset$ whenever $i \neq j$, and

$$\mathbb{Z}^+ \times \mathbb{Z}^+ = \cup_{i \in \mathbb{Z}^+} S_i.$$

Over each S_i, define a map

$$f_i : S_i \to \mathbb{Z}^+, \quad (m, i+1-m) \mapsto (1 + 2 + \cdots + (i-1)) + m.$$

Obviously this map is injective and onto the interval

$$I_i := \{k \in \mathbb{Z} | (1 + 2 + \cdots + (i-1)) + 1 \leq k \leq (1 + 2 + \cdots + (i-1)) + i\}.$$

With noticing that these I_i has no intersectons, and the union of them is $\mathbb{Z}^+$, this constructs the bijective map f defined as

$$f : \mathbb{Z}^+ \times \mathbb{Z}^+ \to \mathbb{Z}^+, \quad f(m, n) = f_i(m, n) \quad \text{if } (m, n) \in S_i.$$

We are done with the proof. □

(This construction of f above seems tricky, but is only a direct translation from the picture we saw from lecture that f is counting points on each diagonal from left to right.)

PROPOSITION 1.7. *Any subset of a countable set is at most countable.*

PROOF. We only need to show any infinite subset of $\mathbb{Z}^+$ is countable.

We first give a lemma whose proof is left to you as an exercise.

LEMMA 1.8. *For any nonempty subset $E \subset \mathbb{Z}^+$,* $\min E$ *exists.*

HINT: Take an arbitary element $n \in E$ and consider the subset

$$\{a \in E | a \leq n\}.$$

It is a nonempty finite subset of E. Since the minimal element of any **finite** set exists, we denote by n_0 the minimal element of this set. Then prove that $n_0 = \min E$. □

Assume $S \subset \mathbb{Z}^+$ is infinite. Define $S_1 = S$, $S_{n+1} = S_n \setminus \{\min S_n\}$, for $n = 1, 2, \cdots$, and a map

$$f : \mathbb{Z}^+ \to S, \quad n \mapsto \min S_n.$$

Then we check this map is bijective.

- It is injective: Assume $f(n) = f(n')$, then $\min S_n = \min S_{n'}$. If $n \neq n'$, and WLOG, assume $n > n'$, then from the construction we can see

$$S_n \subset S_{n'} \setminus \{\min S_{n'}\},$$

  and $S_n \neq \emptyset$ since S is not finite. In particular, $\min S_n \in S_n$ but $\min S_{n'} \notin S_n$, this shows

$$\min S_n \neq \min S_{n'},$$

  and we get contradiction.
- It is surjective: Take any $k \in S$, we show that there must be some $n \in \mathbb{Z}^+$ so that $f(n) = k$.

  Consider the set

$$\{m \in S | m \leq k\}.$$

  It is a finite set, and assume it contains n elements. Then $f(n) = k$ by the construction of f.

We are done with the proof of Proposition 1.7. □

An immediate corollary is

COROLLARY 1.9. *Any set that contains an uncountable set is uncountable.*

Next, we discuss the (un)countability of $\mathbb{R}$.

First, notice Proposition 1.6 can be easily generalized to any finite product, whose proof is an immediate corollary of Proposition 1.6.

COROLLARY 1.10. *The set $X_1 \times X_2 \times X_3 \cdots \times X_N$ is countable, if each X_i is countable.*

However, when we move further to countable cartesian product of (at most) countable sets, the countability is no longer there in general.

PROPOSITION 1.11. *Consider a sequence of sets $X_1, X_2, \cdots$, with each X_i contains at least two elements, then their cartesian product*

$$X_1 \times X_2 \times \cdots := \{(x_1, x_2, \cdots,)|x_i \in X_i\}$$

is uncountable.

PROOF. (1) Using Corollary 1.9, it is enough to prove the proposition for the case that each X_i contains exactly 2 elements.

(2) Now we prove the case of each set containing exactly two elements. Assume

$$X_i = \{x_i^0, x_i^1\}, \quad i = 1, 2, \cdots .$$

Assume $X_1 \times X_2 \times \cdots$ is a countable set and then its elements can be denoted by

$$S_1, S_2, \cdots ,$$

with each S_k, $k = 1, 2, \cdots$ a sequence

$$S_k = (s_{k1}, s_{k2}, s_{k3}, \cdots), \quad s_{ki} \in X_i = \{x_i^0, x_i^1\}.$$

We place these sequence $S_1, S_2, \cdots$ row by row to follow an array, and now use its diagonal to construct a new sequence: Define

$$S^* := (s_1^*, s_2^*, \cdots), \quad s_i^* := \overline{s_{ii}}$$

where $\overline{s_{ii}}$ is the other element in X_i but not s_{ii}.

Notice that S^* is different from any of S_k, $k = 1, 2, \cdots$. That is because for any S_k, the k-th element is s_{kk}, but by construction, the k-th element of S^* is $\overline{s_{kk}}$, which is different from s_{kk}. Then this leads to contradiction with the assumption that $X_1 \times X_2 \times \cdots$ is a countable set.

□

Then recall the decimal representation of $\mathbb{R}$, it follows $\mathbb{R}$ is uncountable. Since $\mathbb{R} \subset \mathbb{C}$, $\mathbb{C}$ is also uncountable.

2. Metric space

2.1. Metric space. Roughly, a metric space is a space (set) together with a metric (a distance function) satisfying several properties.

EXAMPLE 2.1. (1) $(\mathbb{R}, d)$ is a metric space with $d(x, y) := |x - y|$.

(2) $(\mathbb{R}^2, d)$ is a metric space with $d(\vec{x}, \vec{y}) := \sqrt{|x|^2 + |y|^2}$.

In general,

DEFINITION 2.2. A metric space is a set X together with a real value function (called distance function or metric)

$$d : X \times X \to \mathbb{R},$$

which satisfies the following properties:

(1) For any $x, y \in X$, $d(x, y) \geq 0$. Moreover, $d(x, y) = 0$ if and only if $x = y$.
(2) For any $x, y \in X$, $d(x, y) = d(y, x)$.
(3) For any $x, y, z \in X$, $d(x, z) \leq d(x, y) + d(y, z)$.

We call (X, d) a metric space.

In our class, when we consider $\mathbb{R}$ and $\mathbb{R}^2$ ($\mathbb{R}^k$) as metric spaces, we always consider the metric induced from norms. However, they do have other metrics. (See homework problem Rudin Chapter 2 – 10 and 11.)

A metric space (X, d) naturally induces a metric on any of its subsets.

2.2. Open sets and closed sets.

DEFINITION 2.3. For any $x \in X$, $r > 0$,

(1) the subset $B_r(x) := \{y \in X | d(y, x) < r\}$ is called the open ball centered at x with radius r;
(2) the subset $\overline{B}_r(x) := \{y \in X | d(y, x) \leq r\}$ is called the closed ball centered at x with radius r.

An open ball centered at x is also called a neighborhood of x.

EXAMPLE 2.4. An open (closed) ball in $\mathbb{R}$ is equivalent to a finite open (closed) interval, i.e., (a, b) ($[a, b]$), $a, b \in \mathbb{R}$.

DEFINITION 2.5. Assume (X, d) is a metric space. A subset $S \subset X$ is called open, if $S = \emptyset$ or if every $x \in S$, there exists some open ball $B_r(x) \subset S$ for some $r > 0$.

EXAMPLE 2.6. Any open ball is open.

PROOF. Assume $B_r(x)$ is an open ball in a metric space (X, d). Then for any point $y \in B_r(x)$, there is

$$d(y, x) < r.$$

Define $r' := r - d(y, x)$, which is positive.

Consider the ball $B_{r'}(y)$. Let's show it lives in $B_r(x)$. For this, take any point $z \in B_{r'}(y)$. Using the triangle inequality of a metric, we have

$$d(z, x) \leq d(z, y) + d(y, x) < r' + d(y, x) = r.$$

Hence $z \in B_r(x)$, and $B_{r'}(y) \subset B_r(x)$. □

PROPOSITION 2.7. *Assume (X, d) is a metric space.*

(1) *Both $\emptyset$ and X are open.*
(2) *If S_1, S_2 are open, then $S_1 \cap S_2$ is open.*
(3) *For any set Λ so that any $\alpha \in \Lambda$, S_α is an open subset of X, the union $\cup_{\alpha \in \Lambda} S_\alpha$ is open.*

PROOF. (1) Obvious by definition.

(2) Take a point $x \in S_1 \cap S_2$, we need to find an open ball with radius $r > 0$ so that $x \in B_r(x) \subset S_1 \cap S_2$.

To find such $r > 0$, notice that since both S_1, S_2 are open, there are open balls

$$x \in B_{r_i}(x) \subset S_i, \quad i = 1, 2.$$

Take $r := \min\{r_1, r_2\}$. Then $B_r(x) \subset B_{r_i}(x) \subset S_i$, $i = 1, 2$, and hence $B_r(x) \subset S_1 \cap S_2$.

(3) Take a point $x \in \cup_{\alpha \in \Lambda} S_\alpha$, then we can assume x lives in some S_{α_0}, $\alpha_0 \in \Lambda$. Since S_{α_0} is open, take an open ball

$$B_r(x) \subset S_{\alpha_0}.$$

It follows

$$B_r(x) \subset S_{\alpha_0} \subset \cup_{\alpha \in \Lambda} S_\alpha.$$

This shows $\cup_{\alpha\in\Lambda} S_\alpha$ is open.

□

EXAMPLE 2.8. We know $I_n := (-\frac{1}{n}, \frac{1}{n}) \subset \mathbb{R}$ is open for any $n \in \mathbb{Z}^+$. However $\cap_{n\in\mathbb{Z}^+} I_n = \{0\}$ is not open.

DEFINITION 2.9. A subset $S \subset X$ is called closed, if its complement is open.

We use S^c to denote its complement in X.

EXAMPLE 2.10. (1) The closed interval $[a, b]$, $a \leq b$ is closed in $\mathbb{R}$.
(2) A closed ball $\overline{B}_r(x)$, $r > 0$, $x \in X$, is closed. (See Assignment 3 for proof.)

Using the definition of closed sets and Proposition 2.7, we immediately get the following analogue for closed sets.

PROPOSITION 2.11. *Assume (X, d) is a metric space.*
(1) *Both $\emptyset$ and X are closed.*
(2) *If S_1, S_2 are closed, then $S_1 \cup S_2$ is closed.*
(3) *For any set Λ so that any $\alpha \in \Lambda$, S_α is a closed subset of X, the intersection $\cap_{\alpha\in\Lambda} S_\alpha$ is closed.*

PROOF. (1) It follows immediately from $\emptyset = X^c$ and $X = \emptyset^c$.
(2) If follows from Proposition 2.7 (2) that
$$(S_1 \cup S_2)^c = S_1^c \cap S_2^c$$
is open, and hence $S_1 \cup S_2$ is closed.
(3) If follows from Proposition 2.7 (3) that
$$(\cap_{\alpha\in\Lambda} S_\alpha)^c = \cup_{\alpha\in\Lambda} S_\alpha^c$$
is open, and hence $\cap_{\alpha\in\Lambda} S_\alpha$ is closed.

□

EXAMPLE 2.12. Consider a sequence of closed sets $[-1 + \frac{1}{n}, 1 - \frac{1}{n}]$, $n \in \mathbb{Z}^+$, of $\mathbb{R}$. Take their union
$$\cup_{n\in\mathbb{Z}^+}[-1 + \frac{1}{n}, 1 - \frac{1}{n}] = (-1, 1),$$
which is open, not closed.

2.3. Limit points in a metric space.

DEFINITION 2.13. Assume S is a subset of X. A point $x \in X$ is called a limit point of S, if any neighborhood $B_r(x)$ intersects with S contains some point which is not x. (Notice x always lives in $B_r(x)$.)

We use S' to denote the set of limit points of S in X, and use
$$\overline{S} := S \cup S'$$
to denote the union of S with its limit point set and call it the closure of S in X.

Points in $S \setminus S'$ are called isolated points of S.

EXAMPLE 2.14. (1) Consider the metric space $\mathbb{R}$. a and b are limit points of $(a, b]$. The limit point set of $(a, b]$ is $[a, b]$, which is also the closure of $(a, b]$.

(2) Consider the metric space $\mathbb{R}^2$. The limit point set of any open ball $B_r(x)$ is the closed ball $\overline{B}_r(x)$, which is also the closure of $B_r(x)$.
(3) Consider $\mathbb{Q} \subset \mathbb{R}$. $\mathbb{Q}' = \overline{\mathbb{Q}} = \mathbb{R}$. (See Homework.)

DEFINITION 2.15. A subset S of X is called dense, if $\overline{S} = X$.

The above example (3) says that $\mathbb{Q}$ is dense in $\mathbb{R}$. (Why does this definition of "dense" match the one we used before?)

PROPOSITION 2.16. *For any subset $S \subset X$, its closure $\overline{S}$ is closed.*

PROOF. If $\overline{S} = X$, this is automatically true. We now assume $\overline{S} \neq X$.
Then for any $x \in (\overline{S})^c$, we show that there must be some open ball $B_r(x) \subset (\overline{S})^c$.
Assume this is not true, then this says every open ball $B_r(x) \cap \overline{S}$ is not empty. Take a point

$$y \in B_r(x) \cap \overline{S}.$$

Notice $y \neq x$ since $x \notin \overline{S}$.
If y is not in S, then y must be in S'. For this case, further take a smaller ball so that

$$x \notin B_{r'}(y) \subset B_r(x).$$

Since $y \in S'$, there must be some point $y' \neq y$ and in S. Notice $y' \neq x$ again since we can exclude x from $B_{r'}(y)$.
This show, for every open ball $B_r(x) \cap \overline{S}$ is not empty, we can find some y or y' in $S \cap B_r(x)$ and not x. This shows that $x \in S'$, which contradicts with the assumption $x \notin \overline{S}$. □

The proof of the following statements are left to you as homework.

PROPOSITION 2.17. *A subset S of X is closed, if and only if $S = \overline{S}$.*

PROPOSITION 2.18. *The closure of a subset S of X is the smallest closed subset of X that contains S.*

DEFINITION 2.19. A point $x \in X$ is a called a limit of a sequence $\{x_n | n \in \mathbb{Z}^+\}$, if the limit

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

This means, for any $\epsilon > 0$, there exists some $N \in \mathbb{Z}^+$ so that whenever $n > N$,

$$d(x_n, x) < \epsilon.$$

EXAMPLE 2.20. Show that $x = 0$ is a limit of the sequence $\{x_n = \frac{1}{n} | n \in \mathbb{Z}^+\}$ in the metric space $\mathbb{R}$ with the standard distance.

PROOF. Since $d(x_n, x) = |\frac{1}{n} - 0| = \frac{1}{n}$. We only need to check for any $\epsilon > 0$, there exists some $N \in \mathbb{Z}^+$ so that whenever $n > N$, $\frac{1}{n} < \epsilon$.
For it we notice that, given any $\epsilon > 0$, by the Archimedean property for $\mathbb{R}$, there exist some $N \in \mathbb{Z}^+$ so that $N > \frac{1}{\epsilon}$. Then for any $n < N$, there is

$$\frac{1}{n} < \frac{1}{N} < \epsilon.$$

This shows $x = 0$ is a limit of the sequence $\{x_n = \frac{1}{n} | n \in \mathbb{Z}^+\}$. □

LEMMA 2.21. *In a metric space, if a sequence $\{x_n | n \in \mathbb{Z}^+\}$ has a limit, then the limit must be unique.*

PROOF. Assume both x and x' are limits of the sequence $\{x_n | n \in \mathbb{Z}^+\}$. Then for any $\epsilon > 0$, there exists some $N, N' > 0$ so that

$$d(x_n, x) < \frac{\epsilon}{2}, \quad \text{whenever } n > N,$$

and

$$d(x_n, x') < \frac{\epsilon}{2}, \quad \text{whenever } n > N'.$$

Define $N_0 = \max\{N, N'\}$. Then whenever $n > N_0$, both hold.

We estimate by the triangle inequality that

$$0 \le d(x, x') \le d(x, x_n) + d(x_n, x') = d(x_n, x) + d(x_n, x') < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

for any $\epsilon > 0$. This implies $d(x, x') = 0$ and then $x = x'$. □

Because of the uniqueness, we can denote by $\lim_{n\to\infty} x_n = x$, if x is a (and so the by this lemma) limit of sequence $\{x_n | n \in \mathbb{Z}^+\}$.

We prove the following statement, which explain the meaning of limit points.

PROPOSITION 2.22. *Assume (X, d) is a metric space and S is a subset. A point $x \in X$ is a limit point of S, if and only if there exists some sequence $\{x_n \in S \setminus \{x\} | n \in \mathbb{Z}^+\}$ so that* $\lim_{n\to\infty} x_n = x$.

PROOF. (1) Assume x is a limit point of S, then we can consider a sequence of open balls $B_{\frac{1}{n}}(x)$, $n \in \mathbb{Z}^+$. Each one intersects with S at some point which is not x. We pick a such point x_n from each $B_{\frac{1}{n}}(x) \cap S$ for each $n \in \mathbb{Z}^+$.

Notice that

$$d(x_n, x) < \frac{1}{n},$$

then for each $\epsilon > 0$, there exists some $N > 0$ so that $\frac{1}{N} < \epsilon$. (This is due to the archimedean property.) Then for each $n > N$,

$$d(x_n, x) < \frac{1}{n} < \frac{1}{N} < \epsilon,$$

which shows that $\lim_{n\to\infty} x_n = x$.

(2) Conversely, assume there exists a sequence $\{x_n \in S \setminus \{x\} | n \in \mathbb{Z}^+\}$ so that $\lim_{n\to\infty} x_n = x$. Then for each open ball $B_\epsilon(x)$, we can find some $N \in \mathbb{Z}^+$ so that

$$x_n \in B_\epsilon(x), \quad \text{whenever } n > N.$$

Since $x_n \in S \setminus \{x\}$, this shows that x is a limit point of S.

□

3. Compactness

3.1. Equivalence of compactness and sequential compactness in metric spaces.

DEFINITION 3.1. Assume (X, d) is a metric space. A collection of open sets $\{U_\alpha | \alpha \in \Lambda\}$ is called an open cover of a subset S of X, if

$$S \subset \cup_{\alpha\in\Lambda} U_\alpha.$$

For $\Lambda' \subset \Lambda$, if the subcollection $\{U_\alpha | \alpha \in \Lambda'\}$ is also an open cover of S, i.e.,

$$S \subset \cup_{\alpha\in\Lambda'} U_\alpha,$$

then $\{U_\alpha | \alpha \in \Lambda'\}$ is called a subcover. If moreover, Λ' is finite, then it is called a finite subcover.

DEFINITION 3.2. Assume (X, d) is a metric space. A subset S of X is called bounded, if there exists some $C > 0$ and $x_0 \in X$ so that

$$d(x, x_0) < C, \quad \text{for any } x \in S.$$

EXAMPLE 3.3.

(1) Take $\{U_n = (n, n + 1.5) | n \in \mathbb{Z}\}$. It is an open cover of $\mathbb{R}$. Notice that, any finite subset Λ of $\mathbb{Z}$, the subcollection $\{U_n | n \in \Lambda\}$ can not cover the whole $\mathbb{R}$. This shows $\mathbb{R}$ is not compact.
(2) Still take $\{U_n = (n, n+1.5) | n \in \mathbb{Z}\}$, it is also an open cover of any bounded subset A. One can find a finite subcover from it so that it covers A.
(3) Consider $(0, 1) = \cup_{n=1}^{\infty}(\frac{2^{n-1}-1}{2^n}, \frac{2^n-1}{2^n})$. However, any finite subcover from it can not cover $(0, 1)$. (Why?)
(4) Consider $[0.01, 0.99] \subset \cup_{n=1}^{\infty}(\frac{2^{n-1}-1}{2^n}, \frac{2^n-1}{2^n})$. It has finite subcover.

DEFINITION 3.4. Assume (X, d) is a metric space. A subset $K \subset X$ is called a compact subset, if every open cover of K has a finite subcover of K.

EXAMPLE 3.5. Assume (X, d) is a metric space. Any finite set $F \subset X$ is compact.

DEFINITION 3.6. A subset of a sequence $\{x_n | n \in \mathbb{Z}^+\}$ that indexed as

$$\{x_{n_k} | k \in \mathbb{Z}^+, n_1 < n_2 < \cdots\}$$

is called a subsequence of $\{x_n | n \in \mathbb{Z}^+\}$.

DEFINITION 3.7. Assume (X, d) is a metric space. A subset $K \subset X$ is called a sequentially compact subset, if every sequence of points in K has a convergent subsequence converging to a point in K.

THEOREM 3.8. *For any metric space (X, d). A subset K is compact if and only if it is sequentially compact.*

PROOF.

(1) Assume $K \subset X$ is compact. Take any sequence $\{y_n\}$ from K. Assume any point $x \in K$ is not a limit of any subsequence of $\{y_n\}$. Then there exits some open ball $B_{r_x}(x)$ that at most contains one point in $\{y_n\}$ which is x.

   Consider $\{B_{r_x}(x) | x \in K\}$. This is an open cover of K. By the compactness of K, there exists a finite subcover of K. We assume this subcover is

$$\{B_{r_{x_1}}(x_1), \cdots, B_{r_{x_N}}(x_N)\}.$$

   In particular, these open balls cover $\{y_n\}$ and hence there must be some x_i, with $i \in \{1, \cdots, N\}$ so that there are infinitely many $y_j = x_i$. Such $\{y_j = x_i\}$ then is a subsequence of $\{y_n\}$ that converges to $x_i \in K$. This contradicts with the assumption.
(2) Assume $\{U_\alpha | \alpha_\Lambda\}$ is an open cover of K. Assume any finite union can not cover K, then Λ must be an infinite set.

   If Λ is countable, WLOG, assume $\Lambda = \mathbb{Z}^+$. Since any finite

$$U_1 \cup U_2 \cup \cdots \cup U_n$$

can not cover K, we can take some $x_n \in K \setminus (U_1 \cup U_2 \cup \cdots \cup U_n)$, for every $n \in \mathbb{Z}^+$. Then we obtain a sequence $\{x_n\}$ in K and so must have a convergence subsequence $\{x_{n_k}\}$ that converges to some point $x_0 \in K$. It follows there must be some U_N so that $x_0 \in U_N$. Since U_N is open, there exists some small ball

$$B_r(x_0) \subset U_N.$$

On the other hand, since $x_{n_k} \to x_0$ as $k \to \infty$, there must be some N' large, so that

$$x_{n_k} \in B_r(x_0), \qquad \text{whenever } n_k > N'.$$

However, by our way of choosing x_n, whenever $n_k > \max\{N', N\}$, $x_{n_k} \notin U_N$. This leads to contradiction.

In general, one can prove for any infinite Λ, there exits a countable subcover of K, and then the proof is reduced to the case of countable Λ. (Refer Rudin's Chapter 2 Ex 24, 26.)

□

PROPOSITION 3.9. *Assume (X, d) is a metric space. Then any (sequentially) compact subset is bounded and closed.*

PROOF. Assume K is compact subset of X.

(1) We first show K must be closed. For this, we only need to show K^c is open. Assume K^c is not open, then there exists some $x \in K^c$, so that for each $n \in \mathbb{Z}^+$, the open ball $B_{\frac{1}{n}}(x)$ contains some point from K. By this way, we obtain a sequence $\{x_n\}$ in K. Then since K is sequentially compact, there exits a subsequence $\{x_{n_k}\}$ that converges to some $x_0 \in K$. On the other hand, by construction, each with

$$d(x_{n_k}, x) < \frac{1}{n_k}.$$

This shows that $x_{n_k} \to x$. By the uniqueness of limit from Lemma 2.21, we have $x = x_0$ and so lives in K, which contradicts with assumption at the beginning.

Hence K^c is open and so K is closed.

(Another way of proving K is closed is to show $K = \overline{K}$, i.e., to show any limit point of K must live in K.)

(2) We show that for any $x_0 \in X$, there must be some $C > 0$ so that $d(x, x_0) \leq C$ for any $x \in K$.

Assume this is not true. Then for some $x_0 \in X$, we can take a sequence $\{x_n\}$ in K so that

$$d(x_n, x_0) > n.$$

By the sequential compactness of K, there exists a subsequence $\{x_{n_k}\}$ that converges to some $x_\infty \in K$. It follows there exists some N large so that

$$d(x_{n_k}, x_\infty) < \frac{1}{4}, \qquad \text{for any } n_k > N.$$

We pick a such $x_{n_{k_0}}$,

$$d(x_\infty, x_0) \leq d(x_\infty, x_{n_{k_0}}) + d(x_{n_{k_0}}, x_0) \leq \frac{1}{4} + d(x_{n_{k_0}}, x_0).$$

Denote by $C_0 := d(x_{n_{k_0}}, x_0)$. Take a big integer N' so that $N' > C_0 + 1$, then for any $n_k > \max\{N, N'\}$, have

$$C_0 + 1 < N' < n_k < d(x_{n_k}, x_0) \leq d(x_{n_k}, x_\infty) + d(x_\infty, x_0) \leq \frac{1}{4} + (\frac{1}{4} + C_0) = \frac{1}{2} + C_0,$$

which is not possible.

By this way, we prove that such C exists.

□

COROLLARY 3.10. *Assume (X, d) is a metric space. Any convergent sequence $\{x_n\}$ is bounded.*

PROOF. Consider the set $\{x_n\} \cup \{\lim_{n\to\infty} x_n\}$. It is sequentially compact and so closed and bounded. In particular, this shows $\{x_n\}$ is bounded. □

(You can also prove it directly using definition, which is left to you as an exercise.)

PROPOSITION 3.11. *Assume (X, d) is a metric space. K is compact subset and S is a closed subset. Then $K \cap S$ is compact.*

PROOF. For any open cover of $\{U_\alpha | \alpha \in \Lambda\}$ of $K \cap S$, the collection of open sets

$$\{S^c, U_\alpha | \alpha \in \Lambda\}$$

is an open cover of K. Since K is compact, there must be a finite subcover of K.

