Calculus
Lecture notes, Fall 2024 and Spring 2025
(Version: July 1, 2024)

Pu-Zhao Kow

Department of Mathematical Sciences, National Chengchi University
Email address: pzkow@g.nccu.edu.tw

Preface

This lecture note is prepared for the course *Calculus* for undergraduate level during Fall 2024 (113-1, 000713011) and Spring 2025 (113-2, 000713012). The main purpose of this lecture note is to highlight some fundamental facts rather than all full details. In order to avoid too much technical details, the proof of some results in this lecture will be omit, one can see e.g. the monograph [**Apo74, Rud87**] for rigorous proofs of all results. The notations and terminologies in this lecture note, which will be used throughout the course, may differ to other monographs, including other textbooks [**HB10, SCW21**]. This lecture note may updated during the course.

Title. Calculus (Fall 2024, 3 credits)

Lectures (113-1, 000713011). Thursday (13:10–14:00, 14:10–15:00 15:10–16:00). Begins at September 9, 2024 and ends at January 10, 2025.

Language. Chinese and English. Materials will be prepared in English.

Instructor. Pu-Zhao Kow (Email: `pzkow@g.nccu.edu.tw`)

Office hour. Thursday (16:10–17:00)

Teaching Assistant. TBA

Acknowledgments. I would like to give special thanks to students who pointed out my mistakes in this note.

Some difficult materials are included in this lecture note for those interested in mathematics. I understand that it is not possible to remember all details, and I will not going to teach proofs in this course, however, one should at least remember basic definitions and some basic lemmas/propositions/theorems, and know how to utilize them first. In order to do so, I choose some examples and exercises to highlight what you should remember, therefore the quizzes and exams will be prepared based on **Examples** and **Exercises** in this lecture note, not necessarily identical, may slightly change to make the questions interested if necessarily. One should remember the principal rather than the exact formula. You may use methods which I have not taught, but always state the name of the theorem you used and check sufficient conditions carefully.

Contents

Part 1

Fall 2024 (113-1, 000713011)

CHAPTER 1

Set and functions

1.1. A quick and informal introduction of mathematical logic

In logic, a *logical connective* (also called a *logical operator*) is used to connect logical formulas. Some commonly used logical connectives are:

- *negation* (not), denoted as $\neg$
- *conjunction* (and), denoted as $\wedge$
- *disjunction* (or), denoted as $\vee$
- *implication* (if $\cdots$ then), denoted as $\implies$
- *equivalence* (if and only if), denoted as $\iff$.

It is also common to consider the *always true* formula and the *always false* formula to be logical connectives as well:

- *always true* formula, denoted as T
- *always false* formula, denoted as F

Now the above logical connectives can be summarized in the following table:

P	Q	$\neg P$	$P \wedge Q$	$P \vee Q$	$P \implies Q$	$P \iff Q$
T	T	F	T	T	T	T
T	F	F	F	T	F	F
F	T	T	F	T	T	F
F	F	T	F	F	T	T

TABLE 1. Truth table

It is important to see that (which very frequently used in mathematics)

$$P \iff Q \quad \equiv \quad (P \implies Q) \wedge (Q \implies P),$$

as well as

$$\neg(P \implies Q) \quad \equiv \quad \neg Q \implies \neg P. \tag{1.1.1}$$

We now give a simple example to demonstrate the meaning of the truth table in Table 1.

EXAMPLE 1.1.1. We denote P the event "it is raining" and denote Q the event "the floor is wet".

- The negation $\neg P$ becomes the event "it is *not* raining" and the negation $\neg Q$ becomes the event "the floor is *not* wet".
- The conjunction $P \wedge Q$ is the event "it is raining *and* the floor is wet".
- The disjunction $P \vee Q$ is the event "it is raining *or* the floor is wet".
- The implication $P \implies Q$ is the event "*If* it is raining, *then* the floor is wet".
- The equivalence $P \implies Q$ is the event "It is raining, *if and only if* the floor is wet".

Lets give more explanation on the implication $P \implies Q$. Each row can be summarized in the following sentence:

- If it is raining (i.e. $P = \mathsf{T}$), then the floor is wet (i.e. $Q = \mathsf{T}$). This implication is true (i.e. $(P \implies Q) = \mathsf{T}$).
- If it is raining (i.e. $P = \mathsf{T}$), then the floor is not wet (i.e. $Q = \mathsf{F}$). This implication is false (i.e. $(P \implies Q) = \mathsf{F}$).
- If it is not raining (i.e. $P = \mathsf{F}$), then the floor is wet (i.e. $Q = \mathsf{T}$). This implication is true (i.e. $(P \implies Q) = \mathsf{T}$).
- If it is not raining (i.e. $P = \mathsf{F}$), then the floor is not wet (i.e. $Q = \mathsf{F}$). This implication is true (i.e. $(P \implies Q) = \mathsf{T}$).

We see that if it is not raining (i.e. $P = \mathsf{F}$), no matter the floor is wet or not, the implication is always true. This means that, if the assumption is not true, then whatever you say is always true, but the sentence is basically a nonsense. This is exactly reflected in the truth table (Table 1). Therefore, *always check the assumptions before using theorems*.

DEFINITION 1.1.2. If the implication $P \implies Q$ is true, then we say that P is a *sufficient condition* of Q (or P guarantees Q) and in view of the *contrapositive statement* (1.1.1), we also say that Q is a *necessary condition* of P.

EXAMPLE 1.1.3. We denote P the event "you study hard" and denote Q the event "you pass the course". I believe that the implication $Q \implies P$ is true, which means that "for those students who pass the course, they are studied hard". In view of the contrapositive statement (1.1.1), this implication reads $\neg P \implies \neg Q$, which means that "for those students who do not study hard will fail the course". The implication $P \implies Q$ means "if you study hard, then you will pass the exam", *which I believe to be false*. In fact, "there is no any guarantee that you can pass this course even you study hard". In terms of Definition 1.1.2, P is a necessary condition of Q:

if you want to pass this course, you at least have to study hard.

However, P is not a sufficient condition of Q:

even you study hard, there is no any guarantee to pass this course.

In mathematical logic, it is important to mention the following quantifiers:

- universal quantification $\forall$: which is interpreted as "given any", "for all", or "for any".
- existential quantifier $\exists$: which is interpreted as "there exists", "there is at least one", or "for some".

The negation of "the event $P(x)$ holds true for all x" is "there exists x such that the event $P(x)$ does not hold". The negation of "the event $P(x)$ holds true for some x" is "the event $P(x)$ does not hold for all x". Finally, we remind the readers that, one has to be careful about the order of the mathematical argument, analogously, you cannot swap the order of computer program.

1.2. An intuitive introduction of set theory

Set theory, more specifically *Zermelo–Fraenkel set theory*, has been the standard way to provide rigorous foundations for all branches of mathematics since the first half of the 20th century. Rather than explaining details in a rigorous way, here we will only *intuitively* introduce the set theory (`https://en.wikipedia.org/wiki/Set_(mathematics)`), since we will only use the set theory as a mathematical language. One can take a look on a lecture note [**Win10**] for more details on this topic.