Moreover, in this finite subcover, the ones that belong to $\{U_\alpha | \alpha \in \Lambda\}$ form a finite cover of $K \cap S$. This shows that $K \cap S$ is compact. □

COROLLARY 3.12. *Assume (X, d) is a metric space. K is compact subset and S is a closed subset. If $S \subset K$, then S is compact.*

THEOREM 3.13. *Assume (X, d) is a metric space and $\{K_\alpha | \alpha \in \Lambda\}$ is a collection of compact subsets in X. Then the intersection of any finite subcollection of $\{K_\alpha | \alpha \in \Lambda\}$ is nonempty if and only if $\cap_{\alpha\in\Lambda} K_\alpha$ is nonempty.*

PROOF.

(1) If $\cap_{\alpha\in\Lambda} K_\alpha$ is nonempty, obviously, the intersection of any finite subcollection is not empty.

(2) Assume the intersection of any finite subcollection is not empty, but $\cap_{\alpha\in\Lambda} K_\alpha = \emptyset$. Take some K_{α_0} from the collection. Since each K_α is compact, it is closed. Consider the collection

$$\{K_\alpha^c | \alpha \neq \alpha_0\}.$$

Since $\cap_{\alpha\in\Lambda} K_\alpha = \emptyset$, this must be an open cover of K_{α_0}. Because K_{α_0} is compact, it must have a finite subcover, say $K_1, \cdots, K_n$. Then

$$K_{\alpha_0} \cap K_1 \cap \cdots \cap K_n = \emptyset,$$

which contradicts with our assumption.

□

COROLLARY 3.14. *Assume (X, d) is a metric space and $\{K_n | n \in \mathbb{Z}^+\}$ is a sequence of nonempty compact subsets in X that satisfies*

$$K_{n+1} \subset K_n, \quad n = 1, 2, \cdots,$$

then $\cap_{n\in\mathbb{Z}^+} K_n$ is not empty.

EXAMPLE 3.15.

(1) Sequences of open intervals $I_n = (0, \frac{1}{n})$, $n \in \mathbb{Z}^+$. Have

$$I_1 \supset I_2 \supset I_3 \supset \cdots.$$

Then $\cap_{n\in\mathbb{Z}^+} I_n = \emptyset$.

(2) Sequences of closed intervals $\bar{I}_n = [0, \frac{1}{n}]$, $n \in \mathbb{Z}^+$. Have

$$\bar{I}_1 \supset \bar{I}_2 \supset \bar{I}_3 \supset \cdots .$$

Then $\cap_{n \in \mathbb{Z}^+} \bar{I}_n = \{0\} \neq \emptyset$.

3.2. The Heine-Borel theorem. We now understand the metric space $\mathbb{R}$ (and also $\mathbb{R}^n$) in more details.

We have seen that, in $\mathbb{R}$, for any two real numbers $a \leq b$, the closed interval $[a, b]$ is bounded and closed. Similarly, it is not hard to prove that in $\mathbb{R}^2$, for any two pairs of real numbers $a_1 \leq b_1$, $a_2 \leq b_2$, the 2-cell $[a_1, b_1] \times [a_2, b_2]$ is bounded and closed. In this section, we focus on $\mathbb{R}$, but all results work for any $\mathbb{R}^2$ and in fact, for any $\mathbb{R}^n$.

LEMMA 3.16. *For any sequence of nonempty closed intervals in* $\mathbb{R}$ *with*

$$\bar{I}_1 \supset \bar{I}_2 \supset \bar{I}_3 \cdots \supset \cdots ,$$

their intersection $\cap_{n=1}^{\infty} \bar{I}_n$ *is a nonempty closed interval.*

PROOF. Denote by $\bar{I}_n = [a_n, b_n]$, $n \in \mathbb{Z}^+$. Consider the subset

$$L := \{a_n | n \in \mathbb{Z}^+\} \subset \mathbb{R}.$$

It has upper bound b_1 so must has the least upper bound which we denote by a_∞. Similarly, the subset

$$R := \{b_n | n \in \mathbb{Z}^+\} \subset \mathbb{R}$$

has greatest lower bound which we denote by b_∞.

We now show that $a_\infty \leq b_\infty$. If $b_\infty < a_\infty$, then because a_∞ is the least upper bound, there exists some $a_n > b_\infty$. Then because b_∞ is the greatest lower bound, there exists some $b_{n'}$ so that $b_{n'} < a_n$. Further notice $\{b_n\}$ is nondecreasing, we can assume $n' > n$, and then we get contradiction from

$$a_{n'} \leq b_{n'} < a_n.$$

Then the closed interval $[a_\infty, b_\infty]$ is nonempty and lives in the intersection. Further, any point x that lives in the intersection should satisfy

$$a_n \leq x \leq b_n.$$

Then it follows $a_\infty \leq x \leq b_\infty$. This shows that

$$[a_\infty, b_\infty] = \cap_{n=1}^{\infty} [a_n, b_n].$$

□

PROPOSITION 3.17. *For any two real numbers $a < b$, the closed interval $[a, b] \subset \mathbb{R}$ is compact.*

PROOF. Denote by $\bar{I}_0 = [a, b]$. Assume $\bar{I}_0$ is not compact. Then there exists some open cover

$$\{U_\alpha | \alpha \in \Lambda\}$$

which doesn't have any finite subcover of $\bar{I}_0$. Notice that $\bar{I}_0$ is the union of two closed intervals

$$[a, \frac{a+b}{2}] \cup [\frac{a+b}{2}, b],$$

this says at least one of these two closed intervals has no finite subcover. Choose one with no finite subcover and denote it by $\bar{I}_2$.

Then we repeat this construction and obtain a sequence of nonempty closed intervals

$$\bar{I}_1 \supset \bar{I}_2 \supset \bar{I}_3 \cdots \supset \cdots .$$

By Lemma 3.16, their intersection is a nonempty closed interval, which we denote by $[a_\infty, b_\infty]$.

On the other hand, notice that

$$0 \leq b_\infty - a_\infty \leq b_n - a_\infty \leq b_n - a_n = \frac{b-a}{2^n},$$

and $\lim_{n\to\infty} \frac{b-a}{2^n} = 0$. This shows that $a_\infty = b_\infty$ and the intersection is a point $x_0 \in [a,b]$.

Assume $x_0 \in U_{\alpha_0}$ with some $\alpha_0 \in \Lambda$. Since U_{α_0} is open, there exits some $r > 0$ so that

$$(x_0 - r, x_0 + r) \subset U_\alpha.$$

Then since for any point $x \in \bar{I}_n$,

$$|x - x_0| \leq b_n - a_n = \frac{b-a}{2^n},$$

we can take n large enough so that $\frac{b-a}{2^n} < r$, and then it follows

$$\bar{I}_n \subset (x_0 - r, x_0 + r) \subset U_\alpha.$$

This contradicts with the construction of $\bar{I}_n$ that we assume it has no finite subcover from $\{U_\alpha | \alpha \in \Lambda\}$.

□

THEOREM 3.18. *[The Heine-Borel theorem] A subset $K \subset \mathbb{R}$ is (sequentially) compact if and only if it is bounded and closed.*

PROOF. From Proposition 3.9, we only need to show any bounded closed subsets of $\mathbb{R}$ is compact.

First, since K is bounded, there exists some closed interval $[a,b] \supset K$ with $a, b \in \mathbb{R}$, $a \leq b$. Then use Proposition 3.17 and Corollary 3.12, it follows K is compact. □

COROLLARY 3.19. *[The Weierstrass theorem] Each bounded sequence in $\mathbb{R}$ has a convergent subsequence.*

PROOF. Assume $\{x_n\}$ is bounded sequence and bounded by a closed interval $[a,b]$ with $a, b \in \mathbb{R}$, $a \leq b$. Then from the Heine-Borel theorem, $[a,b]$ is compact, and hence sequentially compact. It follows $\{x_n\}$ must have a convergent subsequence. □

EXAMPLE 3.20.

(1) For any sequence $\{x_n\}$ in $\mathbb{R}$ with

$$x_1 \leq x_2 \leq x_3 \leq \cdots,$$

the limit exists if and only if $\{x_n\}$ has an upper bound, and when it has upper bound, the limit is $\sup\{x_n\}$.

(2) For any sequence $\{x_n\}$ in $\mathbb{R}$ with

$$x_1 \geq x_2 \geq x_3 \geq \cdots,$$

the limit exists if and only if $\{x_n\}$ has a lower bound, and when it has lower bound, the limit is $\inf\{x_n\}$.

Before we end this chapter, we mention that the Heine-Borel theorem together with Proposition 3.9 and the first part of the proof of Theorem 3.8 give the complete proof of Theorem 3.8 for $\mathbb{R}$ (actually for $\mathbb{R}^n$), though the proof for the case of general metric space was left to you as homework.

CHAPTER 3

Numerical Sequences and Series

1. Sequences in $\mathbb{R}$

1.1. Convergent sequences in $\mathbb{R}$. We first review some definitions and basic properties of sequences. For our current purpose, we state for $\mathbb{R}$ only, but they work for any metric space with corresponding modifications.

DEFINITION 1.1. (1) A sequence, which we denote by $\{x_n\}$, in $\mathbb{R}$ is a map from $\mathbb{Z}^+$ to $\mathbb{R}$, which maps $n \in \mathbb{Z}^+$ to $x_n \in \mathbb{R}$. The range of the map is called the range of the sequence.

(2) A subsequence of $\{x_n\}$ is defined via an injective map s from $\mathbb{Z}^+$ to a subset of $\mathbb{Z}^+$ satisfying

$$s(k_1) < s(k_2), \quad \text{whenever } k_1, k_2 \in \mathbb{Z}^+, k_1 < k_2,$$

and denoted as $\{x_{n_k}\}$ with $x_{n_k} = x_{s(k)}$.

(3) A sequence $\{x_n\}$ in $\mathbb{R}$ is called convergent, if there exists some $x_0 \in \mathbb{R}$ such that the following holds: For any $\epsilon > 0$, there exists some $N \in \mathbb{Z}^+$ so that whenever $n > N$, $|x_n - x_0| < \epsilon$.

We denote it as $\lim_{n\to\infty} x_n = x_0$ or $x_n \to x_0$, and call $x_0 \in \mathbb{R}$ a limit of the sequence $\{x_n\}$.

(4) A sequence $\{x_n\}$ in $\mathbb{R}$ is called divergent, if it has no limit in $\mathbb{R}$.

PROPOSITION 1.2. (1) *The limit of a convergent sequence in $\mathbb{R}$ is unique.*

(2) *The sequence $\{x_n\}$ converges to $x_0 \in \mathbb{R}$ if and only if every open disk centered at x_0 contains all but finitely many of terms in the sequence.*

(3) *The sequence $\{x_n\}$ converges to $x_0 \in \mathbb{R}$ if and only if every subsequence of it converges to $x_0 \in \mathbb{R}$.*

(4) *If a sequence $\{x_n\}$ in $\mathbb{R}$ is convergent, then it must be bounded.*

(5) *The set of all subsequential limits of a sequence $\{x_n\}$ in $\mathbb{R}$ is closed.*

EXAMPLE 1.3. (1) $\{x_n = c\}$.

(2) $\{x_n = \frac{1}{n}\}$.

(3) $\{x_n = n^2\}$.

(4) $\{x_n = 1 + \frac{(-1)^n}{n}\}$.

(5) $\{x_n = (-1)^n\}$. Subsequences $\{x_{n_k} = x_{2k-1}\}$ and $\{x_{n_k} = x_{2k}\}$.

Next, let's see some properties for sequences in $\mathbb{R}$ (also in $\mathbb{C}$). From now on, in the section, when we mention a sequence we automatically assume it is a sequence in $\mathbb{R}$ unless stated specifically.

PROPOSITION 1.4. *[Weierstrass' theorem 3.19] Every bounded sequence in $\mathbb{R}$ must have a convergent subsequence.*

PROOF. Assume $\{x_n\}$ is a bounded sequence in $\mathbb{R}$, then there exists some $C > 0$ so that the range of the sequence lives in the closed interval $[-C, C]$. Since $[-C, C]$ is sequentially compact, there exists a convergent subsequence of $\{x_n\}$. $\square$

PROPOSITION 1.5. *[The squeeze theorem] Assume sequences $\{x_n\}, \{y_n\}, \{z_n\}$ in $\mathbb{R}$ satisfy $x_n \leq y_n \leq z_n$ after some $N \in \mathbb{Z}^+$, and both $\{x_n\}$ and $\{z_n\}$ converge to the same $a \in \mathbb{R}$. Then $\{y_n\}$ also converges to a.*

PROOF. Since $x_n \leq y_n \leq z_n$ for any $n \in \mathbb{Z}^+$, it follows

$$x_n - a \leq y_n - a \leq z_n - a,$$

and further

$$|y_n - a| \leq \max\{|x_n - a|, |z_n - a|\}.$$

Now for any $\epsilon > 0$, since both $\{x_n\}$ and $\{z_n\}$ converge to some $a \in \mathbb{R}$, there exists $N_1, N_2 \in \mathbb{Z}^+$ so that

$$|x_n - a| \leq \epsilon, \quad \text{whenever } n > N_1,$$

$$|z_n - a| \leq \epsilon, \quad \text{whenever } n > N_2.$$

Take $N = \max\{N_1, N_2\}$. Then for any $n > N$,

$$|y_n - a| \leq \max\{|x_n - a|, |z_n - a|\} \leq \epsilon.$$

This shows $y_n \to a$. □

PROPOSITION 1.6. *Consider convergent sequences $\{x_n\}$ and y_n with $x_n \to x$ and $y_n \to y$. Then*

(1) *$\{x_n \pm y_n\}$ is convergent to $x \pm y$;*
(2) *$\{x_n y_n\}$ is convergent to xy;*
(3) *$\{\frac{x_n}{y_n}\}$ is convergent to $\frac{x}{y}$, if $y_n \neq 0$, $n \in \mathbb{Z}^+$ and $y \neq 0$;*

PROOF. (1) Since $x_n \to x$ and $y_n \to y$, for any $\epsilon > 0$, there exists some N_1, N_2 so that whenever $n > N_1$, $|x_n - x| < \frac{\epsilon}{2}$, and whenever $n > N_2$, $|y_n - y| < \frac{\epsilon}{2}$. Take $N = \max\{N_1, N_2\}$. When $n > N$, have

$$|(x_n \pm y_n) - (x \pm y)| = |(x_n - x) \pm (y_n - y)| \leq |x_n - x| + |y_n - y| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

(2) Since $x_n \to x$, there exits some $C > 0$ so that for every n, $|x_n| \leq C$, and $|y| \leq C$. Then use $x_n \to x$ and $y_n \to y$, we have for any $\epsilon > 0$, there exists some N_1, N_2 so that whenever $n > N_1$, $|x_n - x| < \frac{\epsilon}{2C}$, and whenever $n > N_2$, $|y_n - y| < \frac{\epsilon}{2C}$. Take $N = \max\{N_1, N_2\}$. When $n > N$, have

$$\begin{aligned} |x_n y_n - xy| &= |(x_n y_n - x_n y) + (x_n y - xy)| \\ &\leq |x_n||y_n - y| + |x_n - x||y| \\ &\leq C\frac{\epsilon}{2C} + \frac{\epsilon}{2C}C = \epsilon. \end{aligned}$$

(3) Since $y \neq 0$ and $y_n \to y$, there exists some N_1 so that

$$|y_n - y| \leq \frac{|y|}{2}.$$

It follows $|y_n| \geq \frac{|y|}{2} > 0$, and then there exist some $C > 0$ so that $\frac{1}{|y_n y|} \leq C$.

Now for any $\epsilon > 0$, there exists some $N > N_1$, so that every $n > N$, have

$$|\frac{1}{y_n} - \frac{1}{y}| = \frac{|y_n - y|}{|y_n y|} \leq \frac{\epsilon}{C}C = \epsilon.$$

This shows that $\frac{1}{y_n} \to \frac{1}{y}$.

Then apply (2), we are done with (3).

□

We have mentioned that the real number system $\mathbb{R}$ can be formally extended to $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$. We introduce the following terminology.

DEFINITION 1.7. (1) A sequence $\{x_n\}$ in $\mathbb{R}$ diverges to $+\infty$, if for any $m > 0$, there exists some $N \in \mathbb{Z}^+$, so that whenever $n > N$, $x_n > m$. We denote it by $\lim_{n\to\infty} x_n = +\infty$ or $x_n \to +\infty$.

(2) A sequence $\{x_n\}$ in $\mathbb{R}$ diverges to $-\infty$, if for any $m > 0$, there exists some $N \in \mathbb{Z}^+$, so that whenever $n > N$, $x_n < -m$. We denote it by $\lim_{n\to\infty} x_n = -\infty$ or $x_n \to -\infty$.

Notice that, a sequence that diverges to $+\infty$ or $-\infty$ is divergent in $\mathbb{R}$. In another word, it is not a convergent sequence in $\mathbb{R}$ – Though we write its limit as $+\infty$ or $-\infty$, the limit doesn't not live in $\mathbb{R}$.

PROPOSITION 1.8. *For sequences $\{x_n\}$ and $\{y_n\}$ with $x_n \le y_n$. If $x_n \to +\infty$, then $y_n \to +\infty$.*

PROOF. Exercise for you. □

EXAMPLE 1.9. (1) $n^2 \to +\infty$.

(2) Any polynomial with positive degree diverges to $+\infty$ or $-\infty$.

DEFINITION 1.10. (1) A sequence $x_n \to a+$, if for any $\epsilon > 0$, there exists some $N \in \mathbb{Z}^+$ so that whenever $n > N$, $0 \le x_n - a < \epsilon$.

(2) A sequence $x_n \to a-$, if for any $\epsilon > 0$, there exists some $N \in \mathbb{Z}^+$ so that whenever $n > N$, $0 \le a - x_n < \epsilon$.

From definition, if $x_n \to a+$ or $x_n \to a-$, then $x_n \to a$.

PROPOSITION 1.11. *A sequence $\{x_n\}$ in $\mathbb{R}$ diverges to $\pm\infty$ if and only if $\frac{1}{x_n} \to 0\pm$.*

PROOF. Exercise for you. □

1.2. Some useful examples.

EXAMPLE 1.12. (1) For any $p > 0$, $\frac{1}{n^p} \to 0$.

PROOF. (a) $p \in \mathbb{Z}^+$. Use $0 \le \frac{1}{n^p} \le \frac{1}{n}$.

(b) $p \in \mathbb{Q}^+$. We can write $p = \frac{k}{\ell}$, $k, \ell \in \mathbb{Z}^+$. Then $n^p = (n^k)^{\frac{1}{\ell}}$.

From above, $n^k \to +\infty$, and this shows further $(n^k)^{\frac{1}{\ell}} \to +\infty$, which is equivalent to $\frac{1}{n^p} \to 0$.

(c) For any $p \in \mathbb{R}^+$, $n^p := \sup\{n^q | q \in \mathbb{Q}, q < p\}$. Then $0 \le \frac{1}{n^p} \le \frac{1}{n^q}$, for some $q \in \mathbb{Q}^+$. Apply the above case, this shows $\frac{1}{n^p} \to 0$.

□

(2) For any $|\alpha| < 1$, $\alpha^n \to 0$.

PROOF. Consider $|\frac{1}{\alpha^n}| = (\frac{1}{|\alpha|})^n$. Denote by $a = \frac{1}{|\alpha|}$, it is greater than 1, so can be written as $1 + c$ with $c > 0$. Calculate

$$a^n = (1 + c)^n \ge nc.$$

So $a^n \to +\infty$. This shows $\frac{1}{a^n} \to 0$ and hence $\alpha^n \to 0$. □

(3) $n^{\frac{1}{n}} \to 1$.

PROOF. Denote by $x_n = n^{\frac{1}{n}} - 1$. Notice it is positive whenever $n \geq 2$. We show $x_n \to 0$. Consider the following inequality

$$n = (1 + x_n)^n \geq \frac{n(n-1)}{2} x_n.$$

It follows

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

which shows $x_n \to 0$ by the squeeze theorem. □

(4) For any $\alpha > 0$, $\alpha^{\frac{1}{n}} \to 1$.

PROOF. If $\alpha \geq 1$, for large enough $n > \alpha$, we have

$$1 \leq \alpha^{\frac{1}{n}} \leq n^{\frac{1}{n}}.$$

Then by the squeeze theorem, $\alpha^{\frac{1}{n}} \to 1$.

If $0 < \alpha < 1$, then $\frac{1}{\alpha} > 1$ and $\alpha^{\frac{1}{n}} = \frac{1}{(\frac{1}{\alpha})^{\frac{1}{n}}} \to \frac{1}{1} = 1$. □

(5) For any $p > 0$ and $\alpha \in \mathbb{R}$, $\frac{n^\alpha}{(1+p)^n} \to 0$.

PROOF. If $\alpha < 0$, we have seen that $n^\alpha \to 0$, and $\frac{1}{(1+p)^n} \to 0$. Hence it follows $\frac{n^\alpha}{(1+p)^n} \to 0$. If $\alpha = 0$, $\frac{n^\alpha}{(1+p)^n} = \frac{1}{(1+p)^n} \to 0$.

The real work is for the case $\alpha > 0$. For this case, we want to see $(1 + p)^n$ diverges to $+\infty$ faster than n^α.

Write

$$(1 + p)^n = \Sigma_{k=0}^n C_n^k p^k > C_n^k p^k > \frac{n^k p^k}{2^k k!}.$$

where $C_n^k = \frac{n!}{k!(n-k)!}$. In particular, consider $k > \alpha$, we have

$$(1 + p)^n > \frac{n^k p^k}{2^k k!} = \frac{n^{k-\alpha} p^k}{2^k k!} n^\alpha,$$

and

$$\frac{n^\alpha}{(1 + p)^n} < \frac{2^k k!}{p^k n^{k-\alpha}} \to 0.$$

□

1.3. Cauchy sequences.

DEFINITION 1.13. A sequence $\{x_n\}$ in a metric space (X, d) is called a Cauchy sequence, if for any $\epsilon > 0$, there exists some $N \in \mathbb{Z}^+$ so that whenever $m, n > N$, have

$$d(x_m, x_n) < \epsilon.$$

This is equivalent to say, for any $\epsilon > 0$, there exists some $N \in \mathbb{Z}^+$ so that whenever $n > N$, for any $p > 0$ have

$$d(x_{n+p}, x_n) < \epsilon.$$

LEMMA 1.14. *Any convergent sequence is a Cauchy sequence.*

PROOF. Assume $\{x_n\}$ is a convergent sequence that converges to $x_0 \in X$. Then for any $\epsilon > 0$, there exists some $N \in \mathbb{Z}^+$ so that any $n > N$,

$$d(x_n, x_0) < \frac{\epsilon}{2}.$$

Then for any $m, m' > N$, have

$$d(x_m, x_{m'}) \leq d(x_m, x_0) + d(x_{m'}, x_0) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

This shows $\{x_n\}$ is a Cauchy sequence. □

LEMMA 1.15. *The range of a Cauchy sequence is bounded.*

PROOF. Assume $\{x_n\}$ is a Cauchy sequence. Take $\epsilon = 1$. There exists some $N \in \mathbb{Z}^+$ so that any $n > N$,

$$d(x_n, x_{N+1}) < 1.$$

Take $C := \max_{i=1,\cdots,N}\{d(x_i, x_{N+1}), 1\}$. C is a finite number and for each $n \in \mathbb{Z}^+$, there is

$$d(x_n, x_{N+1}) \leq C,$$

which says the range of $\{x_n\}$ is bounded. □

PROPOSITION 1.16. *For a Cauchy sequence $\{x_n\}$, if it has a convergent subsequence, then $\{x_n\}$ converges to the same limit.*

PROOF. Assume $\{x_n\}$ has a subsequence $\{x_{n_k}\}$ with $x_{n_k} \to x_0 \in X$.

Then for any $\epsilon > 0$, there exists some $N_1 \in \mathbb{Z}^+$ so that any $n_k > N_1$ has

$$d(x_{n_k}, x_0) < \frac{\epsilon}{2}.$$

At the same time, since $\{x_n\}$ is a Cauchy sequence, there exists some $N_2 \in \mathbb{Z}^+$ so that any $m, m' > N$,

$$d(x_m, x_{m'}) < \frac{\epsilon}{2}.$$

Take $N = \max\{N_1, N_2\}$. Then for any $n > N$, automatically $n_n > N$. We then have

$$d(x_n, x_0) \leq d(x_n, x_{n_n}) + d(x_{n_n}, x_0) \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

it says $x_n \to x_0$. □

COROLLARY 1.17. *A Cauchy sequence in a compact subset of a metric space (X, d) is convergent.*

PROOF. Assume $K \subset X$ is compact. Then by Theorem 3.8, K is sequentially compact. Then any sequence $\{x_n\}$ in K has a convergent subsequence. By Proposition 1.16, $\{x_n\}$ is convergent. □

In general, not every Cauchy sequence in a metric space is convergent. For example, sequence

$$\{x_n = \frac{1}{n}\}$$

is not convergent in $\mathbb{R} \setminus \{0\}$.

As another example, consider $\{x_n\}$ as a sequence of increasing rational numbers with $x_n^2 < 2$. It is a Cauchy sequence, but is not convergent in $\mathbb{Q}$.

THEOREM 1.18. *In $\mathbb{R}$, a sequence is convergent, if and only if it is a Cauchy sequence.*

PROOF. We have shown from Lemma 1.14 that every convergent sequence is a Cauchy sequence (in any metric space).

Now assume $\{x_n\}$ is a Cauchy sequence in $\mathbb{R}$, in particular from Lemma 1.15, it is bounded and so we can assume it is a Cauchy sequence in some closed interval $[a, b]$ with some $a, b \in \mathbb{R}$, $a \leq b$.

By the Heine-Borel theorem, $[a, b]$ is compact, and then apply Corollary 1.17, we have shown $\{x_n\}$ is convergent. □

In general, a metric space (X, d) in which every Cauchy sequence is convergent is said to be complete. This theorem states that $\mathbb{R}$ is complete. Our previous examples say that $\mathbb{R}\setminus\{0\}$, $\mathbb{Q}$ are not complete with respect to the metric induced from $\mathbb{R}$. (A general construction, called completion, which construct a complete space from an incomplete space is given in Rudin's book Chapter 3 Ex. 24.)

1.4. Upper and lower limits. We can summarize the methods we met so far to show a sequence in $\mathbb{R}$ is convergent:

(1) A Cauchy sequence in $\mathbb{R}$ must be convergent.
(2) A bounded monotonic sequence in $\mathbb{R}$ is convergent.

In general, we can not expect an arbitrary sequence $\{x_n\}$ in $\mathbb{R}$ is convergent, i.e., we can not expect $\lim_{n\to\infty} x_n$ always exists. But we can introduce upper and lower limits for a sequence as a generalization of limit, so that for any sequence, they make sense, and the limit exists if and only if the upper and lower limits coincide.

Recall in Proposition 1.2 (5), we have shown that the set of all subsequential limits of a sequence $\{x_n\}$ in $\mathbb{R}$ is closed. Now we also include $\pm\infty$ if there is some subsequence diverges to $\pm\infty$, and denote by

$$L := \{a \in \overline{\mathbb{R}} | x_{n_k} \to a \text{ for subsequence } \{x_{n_k}\} \text{ of } \{x_n\}\}.$$

L is never empty now.

DEFINITION 1.19. (1) The upper limit of $\{x_n\}$ is defined as the least upper bound of L, i.e., $\sup L$. If L has no upper bound, define $\sup L = +\infty$. Denote as $\limsup_{n\to\infty} x_n = \sup L$.
(2) The lower limit of $\{x_n\}$ is defined as the greatest lower bound of L, i.e., $\inf L$. If L has no lower bound, define $\inf L = -\infty$. Denote as $\liminf_{n\to\infty} x_n = \inf L$.

PROPOSITION 1.20. *A sequence $x_n \to a$ for some $a \in \overline{\mathbb{R}}$, if and only if*

$$\limsup_{n\to\infty} x_n = \liminf_{n\to\infty} x_n = a.$$

PROOF. Assume $a \in \mathbb{R}$. By definition,

$$\limsup_{n\to\infty} x_n = \liminf_{n\to\infty} x_n = a$$

if and only if $L = \{a\}$ (Need some work here and is left to you.), which is equivalent to say $x_n \to a$.

The cases $a = \pm\infty$ are left to you as exercise.

□

LEMMA 1.21. (1) *$L = \overline{L}$, i.e., L is closed. In particular,* $\sup L, \inf L \in L$*;*
(2) *If $x > \sup L$, then there exists some N so that each $n > N$, $x_n < x$. Similarly, if $x < \inf L$, then there exists some N so that each $n > N$, $x_n > x$.*

PROOF. (1) Take any sequence $\{a_k\}$ from L with $a_k \to a$. We show that $a \in L$. For this, we construct a sequence $\{x_{n_k}\}$ as follows: For $k = 1$, we take n_1 large enough so that $|x_{n_1} - a_1| < 1$, since a_1 is limit of some subsequence of $\{x_n\}$. Assume now for any k, we have found some x_{n_k} so that

$$n_1 < n_2 < \cdots < n_k,$$

and

$$|x_{n_k} - a_k| < \frac{1}{k}.$$

We can further take $n_{k+1} > n_k$ and $|x_{n_{k+1}} - a_{k+1}| < \frac{1}{k+1}$.

We now prove $\lim_{k\to\infty} x_{n_k} = a$. To see this, notice

$$|x_{n_k} - a| \leq |x_{n_k} - a_k| + |a_k - a| \leq \frac{1}{k} + |a_k - a|, \quad k = 1, 2, \cdots.$$

Then the convergence of RHS to zero implies $x_{n_k} \to a$ as $k \to \infty$. Hence $a \in L$.

(2) We prove the $\sup L$ case and the $\inf L$ case is similar.

Assume this is not true, then there must exists a subsequence $\{x_{n_k}\}$ so that each $x_{n_k} \geq x$. If this sequence has no upper bound, then $\sup L = +\infty$, which contradicts with $\sup L < x$. If it has an upper bound, then we can further take a subsequence from it so that this subsequence converges. Then the limit point must be greater than or equal to x, so greater than $\sup L$, which is a contradiction again.

□

The next two propositions are left to you as homework problems.

PROPOSITION 1.22. *Assume two sequences $\{x_n\}$, $\{y_n\}$ satisfy that after some N, $x_n \leq y_n$. Then*

$$\limsup x_n \leq \limsup y_n$$
$$\liminf x_n \leq \liminf y_n.$$

PROPOSITION 1.23.

$$\limsup_{n\to\infty} x_n = \lim_{n\to\infty} (\sup\{x_k | k \geq n\})$$
$$\liminf_{n\to\infty} x_n = \lim_{n\to\infty} (\inf\{x_k | k \geq n\}).$$

2. Series in ℝ (ℂ)

2.1. Definition and basic properties. The results in the section works for ℂ, though we restrict ourselves to ℝ.

Consider a sequence $\{x_n\}$ in ℝ. The summation notation means

$$\Sigma_{k=n}^{n+p} x_k := x_n + x_{n+1} + \cdots + x_{n+p}.$$

In particular, the summation

$$s_n := \Sigma_{k=1}^{n} x_k = x_1 + x_2 + \cdots + x_n$$

is called a partial sum to n.

An (infinite) series is a sequence of partial sums $\{s_n\}$ (for some sequence). When we want to explicitly write down elements that make this series, we can write $\Sigma_{n=1}^{\infty} x_n$ to denote this series. We write it as Σx_n to simplify notations sometimes.

Given a series $\{s_n\}$, we can recover the sequence $\{x_n\}$ by assigning

$$x_1 = s_1, \quad x_n = s_n - s_{n-1}, n \geq 2.$$

Hence the series Σx_n carries equal information as the sequence $\{x_n\}$.

If the limit of the partial sums $\{s_n\}$ exists and is s, we say the series Σx_n converges to s, and write it as

$$\Sigma x_n = s.$$

If the limit of the partial sums $\{s_n\}$ doesn't exist, we say the series Σx_n diverges. If particular, if Σx_n diverges to $\pm\infty$, we write $\Sigma x_n = \pm\infty$.

EXAMPLE 2.1. (1) $\Sigma_{n=1}^{\infty}(-1)^n$ diverges.

Think: Why is the following "calculation" not correct?

$$\begin{aligned} \Sigma(-1)^n &= (-1) + 1 + (-1) + 1 + (-1) + 1 + (-1) + \cdots \\ &= ((-1) + 1) + ((-1) + 1) + ((-1) + 1) + ((-1) + 1) \cdots \\ &= 0 + 0 + 0 + \cdots = 0. \end{aligned}$$

(2)

$$\begin{aligned} \Sigma_{n=1}^{\infty} \frac{1}{2^n} &= \lim_{n\to\infty} S_n \\ &= \lim_{n\to\infty} \frac{1}{2} \cdot \frac{1 - (\frac{1}{2})^n}{1 - \frac{1}{2}} = 1. \end{aligned}$$

(3) Similarly, $\Sigma_{n=0}^{\infty} x^n = \frac{1}{1-x}$ for any $|x| < 1$. Such series is called a geometric series.

(4) $\Sigma_{n=1}^{\infty} \frac{1}{n^p}$ is convergent if and only if $p > 1$. Moreover, if $p \leq 1$, $\Sigma_{n=1}^{\infty} \frac{1}{n^p} = +\infty$. (Will prove later.)

The value of $\Sigma_{n=1}^{\infty} \frac{1}{n^p}$ for $p > 1$ is hard to calculate in general. E.g.,

$$\begin{aligned} \Sigma_{n=1}^{\infty} \frac{1}{n^2} &= \frac{\pi^2}{6} = 1.6449 \cdots \\ \Sigma_{n=1}^{\infty} \frac{1}{n^4} &= \frac{\pi^4}{90} = 1.0823 \cdots . \end{aligned}$$

Similar formula can be derived for even p's. For odd p's, no general formula (e.g., $\Sigma_{n=1}^{\infty} \frac{1}{n^3}$).

Now we derive some tools for testing if a series is convergent.

PROPOSITION 2.2. *[Cauchy's Criterion] A series Σx_n is convergent if and only if for any $\epsilon > 0$, there exists some N so that for any $n > N$ and $p \geq 0$,*

$$|\Sigma_{k=n}^{n+p} x_k| < \epsilon.$$

This implies $\lim_{n\to\infty} |\Sigma_{k=n}^{n+p} x_k| = 0$ *for any $p \geq 0$.*

In particular, by taking $p = 0$, we obtain the following corollary.

COROLLARY 2.3. *For any convergent series Σx_n, the limit of $\{x_n\}$ is zero.*

PROOF OF PROPOSITION 2.2. The series Σx_n is convergent if and only if it is a Cauchy sequence. In particular here, this is

$$|S_{n+p} - S_{n-1}| = |\Sigma_{k=n}^{n+p} x_k| \to 0 \quad \text{as } n \to \infty$$

for any $p \geq 0$.

□

We can see an application of Cauchy's theorem, which in particular shows that the alternating series $\Sigma \frac{(-1)^n}{n}$ is convergent.

PROPOSITION 2.4. *Assume $x_n \geq 0$ and*

(1) $x_1 \geq x_2 \geq \cdots$;
(2) $x_n \to 0$ *as* $n \to \infty$.

Then the series $\Sigma(-1)^n x_n$ is convergent.

PROOF. Consider

$$|\Sigma_{k=n}^{n+p} (-1)^k x_k| = |x_n - x_{n+1} + \cdots + \pm x_{n+p}|$$

It is less than x_n by the first condition. Then the second condition $x_n \to 0$ shows that $|\Sigma_{k=n}^{n+p} (-1)^k x_k| \to 0$ as $n \to \infty$ for any $p \geq 0$. Apply the Cauchy's criterion, this shows $\Sigma(-1)^n x_n$ is convergent. □

PROPOSITION 2.5. *For a series with nonnegative terms, it is convergent if and only if the partial sums form a bounded sequences.*

PROOF. Notice that for a series with nonnegative terms, the sequence of partial sums $\{s_n\}$ is nondecreasing, and then its limit exists if and only if it is bounded. □

For a series Σx_n, if $\Sigma |x_n|$ converges, we say Σx_n converges absolutely.

PROPOSITION 2.6. *If Σx_n converges absolutely, then Σx_n converges.*

PROOF. This follows from the estimate

$$|\Sigma_{k=n}^{n+p} x_k| \leq \Sigma_{k=n}^{n+p} |x_k|$$

and Cauchy's criterion. □

However, not every convergent series is absolutely convergent. For example $\Sigma \frac{(-1)^n}{n}$ is not absolutely convergent.

2.2. Comparison test.

PROPOSITION 2.7. *[Comparison test]*

(1) *If $|x_n| \leq y_n$ for all n and Σy_n converges, then Σx_n converges absolutely (and so converges).*
(2) *If $x_n \leq y_n$ for all n and $\Sigma x_n = +\infty$, then $\Sigma y_n = +\infty$.*

PROOF. (1) It follows from the estimates

$$|\Sigma_{k=n}^{n+p} x_k| \leq \Sigma_{k=n}^{n+p} |x_k| \leq \Sigma_{k=n}^{n+p} y_k = |\Sigma_{k=n}^{n+p} y_k|$$

and Cauchy's criterion.

(2) This holds by looking at the partial sums

$$\Sigma_{k=1}^{n} x_n \leq \Sigma_{k=1}^{n} y_n.$$

Then $\Sigma_{k=1}^{n} x_n \to +\infty$ implies that $\Sigma_{k=1}^{n} y_n \to +\infty$.

□

EXAMPLE 2.8. (1) $\Sigma \frac{1}{n^2+5}$ is convergent by the comparison test:

$$0 \leq \frac{1}{n^2+5} \leq \frac{1}{n^2}$$

and the fact that $\Sigma \frac{1}{n^2}$ is convergent (will prove).

(2) $\Sigma \frac{n}{2^n}$ is convergent by the comparison test:

$$0 \leq \frac{n}{2^n} \leq \frac{1.5^n}{2^n} = 0.75^n$$

and the fact that $\Sigma 0.75^n$ is convergent.

Now we prove the important result for series $\Sigma \frac{1}{n^p}$ using the comparison test.

PROPOSITION 2.9. *The series* $\Sigma \frac{1}{n^p}$ *is convergent if* $p > 1$ *and diverges to* $+\infty$ *if* $p \leq 1$.

To show this result, Rudin's book gives a useful statement.

LEMMA 2.10. *A series* Σx_n *with*

$$x_1 \geq x_2 \geq \cdots \geq 0,$$

is convergent, if and only if the series

$$\Sigma_{k=0}^{\infty} 2^k x_{2^k} = x_1 + 2x_2 + 4x_4 + 8x_8 + \cdots$$

is convergent.

PROOF OF THE LEMMA. (1) Take any N, there exists unique K so that $2^{K-1} < N \leq 2^K$. Then we have

$$\Sigma_{n=1}^{N} x_n \leq \Sigma_{k=1}^{K} 2^k x_{2^k}.$$

So $\Sigma_{k=0}^{\infty} 2^k x_{2^k}$ is convergent implies Σx_n is convergent.

(2) Take any K, and then choose $N > 2^K$. We have

$$\frac{1}{2}\Sigma_{k=1}^{K} 2^k x_{2^k} \leq \Sigma_{n=1}^{N} x_n.$$

So Σx_n is convergent implies $\Sigma_{k=0}^{\infty} 2^k x_{2^k}$ is convergent.

□

PROOF OF THE PROPOSITION. If $p \leq 0$, $\Sigma \frac{1}{n^p}$ is divergent since $\frac{1}{n^p}$ is not convergent to zero.

Now assume $p > 0$. Then the series $\Sigma \frac{1}{n^p}$ satisfies the assumption of the lemma. So we can look at the series $\Sigma_{k=0}^{\infty} 2^k x_{2^k}$ instead. We have

$$\Sigma_{k=0}^{\infty} 2^k x_{2^k} \quad = \quad \Sigma_{k=0}^{\infty} \frac{2^k}{2^{kp}} = \Sigma_{k=0}^{\infty} (\frac{1}{2^{p-1}})^k,$$

which is convergent if and only if $p > 1$. □

2.3. Root and ratio tests.

THEOREM 2.11. *[Root test] For a seires Σx_n, define*

$$\alpha := \limsup \sqrt[n]{|x_n|}.$$

Then we have

(1) *if $\alpha < 1$, Σx_n is (absolutely) convergent;*
(2) *if $\alpha > 1$, Σx_n is divergent.*
(3) *When $\alpha = 1$, root test is failed to detect for convergence or divergence.*

PROOF. (1) If $\alpha < 1$, then there exists some β so that $\alpha < \beta < 1$, and then there exists some $N \in \mathbb{Z}^+$ so that for any $n > N$,

$$\sqrt[n]{|x_n|} < \beta.$$

Then it follows $|x_n| < \beta^n$. Notice that $\Sigma\beta^n$ is convergent, and then by the comparison test, Σx_n is (absolutely) convergent.

(2) If $\alpha > 1$, then there exists subsequence $\{\sqrt[n]{|x_{n_k}|}\}$ with each $\sqrt[n]{|x_{n_k}|} \geq 1$. It follows $|x_{n_k}| \geq 1$, which indicates Σx_n is divergent.

(3) Consider the series $\Sigma\frac{1}{n^p}$. Calculate that

$$\limsup \sqrt[n]{|x_n|} = \limsup \sqrt[n]{\frac{1}{n^p}} = (\frac{1}{\lim \sqrt[n]{n}})^p = 1^p = 1,$$

but we know the convergence of $\Sigma\frac{1}{n^p}$ depends on values of p.

□

EXAMPLE 2.12. (1) $\Sigma\frac{1}{a^n}$. Compute

$$\alpha = \limsup \sqrt[n]{|x_n|} = \frac{1}{|a|}.$$

By root test,

(a) if $|a| > 1$, Σx_n is (absolutely) convergent;
(b) if $|a| < 1$, Σx_n is divergent.
(c) if $|a| = 1$, it is divergent.

(2) $\Sigma\frac{x^n}{n}$. Compute

$$\alpha = \limsup \sqrt[n]{|x_n|} = |x|.$$

By root test,

(a) if $|x| < 1$, Σx_n is (absolutely) convergent;
(b) if $|x| > 1$, Σx_n is divergent.
(c) if $x = 1$, it is divergent.
(d) if $x = -1$, it is convergent.

(3) $\Sigma\frac{x^n}{n^2}$. Compute

$$\alpha = \limsup \sqrt[n]{|x_n|} = |x|.$$

By root test,

(a) if $|x| < 1$, Σx_n is (absolutely) convergent;
(b) if $|x| > 1$, Σx_n is divergent.
(c) if $|x| = 1$, it is convergent.

THEOREM 2.13. *[Ratio test] For a seires Σx_n, define*

$$\alpha := \limsup \frac{|x_{n+1}|}{|x_n|}.$$

Then we have

(1) *if $\alpha < 1$, Σx_n is (absolutely) convergent;*
(2) *if for large enough n, $\frac{|x_{n+1}|}{|x_n|} \geq 1$, Σx_n is divergent. In particular, if*

$$\liminf \frac{|x_{n+1}|}{|x_n|} > 1,$$

Σx_n is divergent.

PROOF. (1) If $\alpha < 1$, then there exists some β so that $\alpha < \beta < 1$, and then there exists some $N \in \mathbb{Z}^+$ so that for any $n \geq N$,

$$\frac{|x_{n+1}|}{|x_n|} < \beta.$$

Then it follows $|x_{n+k}| < \beta^k |x_N|$. Notice that $\Sigma \beta^k$ is convergent, and then by the comparison test, Σx_n is (absolutely) convergent.

(2) If $\alpha > 1$ and $\frac{|x_{n+1}|}{|x_n|} \geq 1$ for $n > N$ for some N, then

$$|x_{n+1}| \geq |x_n|, \quad \text{and} \quad x_n \neq 0.$$

This shows that x_n can not converge to 0 and so Σx_n is divergent.

In particular, if

$$\liminf \frac{|x_{n+1}|}{|x_n|} > 1,$$

then it follows $\frac{|x_{n+1}|}{|x_n|} \geq 1$ for $n > N$ for some N, and so Σx_n is divergent.

□

EXAMPLE 2.14. $\Sigma \frac{x^n}{n!}$. Compute

$$\alpha = \limsup \frac{|x_{n+1}|}{|x_n|} = 0$$

for any x. By ratio test, for any $x \in \mathbb{R}$, Σx_n is convergent.

DEFINITION 2.15. Given $x \in \mathbb{R}$, a series of the type $\Sigma c_n x^n$ is called a power series. (In general, people consider $x, c_n \in \mathbb{C}$.)

Using the root test, we have the following result.

THEOREM 2.16. *[Convergence radius for a power series] For a power series $\Sigma c_n x^n$, define*

$$\alpha := \limsup \sqrt[n]{|c_n|},$$

and $R = \frac{1}{\alpha}$. Here if $\alpha = 0$, define $R = +\infty$, and if $\alpha = +\infty$, define $R = 0$. Then $\Sigma c_n x^n$ is convergent if $|x| < R$; and $\Sigma c_n x^n$ is divergent if $|x| > R$. R is called the convergence radius of $\Sigma c_n x^n$.

We have seen that the power series $\Sigma_{n=0}^{\infty} \frac{x^n}{n!}$ has convergence radius $+\infty$, which says that it is convergent for every $x \in \mathbb{R}$. In particular, take $x = 1$, and we have a convergent sum

$$1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots.$$

Define this sum as e.

We can estimate that $e > 2.5$ and $e < 3$ since

$$2.5 = 1 + 1 + \frac{1}{2!} < 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots \leq 1 + 1 + \Sigma_{n=1}^{\infty}\frac{1}{2^n} = 3.$$

In fact, we can estimate

$$\begin{aligned} e - \Sigma_{n=1}^{N}\frac{1}{n!} &= \Sigma_{n=N+1}^{\infty}\frac{1}{n!} \\ &= \frac{1}{(N+1)!} + \frac{1}{(N+2)!} + \cdots \\ &= \frac{1}{(N+1)!}(1 + \frac{1}{N+2} + \frac{1}{(N+2)(N+3)} + \cdots) \\ &\leq \frac{1}{(N+1)!}(1 + \frac{1}{N+2} + \frac{1}{(N+2)^2} + \cdots) \\ &\leq \frac{1}{N \cdot N!}. \end{aligned}$$

When $N = 10$, $\frac{1}{N \cdot N!} < 10^{-7}$, which says that error between

$$1 + 1 + \frac{1}{2!} + \cdots + \frac{1}{10!}$$

and the precise value of e is smaller than 0.0000001. ($e \approx 2.71828182845904523536028747135 27\cdots$.)

We now give another interpretation for e.

PROPOSITION 2.17. *The sequence* $\{(1 + \frac{1}{n})^n\}$ *converges to e.*

PROOF. Denote by $S_n = \Sigma_{k=0}^{n}\frac{1}{k!}$.

(1) We estimate

$$\begin{aligned} (1 + \frac{1}{n})^n &= 1 + 1 + \frac{n(n-1)}{2!n^2} + \cdots \frac{n(n-1)\cdots(n-(k-1))}{k!n^k} + \cdots + \frac{n!}{n!n^n} \\ &\leq 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{n!} = S_n. \end{aligned}$$

Take $n \to \infty$, we have

$$\limsup_{n\to\infty}(1 + \frac{1}{n})^n \leq \lim_{n\to\infty} S_n = e.$$

(2) On the other hand, for any $m \leq n$, we have

$$\begin{aligned} (1 + \frac{1}{n})^n &= 1 + 1 + \frac{n(n-1)}{2!n^2} + \cdots \frac{n(n-1)\cdots(n-(k-1))}{k!n^k} + \cdots \frac{n!}{n!n^n} \\ &\geq 1 + 1 + \frac{n(n-1)}{2!n^2} + \cdots \frac{n(n-1)\cdots(n-(m-1))}{m!n^m}. \end{aligned}$$

Fix m and let $n \to \infty$, we get

$$\begin{aligned} \liminf_{n\to\infty}(1 + \frac{1}{n})^n &\geq \lim_{n\to\infty}(1 + 1 + \frac{n(n-1)}{2!n^2} + \cdots \frac{n(n-1)\cdots(n-(m-1))}{m!n^m}) \\ &= 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{m!} = S_m. \end{aligned}$$

Then take $m \to \infty$, we have

$$\liminf_{n\to\infty}(1 + \frac{1}{n})^n \geq \lim_{m\to\infty} S_m = e.$$

Combine (1) and (2), we have

$$e \leq \liminf_{n\to\infty}(1+\frac{1}{n})^n \leq \limsup_{n\to\infty}(1+\frac{1}{n})^n \leq e.$$

Hence $\lim_{n\to\infty}(1+\frac{1}{n})^n$ exists and it is e. □

2.4. Addition and multiplication of series.

PROPOSITION 2.18. *Assume series* $\Sigma a_n = A$, $\Sigma b_n = B$. *Then*

(1) $\Sigma(a_n + b_n) = A + B$;
(2) *For any* $c \in \mathbb{R}$, $\Sigma(c \cdot a_n) = cA$.

A question now is how to do multiplications of series.

EXAMPLE 2.19. $\Sigma a_n \cdot \Sigma b_n \neq A \cdot B$. This is obvious since even

$$(a_1 + a_2)\cdot(b_1 + b_2) \neq a_1 b_1 + a_2 b_2$$

in general.

Recall multiplication of polynomials

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots, \quad g(x) = b_0 + b_1 x + b_2 x^2 + \cdots,$$

we have

$$f(x)\cdot g(x) = a_0 b_0 + (a_0 b_1 + a_1 b_0)x + (a_0 b_2 + a_1 b_1 + a_2 b_0)x^2 + \cdots.$$

If we take $x = 1$, then the LHS is $\Sigma a_n \cdot \Sigma b_n$, the RHS should be the same.

By this, we define the series

$$\Sigma c_n, \quad c_n := \Sigma_{k=0}^n a_k b_{n-k}$$

as the multiplication of series. (Notice: the output is a series that may not be convergent even both Σa_n and Σb_n are convergent!)

REMARK 2.20. $c_n \neq \Sigma_{k=0}^n a_k \cdot \Sigma_{k=0}^n b_k$ in general.

EXAMPLE 2.21. Consider the series

$$\Sigma a_n = \Sigma_{n=1}^{\infty} \frac{(-1)^n}{\sqrt{n}}$$

which is convergent since it satisfies

(1) $|a_n| \geq |a_{n+1}|$;
(2) $a_n \to 0$.

However

$$\begin{aligned} & \Sigma a_n \cdot \Sigma a_n = \Sigma c_n \\ = \ & \frac{-1}{\sqrt{1}}\frac{-1}{\sqrt{1}} + (\frac{-1}{\sqrt{1}}\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}\frac{-1}{\sqrt{1}}) + (\frac{-1}{\sqrt{1}}\frac{-1}{\sqrt{3}} + \frac{1}{\sqrt{2}}\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}}\frac{-1}{\sqrt{1}}) \\ & + (\frac{-1}{\sqrt{1}}\frac{1}{\sqrt{4}} + \frac{1}{\sqrt{2}}\frac{-1}{\sqrt{3}} + \frac{-1}{\sqrt{3}}\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{4}}\frac{-1}{\sqrt{1}}) + \cdots \end{aligned}$$

Estimate a term in c_n as follows: Assume n is even, then

$$|c_n| = |\Sigma_{k=1}^n \frac{-1}{\sqrt{k}\sqrt{n-k+1}}| = |\Sigma_{k=1}^n \frac{1}{\sqrt{k}\sqrt{n-k+1}}| \geq \Sigma_{k=1}^n \frac{\sqrt{2}}{n+1} = \frac{\sqrt{2}n}{n+1}$$

which shows that c_n doesn't converge to 0 and hence Σc_n diverges.