DEFINITION 1.2.1. A *set* is a collection of different mathematical objects (e.g. numbers, symbols, points in space, lines, other geometrical shapes, variables, or even other sets); these objects are called *elements* or *members* of the set.

We first explain now to express a set. *Roster* or *enumeration notation* defines a set by listing its elements between curly brackets, separated by commas, for example:

$$A = \{1, 3, 4, a, \text{black}\}.$$

For sets with many elements, especially those following an implicit pattern, the list of members can be abbreviated using an ellipsis "$\cdots$", for example:

$$\{1, 2, 3, \cdots, 100\}, \quad \{a, b, c, \cdots, k\}.$$

To describe an infinite set in roster notation, an ellipsis is placed at the end of the list, or at both ends, to indicated that the list continuous forever. For example:

$$\mathbb{N} := \{1, 2, 3, \cdots\}, \quad \mathbb{Z} := \{\cdots, -3, -2, -1, 0, 1, 2, 3, \cdots\}.$$

Another way to define a set is to use a rule to determine what the elements are, for example:

$$\text{Let } \mathbb{Z}_{\geq 2} \text{ be the set whose members are integers } \geq -2.$$

Such a definition is called a *semantic description*. One also can specify a set as a selection from a larger set, determined by condition(s) on the the elements. For example,

$$\mathbb{Z}_{\geq -2} = \{n : n \in \mathbb{Z}, n \geq -2\} = \{n \in \mathbb{Z} : n \geq -2\}.$$

We usually, unless stated, assuming the following axiom:

AXIOM 1.2.2 (Extensionality). *Two sets that have precisely the same elements are equal. In other words, sets are uniquely characterized by their elements (without repeat counting the same element and without considering the order of elements). For example,* $\{1, 2, 4, 2\} = \{1, 2, 4\} = \{4, 2, 1\}$.

This axiom is just to unify the writing format of the sets, so that we can communicate using the same language. The above extensionality axiom implies the following lemma.

LEMMA 1.2.3. *There exists a unique empty set* $\emptyset$.

REMARK 1.2.4. The empty set $\emptyset$ also can be expressed in Roster notation as $\{\}$. However, one should be careful that $\emptyset$ and $\{\emptyset\}$ are different, since $\{\emptyset\}$ is a set consists of one element, which is called $\emptyset$, therefore $\{\emptyset\}$ is not an empty set.

EXAMPLE 1.2.5. We now introduce some special sets of numbers in mathematics.

- $\mathbb{N}$ is the set of all natural numbers, that is, $\mathbb{N} := \{1, 2, 3, \cdots\}$.
- $\mathbb{Z}$ is the set of all integers, that is, $\mathbb{Z} := \{\cdots, -3, -2, -1, 0, 1, 2, 3, \cdots\}$.
- For each $a \in \mathbb{R}$ and $b \in \mathbb{R}$, it also convenient to define the set $a\mathbb{Z} + b := \{am + b : m \in \mathbb{Z}\}$.
- $\mathbb{Q}$ is the set of all rational numbers, that is, $\mathbb{Q} = \{\frac{a}{b} : a \in \mathbb{Z}, b \in \mathbb{Z}, b \neq 0\}$.

- $\mathbb{R}$ is the set of all real numbers, which is the completion of $\mathbb{Q}$ with respect to the Euclidean norm $|\cdot|$, where
$$|a| := \sqrt{a^2} = \begin{cases} a & \text{if } a \geq 0, \\ -a & \text{if } a < 0. \end{cases}$$
Sometime, we also call $|\cdot|$ the *absolute value of real numbers*. We will give a precise definition in Part 2.
- For each $a, b \in \mathbb{R}$, we denote the *intervals* by
$$\begin{aligned} (a,b) &:= \{x \in \mathbb{R} : a < x < b\} \quad \text{(open interval)}, \\ [a,b] &:= \{x \in \mathbb{R} : a \leq x \leq b\} \quad \text{(closed interval)}, \\ [a,b) &:= \{x \in \mathbb{R} : a \leq x < b\}, \\ (a,b] &:= \{x \in \mathbb{R} : a < x \leq b\}. \end{aligned} \tag{1.2.1}$$
It is also convenient to write
$$(-\infty, b) := \{x \in \mathbb{R} : x < b\}, \quad (a, +\infty) := \{x \in \mathbb{R} : x > a\}, \quad (-\infty, \infty) := \mathbb{R}, \tag{1.2.2}$$
as well as
$$(-\infty, b] := \{x \in \mathbb{R} : x \leq b\}, \quad [a, +\infty) := \{x \in \mathbb{R} : x \geq a\}.$$
- Despite $\pm\infty \notin \mathbb{R}$, we still often abuse the notation by saying that "$I = (a,b)$ for some $-\infty \leq a < b \leq +\infty$", which means that I can be either (1.2.1) or (1.2.2). One can interpret the notions "$I = (a,b)$ for some $-\infty \leq a < b < +\infty$" and "$I = (a,b)$ for some $-\infty < a < b \leq +\infty$" using a similar manner.

Definition 1.2.6. Let A be a set. If a is a member in A, then we denote $a \in A$. If b is not a member in A, then we denote $b \notin A$.

Definition 1.2.7 (Basic operations). Given any two sets A and B:

- we say that A is a *subset* of B, denoted as $A \subset B$, if all elements of A also belongs to B.
- their *union* $A \cup B$ is the set of all elements that are members of A or B or both.
- their *intersection* $A \cap B$ is the set of all things that are members of both A and B. If $A \cap B = \emptyset$, then A and B are said to be *disjoint*.
- the set *difference* $A \setminus B$ is the set of all things that belong to A but not B.
- their *symmetric difference* $A \triangle B$ is the set of all things that belong to A or B but not both, that is, $A \triangle B = (A \setminus B) \cup (B \setminus A)$.
- their *Cartesian product* $A \times B$ is the set of all ordered pairs (a,b) such that $a \in A$ and $b \in B$.

Remark 1.2.8. If $A \subset B$ and $B \subset A$, then by Axiom 1.2.2 we see that A and B are equal, and we denote $A = B$.