THEOREM 2.22. *[By Mertens] Suppose*

(1) $\Sigma a_n = A$ *and* $\Sigma b_n = B$*;*
(2) Σa_n *converges absolutely.*

Then $\Sigma c_n = AB$*, where* $c_n := \Sigma_{k=0}^n a_k b_{n-k}$.

PROOF. We calculate the partial sums for Σc_n.

$$\begin{aligned}\Sigma_{n=0}^m c_n &= \Sigma_{n=0}^m \Sigma_{k=0}^n a_k b_{n-k} \\ &= \Sigma_{k=0}^m \Sigma_{n=0}^{m-k} a_k b_n \\ &= \Sigma_{k=0}^m (a_k \Sigma_{n=0}^{m-k} b_n) \\ &= \Sigma_{k=0}^m a_k S_{m-k}^b.\end{aligned}$$

Here S^b denotes the partial sum for Σb_n. Now we look at

$$\begin{aligned}\Sigma_{n=0}^m c_n - AB &= \Sigma_{k=0}^m a_k S_{m-k}^b - AB \\ &= \Sigma_{k=0}^m a_k (S_{m-k}^b - B) + \Sigma_{k=0}^m a_k B - AB \\ &= \Sigma_{k=0}^m a_k (S_{m-k}^b - B) + (\Sigma_{k=0}^m a_k - A)B.\end{aligned}$$

So we can estimate

$$\begin{aligned}|\Sigma_{n=0}^m c_n - AB| &\leq |\Sigma_{k=0}^m a_k (S_{m-k}^b - B)| + |(\Sigma_{k=0}^m a_k - A)B| \\ &\leq \Sigma_{k=0}^m (|a_k| \cdot |S_{m-k}^b - B|) + |\Sigma_{k=0}^m a_k - A||B|\end{aligned}$$

The second term converges to zero since $\Sigma a_k = A$.

Let's prove the first term $\Sigma_{k=0}^m (|a_k| \cdot |S_{m-k}^b - B|) \to 0$ as $m \to \infty$.

Denote by $\beta_n := S_n^b - B$, we have $\beta_n \to 0$ as $n \to \infty$. Hence for any $\epsilon > 0$, there exists some N so that whenever $n > N$, $|\beta_n| < \epsilon$.

$$\begin{aligned}& \Sigma_{k=0}^m (|a_k| \cdot |S_{m-k}^b - B|) \\ = \;& |a_0||\beta_m| + |a_1||\beta_{m-1}| + \cdots + |a_m||\beta_0| \\ = \;& (|a_0||\beta_m| + |a_1||\beta_{m-1}| + \cdots |a_{m-N-1}||\beta_{N+1}|) + (|a_{m-N}||\beta_N| + \cdots + |a_m||\beta_0|) \\ \leq \;& \epsilon \Sigma_{k=0}^{m-N-1} |a_k| + \max\{|\beta_0|, \cdots, |\beta_N|\} \Sigma_{k=m-N}^m |a_k| \\ \leq \;& \epsilon \Sigma_{k=0}^\infty |a_k| + \max\{|\beta_0|, \cdots, |\beta_N|\} \Sigma_{k=m-N}^m |a_k|.\end{aligned}$$

Notice that $\Sigma_{k=m-N}^m |a_k|$ converges to zero as $m \to \infty$ by the Cauchy's criterion. This shows that for large m, this term is bounded by

$$\epsilon \cdot (\Sigma_{k=0}^\infty |a_k| + \max\{|\beta_0|, \cdots, |\beta_N|\})$$

and hence it has zero as limit as $m \to \infty$. □

Regarding multiplication of series, Abel has the following result and we will prove later using the continuity of power series.

THEOREM 2.23. *[By Abel] If* $\Sigma a_n = A$*,* $\Sigma b_n = B$ *and* $\Sigma c_n = C$*, with*

$$c_n := \Sigma_{k=0}^n a_k b_{n-k},$$

then $C = AB$.

2.5. Rearragnement. Assume $\Sigma_{n=1}^{\infty} x_n$ is a series. Given any bijective map $r : \mathbb{Z}^+ \to \mathbb{Z}^+$, then $\Sigma_{n=1}^{\infty} x_{r(n)}$ is called a rearrangement of $\Sigma_{n=1}^{\infty} x_n$.

EXAMPLE 2.24. Consider the alternating convergent series

$$\Sigma_{n=1}^{\infty} \frac{(-1)^n}{n} = -1 + \frac{1}{2} - \frac{1}{3} + \frac{1}{4} - \frac{1}{5} + \cdots .$$

We know it is convergent but not absolutely convergent. Denote

$$\alpha := \Sigma_{n=1}^{\infty} \frac{(-1)^n}{n}.$$

Then define a rearrangement as

$$\Sigma_{n=1}^{\infty} x_n' := -1 - \frac{1}{3} + \frac{1}{2} - \frac{1}{5} - \frac{1}{7} + \frac{1}{4} - \frac{1}{9} - \frac{1}{11} + \frac{1}{6} - \cdots .$$

One can prove that this rearrangement is also convergent, but to a different number $\beta \neq \alpha$.

In fact, Riemann proved the following general statement.

THEOREM 2.25. *[By Riemann] Assume Σx_n is convergent but not absolutely convergent. Take any $a \leq b$ with a, b possibly being $\pm\infty$. Then there exists a rearrangement $\Sigma_n x_{r(n)}$ so that the partial sums $S_n' := \Sigma_{k=1}^{n} x_{r(n)}$ satisfies*

$$\liminf S_n' = a, \quad \limsup S_n' = b.$$

In particular, this shows that given any $a \in \overline{\mathbb{R}}$, there exists a rearrangement $\Sigma_n x_{r(n)} = a$.

PROOF. See Rudin Theorem 3.54's proof. □

THEOREM 2.26. *Assume Σx_n is absolutely convergent. Then every rearrangement converges absolutely to the same limit.*

PROOF. Since Σx_n is absolutely convergent, for any given $\epsilon > 0$, there exists some $N \in \mathbb{Z}^+$ so that any $n > N$, $p \geq 0$,

$$\Sigma_{k=n}^{n+p} |x_k| < \epsilon.$$

Now consider an arbitrary rearrangement $\Sigma x_{r(n)}$.

Take N' so that for any $n > N'$, $r(n) \geq N + 1$. This is possible since we can take N' as

$$N' = \max\{r(1), r(2), \cdots, r(N), N + 1\}.$$

Then for any $n > N'$, we have

$$\Sigma_{k=n}^{n+p} |x_{r(n)}| \leq \Sigma_{k=N+1}^{N+1+p'} |x_k| < \epsilon,$$

where p' is some nonnegative number so that

$$\{x_{N+1}, x_{N+2}, \cdots, x_{N+1+p'}\}$$

include all numbers

$$\{x_{r(n)}, x_{r(n+1)}, \cdots, x_{r(n+p)}\}.$$

This shows that $\Sigma x_{r(n)}$ converges absolutely.

The details that they converge to the same limit are left to you.

□

CHAPTER 4

Continuity

1. Limits of functions

Assume (X, d_X) is metric space and $U \subset X$ is a subset of X. Then the metric d_X induces a metric on U. We now consider another metric space (Y, d_Y). A map $f : U \to Y$ is also called a function over U with values in Y. In particular, if $Y = \mathbb{R}$, then f is called a real-valued function; and if $Y = \mathbb{C}$, f is called a complex-valued function.

DEFINITION 1.1. Consider a limit point $x_0 \in U'$ and a point $y_0 \in Y$. We say the limit of the function $f(x)$ at x_0 is y_0, denoted as

$$\lim_{x \to x_0} f(x) = y_0 \quad \text{or} \quad f(x) \to y_0 \text{ as } x \to x_0,$$

if for any $\epsilon > 0$, there exists some $\delta > 0$ so that any $x \in U$ with $0 < d_X(x, x_0) < \delta$, there is

$$d_Y(f(x), y_0) < \epsilon.$$

If there is no $y_0 \in Y$ so that $\lim_{x \to x_0} f(x) = y_0$, then we say the limit of $f(x)$ at x_0 doesn't exist.

We make following important remarks for the definition.

REMARK 1.2.

(1) We require $x_0 \in U'$ because this guarantees that there are always some points $x \in U$ with $0 < d_X(x, x_0) < \delta$.
(2) When we take both (X, d_X) and (Y, d_Y) as $(\mathbb{R}, |\cdot|)$, this is just the case we have learnt from calculus, which says $\lim_{x \to x_0} f(x) = y_0$, if for any $\epsilon > 0$, there is some $\delta > 0$ so that any $0 < |x - x_0| < \delta$, we have $|f(x) - y_0| < \epsilon$.
(3) Limit is a local property in the sense that $\lim_{x \to x_0} f(x)$ only needs the information of f on a (small) neighborhood of x_0. Also, the limit $\lim_{x \to x_0} f(x)$ doesn't provide information of $f(x)$ at the point x_0. In fact, $f(x_0)$ may not be defined at all.
(4) We can use $\epsilon - \delta$ language to describe $\lim_{x \to x_0} f(x) \neq y_0$ as follows: There exists some $\epsilon_0 > 0$ so that any $\delta > 0$, there exists some $x \in U$ with $0 < d_X(x, x_0) < \delta$, but

$$d_Y(f(x), y_0) \geq \epsilon_0.$$

This is equivalent to the following sequence interpretation: There exists some $\epsilon_0 > 0$ and a sequence $\{x_n\}$ in $U \setminus \{x_0\}$ that converges to x_0 so that

$$d_Y(f(x_n), y_0) \geq \epsilon_0, \quad n = 1, 2, \cdots.$$

From the above remark (4), we obtain the following sequence interpretation for limits of functions.

PROPOSITION 1.3.

(1) $\lim_{x \to x_0} f(x) = y_0$ *if and only if any sequence* $\{x_n\}$ *in* $U \setminus \{x_0\}$ *that converges to* x_0 *has* $f(x_n) \to y_0$.
(2) $\lim_{x \to x_0} f(x) \neq y_0$ *if and only if there exists some sequence* $\{x_n\}$ *in* $U \setminus \{x_0\}$ *that converges to* x_0 *so that* $f(x_n)$ *doesn't converge to* y_0.

So, use the properties for limits of sequences we can obtain some properties for limits of functions easily.

PROPOSITION 1.4. *If* $\lim_{x\to x_0} f(x)$ *exists, then it is unique.*

PROOF.
(1) (Use definition directly.) Assume $\lim_{x\to x_0} = y_i$, $i = 1, 2$. For any $\epsilon > 0$, there exists some $\delta_i > 0$ so that any $x \in U$ with $0 < d_X(x, x_0) < \delta_i$, we have $d_Y(f(x), y_i) < \frac{\epsilon}{2}$, $i = 1, 2$. Then
$$d_Y(y_1, y_2) \le d_Y(y_1, f(x)) + d_Y(f(x), y_2) < \epsilon$$
for any $\epsilon > 0$. Hence $d_Y(y_1, y_2)$ and so $y_1 = y_2$.
(2) (Use sequences.) Assume $\lim_{x\to x_0} f(x) = y_i$, $i = 1, 2$. Then for any sequence $\{x_n\}$ in $U \setminus \{x_0\}$ that converges to x_0, there is $f(x_n) \to y_i$. By the uniqueness of limits of sequences, $y_1 = y_2$.

□

PROPOSITION 1.5. *If* $Y = \mathbb{R}$ *and* $\lim_{x\to x_0} f(x) = A$, $\lim_{x\to x_0} g(x) = B$, *then*
(1) $\lim_{x\to x_0}(f \pm g)(x) = A \pm B$;
(2) $\lim_{x\to x_0}(f \cdot g)(x) = A \cdot B$;
(3) $\lim_{x\to x_0} \frac{f}{g}(x) = \frac{A}{B}$, *if* $B \neq 0$.
(4) *For any* $c \in \mathbb{R}$, $\lim_{x\to x_0}(cf)(x) = cA$.

PROOF. We prove (1) here and others are left to you as exercises. Take any sequence $\{x_n\}$ in $U\setminus\{x_0\}$ that converges to x_0, we have
$$f(x_n) \to A, \quad \text{and} \quad g(x_n) \to B.$$
Then by the addition/subtraction property for sequences,
$$(f \pm g)(x_n) = f(x_n) \pm g(x_n) \to A \pm B.$$
This shows that $\lim_{x\to x_0}(f \pm g)(x) = A \pm B$. □

EXAMPLE 1.6. Consider the functions $f : \mathbb{R} \to \mathbb{R}$.
(1) $\lim_{x\to x_0} x = x_0$. Use it, we have the limits for any rational functions based on Proposition 1.5 as following examples.
(2) Assume $f : \mathbb{R} \to \mathbb{R}$ is a polynomial function, i.e.,
$$f(x) = a_0 + a_1x + a_2x^2 + \cdots a_nx^n, \quad a_0, a_1, \cdots, a_n \in \mathbb{R}.$$
Then for any $x_0 \in \mathbb{R}$,
$$\lim_{x\to x_0} f(x) = a_0 + a_1x_0 + a_2x_0^2 + \cdots a_nx_0^n = f(x_0).$$
(3) Assume $f, g : \mathbb{R} \to \mathbb{R}$ are two polynomial functions and $D = \{x \in \mathbb{R} | g(x) \neq 0\}$. Then take any $x_0 \in D$,
$$\lim_{x\to x_0} \frac{f(x)}{g(x)} = \frac{f(x_0)}{g(x_0)}.$$
This is due to the quotient property of limits of functions.

EXAMPLE 1.7. We know the series $1 + \Sigma_{n=1}^{\infty} \frac{x^n}{n!}$ is convergent for all $x \in \mathbb{R}$. Define it as e^x. Prove that $\lim_{x\to x_0} e^x = e^{x_0}$.

PROOF. (This is assigned as a homework problem and here is some hint.)

For each n, the partial sum $S_n(x) := 1+\Sigma_{k=1}^{n}\frac{x^k}{k!}$ is a polynomial in x of order n, and then by previous example, we know

$$\lim_{x\to x_0} S_n(x) = S_n(x_0).$$

Notice: Though we have

$$e^{x_0} = \lim_{n\to\infty} S_n(x_0) = \lim_{n\to\infty}\lim_{x\to x_0} S_n(x),$$

in general, we can not directly switch the orders of two limits on RHS and obtain

$$\lim_{n\to\infty}\lim_{x\to x_0} S_n(x) = \lim_{x\to x_0}\lim_{n\to\infty} S_n(x)(=\lim_{x\to x_0} e^x).$$

To prove we can switch orders of these two limits, we look at

$$\begin{aligned} |e^x - e^{x_0}| &= |(e^x - S_n(x)) + (S_n(x) - S_n(x_0)) + (S_n(x_0) - e^{x_0})| \\ &\le |e^x - S_n(x)| + |S_n(x) - S_n(x_0)| + |S_n(x_0) - e^{x_0}|. \end{aligned}$$

Here the thing you need take case is to show that $|e^x - S_n(x)|$ is small for all $n > N$ for some large N which is INDEPENDENT of x (whenever x is close to x_0).

□

Similarly, define

$$\sin x := x - \frac{x^3}{3!} + \frac{x^5}{5!} + \frac{x^7}{7!} - \cdots,$$

and

$$\cos x := 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots.$$

We can easily prove that both series are convergent for any $x \in \mathbb{R}$, so they define functions over x. Moreover, similarly as the above proof for e^x, we can prove that

$$\lim_{x\to x_0} \sin x = \sin x_0, \quad \lim_{x\to x_0} \cos x = \cos x_0.$$

for any $x_0 \in \mathbb{R}$.

REMARK 1.8. Using series, we can see the Euler's formula

$$e^{ix} = \cos x + i\sin x,$$

and all numerical properties of sine and cosine functions. You are invited to think how these work.

2. Continuous functions

Consider $(X, d_X), (Y, d_Y)$ are two metric spaces. U is a subset of X.

DEFINITION 2.1. Call $f : U \to Y$ is continuous at $x_0 \in U$, if either

(1) x_0 is an isolated point of U, i.e., $x \in U \setminus U'$; or
(2) $x_0 \in U \cap U'$ and $\lim_{x\to x_0} f(x) = f(x_0)$.

If f is not continuous at x_0, we say f is discontinous at x_0.

EXAMPLE 2.2. (1) Polynomial functions on $\mathbb{R}$ are continuous everywhere.
(2) Rational functions on $\mathbb{R}$ are continuous on its natural domain.
(3) e^x, $\sin x$, $\cos x$ on $\mathbb{R}$ are continuous everywhere.
(4) x^a, $a > 0$ is continuous on $[0, +\infty) \subset \mathbb{R}$.

When we consider real-valued functions, we have the following properties immediately from Proposition 1.5.

PROPOSITION 2.3. *Assume f, g are two functions from U to Y, and both are continuous at $x_0 \in U$. Then*

(1) $f \pm g$
(2) $f \cdot g$
(3) $\frac{f}{g}$, *assuming that* $g(x_0) \neq 0$
(4) $c \cdot f$, *for any* $c \in \mathbb{R}$

are all continuous at x_0.

PROOF. We prove (1) as an example. Others are left to you. If x_0 is an isolated point, $f + g$ is continuous at x_0 from definition. If x_0 is a limit point, then

$$\begin{aligned}\lim_{x\to x_0}(f+g)(x) &= \lim_{x\to x_0} f(x) + \lim_{x\to x_0} g(x)\\ &= f(x_0) + g(x_0)\\ &= (f+g)(x_0).\end{aligned}$$

□

The continuity has the following equivalent definition.

PROPOSITION 2.4. *f is continuous at x_0 if and only if for any $\epsilon > 0$, there exists some $\delta > 0$ so that any $x \in U$ with $d_X(x, x_0) < \delta$, there is*

$$d_Y(f(x), f(x_0)) < \epsilon.$$

(This is equivalent to say $f(B_\delta(x_0) \cap U) \subset B_{f(x_0)}(\epsilon)$.)

PROOF. We prove the 'only if ' part, and the 'if' part is exactly the same.

(1) If x_0 is an isolated point, then there exists some $\delta > 0$ so that

$$B_{x_0}(\delta) \cap U = \{x_0\}.$$

Then for any $\epsilon > 0$, take δ as the one picked. It follows

$$f(B_{x_0}(\delta) \cap U) = f(\{x_0\}) = \{f(x_0)\} \subset B_{f(x_0)}(\epsilon).$$

(2) If x_0 is a limit point, then this is just the $\epsilon - \delta$ statement for $\lim_{x\to x_0} f(x) = f(x_0)$.

□

Using this proposition, it is easy to prove the following result.

PROPOSITION 2.5. *Consider f is a function from metric space (X, d_X) to metric space (Y, d_Y), and g is a function from the metric space (Y, d_Y) to a metric space (Z, d_Z). For a point $x_0 \in X$, if f is continuous at x_0 and g is continuous at $f(x_0) \in Y$, then $g \circ f$ as a function from X to Z is continuous at x_0.*

PROOF. For any $\epsilon > 0$, since g is continuous at $f(x_0)$, there exists some $\delta_1 > 0$ so that

$$g(B_{\delta_1}(f(x_0))) \subset B_{g(f(x_0))}(\epsilon).$$

Then use the assumption that f is continuous at x_0, there exists some $\delta > 0$ so that

$$f(B_\delta(x_0)) \subset B_{f(x_0)}(\delta_1).$$

It follows

$$g(f(B_\delta(x_0))) \subset g(B_{f(x_0)}(\delta_1)) \subset B_{g(f(x_0))}(\epsilon),$$

which is saying that $g \circ f$ is continuous at x_0. $\square$

COROLLARY 2.6. *The addition, subtraction, multiplication, division, composition of polynomial functions, exponential functions, sine, cosine functions, x^α (with $\alpha > 0$), are all continuous on their natural domains.*

For functions defined on intervals of $\mathbb{R}$, we can discuss different types of discontinuous points.

DEFINITION 2.7. Assume x_0 is a discontinuous point for a function defined on an interval of $\mathbb{R}$.

(1) Call x_0 is of the first kind of discontinuous point, if both $\lim_{x \to x_0 \pm} f(x)$ exist. In particular, if they are the same, x_0 is called a removable discontinuous point.
(2) Call x_0 is of the second kind of discontinuous point, if at least one of $\lim_{x \to x_0 \pm} f(x)$ doesn't exist.

EXAMPLE 2.8. (1)

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

is discontinuous at $x = 0$ and it is first kind discontinuous point.

(2)

$$f(x) = \begin{cases} 0 & x = 0 \\ 1 & x \neq 0 \end{cases}$$

is discontinuous at $x = 0$ and it is removable.

(3)

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

is discontinuous everywhere and every point in $\mathbb{R}$ is of second kind.

(4)

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

is (of second kind) discontinuous everywhere except at 0. It is continuous at $x = 0$.

(5)

$$f(x) = \begin{cases} \sin \frac{1}{x} & x \neq 0 \\ 0 & x = 0 \end{cases}$$

is (of second kind) discontinuous at $x = 0$.

PROPOSITION 2.9. *Assume f is a function from metric space (X, d_X) to metric space (Y, d_Y).*

(1) *f is continuous on X if and only if the preimage of every open subset $V \subset Y$ is an open subset of X.*
(2) *f is continuous on X if and only if the preimage of every closed subset $V \subset Y$ is a closed subset of X.*

PROOF. We prove (1), and (2) is left to you as exercise.

Assume f is continuous. Take $V \subset Y$ as an open subset. We prove $f^{-1}(V)$ is open. For this, take any point $x \in f^{-1}(V)$, then $f(x) \in V$. Since V is open, there exists some $\epsilon > 0$ so that the open ball $B_\epsilon(f(x)) \subset V$.

Now use the assumption that f is continuous at x, we can find some $\delta > 0$ so that

$$f(B_\delta(x)) \subset B_\epsilon(f(x)),$$

which is saying that $B_\delta(x) \subset f^{-1}(V)$, and this shows that f^V is open.

Conversely, assuming the preimage of every open subset $V \subset Y$ is an open subset of X, then in particular, for any $\epsilon > 0$, $f^{-1}(B_\epsilon(f(x)))$ is an open subset of X. Since $x \in f^{-1}(B_\epsilon(f(x)))$, there exists some $\delta > 0$ so that

$$f(B_\delta(x)) \subset B_\epsilon(f(x)),$$

which is saying that f is continuous at x. □

We remark that, in general for a topological space which is not necessarily a metric space, the above proposition is in fact the definition of a continuous map.

3. Continuity and compactness

The same as before, we assume (X, d_X) and (Y, d_Y) are two metric spaces.

THEOREM 3.1. *Assume $f : X \to Y$ is a continuous map. Then for any compact subset $K \subset X$, the image set $f(K)$ is a compact subset of Y.*

PROOF. We prove it by definition. Assume $\{V_\alpha | \alpha \in \Lambda\}$ is an open cover of $f(K)$. By the continuity of f and Proposition 2.9, $\{f^{-1}(V_\alpha) | \alpha \in \Lambda\}$ is an open cover of K. Because K is compact, it must have a finite subcover, which we denote by

$$\{f^{-1}(V_1), f^{-1}(V_2), \cdots, f^{-1}(V_n)\}.$$

Then if follows

$$\{V_1, V_2, \cdots, V_n\}$$

is a finite subcover of $\{V_\alpha | \alpha \in \Lambda\}$, and this shows that $f(K)$ is compact. □

Now we give an alternative proof, which shows that the image of any sequentially compact under a continuous map is sequentially compact. For metric spaces, since compactness is equivalent to sequential compactness, this is just an alternative proof. However, for general topological spaces, compactness and sequential compactness are independent and this is an independent result then.

ALTERNATIVE PROOF. Assume K is sequentially compact, we prove now $f(K)$ is also sequentially compact using definition.

Take an arbitrary sequence $\{f(x_n) | x_n \in K\}$ in $f(K)$, we just need to see it has a convergent subsequence in $f(K)$. For this, notice that $\{x_n\}$ as a sequence in K has a convergent subsequence

$$x_{n_k} \to x_0 \in K$$

since K is sequentially compact. Then by the continuity of f, it follows

$$f(x_{n_k}) \to f(x_0) \in f(K),$$

and we are done.

□

The converse of the statement is not true in general.

DEFINITION 3.2. A function $f : X \to Y$ is called proper if the preimage of any compact subset of Y is compact in X.

EXAMPLE 3.3. Not every continuous function is proper. Consider $f(x) = \frac{1}{x}$ from $(0, +\infty)$ to $\mathbb{R}$. The preimage of $[0, 1]$ is $[1, +\infty)$ which is not compact.

An immediate corollary from the above result and Proposition 3.9 is the following

COROLLARY 3.4. *Assume $f : X \to Y$ is a continuous map. Then for any compact subset $K \subset X$, the image set $f(K)$ is bounded and closed in Y.*

In particular, when $Y = \mathbb{R}$, we have the following important result.

THEOREM 3.5. *A real-valued continuous function defined on a compact subset of a metric space can obtain its supremum and infimum, i.e., has maximum and minimum.*

PROOF. Assume $f : K \to \mathbb{R}$ is a continuous function and Y is a compact subset of X. From Corollary 3.4, we know $f(K)$ is a bounded and closed subset of $\mathbb{R}$.