The simple concept of set has proved enormously useful in mathematics, but paradoxes arise if no restriction are placed on how sets can be constructed, for example, the *Russell's paradox* shows that the "set of all sets that do not contain themselves", i.e.
$$\{X : X \text{ is a set and } X \notin X\} \text{ cannot exist.} \tag{1.2.3}$$
Rather than go through all details how the *Zermelo–Fraenkel set theory* excludes this situation, we will explain this philosophy using some simple examples. In practical, we usually

refer the set consists of other sets as a collection. For example, let $\mathcal{P}$ be the *collection* of all subsets in $\{a, b\}$ means that

$$\mathcal{P} = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}.$$

We intuitively view $\emptyset, \{a\}, \{b\}, \{a, b\}$ as "level-1" objects, and view $\mathcal{P}$ as "level-2" object. The elements in sets also can be viewed as "level-0" objects. We usually refer the set consists of collections (i.e. "level-2" objects) as a *superset*, which is natural to be labeled as "level-3" object. We distinguish between "$\in$" and "$\subset$" as follows:

- We write $x \in X$ for "level-0" object x (point) and for "level-1" object X (set); we write $X \in \mathcal{P}$ for "level-1" object X (set) and for "level-2" object $\mathcal{P}$ (collection), and so on.
- We write $X \subset Y$ for two "level-1" objects X and Y (sets); we write $\mathcal{P} \subset \mathcal{Q}$ for two "level-2" objects $\mathcal{P}$ and $\mathcal{Q}$ (collections).

We now see that (1.2.3) is invalid if we consider the above concept of "levels" (more precisely, the *Zermelo–Fraenkel set theory*). The union, intersection, difference and symmetric difference also can be operated for collections ("level-2" objects) as well as supersets ("level-3" objects). The Cartesian product can be operate for different "levels" of objects.

Finally, we also can explain Remark 1.2.4 in terms of "levels": The empty set $\emptyset$ is a "level-1" object, while $\{\emptyset\}$ is a "level-2" object, which is a collection consists of only one element $\emptyset$. Sometimes we also abuse the notation by denoting the empty collection ("level-2" object) as $\emptyset$, for example the superset ("level-3" object) $\{\{\emptyset\}, \emptyset\}$.

1.3. Functions

DEFINITION 1.3.1. Let X and Y be sets. A *function f from a set X to a set Y*, denoted as $f : X \to Y$, is an assignment of one *unique* element of Y to *each* element of X. In this case, the set X is called the *domain*, while Y is called the *range*. If the element $y \in Y$ is assigned to $x \in X$ by the function, one says that f maps x to y, and this is commonly write $f(x) = y$. Sometimes we also write

$$f : x \mapsto y, \quad \text{or more precisely,} \quad f : x \in X \mapsto y \in Y.$$

One may imagine a function works like a virtual machine, or simply a computer program. We highlight two main points in the above definition:

- When we input an element $x \in X$ into a function, or a "machine" f, we have to make sure that f can accept this element and the output is also valid. For example:
  - **invalid input.** the "machine" $g(x) = \sqrt{x}$ cannot process the input $x < 0$. In this case, g cannot be defined as a function on $\mathbb{R}$.
  - **invalid output.** the "machine" $h(x) = x$ is not well-defined from $\mathbb{R}$ to $\mathbb{R}_{\geq 0}$, since the range $\mathbb{R}_{\geq 0}$ is too small with respect to the domain $\mathbb{R}$.
- After we input an element $x \in X$ into a "machine" f, we must specify a *unique* output (otherwise your computer only suggests you a "fatal error").
  - For example, you input a number 1 and ask the "machine" to solve $x^2 = 1$, then the "machine" will confused it should output $x = 1$ or $x = -1$ if there is no restriction.

Therefore *it is important to mention the domain and range while writing a function* (but unfortunately many textbooks fail to do so). Here is also a reminder for beginners: *Always carefully mention "for all/for each" and "for some/there exists", and the order of sentences is also important* (just like your computer program, you cannot mess up the order).

EXAMPLE 1.3.2. Let $f : \mathbb{R} \setminus \{1\} \to \mathbb{R}$ given by $f(x) = \frac{x^2-x}{x-1}$ for all $x \in \mathbb{R} \setminus \{1\}$, and let $g : \mathbb{R} \to \mathbb{R}$ given by $g(x) = x$ for all $x \in \mathbb{R}$. One sees that

$$f(x) = g(x) \quad \text{for all } x \in \mathbb{R} \setminus \{1\},$$

which means that the functions $f : \mathbb{R} \setminus \{1\} \to \mathbb{R}$ and $g : \mathbb{R} \setminus \{1\} \to \mathbb{R}$ are identical, but the functions $f : \mathbb{R} \setminus \{1\} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ are different, since $f(1)$ is not well-defined.

We now consider the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ for all $x \in \mathbb{R}$. We see that $f(x) \geq 0$ for all $x \in \mathbb{R}$, which suggests that the range $f : \mathbb{R} \to \mathbb{R}$ is redundant, that is, $f : \mathbb{R} \to \mathbb{R}_{\geq 0}$ is also a well-defined function. This suggests us the following definition:

DEFINITION 1.3.3. A function $f : X \to Y$ is said to be *onto* or *surjective*, if for each $y \in Y$, there exists a $x \in X$ such that $f(x) = y$.

Here the choice $x \in X$ is not necessarily unique, for example the function $f : \mathbb{R} \to \mathbb{R}_{\geq 0}$ is surjective since for each $y \in \mathbb{R}_{\geq 0}$ one sees that

$$f(\sqrt{y}) = y \quad \text{and} \quad f(-\sqrt{y}) = y.$$

Let $f : X \to Y$be a function, we define its *image*

$$f(X) := \{f(x) : x \in X\}. \tag{1.3.1}$$

One see that $f : X \to f(X)$ is surjective.

On the other hand, if $f : X \to Y$ is a function, then so is $f : X_0 \to Y$ for any $X_0 \subset X$. This strongly suggests the following notion:

DEFINITION 1.3.4. A function $f : X \to Y$ is said to be *one-to-one* or *injective*, if $f(x_1) = f(x_2)$ implies $x_1 = x_2$.

REMARK. We recall a logic facts: "P implies Q" equivalent to "negative-Q" implies "negative-P". Therefore, the above definition is also equivalent to: $x_1 \neq x_2$ implies $f(x_1) \neq f(x_2)$.

DEFINITION 1.3.5. If $f : X \to Y$ is both injective and surjective, then we say that $f : X \to Y$ is *bijective*.

DEFINITION 1.3.6. Let $f : X \to Y_1$ and $g : Y_2 \to Z$ be functions. If $Y_1 \subset Y_2$, then we denote the function $g \circ f : X \to Z$ by

$$(g \circ f)(x) = g(f(x)) \quad \text{for all } x \in X,$$

which is called the *composition of f and g*.

DEFINITION 1.3.7. We say that a function $f : X \to Y$ is *invertible* if there exists a function $f^{-1} : Y \to X$ such that $(f^{-1} \circ f)(x) = x$ for all $x \in X$ and $(f \circ f^{-1})(y) = y$ for all $y \in Y$.

THEOREM 1.3.8 ([**Win10**, Lemma 3.9]). *Let $f : X \to Y$ be a function. Then it is bijective if and only if it is invertible.*

EXAMPLE 1.3.9. This example also explains the importance of stating the domain and range of functions.

- The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ for all $x \in \mathbb{R}$ is neither injective nor surjective.

- The function $f : \mathbb{R} \to \mathbb{R}_{\geq 0}$ given by $f(x) = x^2$ for all $x \in \mathbb{R}$ is surjective but not injective.
- The function $f : \mathbb{R}_{\geq 0} \to \mathbb{R}$ given by $f(x) = x^2$ for all $x \in \mathbb{R}_{\geq 0}$ is injective but not surjective.
- The function $f : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ given by $f(x) = x^2$ for all $x \in \mathbb{R}_{\geq 0}$ is bijective, with inverse function $f^{-1} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ given by $f^{-1}(y) = \sqrt{y}$ for all $y \in \mathbb{R}_{\geq 0}$.

REMARK 1.3.10. Let $f : X \to Y$ be a bijective function. If $f(x) = y$, then $f^{-1}(y) = x$. Hence it is recommend to use different variable for f and f^{-1}. For example, it is recommend to write the inverse function $f^{-1} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ of $f : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ given by $f(x) = x^2$ for all $x \in \mathbb{R}_{\geq 0}$ as $f^{-1}(y) = \sqrt{y}$ for all $y \in \mathbb{R}_{\geq 0}$ (rather than $f^{-1}(x) = \sqrt{x}$ for all $x \in \mathbb{R}_{\geq 0}$ even though this is true).

DEFINITION 1.3.11. Let I be an interval in $\mathbb{R}$.

(a) A function $f : I \to \mathbb{R}$ is called *non-decreasing* if $f(x_1) \leq f(x_2)$ for all $x_1, x_2 \in I$ with $x_1 < x_2$.
(b) A function $f : I \to \mathbb{R}$ is called *strictly increasing* if $f(x_1) < f(x_2)$ for all $x_1, x_2 \in I$ with $x_1 < x_2$.
(c) A function $f : I \to \mathbb{R}$ is called *non-increasing* if $f(x_1) \geq f(x_2)$ for all $x_1, x_2 \in I$ with $x_1 < x_2$.
(d) A function $f : I \to \mathbb{R}$ is called *strictly decreasing* if $f(x_1) > f(x_2)$ for all $x_1, x_2 \in I$ with $x_1 < x_2$.

LEMMA 1.3.12. *If $f : I \to \mathbb{R}$ is either strictly increasing or strictly decreasing, then $f : I \to f(I)$ is bijective, with inverse $f^{-1} : f(I) \to I$.*

It is easy to construct a bijective function which is neither non-decreasing nor non-increasing:

EXAMPLE 1.3.13. One sees that $f : [-1, 1) \to [-1, 1)$ by

$$f(x) = \begin{cases} -x - 1 & \text{if } -1 \leq x \leq 0, \\ -x + 1 & \text{if } 0 < x < 1, \end{cases}$$

is bijective, but neither non-decreasing nor non-increasing.

EXAMPLE 1.3.14. Here we exhibit some basic functions.

(1) A function $P : \mathbb{R} \to \mathbb{R}$ is called a *polynomial* if

$$P(x) = \sum_{j=0}^{n} a_j x^j = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 \quad \text{for all } x \in \mathbb{R}.$$

The numbers $a_0, a_1, \cdots, a_n$ are called the *coefficients*. If the leading coefficient $a_n \neq 0$, then the *degree* of the polynomial P is n, and we simply denote $\deg(P) = n$.

(2) Suggested by the polynomial, we are now interested in the power function of the form $f_p(x) = x^p$ for $p \in \mathbb{R}$, where the domain to be specify later.

 (a) For each $n \in \mathbb{N}$, one sees that $f_n : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is simply the polynomial. Since $f_n : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is strict increasing and bijective. In view of the exponential rule, we define

$$f_{\frac{1}{n}}(x) \equiv x^{\frac{1}{n}} := f_n^{-1}(x) \quad \text{for all } x \in \mathbb{R}_{\geq 0}.$$

Accordingly, we define $f_p : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ for all $p \in \mathbb{Q}_{>0}$ by

$$f_{\frac{m}{n}}(x) \equiv x^{\frac{m}{n}} := (f_{\frac{1}{n}} \circ f_m)(x) \quad \text{for all } x \in \mathbb{R}_{\geq 0}.$$

Finally, the function $f_p : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ can be defined for all $p \in \mathbb{R}_{>0}$ via the completion (will not rigorously explain this at this point).

(b) For $p < 0$, we simply define $f_p : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ by

$$f_p(x) := \frac{1}{f_{|p|}(x)} \quad \text{for all } x \in \mathbb{R}_{>0}.$$

In view of the exponential rule, we use the convention $f_0(x) = 1$ for all $x \in \mathbb{R}_{\geq 0}$.

(3) A function f is called a *algebraic function* if it can be constructed using algebraic operations (such as addition, subtraction, multiplication, division, and composite with the power function above). The functions that are not algebraic are called *transcendental*; these include the trigonometric, exponential and logarithmic functions mentioned below.

EXAMPLE 1.3.15 (Trigonometric functions). Let B be the unit ball in $\mathbb{R}^2$ with radius 1 centered at 0, and let ∂B be its boundary, i.e. the unit circle. Let π be the area of B. It is well-known that $\pi = 3.14159\cdots$, and the length of ∂B (also known as *circumference* or the *perimeter* of the unit circle). Let L_1 and L_2 are two straight line both passing through the origin, and let $P_1 := L_1 \cap \partial B$ and $P_2 := L_2 \cap \partial B$, and we see that the circle ∂B is partitioned into two parts, says Γ_1 and Γ_2, by the points P_1 and P_2. Intuitively, it is natural to define the angle between L_1 and L_2 by the length of Γ_1 or Γ_2, but however this may cause some trouble in mathematics, since this is not a function, since both choices Γ_1 and Γ_2 correspond to the same geometry. In order to make the definition rigorous, we define angle with orientation. Starting from the point P_1, which corresponds to line L_1, we rotate *counter-clockwise* and stop at P_2 (not necessarily stop at the first meeting), which corresponds to line L_2. Let Γ be the portion of ∂B during the rotation. Then we say that[1]:

the angle θ (in radian) from L_1 to L_2 is defined by the length of Γ.

Now the angle is oriented, one sees that the angle from L_2 to L_1 is $-\theta$. In addition, θ can be any value in $\mathbb{R}$. In some occasion, we sometimes refer the $|\theta|$ the (phaseless) angle between L_1 and L_2, even it is not so rigorous.

With the oriented angle at hand, we now can define the trigonometric functions, as in Figure 1.3.1 below:

[1]Since π is transcendental, it is not so convenient in some application (e.g. aviation). We usually normalize the angle as follows:

the angle $\tilde{\theta}$ (in degree) is defined by $\tilde{\theta} := \frac{360}{2\pi}\theta$, where θ is angle in radian.

The reason we choose 360 is it is dividable by many integers, including $2, 3, 4, 5, 6, 8, 9, 10, \cdots$.