Then by the completeness of $\mathbb{R}$, $\sup f(K), \inf f(K) \in \mathbb{R}$. Since $\sup f(K), \inf f(K)$ are limit points of $f(K)$, further by the closeness of $f(K)$, they live in $f(K)$. This is saying that there are points $x_1, x_2 \in K$ so that

$$f(x_1) = \sup f(K), \quad f(x_2) = \inf f(K).$$

□

EXAMPLE 3.6. The assumption of compactness is important. For example $f(x) = \frac{1}{x}$ has neither max nor min over $(0, +\infty) \subset \mathbb{R}$, but it has both over any closed interval $[a, b] \subset (0, +\infty)$, for $a \le b$.

Next, we consider the continuity of the inverse function of a continuous function. We know that if f is an injective function from X to Y, then we can define the inverse function

$$f^{-1} : f(X) \to X,$$

which maps $f(x)$ back to x. Here comes the question that if f is continuous, then is f^{-1} continuous on $f(X)$? The answer is no in general.

EXAMPLE 3.7. Consider

$$f(x) = \begin{cases} 1 - x & 0 \le x < 1 \\ 2 - x & 2 \le x \le 3. \end{cases}$$

The inverse function is

$$f(x) = \begin{cases} 2 - x & -1 \le x \le 0 \\ 1 - x & 0 < x \le 1. \end{cases}$$

It is NOT continuous at $x = 0$.

In fact, we can prove the following result.

THEOREM 3.8. *Assume f is an bijective function from X to Y and f is continuous. Then if X is compact, its inverse is also continuous.*

PROOF. We give two proofs. The first one uses open sets interpretation and the second proof uses sequence interpretation.

(1) This is enough to show f is an open map, i.e., f maps any open subset U of X to an open set of Y.

To show $f(U)$ is open, it is enough to show $f(U)^c$ is closed. Notice that

$$f(U)^c = f(U^c).$$

Since U^c is closed subset of X, which is compact, U^c is compact, and hence $f(U^c)$ is also compact by Theorem 3.1. Then $f(U)^c$ is closed and we are done.

(2) Taking an arbitrary sequence $\{f(x_n) | x_n \in X\}$ in Y that converges to some $f(x_0)$, we prove that the sequence $\{f^{-1}(f(x_n)) = x_n\}$ converges to x_0.

Assume this is not the case, then there exists some $\epsilon_0 > 0$ and a subsequence $\{x_{n_k}\}$ so that

$$d_X(x_{n_k}, x_0) \geq \epsilon_0.$$

Assuming X is sequentially compact, we can further find a subsequence $\{x_{n_{k_\ell}}\}$ so that it converges to some $x_0' \in X$, and

$$d_X(x_0', x_0) \geq \epsilon_0.$$

At the same time, by the continuity of f, $\{f(x_{n_{k_\ell}})\}$ converges to $f(x_0')$. By the uniqueness of limit, there must be

$$f(x_0) = f(x_0'),$$

and thus $x_0 = x_0'$, which contradicts with $d_X(x_0', x_0) \geq \epsilon_0$.

□

REMARK 3.9. Two metric spaces X and Y are called homeomorphic, if there exists some bijective map $f : X \to Y$ with both f and f^{-1} continuous. The above result shows that assuming X is compact, for a homeomorphism, it is enough to check f is continuous.

The last property we want to introduce for continuous functions defined on a compact space is about continuity itself.

DEFINITION 3.10. A function $f : X \to Y$ is called uniformly continuous, if for any given $\epsilon > 0$, there exists some $\delta > 0$ so that any $x_1, x_2 \in X$ with $d_X(x_1, x_2) < \delta$, there is

$$d_Y(f(x_1), f(x_2)) < \epsilon.$$

If we fix x_1, then the definition is stating that f is continuous at x_1, so we immediately get the following result.

PROPOSITION 3.11. *If $f : X \to Y$ is uniformly continuous, then it is continuous.*

The converse is not true in general. Consider the function

$$f(x) = \frac{1}{x}, \quad x \in (0, +\infty)$$

is continuous but not uniformly continuous. This can be seen from the equalities that

$$|\frac{1}{x_1} - \frac{1}{x_2}| = \frac{|x_1 - x_2|}{|x_1 x_2|},$$

from which we see that even one requires $|x_1 - x_2| < \delta$, $|\frac{1}{x_1} - \frac{1}{x_2}|$ can be arbitrarily big since $|x_1 x_2|$ can be arbitrarily small.

On the other hand, if we restrict the domain to a closed interval, e.g., $[1, 2]$, the above situation that $|x_1 x_2|$ can be arbitrarily small will not happen anymore. This indicates the following general statement.

THEOREM 3.12. *For a function defined on a compact domain, it is continuous if and only if it is uniformly continuous.*

PROOF. We only need to prove that under the assumption of compactness, continuity implies uniform continuity. Again we provide two proofs, one is using compactness, the other uses sequential compactness.

(1) Since f is continuous everywhere on X, for any given $\epsilon > 0$, for each $x \in X$, there exists some $\delta(x) > 0$ (may DEPEND on x) so that

$$f(B_{\delta(x)}(x)) \subset B_\epsilon(f(x)).$$

Now notice that

$$\{B_{\frac{1}{2}\delta(x)}(x) | x \in X\}$$

form an open cover of X (here to shrink $\delta(x)$ to $\frac{1}{2}\delta(x)$ is technically important.) and X is compact, we can pick finite points $x_1, \cdots, x_n$ from X so that the balls

$$B_{\frac{1}{2}\delta(x_1)}(x_1), B_{\frac{1}{2}\delta(x_2)}(x_2), \cdots, B_{\frac{1}{2}\delta(x_n)}(x_n)$$

cover X. Define

$$\delta := \min\{\frac{1}{2}\delta(x_1), \cdots, \frac{1}{2}\delta(x_n)\}.$$

Now for any two points $p_1, p_2 \in X$ with distance

$$d_X(p_1, p_2) < \delta,$$

we estimate $d_Y(f(p_1), f(p_2))$.

First, p_1 must live in a ball $B_{\frac{1}{2}\delta(x_i)}(x_i)$ for some $i = 1, \cdots, n$. We claim that $p_2 \in B_{\delta(x_i)}(x_i)$. This is because

$$\begin{aligned} & d_X(x_i, p_2) \\ \leq\ & d_X(x_i, p_1) + d_X(p_1, p_2) \\ \leq\ & \frac{1}{2}\delta(x_i) + \delta \\ \leq\ & \frac{1}{2}\delta(x_i) + \frac{1}{2}\delta(x_i) = \delta(x_i). \end{aligned}$$

Then

$$\begin{aligned} & d_Y(f(p_1), f(p_2)) \\ \leq \ & d_Y(f(p_1), f(x_i)) + d_Y(f(x_i), f(p_2)) \\ \leq \ & \epsilon + \epsilon = 2\epsilon. \end{aligned}$$

This proves f is uniformly continuous.

(2) We now give a proof using sequences.

Assume f is not uniformly continuous. Then there exists some $\epsilon > 0$ and a pair of sequences $\{x_n\}, \{x'_n\}$ with

$$d_X(x_n, x'_n) < \frac{1}{n}$$

but

$$d_Y(f(x_n), f(x'_n)) \geq \epsilon_0.$$

Because X is compact, after passing to subsequences, we can assume $x_n \to x_0$ and $x'_n \to x'_0$ for some $x_0, x'_0 \in X$. (I abuse notations here that I still use $\{x_n\}$ to denote a subsequence of it, but this way is very commonly used in literatures.) By the continuity of the distance function $d_X : X \times X \to \mathbb{R}$ (This is supposed to be proved in your homework.) implies

$$d_X(x_0, x'_0) \leq \lim_{n\to\infty} \frac{1}{n} = 0, \quad \text{thus, } x_0 = x'_0.$$

On the other hand, apply the continuity of f and the distance function $d_Y : Y \times Y \to \mathbb{R}$, it follows

$$d_Y(f(x_0), f(x'_0)) \geq \epsilon_0,$$

which then contradicts with $x_0 = x'_0$.

This proves f must be uniformly continuous.

□

4. Continuity and connectedness

DEFINITION 4.1. A metric space X is called connected, if any subset S of X if it is both open and closed, then it is either X or the empty set.

PROPOSITION 4.2. *X is connected, if and only $X = U \cup V$, $U \cap V = \emptyset$, with both U, V open implies that either U or V is empty set.*

PROOF. If X is connected, and $X = U \cup V$ with both U, V open, then $U = V^c$ is both open and closed. If $U \neq \emptyset$, then $U = X$, and then $V = U^c = \emptyset$.

Conversely, take any $S \subset X$, we can write $X = S \cup S^c$. Assume S is open and closed, then we must have S or S^c is empty, which implies S is either empty or X. □

EXAMPLE 4.3. One ball in $\mathbb{R}^n$ is connected. Two disjoint balls in $\mathbb{R}^n$ forms a disconnected metric space. (We are going to see a proof later.)

THEOREM 4.4. *If $f : X \to Y$ is a continuous function and X is connected, then $f(X)$ is connected as a metric space with the metric from Y.*

PROOF. Assume $f(Y) = U \cup V$, with both U, V are open and $U \cap V = \emptyset$. Then since f is continuous, $f^{-1}(U)$ and $f^{-1}(V)$ are both open and have no intersection. Moreover,

$$f^{-1}(U) \cup f^{-1}(V) = X.$$

Then because X is connected, either $f^{-1}(U)$ or $f^{-1}(V)$ is empty. This proves that U or V is empty and hence $f(X)$ is connected.

□

EXAMPLE 4.5. The subset $\mathbb{R}$ is connected, but $\mathbb{R} \setminus \{0\}$ is not connected. Intuitionally, a set with a 'gap' is not connected. However, the following example may obey your intuition. Consider the subset

$$S := \{(x, \sin \frac{1}{x}) | x \neq 0\} \cup (\{0\} \times [0, 1]).$$

Exercise: Prove S is connected as a metric space with the metric from $\mathbb{R}^2$. (This is named topologists' sine function.)

To exclude such pathological cases, we introduce another definition to describe connectedness.

DEFINITION 4.6. A metric space X is called path-connected, if any two points $x_0, x_1 \in X$, there is a continuous function $f : [0, 1] \to X$ with $f(0) = x_0$, $f(1) = x_1$.

THEOREM 4.7. $[0, 1] \subset \mathbb{R}$ *is both connected and path-connected.*

PROOF. It is clear from definition that $[0, 1]$ is path-connected. We prove it is also connected.

Assume $[0, 1] = A \cup B$ with both A and B are open in $[0, 1]$ and A, B are disjoint. Then we can pick $a \in A$ and $b \in B$, and WLOG, assume $a < b$. Consider the interval $[a, b] \subset [0, 1]$.

Denote by $x_0 := \sup([a, b] \cap A)$. Such x_0 exists and lives in $[0, 1]$, but x_0 can not be in A because A is open. (Why?) On the other hand, x_0 can not be in B because B is open. So either A or B is empty then, which shows $[0, 1]$ is connected. □

Using this, we now prove

THEOREM 4.8. *Any path-connected metric space must be connected.*

PROOF. Assume X is a path-connected metric space, we prove it must be connected. For this, assume

$$X = U \cup V$$

with both U, V open and U, V are disjoint, then if neither is empty, we pick $x_0 \in U$ and $x_1 \in V$.

Since X is path-connected, there exists a continuous function

$$f : [0, 1] \to X, \quad f(0) = x_0, f(1) = x_1.$$

Define $A := f^{-1}(U)$ and $B := f^{-1}(V)$. Both A, B are open in $[0, 1]$ since f is continuous, but neither is empty since

$$0 \in A, \quad 1 \in B.$$

Hence $[0, 1]$ is not connected, which contradicts with the fact we just proved in Theorem 4.7. □

However, the verse vice is not true as we just see from the topologists' sine function.

EXAMPLE 4.9. Any open or closed ball B in $\mathbb{R}^n$ is (path) connected. The proof is left to you. In particular, for any closed interval $[a,b]$, we can construct a continuous path

$$f : [0,1] \to [a,b], \quad f(t) = (1-t)a + tb.$$

In fact, such f is a bijection, and the inverse map is also continuous. In another word, any nonempty finite closed interval is homeomorphic to $[0,1]$.

Though in general, the path-connectedness is stronger than connectedness, for subsets in $\mathbb{R}$, the connectedness and path-connectedness are exactly the same thing.

THEOREM 4.10. *In $\mathbb{R}$, a subset is connected if and only it is path-connected.*

PROOF. We only need to prove any connected subset of $\mathbb{R}$ is path-connected. Take $S \subset \mathbb{R}$ which is connected. For any two points $x_0, x_1 \in S$, we show that $[x_0, x_1] \subset S$.

If this is not the case, there exists some $x \in [x_0, x_1]$ but not in S. Then define

$$A := S \cap (-\infty, x), \quad B := S \cap (x, +\infty).$$

They are disjoint and both are open in S. Moreover, $S = A \cup B$, but neither A nor B is empty. This contradicts with the connectedness of S. □

Again, we have seen from topologists' sine function that in R^2, this is not the case.

THEOREM 4.11. *[Intermediate value theorem]Assume $[a,b] \subset \mathbb{R}$ is a closed interval, and $f : [a,b] \to \mathbb{R}$ is a continuous function. If $f(a) < f(b)$, then for each*

$$f(a) < y_0 < f(b),$$

there exists some $x_0 \in (a,b)$ so that $f(x_0) = y_0$.

PROOF. Since $[a,b]$ is connected, using Theorem 4.4, $f([a,b])$ is connected. Moreover, $f(a), f(b) \in f([a,b]) \subset \mathbb{R}$. From the proof of Theorem 4.10, the closed interval

$$[f(a), f(b)] \subset f([a,b]).$$

Now for any $f(a) < y_0 < f(b)$, there is

$$y_0 \in f([a,b]).$$

It follows there must be some $x_0 \in (a,b)$ with $f(x_0) = y_0$.

□

REMARK 4.12. This condition that f is a continuous on the CLOSED interval $[a,b]$ is essential. If it is only continuous on (a,b), the conclusion no longer holds.

5. Monotonic functions over segments in $\mathbb{R}$

DEFINITION 5.1. A real valued function $f : (a,b) \to \mathbb{R}$ is called

(1) increasing, if any $a < x_1 \le x_2 < b$, there is $f(x_1) \le f(x_2)$.
(2) decreasing, if any $a < x_1 \le x_2 < b$, there is $f(x_1) \ge f(x_2)$.
(3) monotonic, if it is increasing or decreasing.

THEOREM 5.2. *Assume $f : (a, b) \to \mathbb{R}$ is an increasing function. Then both*

$$f(x_0+) := \lim_{x \to x_0+} f(x) \quad \textit{and} \quad f(x_0-) := \lim_{x \to x_0-} f(x)$$

exist for any $x_0 \in (a, b)$. More precisely,

$$\sup_{a<t<x_0} f(t) = f(x_0-) \leq f(x_0) \leq f(x_0+) = \inf_{x_0<t<b} f(t).$$

Further more, if $a < x < y < b$, then

$$f(x+) \leq f(y-).$$

PROOF. (1) Take any $x_0 \in (a, b)$. Consider the set $\{f(t) | a < t < x_0\}$. Since f is increasing, it has upper bound $f(x_0)$. Hence it must has the l.u.b., which we denote by

$$A := \sup_{a<t<x_0} f(t).$$

Now given any $\epsilon > 0$, there exists some $t_0 \in (a, x_0)$ so that

$$f(t_0) > A - \epsilon.$$

Notice f is increasing, and so any $t \in [t_0, x_0)$ has

$$A - \epsilon < f(t_0) \leq f(t) \leq A,$$

which exactly states that $f(x_0-) = A$.

For the $f(x_0+)$ part the proof is similar and is left to you as exercise.

Then the inequality

$$\sup_{a<t<x_0} f(t) = f(x_0-) \leq f(x_0) \leq f(x_0+) = \inf_{x_0<t<b} f(t).$$

follows from the monotonicity of the the function f.

(2) Take some number $z \in (x, y)$, from the above together with f is increasing, there is

$$f(x+) = \inf_{x<t<b} f(t) \leq \inf_{x<t<z} f(t) \leq f(z) \leq \sup_{z<t<y} f(t) \leq \sup_{a<t<y} f(t) = f(y-).$$

□

COROLLARY 5.3. *Monotonic functions have no discontinuity of the second kind.*

PROOF. This is because we have proved that for any point $x \in (a, b)$, both $f(x-), f(x+)$ exist. □

THEOREM 5.4. *Assume $f : (a, b) \to \mathbb{R}$ is monotonic. Then the set of discontinuous points is at most countable.*

PROOF. WLOG, assuming f is increasing. For each discontinuous point $x \in (a, b)$, there must be $f(x-) < f(x+)$. We choose some rational number $r_x \in \mathbb{Q}$ so that

$$f(x-) < r_x < f(x+).$$

By this way, we set up an injective (why?) map from discontinuous points in (a, b) to $\mathbb{Q}$, and so the set of discontinuous points must be at most countable.

□

CHAPTER 5

Differentiation

We focus on real valued functions defined on open or closed intervals.

1. The derivative of a real function

DEFINITION 1.1. A function $f : [a, b] \to \mathbb{R}$ is called differentiable at $x_0 \in [a, b]$, if the limit of the function

$$\phi_{x_0}(t) := \frac{f(t) - f(x_0)}{t - x_0}, \quad a < t < b, t \neq x_0$$

exists as $t \to x_0$. For this case, we write

$$f'(x_0) = \lim_{t \to x_0} \phi_{x_0}(t) = \lim_{t \to x_0} \frac{f(t) - f(x_0)}{t - x_0}.$$

The function f is called differentiable over $[a, b]$ if it is differentiable for each $x \in [a, b]$. It induces the function

$$\frac{df}{dx} = f' : [a, b] \to \mathbb{R},$$

which is called the derivative of f.

PROPOSITION 1.2. *If $f : [a, b] \to \mathbb{R}$ is differentiable at $x_0 \in [a, b]$, then it must be continuous at x_0.*

PROOF. We have

$$|f(x) - f(x_0)| = \frac{|f(x) - f(x_0)|}{|x - x_0|} \cdot |x - x_0|, \quad a \leq x \leq b, x \neq x_0.$$

The limit exists and it is zero as $x \to x_0$ when $f'(x_0)$ exists. $\square$

Usually, people use $C^1([a, b])$ to denote the set of differentiable functions over $[a, b]$ whose derivative is continuous. More general, people use $C^k([a, b])$ to denote the set of functions whose kth ordered derivative is continuous. In particular, $C^0([a, b])$ is the set of continuous functions over $[a, b]$.

THEOREM 1.3. *Suppose $f, g : [a, b] \to \mathbb{R}$ are differentiable at $x_0 \in [a, b]$. Then $f \pm g$, fg and f/g (when $g(x_0) \neq 0$) are differentiable at x_0. Moreover,*

(1) $(f \pm g)'(x_0) = f'(x_0) \pm g'(x_0)$;
(2) $(fg)'(x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0)$;
(3) $(f/g)'(x_0) = (f'(x_0)g(x_0) - f(x_0)g'(x_0))/g(x_0)^2$.

PROOF. We take (2) as an example.

We calculate

$$\begin{aligned}
\frac{(f(x)g(x) - f(x_0)g(x_0)}{x - x_0} &= \frac{(f(x) - f(x_0))g(x) + f(x_0)(g(x) - g(x_0))}{x - x_0} \\
&= \frac{f(x) - f(x_0)}{x - x_0} \cdot g(x) + f(x_0) \cdot \frac{g(x) - g(x_0)}{x - x_0} \\
&\to f'(x_0)g(x_0) + f(x_0)g'(x_0), \quad \text{as } x \to x_0,
\end{aligned}$$

where we use f and g are differentiable at x_0 and Proposition 1.2. $\square$

THEOREM 1.4. *[The chain rule] Let $f : [a,b] \to \mathbb{R}$ be a real-valued function that is differentiable at $x_0 \in [a,b]$. Let g be a real-valued function defined on an interval that contains $f([a,b])$, and g is differentiable at $f(x_0)$. Then, the composition*

$$h(x) := g \circ f(x) := g(f(x)) : [a,b] \to \mathbb{R}$$

is differentiable at x_0, and the derivative at x_0 can be calculated as

$$h'(x_0) = g'(f(x_0))f'(x_0).$$

PROOF. We introduce some useful method for proofs regarding limits.

In general, for a function say $p(x)$, if

$$\lim_{x \to x_0} p(x) = 0,$$

we write $p(x) = o(|x - x_0|)$. Here you can think $o(|x - x_0|)$ denotes a function that defined in a sufficiently small neighborhood of x_0 excluding x_0, which goes to 0 as $|x - x_0|$ goes to 0.

Using this notation, we can write

$$\begin{aligned} h(x) - h(x_0) &= g(f(x)) - g(f(x_0)) \\ &= (g'(f(x_0)) + o(|f(x) - f(x_0)|))(f(x) - f(x_0)) \\ &= (g'(f(x_0)) + o(|f(x) - f(x_0)|))(f'(x_0) + o(|x - x_0|))(x - x_0), \end{aligned}$$

and so

$$\frac{h(x) - h(x_0)}{x - x_0} = (g'(f(x_0)) + o(|f(x) - f(x_0)|))(f'(x_0) + o(|x - x_0|)).$$

By Proposition 1.2, the differentiability of $f(x)$ at x_0 implies $f(x)$ is continuous at x_0, i.e.,

$$|f(x) - f(x_0)| = o(|x - x_0|).$$

We can then write

$$\frac{h(x) - h(x_0)}{x - x_0} = (g'(f(x_0)) + o(|x - x_0|))(f'(x_0) + o(|x - x_0|)).$$

Take limit $x \to x_0$, it follows

$$\begin{aligned} \lim_{x \to x_0} \frac{h(x) - h(x_0)}{x - x_0} &= \lim_{x \to x_0} (g'(f(x_0)) + o(|x - x_0|))(f'(x_0) + o(|x - x_0|)) \\ &= g'(f(x_0))f'(x_0). \end{aligned}$$

$\square$

(For the chain rule, Rudin's book (Thm 5.5) also requires f is continuous over the whole interval, but in fact this is not necessary. We only need continuity at x_0 which can be derived from the differentiability of f at x_0.)

EXAMPLE 1.5. (1)

$$f(x) = \begin{cases} x^2 & x \le 0 \\ 0 & x > 0 \end{cases}$$

The function $f(x)$ is differentiable over $\mathbb{R}$, but its derivative $f'(x)$ is not differentiable but only continuous.

(2) The derivative of a differentiable function may not be continuous.

$$f(x) = \begin{cases} x^2 \sin\frac{1}{x} & x \neq 0 \\ 0 & x = 0 \end{cases}$$

(Later, we will prove that if $f(x)$ is differentiable over $[a, b]$, its derivative $f'(x)$ has no first kind discontinuity.)

(3) A continuous function may not be differentiable.

$$f(x) = \begin{cases} x \sin\frac{1}{x} & x \neq 0 \\ 0 & x = 0 \end{cases}$$

(Regarding continuous but not differentiable functions, a more pathological example is the Weierstrass function, which is continuous everywhere over $\mathbb{R}$ but differentiable nowhere. I didn't mention this example in class, but you can have a look at `https://en.wikipedia.org/wiki/Weierstrass_function` to get some intuition on it.)

2. Mean value theorem

DEFINITION 2.1. Let f be a real valued function defined over a metric space X. We say f has a local maximum at a point $x_0 \in X$, if there is an open ball $B_\delta(x_0)$ for some $\delta > 0$, so that

$$f(x_0) \ge f(x), \quad \text{for any } x \in B_\delta(x_0).$$

We say f has a local minimum at a point $x_0 \in X$, if there is an open ball $B_\delta(x_0)$ for some $\delta > 0$, so that

$$f(x_0) \le f(x), \quad \text{for any } x \in B_\delta(x_0).$$

DEFINITION 2.2. For a function $f : (a, b) \to \mathbb{R}$, a point $x_0 \in [a, b]$ is called a critical point if f is not differentiable at x_0 or $f'(x_0) = 0$.

THEOREM 2.3. *Assume f is defined over $[a, b]$. If f has a local maximum or local minimum at some $x_0 \in (a, b)$, then x_0 is a critical point of f.*

PROOF. If f is not differentiable at x_0, we are done. Assume now f is differentiable at x_0 and x_0 is a local maximum.

Then there exists some $\delta > 0$ so that

$$f(x_0) \ge f(x), \quad \text{for any } x \in B_\delta(x_0).$$

It follows

$$\frac{f(x) - f(x_0)}{x - x_0} \begin{cases} \ge 0 & x_0 - \delta < x < x_0 \\ \le 0 & x_0 < x < x_0 + \delta. \end{cases}$$

Further because $f'(x_0)$ exists, there is

$$f'(x_0-) \geq 0, \quad f'(x_0+) \leq 0,$$

but $f'(x_0-) = f'(x_0+) = f'(x_0)$. Hence, $f'(x_0) = 0$. $\square$

We make the following remark for the above theorem: One can not make the conclusion for $x_0 = a$ or b.

THEOREM 2.4 (Rolle's theorem). *Assume $f(x)$ is continuous over $[a, b]$, differentiable over (a, b), and $f(a) = f(b)$. Then there exists some $x_0 \in (a, b)$ so that*

$$f'(x_0) = 0.$$

PROOF. If $f(x)$ is constant over $[a, b]$, then by definition $f'(x) = 0$ for every $x \in [a, b]$. We are done. Otherwise, there must be some $t_0 \in (a, b)$ so that $f(t_0) \neq f(a) = f(b)$. WLOG, let's assume $f(t_0) > f(a) = f(b)$.

At the same time, because $f(x)$ is continuous over $[a, b]$ which is a compact subset of $\mathbb{R}$, by Theorem 3.5, there exists some $x_0 \in [a, b]$ so that

$$f(x_0) = \max_{[a,b]} f(x).$$

Notice that $f(x_0) \geq f(t_0) > f(a) = f(b)$, x_0 must live in the interior (a, b).