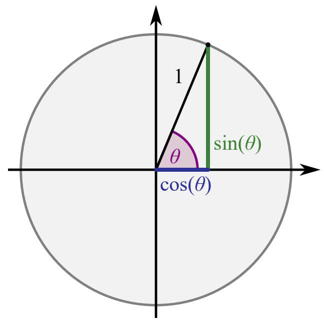

FIGURE 1.3.1. Definition of sine function and cosine function: Stephan Kulla (User:Stephan Kulla), CC0, via Wikimedia Commons

Since the angle is oriented, thus the trigonometric also has sign, for example, $\sin\theta < 0$ when $\frac{\pi}{2} < \theta < \frac{3}{2}\pi$. According to the above definition, we also see that

$$\sin : \mathbb{R} \to [-1,1] \quad \text{and} \quad \cos : \mathbb{R} \to [-1,1]$$

are both surjective functions, but not injective. The definition of sine and cosine function immediately gives

$$(\cos\theta)^2 + (\sin\theta)^2 = 1 \quad \text{for all } \theta \in \mathbb{R}.$$

Some special values are showed in Figure 1.3.2 below:

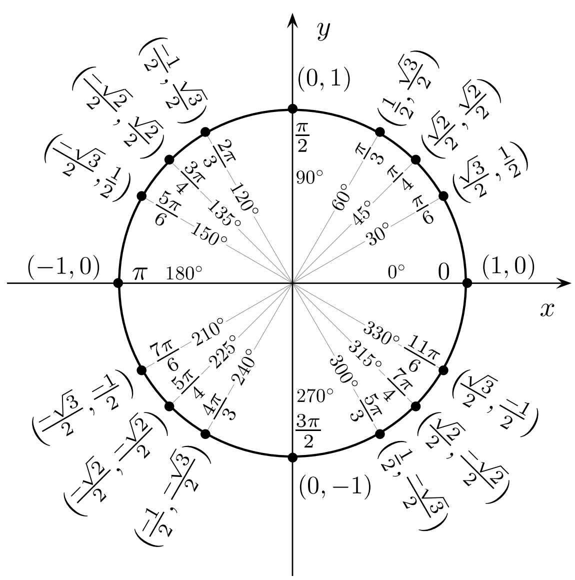

FIGURE 1.3.2. Some special values of $(\cos\theta, \sin\theta)$: Gustavb (talk · contribs), Public domain, via Wikimedia Commons

It is remarkable to mention that

$$\sin(-\theta) = -\sin\theta \quad \text{for all } \theta \in \mathbb{R},$$
$$\cos(-\theta) = \cos\theta \quad \text{for all } \theta \in \mathbb{R},$$

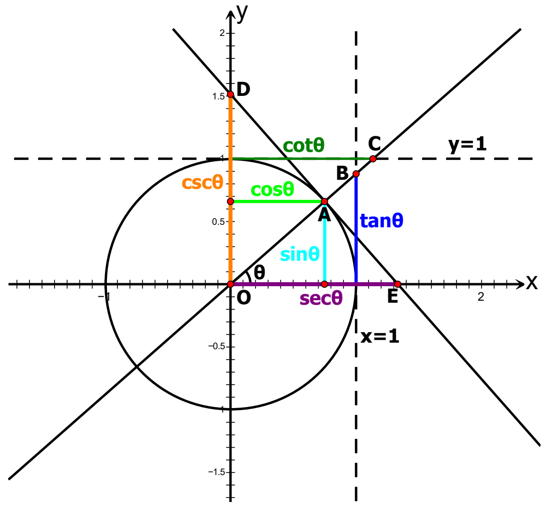

FIGURE 1.3.3. Six trigonometric functions: Onmaditque, CC BY-SA 4.0, via Wikimedia Commons

that is, the sine function is odd while the cosine function is even. The other trigonometric functions are defined as follows (see Figure 1.3.3):

$$\begin{aligned}
\tan\theta &:= \frac{\sin\theta}{\cos\theta} \quad \text{for all } \theta \in \mathbb{R} \setminus \left(\pi\mathbb{Z} + \frac{\pi}{2}\right), \\
\sec\theta &:= \frac{1}{\cos\theta} \quad \text{for all } \theta \in \mathbb{R} \setminus \left(\pi\mathbb{Z} + \frac{\pi}{2}\right), \\
\cot\theta &:= \frac{\cos\theta}{\sin\theta} \quad \text{for all } \theta \in \mathbb{R} \setminus \pi\mathbb{Z}, \\
\mathrm{cosec}\theta \equiv \csc\theta &:= \frac{1}{\sin\theta} \quad \text{for all } \theta \in \mathbb{R} \setminus \pi\mathbb{Z}.
\end{aligned}$$

One see that

$$\cot\theta = \frac{1}{\tan\theta} \quad \text{for all } \theta \in \mathbb{R} \setminus \left(\pi\mathbb{Z} \cup \left(\pi\mathbb{Z} + \frac{\pi}{2}\right)\right) = \mathbb{R} \setminus \frac{\pi}{2}\pi,$$

which means that the identity only holds true in restricted domain.

EXERCISE 1.3.16. We say that $f : \mathbb{R} \to \mathbb{R}$ is an *odd function* if $f(-x) = -f(x)$ for all $x \in \mathbb{R}$, and we say that $f : \mathbb{R} \to \mathbb{R}$ is an *even function* if $f(-x) = f(x)$ for all $x \in \mathbb{R}$. Given any function $g : \mathbb{R} \to \mathbb{R}$, show that there exists an odd function $g_{\mathrm{odd}} : \mathbb{R} \to \mathbb{R}$ and an even function $g_{\mathrm{even}} : \mathbb{R} \to \mathbb{R}$ such that

$$g(x) = g_{\mathrm{odd}}(x) + g_{\mathrm{even}}(x) \quad \text{for all } x \in \mathbb{R}.$$

In addition, if $g \not\equiv 0$ and there exists an odd function $h_{\mathrm{odd}} : \mathbb{R} \to \mathbb{R}$ and an even function $h_{\mathrm{even}} : \mathbb{R} \to \mathbb{R}$ such that

$$g_{\mathrm{odd}}(x) + g_{\mathrm{even}}(x) = h_{\mathrm{odd}}(x) + h_{\mathrm{even}}(x) \quad \text{for all } x \in \mathbb{R},$$

show that $g_{\mathrm{odd}}(x) = h_{\mathrm{odd}}(x)$ and $g_{\mathrm{even}}(x) = h_{\mathrm{even}}(x)$ for all $x \in \mathbb{R}$.