Then apply Theorem 2.3, we obtain $f'(x_0) = 0$.

$\square$

THEOREM 2.5 (The Mean Value Theorem). *If $f(x)$ is continuous over $[a, b]$, differentiable over (a, b), then there exists some $x_0 \in (a, b)$ so that*

$$f(b) - f(a) = f'(x_0)(b - a).$$

PROOF. Consider the function

$$h(x) = f(x) - \frac{f(b) - f(a))}{b - a} \cdot x.$$

It is continuous over $[a, b]$, differentiable over (a, b) and

$$h(a) = h(b).$$

Then From the Rolle's theorem, there exists some $x_0 \in (a, b)$ so that

$$h'(x_0) = f'(x_0) - \frac{f(b) - f(a))}{b - a} = 0,$$

which is equivalent to

$$f(b) - f(a) = f'(x_0)(b - a).$$

$\square$

We can generalize the mean value theorem to the general mean value theorem.

THEOREM 2.6 (The Cauchy's Mean Value Theorem). *Assume f, g are two real-valued functions that are continuous over $[a, b]$ and differentiable over (a, b). Then there exists some $x_0 \in (a, b)$ so that*

$$(f(b) - f(a))g'(x_0) = (g(b) - g(a))f'(x_0).$$

PROOF. (From the geometric meaning, $f(x)$ here plays the role of coordinate x in the mean value theorem. This gives the motivation of constructing the following function h.) Consider

$$h(x) := (f(b) - f(a))g(x) - (g(b) - g(a))f(x).$$

Then $h \in C^0([a, b])$, differentiable over (a, b) and $h(a) = h(b)$. Apply the Rolle's theorem, there exists some $x_0 \in (a, b)$ so that

$$h'(x_0) = (f(b) - f(a))g'(x_0) - (g(b) - g(a))f'(x_0) = 0.$$

We are done. □

An immediately corollary from the mean value theorem is the following.

THEOREM 2.7. *If $f(x)$ is differentiable over (a, b), then*

(1) *$f'(x) \geq 0$ implies $f(x)$ is increasing;*
(2) *$f'(x) \leq 0$ implies $f(x)$ is decreasing;*
(3) *$f'(x) = 0$ implies $f(x)$ is constant.*

PROOF. Take any $a < x_1 \leq x_2 < b$, and apply the mean value theorem over $[x_1, x_2]$. There exists some $x_1 < x_0 < x_2$ so that

$$f(x_2) - f(x_1) = f'(x_0)(x_2 - x_1).$$

Then

(1) if $f' \geq 0$, then $f(x_2) \geq f(x_1)$. This shows f is increasing.
(2) if $f' \leq 0$, then $f(x_2) \leq f(x_1)$. This shows f is decreasing.
(3) if $f' = 0$, then $f(x_2) = f(x_1)$. This shows f is constant.

□

Another application of the mean value theorem is the following about uniform continuity. The proof is left to you.

PROPOSITION 2.8. *Assume $f : (a, b) \to \mathbb{R}$ is differentiable and $f'(x)$ is bounded (which means there exists some $M > 0$ so that $|f'(x)| \leq M$ for any $x \in (a, b)$). Then f is uniformly continuous over (a, b).*

We make a remark that a continuous function defined over a closed interval must be uniformly continuous from Theorem 3.12. Since a differentiable function is continuous, we immediately know that a differentiable function over a closed interval is uniformly continuous.

3. The intermediate value property of derivatives

Recall from Theorem 4.11 that a continuous function over a closed interval can achieve any value between the values of its endpoints. Such property is referred as the intermediate value property.

On the other hand, we have seen from example before that not every derivative function is continuous. However, every derivative function still has the intermediate value property.

THEOREM 3.1 (The Intermediate Value Property for Derivatives). *Assume $f(x)$ is differentiable over $[a, b]$ and $f'(a) < f'(b)$. Then for each $f'(a) < \mu < f'(b)$, there exists some $x_0 \in (a, b)$ so that $f'(x_0) = \mu$.*

Similarly result applies to the case $f'(a) > f'(b)$.

PROOF. We consider the function

$$g(x) = f(x) - \mu x.$$

Clearly, it is differentiable and so continuous over $[a, b]$ and

$$g'(a) = f'(a) - \mu < 0, \quad g'(b) = f'(b) - \mu > 0.$$

Notice that $g'(a) < 0$ implies that there exists some $a < x_1 < b$ so that $g(x_1) < g(a)$; and $g'(b) > 0$ implies that there exists some $a < x_2 < b$ so that $g(x_2) < g(b)$. This says a, b can not be the global minimal of g over $[a, b]$. But such global minimum point exists, say x_0, since g is continuous. It follows $x_0 \in (a, b)$.

Then apply Theorem 2.3, we have

$$g'(x_0) = 0,$$

which is equivalent to $f'(x_0) = \mu$.

$\square$

COROLLARY 3.2. *Assume $f(x)$ is differentiable over $[a, b]$. Then $f'(x)$ can only have second kind of discontinuous points on $[a, b]$.*

PROOF. Assume $f'(x)$ is discontinuous at $x_0 \in [a, b]$ with both $f'(x_0-), f'(x_0+)$ exist but Either

(1) $f'(x_0-) \neq f'(x_0+)$; or
(2) $f'(x_0-) = f'(x_0+) \neq f'(x_0)$.

For the case (1), WLOG, let's assume

$$f'(x_0-) < f'(x_0+), \quad \text{and write } \ell := f'(x_0+) - f'(x_0-).$$

Then there exists some $\delta > 0$ so that

- for any $x_0 - \delta < x < x_0$, there is
$$f'(x) < f'(x_0-) + \frac{1}{3}\ell.$$
- for any $x_0 < x < x_0 + \delta$,
$$f'(x) > f'(x_0+) - \frac{1}{3}\ell.$$

Then in the interval $[x_0 - \frac{1}{2}\delta, x_0 + \frac{1}{2}\delta]$, there is at most one value in

$$(f'(x_0-) + \frac{1}{3}\ell, f'(x_0+) - \frac{1}{3}\ell)$$

can be taken by $f'(x)$. This contradicts with Theorem 3.1.

For the case (2), the proof is similar details are left to you.

$\square$

We have proved that a monotonic function doesn't have second kind discontinuous points. Hence, if $f'(x)$ is not continuous, it can not be monotonic either.

4. L'Hospital's Rule

In calculus, we have learnt a very useful way of calculating limits of functions of the types '$\frac{0}{0}$' or '$\frac{\cdot}{\infty}$', which is called the L'Hospital's Rule.

EXAMPLE 4.1. (1) $\lim_{x\to 0}\frac{\sin x}{x}$. There are

$$\lim_{x\to 0}\sin x=0,\quad \lim_{x\to 0}x=0,$$

so we call it of type '$\frac{0}{0}$' . The L'Hospital's Rule tells us, we can calculate the limit as

$$\lim_{x\to 0}\frac{\sin x}{x}=\lim_{x\to 0}\frac{(\sin x)'}{x'}=\lim_{x\to 0}\frac{\cos x}{1}=1.$$

(2) $\lim_{x\to+\infty}\frac{\log x}{x}$. There are

$$\lim_{x\to+\infty}\log x=\infty,\quad \lim_{x\to+\infty}x=+\infty,$$

so we call it of type '$\frac{\infty}{\infty}$' . The L'Hospital's Rule tells us, we can calculate the limit as

$$\lim_{x\to+\infty}\frac{\log x}{x}=\lim_{x\to+\infty}\frac{(\log x)'}{x'}=\lim_{x\to+\infty}\frac{\frac{1}{x}}{1}=0.$$

Now we give a proof of the L'Hospital's Rule. Since the limit at a point can be calculated using one-side limits, we only need to understand the corresponding L'Hospital's Rule for one-side limits.

THEOREM 4.2. *Assume f, g are differentiable over (a,b) with $g(x)\neq 0$. If either*

(1) $\lim_{x\to a}f(x)=0$ *and* $\lim_{x\to a}g(x)=0$*; or*
(2) $\lim_{x\to a}|g(x)|=+\infty$,

and

$$\lim_{x\to a}\frac{f'(x)}{g'(x)}=A\in[-\infty,+\infty],\quad \textit{assuming } g'(x)\neq 0 \textit{ over } (a,b),$$

then

$$\lim_{x\to a}\frac{f(x)}{g(x)}=A.$$

PROOF. I prove the case when $a\in\mathbb{R}$ and $A\in\mathbb{R}$. Others are left to you as exercise. Since $\lim_{x\to a}\frac{f'(x)}{g'(x)}=A$, then for any $\epsilon>0$, there exists some $\delta>0$ so that any

$$a<x<a+\delta,$$

there is

$$A-\epsilon<\frac{f'(x)}{g'(x)}<A+\epsilon.$$

Now we take any $x, y\in(a,a+\delta)$, by Cauchy's mean value theorem, there exists some $\xi\in(a,a+\delta)$, which may depend on x, y, so that

$$(f(x)-f(y))g'(\xi)=(g(x)-g(y))f'(\xi).$$

It follows

$$\frac{f'(\xi)}{g'(\xi)}=\frac{f(x)-f(y)}{g(x)-g(y)}\in(A-\epsilon,A+\epsilon). \tag{4.1}$$

(1) If $\lim_{x\to a} f(x) = 0$ and $\lim_{x\to a} g(x) = 0$, for each fixed y, we take $x \to a$, and it follows
$$\frac{f(y)}{g(y)} = \lim_{x\to a} \frac{f(x) - f(y)}{g(x) - g(y)} \in [A - \epsilon, A + \epsilon].$$
By definition of limit, this says
$$\lim_{y\to a} \frac{f(y)}{g(y)} = A.$$
(2) If $\lim_{x\to a} |g(x)| = +\infty$, then for each fixed y, we can make x be close enough to a so that $a < x < y < b$ and
$$\frac{g(x) - g(y)}{g(x)} > 0.$$
Multiplying it to (4.1), we obtain
$$(A - \epsilon) \cdot \frac{g(x) - g(y)}{g(x)} < \frac{f(x) - f(y)}{g(x)} < (A + \epsilon) \cdot \frac{g(x) - g(y)}{g(x)}.$$
Take $x \to a$, we obtain
$$A - \epsilon \le \liminf_{x\to a} \frac{f(x)}{g(x)} \le \limsup_{x\to a} \frac{f(x)}{g(x)} \le A + \epsilon.$$
At last, take $\epsilon \to 0$, we are done.

□

Now let's see more examples on L'Hospital's Rule.

EXAMPLE 4.3. (1) $\lim_{x\to 0} \frac{1-\cos x}{x^2} = \frac{1}{2}$.
(2) $\lim_{x\to +\infty} \frac{x^2}{e^{3x}} = 0$.
(3) $\lim_{x\to 0+} x \log x = 0$.
(4) $\lim_{x\to 0+} \frac{\log x}{x} = -\infty$. L'Hospital Rule doesn't work!
(5) $\lim_{x\to 0+} x^x = 1$. Use $x^x = e^{x\log x}$. (This also implies $\lim_{n\to\infty} n^{\frac{1}{n}} = 1$.)
(6) $\lim_{x\to\infty} (1 + \frac{1}{x})^x = e$. Use $(1 + \frac{1}{x})^x = e^{x\log(1+\frac{1}{x})}$.

5. Taylor expansion

5.1. The statement of Taylor expansion. The Taylor expansion can be considered as a generalization of the mean value theorem to higher order cases.

Consider a function $f : [a, b] \to \mathbb{R}$. We first look at the mean value theorem from the viewpoint of approximations for $f(x)$ near a point a. We can regard the constant function
$$f_0(x) = f(a)$$
as the zero order approximation of $f(x)$. Then we ask if we can understand the remainder
$$R_1(x) := f(x) - f(a), \quad x \in [a, b].$$
for this approximation. For this, if we assume $f \in C^0([a, b])$ and f' exists over (a, b), then the mean value theorem tells us, there exists some $a < \xi_x < x$ (here ξ_x emphasizes that ξ depends on x) so that we can write R_1 as
$$R_1(x) = f'(\xi_x)(x - a).$$
This is saying that the derivative of f can control the remainder $R_1(x)$ as an order 1 monomial.

Next we consider the first order approximation, i.e., the linear approximation at a. This is the linear function

$$f_1(x) = f(a) + f'(a)(x-a), \quad x \in [a,b].$$

The corresponding remainder is defined as

$$R_2(x) := f(x) - f_1(x), \quad x \in [a,b].$$

The approximation guarantees that

$$R_2(a) = 0, \quad R_2'(a) = 0.$$

If we assume $f \in C^1([a,b])$ and f'' exists over (a,b), then we are able to prove the following lemma, which says that R_2 can be controlled by the second derivative of f as a monomial order 2.

LEMMA 5.1. *For each $x \in (a,b]$, there exists some $a < \xi_x < x$ so that*

$$R_2(x) = \frac{f''(\xi_x)}{2}(x-a)^2.$$

PROOF. For each fixed $x \in (a,b]$, consider the function

$$h(t) := f(t) - f(a) - f'(a)(t-a) - M(x)(t-a)^2,$$

defined for $t \in [a,x]$, with

$$M(x) := \frac{R_2(x)}{(x-a)^2} = \frac{f(x) - f(a) - f'(a)(x-a)}{(x-a)^2}.$$

Then $h \in C^1([a,x])$, h'' exists over (a,x), and

$$h(a) = h(x) = 0, \quad h'(a) = 0.$$

Apply Rolle's theorem, there exists some $\xi_x' \in (a,x)$ so that $h'(\xi_x') = 0$. Then apply Rolle's theorem again for $h'(t)$ over $[a,\xi_x']$, there exists some $\xi_x \in (a,\xi_x') \subset (a,x)$ so that

$$h''(\xi_x) = f''(\xi_x) - 2M(x) = 0.$$

Hence $M(x) := \frac{R_2(x)}{(x-a)^2} = \frac{f''(\xi_x)}{2}$, and so

$$R_2(x) = \frac{f''(\xi_x)}{2}(x-a)^2.$$

□

Now we generalize the above two examples to higher orders. Assume $f : [a,b] \to \mathbb{R}$ has up to n-th ordered derivatives at a, i.e., $f'(a), \cdots, f^{(n)}(a)$ exist. Denote by

$$\begin{aligned} f_n(x) &:= f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n \\ &= \Sigma_{k=0}^n \frac{f^{(k)}(a)}{k!}(x-a)^k, \quad x \in [a,b], \end{aligned}$$

the n-th order approximation, whose error is defined as

$$R_{n+1}(x) := f(x) - f_n(x).$$

Then there is the following Taylor's theorem.

THEOREM 5.2 (Taylor's Theorem). *Assume $f \in C^n([a,b])$ and $f^{(n+1)}$ exists over (a,b), where $n \in \{0,1,2,\cdots\}$. Then for each $x \in (a,b]$, there exists some $a < \xi_x < x$ so that*

$$R_{n+1}(x) = \frac{f^{(n+1)}(\xi_x)}{(n+1)!}(x-a)^{n+1}.$$

PROOF. The proof is a higher order analogue to Lemma 5.1 (which is for $n = 1$ case). For each fixed $x \in (a,b]$, consider the function

$$h(t) := f(t) - f_n(t) - M(x)(t-a)^{n+1},$$

defined for $t \in [a,x]$, with

$$M(x) := \frac{R_{n+1}(x)}{(x-a)^{n+1}} = \frac{f(x) - f_n(x)}{(x-a)^{n+1}}.$$

Then $h \in C^n([a,x])$, $h^{(n+1)}$ exists over (a,x), and

$$h(a) = h(x) = 0, \quad h'(a) = 0, \cdots, h^n(a) = 0.$$

Apply Rolle's theorem for h over $[a,x]$, there exists some $\xi_{1,x} \in (a,x)$ so that $h'(\xi_{1,x}) = 0$. Then apply Rolle's theorem for $h'(t)$ over $[a,\xi_{1,x}]$, there exists some $\xi_{2,x} \in (a,\xi_{1,x}) \subset (a,x)$ so that $h''(\xi_{2,x}) = 0$. Keep repeating this procedure, that after applying Rolle's theorem for $h^{(n)}(t)$ over $[a,\xi_{n,x}]$, there exists some $\xi_{n+1,x} \in (a,\xi_{n,x}) \subset (a,x)$ so that

$$h^{(n+1)}(\xi_{n+1,x}) = f^{(n+1)}(\xi_{n+1,x}) - (n+1)!M(x) = 0.$$

Denote by $\xi_x := \xi_{n+1,x}$. Hence $M(x) := \frac{R_{n+1}(x)}{(x-a)^{n+1}} = \frac{f^{(n+1)}(\xi_x)}{(n+1)!}$, and so

$$R_{n+1}(x) = \frac{f^{(n+1)}(\xi_x)}{(n+1)!}(x-a)^{n+1}.$$

□

DEFINITION 5.3. Assume $f \in C^\infty((a,b))$ and $c \in (a,b)$. We call the series

$$\Sigma_{n=0}^{\infty} \frac{f^{(n)}(x)}{n!}(x-c)^n$$

the Taylor series for f about c.

In general, f may be different from its Taylor series, and from definitions, we have that for any $x \in (a,b)$,

$$f(x) = \Sigma_{n=0}^{\infty} \frac{f^{(n)}(x)}{n!}(x-c)^n$$

if and only if

$$\lim_{n\to\infty} R_{n+1}(x) = 0.$$

Taylor's theorem gives us a way to estimate $R_{n+1}(x)$ by which one is able to tell if $f(x)$ is the same as its Taylor's series.

5.2. The example of power series. More general, consider the power series

$$f(x) = \Sigma_{n=0}^{\infty} c_n x^n, \quad c_n \in \mathbb{R}.$$

The convergence radius R can be calculated as

$$R = \frac{1}{\limsup_{n\to\infty} \sqrt[n]{|c_n|}} \in [0, \infty].$$

Notice that if we take derivative for each term,

$$(c_n x^n)' = n c_n x^{n-1}, \quad n \geq 1.$$

The new power series

$$\Sigma_{n=1}^{\infty} n c_n x^{n-1}$$

has the same convergence radius R.

In fact, we can prove (later after we introduce the concept of uniformly convergence, this follows from some more general statement),

THEOREM 5.4. *Assume $R \in [0, +\infty]$ is the convergence radius of the power series*

$$f(x) = \Sigma_{n=0}^{\infty} c_n x^n, \quad c_n \in \mathbb{R}.$$

Then f is differentiable over $(-R, R)$ and

$$f'(x) = \Sigma_{n=1}^{\infty} n c_n x^{n-1}.$$

Further, f is $C^\infty((-R, R))$, i.e., any order derivative of f exists over $(-R, R)$, with

$$f^{(k)}(x) = \Sigma_{n=k}^{\infty} n(n-1)\cdots(n-k+1) c_n x^{n-k} = \Sigma_{n=k}^{\infty} \frac{n!}{(n-k)!} c_n x^{n-k}, \tag{5.1}$$

where $k = 0, 1, \cdots$. In particular,

$$f^{(k)}(0) = k! c_k.$$

It follows from this theorem that a power series about 0 is the same as its Taylor series about 0.

EXAMPLE 5.5. Consider a polynomial $f(x) = 2 + 2x + 3x^2 + x^3$. We can calculate its Taylor series at $a = 1$ as follows.

$$\begin{aligned} f(x) &= 2 + 2((x-1)+1) + 3((x-1)+1)^2 + ((x-1)+1)^3 \\ &= 8 + 11(x-1) + 6(x-1)^2 + (x-1)^3. \end{aligned}$$

From it, we have

$$f(1) = 8, f'(1) = 11, f''(1) = 12, f'''(1) = 6.$$

This example can be generalized to arbitrary power series, whose proof is based on the possibility of changing orders of summations (we leave its proof to later).

THEOREM 5.6 (Taylor's Theorem for power series). *Assume $R \in [0, +\infty]$ is the convergence radius of the power series*

$$f(x) = \Sigma_{n=0}^{\infty} c_n x^n, \quad c_n \in \mathbb{R}.$$

For any $a \in (-R, R)$, define

$$R_a = R - |a|.$$

Then we can write f as a power series over $(a - R_a, a + R_a)$ as

$$f(x) = \Sigma_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n.$$

PROOF. We write

$$\begin{aligned} f(x) &= \Sigma_{n=0}^{\infty} c_n x^n \\ &= \Sigma_{n=0}^{\infty} c_n ((x-a)+a)^n \\ &= \Sigma_{n=0}^{\infty} c_n \Sigma_{k=0}^{n} \frac{n!}{k!(n-k)!}(x-a)^k a^{n-k} \\ &= \Sigma_{k=0}^{\infty} (\Sigma_{n=k}^{\infty} \frac{n!}{k!(n-k)!} c_n a^{n-k})(x-a)^k \\ &= \Sigma_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x-a)^k. \end{aligned}$$

The fourth equality above is NOT trivial, but follows from the convergence of the double summation (we leave it proof to later)

$$\Sigma_{n=0}^{\infty} \Sigma_{k=0}^{n} |c_n \frac{n!}{k!(n-k)!}(x-a)^k a^{n-k}|.$$

This series is convergent because

$$\Sigma_{k=0}^{n} |c_n \frac{n!}{k!(n-k)!}(x-a)^k a^{n-k}| = |c_n|(|x-a| + |a|)^n < |c_n| R^n,$$

and

$$\Sigma_{n=0}^{\infty} |c_n| R^n$$

is convergent.

□

CHAPTER 6

The Riemann–Stieltjes Integral

1. Definition of Riemann–Stieltjes Integral

Assume $[a, b]$ is a closed interval in $\mathbb{R}$. By a partition $\mathcal{P}$, we mean a finite set of points

$$a = x_0 \leq x_1 \leq \cdots \leq x_{n-1} \leq x_n = b.$$

Assume f is a bounded real-valued function over $[a, b]$ and α is an increasing function over $[a, b]$. Denote by

$$M_i = \sup_{[x_{i-1}, x_i]} f(x), \quad m_i = \inf_{[x_{i-1}, x_i]} f(x),$$

and by

$$\Delta\alpha_i = \alpha(x_i) - \alpha(x_{i-1}).$$

Define the upper sum of f with respect to the partition and α as

$$U(f, \alpha; \mathcal{P}) := \Sigma_{i=1}^n M_i \Delta\alpha_i,$$

and the lower sum of f with respect to the partition and α as

$$L(f, \alpha; \mathcal{P}) := \Sigma_{i=1}^n m_i \Delta\alpha_i.$$

Define the upper Riemann–Stieltjes integral as

$$\overline{\int_a^b} f(x) d\alpha(x) := \inf_{\mathcal{P}} U(f, \alpha; \mathcal{P})$$

and the lower Riemann–Stieltjes integral as

$$\underline{\int_a^b} f(x) d\alpha(x) := \sup_{\mathcal{P}} L(f, \alpha; \mathcal{P}).$$

It is easy to see from definition that

$$\underline{\int_a^b} f(x) d\alpha(x) \leq \overline{\int_a^b} f(x) d\alpha(x).$$

Definition 1.1. Call a function f is Riemann–Stieltjes integrable with respect to α over $[a, b]$, if

$$\overline{\int_a^b} f(x) d\alpha(x) = \underline{\int_a^b} f(x) d\alpha(x).$$

We use $\int_a^b f(x) d\alpha(x)$ to denote the common value and call it the Riemann–Stieltjes of f with respect to α over $[a, b]$.

We use the notation $R(\alpha)([a, b])$ to denote the set of Riemann–Stieltjes integrable functions with respect to α over $[a, b]$.

In particular, when $\alpha(x) = x$, we call the corresponding Riemann–Stieltjes integration the Riemann integration and use $R([a, b])$ to denote the set of Riemann integrable functions.

The first problem we need to understand is what kind of functions are Riemann–Stieltjes integrable.

THEOREM 1.2. *$f \in R(\alpha)([a,b])$ if and only if for each $\epsilon > 0$, there exists some partition $\mathcal{P}$ so that*

$$U(f,\alpha;\mathcal{P}) - L(f,\alpha;\mathcal{P}) < \epsilon.$$

PROOF. (1) Assume $f \in R(\alpha)([a,b])$, by definition

$$\inf_{\mathcal{P}} U(f,\alpha;\mathcal{P}) = \int_a^b f(x)d\alpha(x) = \sup_{\mathcal{P}} L(f,\alpha;\mathcal{P}).$$

For any $\epsilon > 0$, there exists partitions $\mathcal{P}_1, \mathcal{P}_2$ so that

$$U(f,\alpha;\mathcal{P}_1) < \int_a^b f(x)d\alpha(x) + \frac{1}{2}\epsilon$$

and

$$L(f,\alpha;\mathcal{P}_2) > \int_a^b f(x)d\alpha(x) - \frac{1}{2}\epsilon.$$

Consider the common refinement $\mathcal{P}$ by making union of points in $\mathcal{P}_1$ and $\mathcal{P}_2$. It follows

$$L(f,\alpha;\mathcal{P}_2) \le L(f,\alpha;\mathcal{P}) \le U(f,\alpha;\mathcal{P}) \le U(f,\alpha;\mathcal{P}_1),$$

and then

$$\begin{aligned}
& U(f,\alpha;\mathcal{P}) - L(f,\alpha;\mathcal{P}) \\
\le\ & U(f,\alpha;\mathcal{P}_1) - L(f,\alpha;\mathcal{P}_2) \\
\le\ & (\int_a^b f(x)d\alpha(x) + \frac{1}{2}\epsilon) - (\int_a^b f(x)d\alpha(x) - \frac{1}{2}\epsilon) \\
=\ & \epsilon.
\end{aligned}$$

(2) Assume that for each $\epsilon > 0$, there exists some partition $\mathcal{P}$ so that

$$U(f,\alpha;\mathcal{P}) - L(f,\alpha;\mathcal{P}) < \epsilon,$$

then it follows

$$0 \le \overline{\int_a^b} f(x)d\alpha(x) - \underline{\int_a^b} f(x)d\alpha(x) \le \epsilon.$$

Take $\epsilon \to 0$, we are done.