EXAMPLE 1.3.17 (Inverse trigonometric functions). In order to define inverse functions, we usually (unless stated) consider

$$\begin{aligned}\sin &: [-\pi/2, \pi/2] \to [-1, 1]\\ \cos &: [0, \pi] \to [-1, 1],\\ \tan &: (-\pi/2, \pi/2) \to \mathbb{R},\\ \cot &: (0, \pi) \to \mathbb{R},\\ \sec &: [0, \pi] \setminus \{\pi/2\} \to \mathbb{R}_{\leq -1} \cup \mathbb{R}_{\geq 1},\\ \mathrm{cosec} \equiv \csc &: [-\pi/2, \pi/2] \setminus \{0\} \to \mathbb{R}_{\leq -1} \cup \mathbb{R}_{\geq 1},\end{aligned}$$

which are bijective, and hence the corresponding inverse functions, called the *inverse trigonometric functions*, are defined as :

$$\begin{aligned}\arcsin \equiv \sin^{-1} &: [-1, 1] \to [-\pi/2, \pi/2],\\ \arccos \equiv \cos^{-1} &: [-1, 1] \to [0, \pi],\\ \arctan \equiv \tan^{-1} &: \mathbb{R} \to (-\pi/2, \pi/2),\\ \cot^{-1} &: \mathbb{R} \to (0, \pi),\\ \sec^{-1} &: \mathbb{R}_{\leq -1} \cup \mathbb{R}_{\geq 1} \to [0, \pi] \setminus \{\pi/2\},\\ \csc^{-1} &: \mathbb{R}_{\leq -1} \cup \mathbb{R}_{\geq 1} \to [-\pi/2, \pi/2] \setminus \{0\}.\end{aligned}$$

Since all other trigonometric functions can be generated by sine and cosine function, *throughout this course, we will only focus on sine and cosine functions*.

EXERCISE 1.3.18. Sketch the function $\sin^{-1} \circ \sin : \mathbb{R} \to \mathbb{R}$ and $\cos^{-1} \circ \cos : \mathbb{R} \to \mathbb{R}$.

EXAMPLE 1.3.19 (Euler formula and trigonometric functions). Here we also explain a simple way to derive trigonometric identities. We *formally* write the *imaginary number* $\mathbf{i} := \sqrt{-1}$, one can see e.g. my other lecture note [**Kow23**], which is much more advance, for a precise definition. The Euler formula reads:

$$e^{\mathbf{i}\theta} := \cos\theta + \mathbf{i}\sin\theta \quad \text{for all } \theta \in \mathbb{R}.$$

Performing some *formal* computations lead (more precisely, the *de Moivre theorem*)

$$\begin{aligned}&\cos(\theta_1 + \theta_2) + \mathbf{i}\sin(\theta_1 + \theta_2) = e^{\mathbf{i}(\theta_1+\theta_2)}\\ &\quad = e^{\mathbf{i}\theta_1}e^{\mathbf{i}\theta_2} = (\cos\theta_1 + \mathbf{i}\sin\theta_1)(\cos\theta_2 + \mathbf{i}\sin\theta_2)\\ &\quad = \cos\theta_1\cos\theta_2 + \mathbf{i}^2\sin\theta_1\sin\theta_2 + \mathbf{i}(\cos\theta_1\sin\theta_2 + \sin\theta_1\cos\theta_2)\\ &\quad = (\cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2) + \mathbf{i}(\cos\theta_1\sin\theta_2 + \sin\theta_1\cos\theta_2).\end{aligned}$$

Comparing the real and imaginary parts lead us to the *sum-to-product formula*:

$$\begin{aligned}\cos(\theta_1 + \theta_2) &= \cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2,\\ \sin(\theta_1 + \theta_2) &= \cos\theta_1\sin\theta_2 + \sin\theta_1\cos\theta_2.\end{aligned}$$

Choosing $\theta_1 = \theta_2 = \theta$ leads to the *multiple angle formula*. It is easy to obtain further generalization by consider $e^{\mathbf{i}(\theta_1+\theta_2+\theta_3)} = e^{\mathbf{i}\theta_1}e^{\mathbf{i}\theta_2}e^{\mathbf{i}\theta_3}$ and so on. From this, it is easy to derive the *product-to-sum formula*, for example,

$$\begin{aligned}&\cos(\theta_1+\theta_2)+\cos(\theta_1-\theta_2)\\&\quad=\cos\theta_1\cos\theta_2-\sin\theta_1\sin\theta_2+\cos\theta_1\cos(-\theta_2)-\sin\theta_1\sin(-\theta_2)\\&\quad=\cos\theta_1\cos\theta_2-\cancel{\sin\theta_1\sin\theta_2}+\cos\theta_1\cos\theta_2+\cancel{\sin\theta_1\sin\theta_2}\\&\quad=2\cos\theta_1\cos\theta_2.\end{aligned}$$

The other three product-to-sum formula can be easily obtained by considering

$$\cos(\theta_1+\theta_2)-\cos(\theta_1-\theta_2),\quad \sin(\theta_1+\theta_2)+\sin(\theta_1-\theta_2),\quad \sin(\theta_1+\theta_2)-\sin(\theta_1-\theta_2),$$

here we left the details for readers as an exercise.

Here we also exhibit some interesting functions, which will serves as counterexample in the future.

(a) $f:\mathbb{R}\setminus\{0\}\to\mathbb{R}$, $f(x)=\sin(1/x)$ for all $x\in\mathbb{R}\setminus\{0\}$.

(b) Given a parameter $p>0$, we consider the function $f:\mathbb{R}\setminus\{0\}\to\mathbb{R}$ is given by

$$f(x)=|x|^p\sin(1/x)\quad\text{for all } x\in\mathbb{R}\setminus\{0\}.$$

(c) $f:\mathbb{R}\to\mathbb{R}$, $f(x)=\begin{cases}1 & ,\text{if } x\in\mathbb{Q},\\ 0 & ,\text{if } x\in\mathbb{R}\setminus\mathbb{Q}.\end{cases}$

(d) $f:\mathbb{R}\to\mathbb{R}$, $f(x)=\begin{cases}\frac{1}{q} & ,\text{if } x=\frac{p}{q}\in\mathbb{Q}, q>0, \gcd(p,q)=1,\\ 0 & ,\text{if } x\in(\mathbb{R}\setminus\mathbb{Q})\cup\{0\}.\end{cases}$

The above examples demonstrate that the functions may oscillating intensely. In some case, it is even not possible to sketch, see Example 1.3.19(c)(d) above.

EXERCISE 1.3.20. Sketch Example 1.3.19(a)(b).

EXAMPLE 1.3.21 (Exponential function and logarithmic function). Given any $a>0$ with $a\neq 1$ (this case is trivial), we already showed in Example 1.3.14 that

$$f:\mathbb{R}\to\mathbb{R}_{>0},\quad f(x):=a^x \tag{1.3.2}$$

is a well-defined function. One sees that the function is strictly increasing when $a>1$, and strictly decreasing when $0<a<1$, and in fact (1.3.2) is bijective. One sees that

$$f(0)=1\quad\text{for all } a>0 \text{ with } a\neq 1.$$

We now consider the tangent line of the graph of (1.3.2). One sees that the tangent line has slope about $0.693147\cdots$ when $a=2$, and about $1.0986\cdots$ when $a=3$. In fact, there exists a unique number e, which is called the *natural exponent*, with value about $2.71828\cdots$, such that the tangent line has slope exactly 1. In this case, we usually denote the function

$$\exp:\mathbb{R}\to\mathbb{R}_{>0},\quad \exp(x):=e^x,$$

which is bijective, with inverse function

$$\ln:\mathbb{R}_{>0}\to\mathbb{R}. \tag{1.3.3}$$

called the *natural logarithmic function*. For each $p\in\mathbb{R}$ and $q\in\mathbb{R}$, since

$$\exp(\ln(x^py^p))=x^py^p=(\exp(\ln x))^p(\exp(\ln y))^q=\exp(p\ln x+q\ln y)\quad\text{for all } x,y>0,$$

then we reach the following *fundamental identity for logarithmic function*:

$$\ln(x^py^q)=p\ln x+q\ln y\quad\text{for all } x>0, y>0, p\in\mathbb{R}, q\in\mathbb{R}. \tag{1.3.4}$$

For each $a > 0$, one also may define the *logarithmic function with base a* as

$$\log_a : \mathbb{R}_{>0} \to \mathbb{R}, \quad \log_a(x) := \frac{\ln x}{\ln a} \quad \text{for all } a > 0, \tag{1.3.5}$$

which is clearly a bijective function. It is easy to extend the fundamental identity (1.3.4) for (1.3.5). In addition, one sees that

$$\log_a(a^y) = \frac{\ln a^y}{\ln a} = \frac{y \ln a}{\ln a} = y \quad \text{for all } y \in \mathbb{R}, \tag{1.3.6}$$

together with the bijection (1.3.5), we conclude that (1.3.5) is exactly the inverse function of (1.3.2). In view of (1.3.5), it is not interesting to consider arbitrary base a, throughout this course, *we will only focus on the natural logarithmic function (i.e. the logarithmic function with base e)*.

CHAPTER 2

Limits and continuity

2.1. Limit and limit superior in $\mathbb{R}$

In Example 1.3.2, we have explained that the functions $f : \mathbb{R} \setminus \{1\} \to \mathbb{R}$ given by $f(x) = \frac{x^2-x}{x-1}$ for all $x \in \mathbb{R} \setminus \{1\}$ and $g : \mathbb{R} \to \mathbb{R}$ given by $g(x) = x$ for all $x \in \mathbb{R}$ are different *in the sense of functions*, despite they looks similar *intuitively*. Despite it is not possible to define f *at* 1, but it is possible to discuss the behavior of f *near* 1. This situation suggests us the following definition:

Definition 2.1.1. Let $a, x_0, b \in \mathbb{R}$ with $a < x_0 < b$ and let $f : (a,b) \setminus \{x_0\} \to \mathbb{R}$ be a function. We say that the limit $\lim_{x\to x_0} f(x) = L \in \mathbb{R}$ exists if: Given any $\epsilon > 0$, there exists $\delta = \delta(\epsilon) > 0$ such that

$$0 < |x - x_0| < \delta \text{ implies } |f(x) - L| < \epsilon. \tag{2.1.1}$$

The implication (2.1.1) roughly means that, if $x \neq x_0$ is "sufficiently close" to x, then $f(x)$ is also "close" to the number $L \in \mathbb{R}$. Here is the main point: what is the precise meaning of "sufficient close"? The idea is: Lets find a third-party judge, which is absolutely fair, give a *tolerance level* $\epsilon > 0$ (ϵ is the Greek letter corresponding to English character "e", which represents the *error*), we then decide a *reasonable distance* $\delta > 0$ depends on the tolerance level ϵ so that (2.1.1) works. Since δ depends on ϵ (and in fact also depends on x_0), it is recommend to write $\delta = \delta(\epsilon)$ in order to emphasize (and remind yourself) the dependence on ϵ. According to the above definition, it is not difficult to see that

$$\lim_{x\to 1} f(x) = 1 \quad \text{despite } f(1) \text{ is not well-defined.}$$

This example also reminds the reader that the definition of limit *does not involve the value of* $f(x_0)$, so the function f in Definition 2.1.1 may not well-defined at x_0. One sees that the absolute value function $|\cdot|$ is simply the Euclidean norm, therefore it is natural to write the ball $B_r(x) := \{x \in \mathbb{R} : |x| < r\}$. We can rewrite (2.1.1) as:

$$x \in B_\delta(x_0) \setminus \{x_0\} \text{ implies } f(x) \in B_\epsilon(L), \tag{2.1.2}$$

or in terms of image (1.3.1), we even can write

$$f\left(B_\delta(x_0) \setminus \{x_0\}\right) \subset B_\epsilon(L).$$

Now we consider the Heaviside function $H : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ defined by

$$H(x) = \begin{cases} 1 & , x > 0, \\ 0 & , x < 0. \end{cases} \tag{2.1.3}$$

Intuitively, we may expect the limit $\lim_{x\to 0} H(x)$ does not exist. However, this suggests the following definition.

DEFINITION 2.1.2. Let $a, x_0, b \in \mathbb{R}$ with $a < x_0 < b$ and let $f : (x_0, b) \to \mathbb{R}$ and $g : (a, x_0) \to \mathbb{R}$ be functions. We say that the *right limit* $\lim_{x\to x_0+} f(x) = L \in \mathbb{R}$ exists if: Given any $\epsilon > 0$, there exists $\delta = \delta(\epsilon) > 0$ such that

$$0 < |x - x_0| < \delta \text{ and } x > x_0 \text{ together imply } |f(x) - L| < \epsilon. \tag{2.1.4}$$

Similarly, we say that the left limit $\lim_{x\to x_0-} g(x) = L \in \mathbb{R}$ exists if: Given any $\epsilon > 0$, there exists $\delta = \delta(\epsilon) > 0$ such that

$$0 < |x - x_0| < \delta \text{ and } x < x_0 \text{ together imply } |g(x) - L| < \epsilon. \tag{2.1.5}$$

It is easy to see that (2.1.4) is equivalent to

$$0 < x - x_0 < \delta \text{ implies } |f(x) - L| < \epsilon,$$

and similarly (2.1.5) is equivalent to

$$-\delta < x - x_0 < 0 \text{ implies } |f(x) - L| < \epsilon.$$

From the definition it is not difficult to see that:

LEMMA 2.1.3. *Let $a, x_0, b \in \mathbb{R}$ with $a < x_0 < b$ and let $f : (a, b) \setminus \{x_0\} \to \mathbb{R}$ be a function.*

- *If $\lim_{x\to x_0} f(x) \in \mathbb{R}$ exists, then both $\lim_{x\to x_0+} f(x) \in \mathbb{R}$ and $\lim_{x\to x_0-} f(x) \in \mathbb{R}$ exist and*

$$\lim_{x\to x_0} f(x) = \lim_{x\to x_0+} f(x) = \lim_{x\to x_0-} f(x). \tag{2.1.6}$$

- *If both $\lim_{x\to x_0+} f(x) \in \mathbb{R}$ and $\lim_{x\to x_0-} f(x) \in \mathbb{R}$ exist and $\lim_{x\to x_0+} f(x) = \lim_{x\to x_0-} f(x)$, then $\lim_{x\to x_0} f(x) \in \mathbb{R}$ exists and satisfy (2.1.6).*

In view of the above notions, we now see that the Heaviside function (2.1.3) satisfies

$$\lim_{x\to 0+} H(x) = 1 \quad \text{and} \quad \lim_{x\to 0-} H(x) = 0,$$

thus according to Lemma 2.1.3 we conclude that $\lim_{x\to 0} H(x)$ does not exist since the left and right limits are not identical. As an immediate consequence, we also see that, if either left limit or right limit does not exist, then we immediately know that the limit does not exist.