□

Now we use this criterion to prove the following several theorems.

THEOREM 1.3. $C^0([a,b]) \subset R(\alpha)([a,b])$.

PROOF. If α is constant, then by definition, $\int_a^b f(x)d\alpha(x) = 0$. In the following, we assume $\alpha(a) < \alpha(b)$.

For any $f \in C^0([a,b])$, it is uniformly continuous over $[a,b]$. Hence for any $\epsilon > 0$, there exists some $\delta > 0$ so that

$$|f(t_1) - f(t_2)| < \frac{1}{2(\alpha(b) - \alpha(a))}\epsilon$$

whenever $|t_1 - t_2| < \delta$.

Now we take a partition $\mathcal{P}$ so that $\Delta_i = x_i - x_{i-1} < \delta$. Then it follows

$$M_i - m_i < \frac{1}{\alpha(b) - \alpha(a)}\epsilon.$$

For this partition $\mathcal{P}$,

$$\begin{aligned} & U(f,\alpha;\mathcal{P}) - L(f,\alpha;\mathcal{P}) \\ = \ & \Sigma_i (M_i - m_i)\Delta\alpha_i \\ \leq \ & \Sigma_i \frac{1}{\alpha(b)-\alpha(a)}\epsilon\Delta\alpha_i \\ = \ & \epsilon. \end{aligned}$$

Apply Theorem 1.2, we proved $f \in R(\alpha)([a,b])$.

□

THEOREM 1.4. *Assume f is monotonic and α is continuous on $[a,b]$, then $f \in R(\alpha)([a,b])$.*

PROOF. If f is constant C, then by definition, we can see $\int_a^b f(x)d\alpha(x) = C(\alpha(b)-\alpha(a))$. Otherwise, WLOG, we can assume

$$f(b) - f(a) > 0.$$

Sine $\alpha \in C^0([a,b])$ implies α is uniformly continuous over $[a,b]$. Then for any $\epsilon > 0$, there exists some $\delta > 0$ so that

$$|\alpha(t_1) - \alpha(t_2)| < \frac{1}{2(f(b)-f(a))}\epsilon$$

whenever $|t_1 - t_2| < \delta$.

Now we take a partition $\mathcal{P}$ so that $\Delta_i = x_i - x_{i-1} < \delta$. Then it follows

$$\Delta\alpha_i < \frac{1}{f(b)-f(a)}\epsilon.$$

For this partition $\mathcal{P}$, there is

$$\begin{aligned} & U(f,\alpha;\mathcal{P}) - L(f,\alpha;\mathcal{P}) \\ = \ & \Sigma_i (f(x_i) - f(x_{i-1}))\Delta\alpha_i \\ \leq \ & \Sigma_i (f(x_i) - f(x_{i-1}))\frac{1}{f(b)-f(a)}\epsilon \\ = \ & \epsilon. \end{aligned}$$

Apply Theorem 1.2, we have $f \in R(\alpha)([a,b])$.

□

THEOREM 1.5. *Assume f is bounded over $[a,b]$ with only finitely many discontinuous points and α is continuous on these points. Then $f \in R(\alpha)([a,b])$.*

PROOF. Assume f discontinuous at $p_1 < p_2 < \cdots < p_m \in [a,b]$. For each $i = 1,\cdots,m$, since α is continuous at p_i, for any $\epsilon > 0$ there exists $\delta_i > 0$, so that any $|x - p_i| < \delta_i$,

$$|\alpha(x) - \alpha(p_i)| < \epsilon.$$

Moreover, we can shrink δ_i's so that these intervals $[p_i - \delta_i, p_i + \delta_i]$ have no intersection.

On the other hand, f is continuous over the complement of the union of interiors of these intervals, which we denote by $K \subset [a,b]$. In fact, K is the union of finite closed intervals as

$$K = [a, p_1 - \delta_1] \cup [p_1 + \delta_1, p_2 - \delta_2] \cup \cdots \cup [p_{m-1} + \delta_{m-1}, p_m - \delta_m] \cup [p_m + \delta_m, b].$$

We denote them by $K_0, K_1, \cdots, K_{m-1}, K_m$ one by one from left to right.

It follows from Theorem 1.3 and Theorem 1.2 that, for each K_j, $j = 0, 1, \cdots, m$, there exists a partition $\mathcal{P}_j$ so that

$$U(f|_{K_j}, \alpha; \mathcal{P}_j) - L(f|_{K_j}, \alpha; \mathcal{P}_j) < \epsilon.$$

Now we consider a partition $\mathcal{P}$ for $[a, b]$ whose points are the union of partitions $\mathcal{P}_0, \cdots, \mathcal{P}_m$. It follows from our construction that

$$\begin{aligned}
& U(f, \alpha; \mathcal{P}) - L(f, \alpha; \mathcal{P}) \\
= \; & \Sigma_{j=0}^{m}(U(f|_{K_j}, \alpha; \mathcal{P}_j) - L(f|_{K_j}, \alpha; \mathcal{P}_j)) + \Sigma_{i=1}^{m}(\sup_{[p_i-\delta_i, p_i+\delta_i]} f - \inf_{[p_i-\delta_i, p_i+\delta_i]} f)(\alpha(p_i + \delta_i) - \alpha(p_i - \delta_i)) \\
\le \; & (m+1)\epsilon + 2Mm\epsilon \\
= \; & (m+1+2Mm)\epsilon.
\end{aligned}$$

Here M is a fixed number so that $|f(x)| \le M$ for any $x \in [a, b]$.

Apply Theorem 1.2, we have $f \in R(\alpha)([a, b])$.

□

THEOREM 1.6. *Assume $f \in R(\alpha)([a, b])$ with $m \le f \le M$ and $g \in C^0([m, M])$. Then $g \circ f \in R(\alpha)([a, b])$.*

PROOF. First, $g \in C^0([m, M])$ implies g is uniformly continuous over $[m, M]$. Hence for any $\epsilon > 0$, there exists $\delta > 0$ so that any

$$|g(y_1) - g(y_2)| < \epsilon \quad \text{for any} \quad |y_1 - y_2| < \delta.$$

Next, use $f \in R(\alpha)([a, b])$, there exists a partition $\mathcal{P}$ for $[a, b]$ so that

$$U(f, \alpha; \mathcal{P}) - L(f, \alpha; \mathcal{P}) < \delta^2.$$

Each $[x_{i-1}, x_i]$ determined by the partition $\mathcal{P}$ belongs to either one of the following two cases:

(1) $\sup_{[x_{i-1}, x_i]} f - \inf_{[x_{i-1}, x_i]} f < \delta$. We denote by $\mathcal{P}_1$ the sub-partition that contains such intervals. Over such intervals,

$$\sup_{[x_{i-1}, x_i]} g \circ f - \inf_{[x_{i-1}, x_i]} g \circ f < \epsilon.$$

(2) $\sup_{[x_{i-1}, x_i]} f - \inf_{[x_{i-1}, x_i]} f \ge \delta$. We denote by $\mathcal{P}_2$ the sub-partition that contains such intervals. Over $\mathcal{P}_2$, we have

$$\delta \Sigma_{i \in \mathcal{P}_2} \Delta\alpha_i \le U(f, \alpha; \mathcal{P}_2) - L(f, \alpha; \mathcal{P}_2) \le U(f, \alpha; \mathcal{P}) - L(f, \alpha; \mathcal{P}) < \delta^2,$$

and hence

$$\Sigma_{i \in \mathcal{P}_2} \Delta\alpha_i < \delta.$$

We can further shrink δ so that $\delta < \epsilon$.

Now we consider $\mathcal{P}_1, \mathcal{P}_2$ together and obtain

$$\begin{aligned}
& U(g \circ f, \alpha; \mathcal{P}) - L(g \circ f, \alpha; \mathcal{P}) \\
= \; & \Sigma_{i \in \mathcal{P}_1}(\sup_{[x_{i-1}, x_i]} g \circ f - \inf_{[x_{i-1}, x_i]} g \circ f)\Delta\alpha_i + \Sigma_{i \in \mathcal{P}_2}(\sup_{[x_{i-1}, x_i]} g \circ f - \inf_{[x_{i-1}, x_i]} g \circ f)\Delta\alpha_i \\
\le \; & \epsilon \Sigma_{i \in \mathcal{P}_1} \Delta\alpha_i + 2C \Sigma_{i \in \mathcal{P}_2} \Delta\alpha_i \\
\le \; & \epsilon(\Sigma_{i \in \mathcal{P}_1} \Delta\alpha_i + 2C) \\
\le \; & \epsilon(\alpha(b) - \alpha(a) + 2C).
\end{aligned}$$

Here C is an upper bound of $|g|$ over $[m, M]$.

At last, apply Theorem 1.2, we obtain $g \circ f \in R(\alpha)([a, b])$.

□

EXAMPLE 1.7. If $f \in R(\alpha)([a, b])$, then $f^2 \in R(\alpha)([a, b])$.

2. Properties of the integral

THEOREM 2.1. (1) *Linearity of f:*

(a) *If $f_1, f_2 \in R(\alpha)([a, b])$, then $f_1 + f_2 \in R(\alpha)([a, b])$ and*

$$\int_a^b (f_1 + f_2) d\alpha = \int_a^b f_1 d\alpha + \int_a^b f_2 d\alpha.$$

(b) *If $f \in R(\alpha)([a, b])$ and $c \in \mathbb{R}$, then $cf \in R(\alpha)([a, b])$ and*

$$\int_a^b (cf) d\alpha = c \int_a^b f d\alpha.$$

(2) *Linearity of α:*

(a) *If $f \in R(\alpha_1)([a, b]) \cap R(\alpha_2)([a, b])$, then $f \in R(\alpha_1 + \alpha_2)([a, b])$ and*

$$\int_a^b f d(\alpha_1 + \alpha_2) = \int_a^b f d\alpha_1 + \int_a^b f d\alpha_2.$$

(b) *If $f \in R(\alpha)([a, b])$ and $c \geq 0$, then $f \in R(c\alpha)([a, b])$ and*

$$\int_a^b f d(c\alpha) = c \int_a^b f d\alpha.$$

(3) *If $f_1, f_2 \in R(\alpha)([a, b])$ and $f_1 \leq f_2$, then*

$$\int_a^b f_1 d\alpha \leq \int_a^b f_2 d\alpha.$$

(4) *If $f \in R(\alpha)([a, b])$ and $c \in [a, b]$, then $f \in R(\alpha)([a, c]) \cap R(\alpha)([c, b])$ and*

$$\int_a^b f d\alpha = \int_a^c f d\alpha + \int_c^b f d\alpha.$$

(5) *If $f \in R(\alpha)([a, b])$ and $|f| \leq M$, then*

$$|\int_a^b f d\alpha| \leq M(\alpha(b) - \alpha(a)).$$

(6) *If $f, g \in R(\alpha)([a, b])$, then $fg \in R(\alpha)([a, b])$.*

(7) *If $f \in R(\alpha)([a, b])$, then $|f| \in R(\alpha)([a, b])$ and*

$$|\int_a^b f d\alpha| \leq \int_a^b |f| d\alpha.$$

PROOF. (1) (a) First we notice that for any partition $\mathcal{P}$ over $[a, b]$, we have the following inequality

$$\begin{aligned} L(f_1, \alpha; \mathcal{P}) + L(f_2, \alpha; \mathcal{P}) &\leq L(f_1 + f_2, \alpha; \mathcal{P}) \\ &\leq U(f_1 + f_2, \alpha; \mathcal{P}) \leq U(f_1, \alpha; \mathcal{P}) + U(f_2, \alpha; \mathcal{P}). \end{aligned}$$

Hence

$$\begin{aligned} & U(f_1 + f_2, \alpha; \mathcal{P}) - L(f_1 + f_2, \alpha; \mathcal{P}) \\ \leq\ & (U(f_1, \alpha; \mathcal{P}) - L(f_1, \alpha; \mathcal{P})) + (U(f_2, \alpha; \mathcal{P}) - L(f_2, \alpha; \mathcal{P})). \end{aligned} \tag{2.1}$$

Now since $f_1, f_2 \in R(\alpha)([a,b])$, using Theorem 1.2, for any $\epsilon > 0$, there exists partitions $\mathcal{P}_1, \mathcal{P}_2$ for $[a,b]$ so that

$$\begin{aligned} U(f_1,\alpha;\mathcal{P}_1) - L(f_1,\alpha;\mathcal{P}_1) &< \epsilon/2 \\ U(f_2,\alpha;\mathcal{P}_2) - L(f_2,\alpha;\mathcal{P}_2) &< \epsilon/2. \end{aligned}$$

Take $\mathcal{P}$ as a common refinement of $\mathcal{P}_1$ and $\mathcal{P}_2$, it follows

$$\begin{aligned} U(f_1,\alpha;\mathcal{P}) - L(f_1,\alpha;\mathcal{P}) \le U(f_1,\alpha;\mathcal{P}_1) - L(f_1,\alpha;\mathcal{P}_1) < \epsilon/2 \\ U(f_2,\alpha;\mathcal{P}) - L(f_2,\alpha;\mathcal{P}) \le U(f_2,\alpha;\mathcal{P}_2) - L(f_2,\alpha;\mathcal{P}_2) < \epsilon/2. \end{aligned}$$

Connect it with (2.1), we get

$$U(f_1+f_2,\alpha;\mathcal{P}) - L(f_1+f_2,\alpha;\mathcal{P}) < \epsilon$$

and then Theorem 1.2 implies $f_1 + f_2 \in R(\alpha)([a,b])$.

Now for any $\epsilon > 0$, take a partition $\mathcal{P}_i$, $i = 1,2$, so that

$$\int_a^b f_i d\alpha \le U(f_i,\alpha;\mathcal{P}_i) \le \int_a^b f_i d\alpha + \epsilon.$$

Assume $\mathcal{P}$ is a common refinement of $\mathcal{P}_i$. It follows the following inequality

$$U(f_1+f_2,\alpha;\mathcal{P}) \le U(f_1,\alpha;\mathcal{P}) + U(f_2,\alpha;\mathcal{P}) \le U(f_1,\alpha;\mathcal{P}_1) + U(f_2,\alpha;\mathcal{P}_2),$$

and then

$$\int_a^b (f_1+f_2) d\alpha \le U(f_1+f_2,\alpha;\mathcal{P}) \le \int_a^b f_1 d\alpha + \int_a^b f_2 d\alpha + 2\epsilon.$$

Since $\epsilon > 0$ can be arbitrarily small, this proves

$$\int_a^b (f_1+f_2) d\alpha \le \int_a^b f_1 d\alpha + \int_a^b f_2 d\alpha.$$

Similarly, using lower sums, we will obtain

$$\int_a^b (f_1+f_2) d\alpha \ge \int_a^b f_1 d\alpha + \int_a^b f_2 d\alpha.$$

(b) Similar to (a) and is left to you.

(2) Exercise.

(3) Exercise.

(4) Exercise.

(5) Now for any partition $\mathcal{P}$,

$$U(f,\alpha;\mathcal{P}) \le U(M,\alpha;\mathcal{P}) = M(\alpha(b) - \alpha(a)).$$

Take supremum of $\mathcal{P}$, since $f \in R(\alpha)([a,b])$, we have

$$\int_a^b f d\alpha = \overline{\int}_a^b f d\alpha = \sup_{\mathcal{P}} U(f,\alpha;\mathcal{P}) \le M(\alpha(b) - \alpha(a)).$$

Similarly,

$$L(-f,\alpha;\mathcal{P}) \ge L(-M,\alpha;\mathcal{P}) = -M(\alpha(b) - \alpha(a)).$$

Take infimum of $\mathcal{P}$, since $-f \in R(\alpha)([a,b])$ (from 1(b)), we have

$$-\int_a^b f d\alpha = \int_a^b (-f) d\alpha \underline{\int}_a^b f d\alpha = \inf_{\mathcal{P}} L(f,\alpha;\mathcal{P}) \ge -M(\alpha(b) - \alpha(a)).$$

Together this proves

$$|\int_a^b f d\alpha| \le M(\alpha(b) - \alpha(a)).$$

(6) Notice the following equality

$$fg = \frac{1}{4}((f+g)^2 - (f-g)^2).$$

Then $fg \in R(\alpha)([a,b])$ follows from Property 1 and Theorem 1.6.

(7) $|f| \in R(\alpha)([a,b])$ follows from the continuity of the absolute value function and Theorem 1.6. Then inequality can be proved similarly as for (5).

□

REMARK 2.2. For the property (7) above, notice that $|f| \in R(\alpha)([a,b])$ doesn't imply $f \in R(\alpha)([a,b])$. Consider the following example

$$f(x) = \begin{cases} 1 & x \in \mathbb{Q} \cap [0,1] \\ -1 & x \in \mathbb{Q}^c \cap [0,1]. \end{cases}$$

It is not Riemann integrable over $[0,1]$ (Excise: prove this) but $|f|$ is Riemann integrable.

The following three theorems are related to the useful formula of "substitution" for integration in calculus.

THEOREM 2.3. *Assume α' exists and $\alpha' \in R([a,b])$. Assume f is bounded over $[a,b]$. Then $f \in R(\alpha)([a,b])$ if and only if $f\alpha' \in R([a,b])$. In that case,*

$$\int_a^b f(x) d\alpha(x) = \int_a^b f(x)\alpha'(x) dx.$$

PROOF. Since $\alpha' \in R([a,b])$, for any $\epsilon > 0$, there exists a partition $\mathcal{P}$ for $[a,b]$ so that

$$U(\alpha';\mathcal{P}) - L(\alpha';\mathcal{P}) < \epsilon. \tag{2.2}$$

At the same time, apply the mean value theorem over each interval $[x_{i-1}, x_i]$ from the partition $\mathcal{P}$, there exist points $t_i \in (x_{i-1}, x_i)$ so that

$$\Delta\alpha_i = \alpha(x_i) - \alpha(x_{i-1}) = \alpha'(t_i)(x_i - x_{i-1}) = \alpha'(t_i)\Delta_i.$$

If follows

$$U(f,\alpha;\mathcal{P}) - L(f,\alpha;\mathcal{P}) = \Sigma_i(M_i - m_i)\Delta\alpha_i = \Sigma_i(M_i - m_i)\alpha'(t_i)\Delta_i = \Sigma_i(M_i\alpha'(t_i) - m_i\alpha'(t_i))\Delta_i.$$

On the other hand, notice that there exist points $s_i, s_i' \in (x_{i-1}, x_i)$

$$\begin{aligned}
& \Sigma_i(\sup_{[x_{i-1},x_i]}(f\alpha') - \inf_{[x_{i-1},x_i]}(f\alpha'))\Delta_i \\
\le\ & \Sigma_i(M_i \sup_{[x_{i-1},x_i]}\alpha' - m_i \inf_{[x_{i-1},x_i]}\alpha')\Delta_i \\
\le\ & \Sigma_i(M_i\alpha'(s_i) - m_i\alpha'(s_i') + \frac{\epsilon}{b-a})\Delta_i \\
=\ & \Sigma_i(M_i\alpha'(s_i) - m_i\alpha'(s_i'))\Delta_i + \epsilon.
\end{aligned}$$

Now if $f \in R(\alpha)([a,b])$, then $|f|$ is bounded by some $C > 0$ and

$$\begin{aligned}
& \Sigma_i(\sup_{[x_{i-1},x_i]}(f\alpha') - \inf_{[x_{i-1},x_i]}(f\alpha'))\Delta_i \\
\leq\ & \Sigma_i(M_i\alpha'(s_i) - m_i\alpha'(s_i'))\Delta_i + \epsilon \\
\leq\ & \Sigma_i(M_i\alpha'(t_i) - m_i\alpha'(t_i))\Delta_i + \epsilon + \Sigma_i(M_i|\alpha'(s_i) - \alpha'(t_i)| - m_i|\alpha'(s_i') - \alpha'(t_i)|)\Delta_i \\
\leq\ & \Sigma_i(M_i\alpha'(t_i) - m_i\alpha'(t_i))\Delta_i + \epsilon + 2C(U(\alpha';\mathcal{P}) - L(\alpha';\mathcal{P})) \\
=\ & U(f,\alpha;\mathcal{P}) - L(f,\alpha;\mathcal{P}) + \epsilon + 2C\epsilon.
\end{aligned}$$

By taking a refinement of $\mathcal{P}$ which we still denote by $\mathcal{P}$, we can make $U(f,\alpha;\mathcal{P}) - L(f,\alpha;\mathcal{P}) < \epsilon$. Then using Theorem 1.2, we proved $f\alpha' \in R([a,b])$.

Now assume $f\alpha' \in R([a,b])$. There exist $r_i, r_i' \in [x_{i-1}, x_i]$ so that the following estimates hold

$$\begin{aligned}
& U(f,\alpha;\mathcal{P}) - L(f,\alpha;\mathcal{P}) \\
=\ & \Sigma_i(M_i - m_i)\Delta\alpha_i \\
\leq\ & \Sigma_i(f(r_i) - f(r_i'))\Delta\alpha_i + \epsilon(\alpha(b) - \alpha(a)) \\
=\ & \Sigma_i(f(r_i) - f(r_i'))\alpha'(t_i)\Delta_i + \epsilon(\alpha(b) - \alpha(a)) \\
=\ & \Sigma_i(f(r_i)\alpha'(t_i) - f(r_i')\alpha'(t_i))\Delta_i + \epsilon(\alpha(b) - \alpha(a)) \\
\leq\ & \Sigma_i(f(r_i)\alpha'(r_i) - f(r_i')\alpha'(r_i'))\Delta_i + \Sigma_i(f(r_i)|\alpha'(r_i) - \alpha'(t_i)| - f(r_i')|\alpha'(r_i') - \alpha'(t_i)|)\Delta_i + \epsilon(\alpha(b) - \alpha(a)) \\
\leq\ & \Sigma_i(\sup_{[x_{i-1},x_i]}(f\alpha') - \inf_{[x_{i-1},x_i]}(f\alpha'))\Delta_i + 2C\epsilon + \epsilon(\alpha(b) - \alpha(a)).
\end{aligned}$$

If then follows from Theorem 1.2 that $f \in R(\alpha)([a,b])$.

At last, the equality follows from the same estimates as above but for U and L separately: For example, we have

$$\begin{aligned}
& U(f,\alpha;\mathcal{P}) \\
=\ & \Sigma_i M_i\Delta\alpha_i \\
\leq\ & \Sigma_i f(r_i)\Delta\alpha_i + \epsilon(\alpha(b) - \alpha(a)) \\
=\ & \Sigma_i f(r_i)\alpha'(t_i)\Delta_i + \epsilon(\alpha(b) - \alpha(a)) \\
=\ & \Sigma_i f(r_i)\alpha'(t_i)\Delta_i + \epsilon(\alpha(b) - \alpha(a)) \\
\leq\ & \Sigma_i f(r_i)\alpha'(r_i)\Delta_i + \Sigma_i f(r_i)|\alpha'(r_i) - \alpha'(t_i)|\Delta_i + \epsilon(\alpha(b) - \alpha(a)) \\
\leq\ & \Sigma_i \sup_{[x_{i-1},x_i]}(f\alpha')\Delta_i + C\epsilon + \epsilon(\alpha(b) - \alpha(a)).
\end{aligned}$$

Take infimum of $\mathcal{P}$ and let $\epsilon \to 0$, it follows

$$\int_a^b f(x)d\alpha(x) \leq \int_a^b f(x)\alpha'(x)dx.$$

The other direction can be obtained similarly and we skip details. $\square$

THEOREM 2.4 (Change of variable). *Assume $f \in R(\alpha)([a,b])$. Assume φ is strictly increasing and continuous that maps interval $[A,B]$ to $[a,b]$. Define*

$$\beta = \alpha\circ\varphi$$

which is increasing on $[A, B]$*, and define*

$$g = f \circ \varphi.$$

Then $g \in R(\beta)([A, B])$ *and*

$$\int_A^B g d\beta = \int_a^b f d\alpha.$$

PROOF. By the strictly increasing property of φ, each partition $\mathcal{P} = \{x_i\}$ for $[a, b]$ corresponds to a partition $\mathcal{P}' = \{y_i\}$ for $[A, B]$ with $x_i = \varphi(y_i)$, and

$$U(f, \alpha; \mathcal{P}) = U(g, \beta; \mathcal{P}'), \quad L(f, \alpha; \mathcal{P}) = L(g, \beta; \mathcal{P}').$$

The conclusion then immediately follow from definition of integration. $\square$

Using it, we obtain the following important formula for change of variables.

THEOREM 2.5. *Assume* $f \in R([a, b])$*. Assume* φ *is strictly increasing that maps interval* $[A, B]$ *to* $[a, b]$ *and* $\varphi' \in R([A, B])$*. Then*

$$\int_a^b f(x)dx = \int_A^B f(\varphi(y))\varphi'(y)dy.$$

3. Fundamental theorem of calculus

THEOREM 3.1. *Assume* $f \in R([a, b])$*. For* $a \le x \le b$*, define*

$$F(x) := \int_a^x f(t)dt.$$

Then $F \in C^0([a, b])$*. Furthermore, if* f *is continuous at a point* $x_0 \in [a, b]$*, then* F *is differentiable at* x_0*, and*

$$F'(x_0) = f(x_0).$$

PROOF.

(1) Because $f \in R([a, b])$, it must be bounded. (Why?) Assume $|f| \le M$. Consider any $x, y \in [a, b]$,

$$|F(y) - F(x)| = |\int_a^y f(t)dt - \int_a^x f(t)dt| = |\int_x^y f(t)dt| \le M|y - x|.$$

This proves that F is uniformly continuous over $[a, b]$, hence continuous.

(2) Assume f is continuous at a point $x_0 \in [a, b]$, then for any $\epsilon > 0$, there exists $\delta > 0$ so that any $|x - x_0| < \delta$, there is

$$|f(x) - f(x_0)| < \epsilon.$$

Now consider any $x \in (x_0 - \delta, x_0 + \delta) \cap [a, b]$ and $x \neq x_0$. We have

$$\begin{aligned}
|\frac{F(x) - F(x_0)}{x - x_0} - f(x_0)| &= |\frac{1}{x - x_0}\int_{x_0}^x f(t)dt - f(x_0)| \\
&= |\frac{1}{x - x_0}\int_{x_0}^x f(t)dt - \frac{1}{x - x_0}\int_{x_0}^x f(x_0)dt| \\
&= |\frac{1}{x - x_0}\int_{x_0}^x (f(t) - f(x_0))dt| \\
&\le \frac{\int_{x_0}^x |f(t) - f(x_0)|dt}{|x - x_0|} \\
&< \frac{\epsilon|x - x_0|}{|x - x_0|} = \epsilon.
\end{aligned}$$

□

Usually, we call such F an antiderivative of f.

THEOREM 3.2 (The fundamental theorem of calculus). *If $f \in R([a,b])$ and if there is a differentiable function F on $[a,b]$ so that*

$$F' = f,$$

then $\int_a^b f(x)dx = F(b) - F(a)$.

PROOF. Since $f \in R([a,b])$, for any $\epsilon > 0$, there exists a partition $\mathcal{P}$ for $[a,b]$ so that

$$U(f;\mathcal{P}) - L(f;\mathcal{P}) < \epsilon$$

and

$$|U(f;\mathcal{P}) - \int_a^b f(x)dx| < \epsilon.$$

In each interval $[x_{i-1}, x_i]$, apply the mean value theorem to F: There exists some $t_i \in (x_{i-1}, x_i)$ so that

$$F(x_i) - F(x_{i-1}) = F'(t_i)(x_i - x_{x-1}) = f(t_i)\Delta_i.$$

Take summation over all such intervals, we obtain

$$F(b) - F(a) = \Sigma_i f(t_i)\Delta_i.$$

It follows

$$\begin{aligned} |F(b) - F(a) - \int_a^b f(x)dx| &= |\Sigma_i f(t_i)\Delta_i - \int_a^b f(x)dx| \\ &\leq |\Sigma_i f(t_i)\Delta_i - U(f;\mathcal{P})| + |U(f;\mathcal{P}) - \int_a^b f(x)dx| \\ &\leq \epsilon + \epsilon = 2\epsilon. \end{aligned}$$

Take $\epsilon \to 0$, we obtain $\int_a^b f(x)dx = F(b) - F(a)$.

□

A useful corollary for calculation is the following formula of integration by parts.

THEOREM 3.3 (Integration by parts). *Assume F, G are differentiable on $[a,b]$ with $F' = f \in R([a,b])$ and $G' = g \in R([a,b])$. Then*

$$\int_a^b F(x)g(x)dx = F(b)G(b) - F(a)G(a) - \int_a^b f(x)G(x)dx.$$

PROOF. Consider the function FG over $[a,b]$. It is differentiable and

$$(FG)' = F'G + FG' = fG + Fg.$$

Notice that $fG + Fg$ is Riemann integrable since both $f, g \in R([a,b])$ and $F, G \in C^0([a,b])$. Apply the Fundamental Theorem of Calculus 3.2,

$$F(b)G(b) - F(a)G(a) = \int_a^b (fG + Fg)dx = \int_a^b fGdx + \int_a^b Fgdx.$$

□

CHAPTER 7

Sequence and series of functions

1. Uniform Convergence

If $f_n : [a, b] \to \mathbb{R}$, $n = 1, 2, \cdots$, is a sequence of functions, and for each $x \in [a, b]$, $f_n(x) \to f(x)$ as $n \to \infty$. The main question we are trying to answer in this chapter is whether we can obtain certain properties of f from corresponding properties of $\{f_n\}$. For example, if we know each f_n is continuous, can we conclude that f is also continuous over $[a, b]$? For this question, the answer is no for general cases. An example is

$$f_n(x) = x^n, \quad x \in [0, 1].$$

It pointwise converges to

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

which is discontinuous at 1.

But if we cut the domain to $[0, \frac{1}{2}]$, then the limiting function is continuous.

Let's check what it needs to make $f(x)$ be continuous at x_0. Notice that the continuity at x_o is equivalent to

$$\lim_{x\to x_0} \lim_{n\to\infty} f_n(x) = \lim_{x\to x_0} f(x) = f(x_0) = \lim_{n\to\infty} f_n(x_0) = \lim_{n\to\infty} \lim_{x\to x_0} f_n(x).$$

In another word, the continuity is in fact a problem about changing the orders of two limiting process. For the above example, such change is not allowed for $x_0 = 1$.

It turns out that if we require a stronger convergence than only the pointwise convergence, then we can prove the continuity of each f_n implies the continuity of f. Such stronger convergence is called uniform convergence that we are going to introduce now.

In this chapter, we restrict to real-valued functions, but all results holds for complex-valued functions too.

DEFINITION 1.1. Assume $\{f_n\}$ is a sequence of functions defined over a set X and f is also a function defined over X. We say $\{f_n\}$ uniformly converges to f over X, if for any $\epsilon > 0$, there exists $N > 0$ (which is independent of x), so that

$$|f_n(x) - f(x)| < \epsilon$$

for any $x \in X$.

We use notation $f_n \overset{X}{\rightrightarrows} f$ to denote this uniform convergence over X.

From definition, if we want to show f_n doesn't converge to f uniformly, we only need to find some $\epsilon_0 > 0$ and a subsequence $\{n_k\}$ of $\{n\}$ together with sequence of points $\{x_{n_k}\}$ in X so that

$$|f_{n_k}(x_{n_k}) - f(x_{n_k})| \ge \epsilon_0.$$

EXAMPLE 1.2. (1) $f_n(x) = x^n$, $x \in [0, 1]$, doesn't uniformly converge to its pointwise limit

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

For this, just notice that for each n, there exists some $x_n \in (0, 1)$ so that $(x_n)^n = \frac{1}{2}$. Then it follows

$$|f_n(x_n) - f(x_n)| = |\frac{1}{2} - 0| = \frac{1}{2}.$$

(2)

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

THEOREM 1.3. *Consider a sequence of functions $\{f_n\}$ defined over X which pointwise converges to f. Define*

$$\epsilon_n := \sup_{x \in X} |f_n(x) - f(x)|.$$

Then $f_n \overset{X}{\rightrightarrows} f$ if and only if $\lim_{n\to\infty} \epsilon_n = 0$.

THEOREM 1.4 (Cauchy's Criterion). *Consider sequence of functions $\{f_n\}$ defined over X. It uniformly converges to some function over X, if and only if it satisfies the following Cauchy sequence condition: For any $\epsilon > 0$, there exists some $N > 0$ (which is independent of $x \in X$) so that any $m, n > N$,*

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

for any $x \in X$.

PROOF. (1) Uniform convergence sequence satisfies the Cauchy condition, since the following inequality

$$|f_m(x) - f_n(x)| \le |f_m(x) - f(x)| + |f(x) - f_n(x)|.$$

(2) Assume $\{f_n\}$ satisfies the Cauchy sequence condition. Then for each $x \in X$, $\{f_n(x)\}$ is a Cauchy sequence in $\mathbb{R}$. By the completeness of $\mathbb{R}$, there exists a function $f : X \to \mathbb{R}$ so that $\{f_n\}$ converges to f pointwisely. Then we only need to show the convergence is in fact uniformly.

For this, for any $\epsilon > 0$, take $N > 0$ so that any $m, n > N$,

$$|f_m(x) - f_n(x)| < \epsilon.$$

Then

$$|f_n(x) - f(x)| \le |f_n(x) - f_m(x)| + |f_m(x) - f(x)| < \epsilon + |f_m(x) - f(x)|.$$

It follows

$$|f_n(x) - f(x)| \le \lim_{m\to\infty} (\epsilon + |f_m(x) - f(x)|) = \epsilon.$$

We are done.

□

THEOREM 1.5 (Weierstrass-M test). *Assume $\{a_n(x)\}$ is a sequence of functions defined over X and assume that there exists some sequence $\{M_n\}$ in $\mathbb{R}$ so that*

$$|a_n(x)| \leq M_n, \quad \text{for any } x \in X.$$

Then $\Sigma a_n(x)$ uniformly converges over X, if $\Sigma_n M_n$ is convergent.

PROOF. This immediately follows from Cauchy criterion for uniform convergence. □

Use the Weierstrass-M test, we obtain the following important result for power series.

THEOREM 1.6. *Assume $\Sigma_{n=0} c_n x^n$ is a power series with convergence radius $R \in [0, \infty]$. Then for any $0 < \epsilon < R$, the series convergences uniformly over $[-R+\epsilon, R-\epsilon]$.*

PROOF. It immediately follows from the Weierstrass-M test based on the estimates

$$|c_n x^n| \leq |c_n|(R - \frac{\epsilon}{2})^n \quad \text{for any} \quad x \in [-R+\epsilon, R-\epsilon],$$

and the convergence of the series $\Sigma|c_n|(R-\frac{\epsilon}{2})^n$ (why is this series convergent?).

□

2. Uniform convergence and continuity

THEOREM 2.1. *Assume X is a metric space and $f_n \overset{S}{\rightrightarrows} f$, where S is a subset of X. If x_0 is a limit point of E, and at $x_0 \in X$, for each n, the limit*

$$\lim_{x \to x_0} f_n(x) =: A_n$$

exist, then $\{A_n\}$ converges and

$$\lim_{n\to\infty} A_n = \lim_{x \to x_0} f(x).$$

In another word,

$$\lim_{n\to\infty} \lim_{x\to x_0} f_n(x) = \lim_{x\to x_0} \lim_{n\to\infty} f_n(x).$$

An immediate corollary is that a uniformly convergent sequence can pass the continuity to the limiting function.

COROLLARY 2.2. *Assume X is a metric space and $f_n \overset{X}{\rightrightarrows} f$. If each f_n is continuous, then f is also continuous.*

PROOF. For this case, A_n in Theorem 2.1 is taken as $f_n(x_0)$. □

PROOF OF THEOREM 2.1. (1) To show $\{A_n\}$ is convergent, it is enough to show it is a Cauchy sequence. Take any $\epsilon > 0$, since $f_n \overset{S}{\rightrightarrows} f$, it follows from Theorem 1.4 that there exists some $N > 0$ so that any $m, n > N$

$$|f_m(x) - f_n(x)| < \epsilon, \quad \text{for any } x \in S.$$

At the same time, because

$$\lim_{x\to x_0} f_m(x) = A_m, \quad \lim_{x\to x_0} f_n(x) = A_n,$$

there exists some $\delta > 0$ (may depend on m, n) so that any $x \in S$ with $0 < |x - x_0| < \delta$, there are

$$|f_m(x) - A_m| < \epsilon, \quad |f_n(x) - A_n| < \epsilon.$$

Now apply the triangle inequality as follows, where $x \in S$ with $0 < |x - x_0| < \delta$ is fixed,

$$|A_m - A_n| \le |A_m - f_m(x)| + |f_m(x) - f_n(x)| + |f_n(x) - A_n| \le 3\epsilon.$$

We have proved that $\{A_n\}$ is a Cauchy sequence.

(2) Denote by $A = \lim_{n\to\infty} A_n$. We now show that $\lim_{x\to x_0} f(x) = A$.

For any $\epsilon > 0$, take $N > 0$ so that $|A_N - A| < \epsilon$. Then for such N, take $x \in S$ in a neighborhood of x_0 so that

$$|f_N(x) - f(x)| < \epsilon \quad \text{and } |f_N(x) - A_N| < \epsilon.$$

Then the convergence follows from the estimates:

$$|f(x) - A| \le |f(x) - f_N(x)| + |f_N(x) - A_N| + |A_N - A| < 3\epsilon.$$

□

The above theorems have a topological interpretation.

Consider a compact metric space X. Define $C^0(X)$ as the set of real-valued continuous functions over X. This set is closed under addition, subtraction and scaler multiplication.

Over $C^0(X)$, there is a natural norm defined as

$$\|f\|_{C^0} := \sup_{x\in X} |f(x)|.$$

(The well-definedness of the this norm is by Theorem 3.5 from Chapter 2.) Using $\|\cdot\|_{C^0}$, we define a metric

$$d_{C^0}(f,g) := \|f - g\|_{C^0}.$$

(Check this is a metric.) From Theorem 1.3, we know a sequence of points $\{f_n\}$ in the metric space $(C^0(X), d_{C^0})$ which converges to some $f \in C^0(X)$, if and only if it uniformly converges to f over X.

Moreover, in fact we have proved that

THEOREM 2.3. *The metric space $(C^0(X), d_{C^0})$ is complete, i.e., every Cauchy sequence in this metric space is convergent.*

PROOF. Assume $\{f_n\}$ is a Cauchy sequence in the metric space $(C^0(X), d_{C^0})$. Then it is a uniform Cauchy sequence as a sequence of functions over X. By the Cauchy's Criterion 1.4, it uniformly converges to some function $f : X \to \mathbb{R}$. Then from Corollary 2.2, $f \in C^0(X)$.

This proves the completeness of $(C^0(X), d_{C^0})$. □

Different from the metric space $\mathbb{R}^n$ that every bounded sequence has a convergent subsequence, bounded sequences in $(C^0(X), d_{C^0})$ may not have any convergent subsequence, as shown from Example 1.2(1). However, by adding the condition of equicontinuity, a bounded closed subset in $C^0(X)$ must be (sequentially) compact.

DEFINITION 2.4. A subset $F \subset C^0(X)$ is called an equicontinuous family, if for any $\epsilon > 0$ there exists $\delta > 0$ so that any $x, y \in X$ with $d_X(x,y) < \delta$, there is

$$|f(x) - f(y)| < \epsilon.$$

EXAMPLE 2.5. (1) Try to prove the sequence in $\{f_n\}$ in Example 1.2(1) is not an equicontinuous family.

(2) For a sequence of differentiable functions $\{f_n\}$ over $[a,b]$, if is derivative sequence $\{f_n'\}$ is uniformly bounded, then it is equicontinuous.

PROOF. This follows from the mean value theorem that for any $x, y \in [a,b]$, there exists some ξ between x, y so that

$$|f_n(x) - f_n(y)| = |f'(\xi)||x-y| \le C|x-y|.$$

□

The following theorem is referred as Arzela–Ascoli lemma, which is very useful in analysis.

THEOREM 2.6 (Arzela–Ascoli Lemma). *Assume X is compact metric space. Then any uniformly bounded, equicontinuous sequence $\{f_n\}$ of functions over X has a uniformly convergent subsequence.*

(We skip the proof but it is not hard and you are welcome to have a try.)

As a corollary, we immediately get the following statement.

COROLLARY 2.7. *Assume X is compact metric space. Then a bounded closed subset F in the metric space $(C^0(X), d_{C^0})$ is (sequentially) compact, if it is equicontinuous.*

3. Uniform convergence and integration

THEOREM 3.1. *Assume $\{f_n\}$ is a sequence of functions defined over $[a,b]$ and each $f_n \in R(\alpha)([a,b])$. If $f_n \overset{[a,b]}{\rightrightarrows} f$, then $f \in R(\alpha)([a,b])$, and*

$$\lim_{n\to\infty} \int_a^b f_n d\alpha = \int_a^b f d\alpha.$$

PROOF. Define

$$\epsilon_n := \sup_{[a,b]} |f_n(x) - f(x)|.$$

If follows

$$f_n(x) - \epsilon_n \le f(x) \le f_n(x) + \epsilon_n, \qquad \text{for any } x \in [a,b]. \tag{3.1}$$

Since $f_n \overset{[a,b]}{\rightrightarrows} f$, there is

$$\lim_{n\to\infty} \epsilon_n = 0.$$

(1) We first prove $f \in R(\alpha)([a,b])$. For this, it is enough to show

$$\underline{\int_a^b} f d\alpha = \overline{\int_a^b} f d\alpha.$$

Notice, from (3.1) we have

$$\int_a^b (f_n - \epsilon_n) d\alpha \le \underline{\int_a^b} f d\alpha \le \overline{\int_a^b} f d\alpha \le \int_a^b (f_n + \epsilon_n) d\alpha, \tag{3.2}$$

then it follows

$$0 \le \overline{\int_a^b} f d\alpha - \underline{\int_a^b} f d\alpha \le 2\epsilon_n(\alpha(b) - \alpha(a)).$$

Let $\epsilon_n \to 0$, what we need follows.

(2) Now we go back to (3.2) and get

$$\int_a^b (f_n - \epsilon_n) d\alpha \le \int_a^b f d\alpha \le \int_a^b (f_n + \epsilon_n) d\alpha$$

and then

$$|\int_a^b f d\alpha - \int_a^b f_n d\alpha| \le \epsilon_n(\alpha(b) - \alpha(a)).$$

Take $n \to \infty$, we obtain

$$\lim_{n\to\infty} \int_a^b f_n d\alpha = \int_a^b f d\alpha.$$

□

COROLLARY 3.2. *Assume $a_n \in R(\alpha)([a,b])$ and*

$$f(x) := \Sigma_{n=0}^{\infty} a_n(x)$$

converges uniformly. Then it follows

$$\int_a^b f d\alpha = \Sigma_{n=0}^{\infty} \int_a^b a_n d\alpha.$$

PROOF. Consider the sequence of partial sums

$$f_n(x) := \Sigma_{k=0}^{n} a_k(x), \quad n = 0, 1, \cdots.$$

It follows $f_n \in R(\alpha)([a,b])$ and $f_n \overset{[a,b]}{\rightrightarrows} f$. Apply Theorem 3.1 to $\{f_n\}$, the conclusion follows. □

EXAMPLE 3.3. Prove that $\int_0^a e^x dx = e^a - 1$, $a \ge 0$, from the definition of e^x as $e^x := \Sigma_{n=0}^{\infty} \frac{x^n}{n!}$.

PROOF. The convergence radius of $e^x := \Sigma_{n=0}^{\infty} \frac{x^n}{n!}$ is $R = +\infty$. Hence by Theorem 1.6, this series converges uniformly over $[0, a]$. Each term $\frac{x^n}{n!}$ is Riemann integrable. Then the integration follows from direct calculation term by term and Corollary 3.2. □

4. Uniform convergence and differentiation

THEOREM 4.1. *Assume $\{f_n\}$ is a sequence of functions defined over $[a,b]$ and differentiable. If $\{f_n'\}$ uniformly converges on $[a,b]$ and $\{f_n\}$ converges at some point $x_0 \in [a,b]$, then $\{f_n\}$ uniformly converges on $[a,b]$ to some function f. Moreover, f is differentiable and*

$$f'(x) = \lim_{n\to\infty} f_n'(x), \quad \textit{for any } x \in [a,b].$$

PROOF. (1) Using the Cauchy criterion, we prove the convergence of $\{f_n\}$ by showing it is a uniform Cauchy sequence.

First, the convergence of $\{f_n(x_0)\}$ implies it is a Cauchy sequence in $\mathbb{R}$. Then for any $\epsilon > 0$, there exists $N > 0$ so that

$$|f_n(x_0) - f_m(x_0)| < \epsilon, \quad \text{for any} \quad n, m > N.$$

Second, for each such $n, m > N$, the function $f_n - f_m$ is differentiable over $[a,b]$, hence apply the mean value theorem, we obtain

$$(f_n(x) - f_m(x)) - (f_n(x_0) - f_m(x_0)) = (f_n'(\xi) - f_m'(\xi))(x - x_0)$$

for some ξ living between x_0 and x, where $x \in [a,b]$ is an arbitrary point.

The uniform convergence of $\{f_n'\}$ implies it is a uniform Cauchy sequence. Hence, we can make N bigger if necessary so that

$$|f_n'(\xi) - f_m'(\xi)| \leq \epsilon.$$

Then using the triangle inequality

$$\begin{aligned} |f_n(x) - f_m(x)| &\leq |(f_n(x) - f_m(x)) - (f_n(x_0) - f_m(x_0))| + |f_n(x_0) - f_m(x_0)| \\ &\leq \epsilon|x - x_0| + \epsilon \leq \epsilon((b-a)+1). \end{aligned}$$

We are done in proving $\{f_n\}$ is a uniform Cauchy sequence.

(2) Denote by f the limiting of $\{f_n\}$. Now we prove it is differentiable and

$$f'(x) = \lim_{n\to\infty} f_n'(x), \qquad \text{for any } x \in [a,b].$$

Take any $x \in [a,b]$, denote by

$$\phi_n(y) := \frac{f_n(y) - f_n(x)}{y - x}, \quad y \in [a,b] \setminus \{x\},$$

and by

$$\phi(y) := \frac{f(y) - f(x)}{y - x}, \quad y \in [a,b] \setminus \{x\}.$$

Notice that by the mean value theorem

$$\begin{aligned} |\phi_n(y) - \phi_m(y)| &= \frac{|(f_n(y) - f_m(y)) - (f_n(x) - f_m(x))|}{|y-x|} \\ &= \frac{|f_n'(\xi) - f_m'(\xi)||y-x|}{|y-x|} \\ &= |f_n'(\xi) - f_m'(\xi)|, \end{aligned}$$

for some ξ living between x and y. Then we obtain that $\{\phi_n\}$ is a uniform Cauchy sequence over $[a,b]\setminus\{x\}$ from the assumption that $\{f_n'\}$ is a uniform Cauchy sequence (since it uniformly convergent).

Moreover, from the convergence $f_n \overset{[a,b]}{\rightrightarrows} f$, it follows $\phi_n \overset{[a,b]\setminus\{x\}}{\rightrightarrows} \phi$. Also notice that by assumption

$$\lim_{y\to x} \phi_n(y) = f_n'(x).$$

We can now apply Theorem 2.1 to the sequence $\{\phi_n\}$ over $[a,b] \setminus \{x\}$, and thus obtain the change of limiting processes as

$$\lim_{n\to\infty} f_n'(x) = \lim_{y\to x} \phi(y).$$

This proves that

$$\phi'(x) := \lim_{y\to x} \phi(y)$$

exists and the same as $\lim_{n\to\infty} f_n'(x)$.

$\square$

Again, apply this theorem to series, we obtain the following corollary.

COROLLARY 4.2. *Assume $\{a_n(x)\}$ is a sequence of functions defined over $[a, b]$ and differentiable. If $\Sigma a_n'(x)$ uniformly converges on $[a, b]$ and $\Sigma a_n(x_0)$ converges at some point $x_0 \in [a, b]$, then $\Sigma a_n(x)$ uniformly converges on $[a, b]$ to some function f. Moreover, f is differentiable and*

$$f'(x) = \Sigma a_n'(x), \quad \text{for any } x \in [a, b].$$

EXAMPLE 4.3. Prove that $(e^x)' = e^x$ for any $x \in \mathbb{R}$.

PROOF. Take any $x_0 \in \mathbb{R}$ and an interval $[a, b]$ that containing x_0 as an interior point. Take

$$a_n(x) = \frac{x^n}{n!}, \quad n = 0, 1, 2, \cdots,$$

and then

$$a_n'(x) = \frac{x^{n-1}}{(n-1)!}, \quad n = 1, 2, \cdots, \quad a_0'(x) = 0,$$

and

$$\Sigma a_n'(x_0) = e^{x_0}.$$

Since the series $\Sigma_{n=0} a_n'(x)$ has convergent radius $R = +\infty$, it is uniformly convergent over $[a, b]$ by Theorem 1.6.

Apply Corollary 4.2 with so defined $\{a_n(x)\}$, we obtain

$$\frac{de^x}{dx}|_{x=x^0} = \Sigma a_n'(x_0) = e^{x_0}$$

for any $x_0 \in \mathbb{R}$.

□

REMARK 4.4. The uniform convergence of $\{f_n\}$ doesn't imply the uniform convergence of $\{f_n'\}$, even assuming each f_n is differentiable.

For example, the sequence $\{\frac{\sin nx}{n}\}$ uniformly converges to 0 over $\mathbb{R}$ and each $\frac{\sin nx}{n}$ is differentiable with

$$(\frac{\sin nx}{n})' = \cos nx.$$

The sequence $\{\cos nx\}$ is not convergent even in pointwise sense.

5. The Stone-Weierstrass approximation theorem (*This is not required for the final.)

The following result which was first shown by Weierstrass and then generalized by Stone to a more general statement is a very useful result in understanding (complex or real valued) continuous functions over a closed interval. We state the result and sketch the proof. For details and Stone's generalization, you may refer Rudin's book P159–165.

THEOREM 5.1 (Stone–Weierstrass approximation theorem). *For any $f \in C^0([a, b])$, there exists a sequence of polynomials functions $\{P_n\}$ over $[a, b]$ which uniformly converges to f.*

SKETCH OF THE PROOF. WLOG, we can assume $[a, b] = [0, 1]$ and $f(0) = f(1) = 0$. We continuously extend such f to $\mathbb{R}$ by defining

$$f(x) = 0, \quad x \notin [0, 1].$$

A sequence of functions called Landau kernel functions defined as

$$Q_n(x) = c_n(1 - x^2)^n, \quad \int_{-1}^{1} Q_n(x)dx = 1, \quad n = 1, 2, \cdots,$$

play an essential role in the construction of $\{P_n\}$.

To be concrete, define P_n as the convolution of f with Q_n, i.e.,

$$P_n(x) = (f * Q_n)(x) := \int_{-1}^{1} f(x+t)Q_n(t)dt, \quad x \in [0,1].$$

A key advantage of introducing such Q_n is that every P_n is a polynomial. Moreover, $\{P_n\}$ uniformly converges to f as we want.

□

Because polynomials are among the simplest functions, and because computers can directly evaluate polynomials, this theorem has both practical and theoretical relevance, especially in polynomial interpolation.