LEMMA 2.1.4 (Basic properties of limits). *Let $a, x_0, b \in \mathbb{R}$ with $a < x_0 < b$ and let $g_1 : (a, b) \setminus \{x_0\} \to \mathbb{R}$ as well as $g_2 : (a, b) \setminus \{x_0\} \to \mathbb{R}$. If both limits $\lim_{x\to x_0} g_1(x)$ and $\lim_{x\to x_0} g_2(x)$ exist in $\mathbb{R}$, then the following holds true:*

(a) *for each $c_1 \in \mathbb{R}$ and $c_2 \in \mathbb{R}$the limit $\lim_{x\to x_0}(c_1 g_1(x) + c_2 g_2(x))$ exists in $\mathbb{R}$ and satisfies*

$$\lim_{x\to x_0} (c_1 g_1(x) + c_2 g_2(x)) = c_1 \lim_{x\to x_0} g_1(x) + c_2 \lim_{x\to x_0} g_2(x) \quad \textit{(linearity)}.$$

(b) *if $g_1(x) \le g_2(x)$ for all $x \in (a, b) \setminus \{x_0\}$, then*

$$\lim_{x\to x_0} g_1(x) \le \lim_{x\to x_0} g_2(x) \quad \textit{(monotonicity)}.$$

(c) *the limit $\lim_{x\to x_0}(g_1(x)g_2(x))$ exists in $\mathbb{R}$ and satisfies*

$$\lim_{x\to x_0} (g_1(x)g_2(x)) = \left(\lim_{x\to x_0} g_1(x)\right)\left(\lim_{x\to x_0} g_2(x)\right).$$

(d) *if we additionally assume that* $\lim_{x\to x_0} g_2(x) \neq 0$*, then the limit* $\lim_{x\to x_0} \frac{g_1(x)}{g_2(x)}$ *exists in* $\mathbb{R}$ *and satisfies*

$$\lim_{x\to x_0} \frac{g_1(x)}{g_2(x)} = \frac{\lim_{x\to x_0} g_1(x)}{\lim_{x\to x_0} g_2(x)}.$$

Similar result also holds true for left and right limits.

In general, it is not easy to check that whether the limit exists or not using a rigorous mathematical formula, one simple way is to proof the existence by using the existence of other functions. The following lemma is an immediate consequence of Definition 2.1.1.

LEMMA 2.1.5 (Squeeze theorem). *Let* $a, x_0, b \in \mathbb{R}$ *with* $a < x_0 < b$ *and let* $g_1 : (a,b) \setminus \{x_0\} \to \mathbb{R}_{\geq 0}$*,* $g_2 : (a,b) \setminus \{x_0\} \to \mathbb{R}_{\geq 0}$ *as well as* $f : (a,b) \setminus \{x_0\} \to \mathbb{R}_{\geq 0}$*. If*

$$g_1(x) \leq f(x) \leq g_2(x) \text{ for all } x \in (a,b) \setminus \{x_0\},$$

both $\lim_{x\to x_0} g_1(x)$ *and* $\lim_{x\to x_0} g_1(x)$ *exist in* $\mathbb{R}$ *satisfying*

$$\lim_{x\to x_0} g_1(x) = \lim_{x\to x_0} g_1(x),$$

then $\lim_{x\to x_0} f(x)$ *exists and*

$$\lim_{x\to x_0} f(x) = \lim_{x\to x_0} g_1(x) = \lim_{x\to x_0} g_1(x).$$

Similar result also holds true for right and left limits.

However, in many cases, for example the function

$$f : \mathbb{R} \to \mathbb{R}, \quad f(x) = \begin{cases} 1 & , x \in \mathbb{Q}, \\ 0 & , x \in \mathbb{R} \setminus \mathbb{Q}, \end{cases} \tag{2.1.7}$$

consider in Example 1.3.19, it is not easy to check that the left and right limits $\lim_{x\to 0+} f(x)$ and $\lim_{x\to 0-} f(x)$ does not exist using a rigorous mathematical formulation, despite it is not easy to guess intuitively. From Definition 2.1.1, one sees that

$$\lim_{x\to x_0} f(x) = L \in \mathbb{R} \text{ exists if and only if } \lim_{x\to x_0} |f(x) - L| = 0,$$

that is, given any $\epsilon > 0$, there exists a $\delta = \delta(\epsilon) > 0$ such that

$$x \in B_\delta(x_0) \setminus \{x_0\} \text{ implies } |f(x) - L| < \epsilon.$$

This observation suggests us the following definition:

DEFINITION 2.1.6. Let $a, x_0, b \in \mathbb{R}$ with $a < x_0 < b$ and let $g : (a,b) \setminus \{x_0\} \to \mathbb{R}_{\geq 0}$ be a function. We define the *limit superior* or *upper limit* by

$$\limsup_{x\to x_0} g(x) := \lim_{r\to 0+} \left(\sup_{B_r(x_0)\setminus\{x_0\}} g \right), \tag{2.1.8}$$

where

$$\sup_{B_r(x_0)\setminus\{x_0\}} g = \inf \{ M : M > g(x) \text{ for all } x \in B_r(x_0) \setminus \{x_0\} \},$$

and the infimum is understood in the limit sense. The *limit superior from right* or *upper limit from right* is defined by

$$\limsup_{x\to x_0+} g(x) := \lim_{r\to 0+} \left(\sup_{(x_0, x_0+r)} g \right),$$

and similarly the *limit superior from left* or *upper limit from left* is defined by

$$\limsup_{x\to x_0-} g(x) := \lim_{r\to 0+}\left(\sup_{(x_0-r,x_0)} g\right).$$

One sees that the function

$$\phi : \mathbb{R}_{>0} \to \mathbb{R}_{\geq 0}, \quad \phi(r) = \sup_{B_r(x_0)\setminus\{x_0\}} g$$

is monotone non-increasing, therefore the limit (2.1.1) always exist in $\mathbb{R}_{\geq 0}$ in the sense of Definition 2.1.1. We now introduce the following powerful lemma (we will extend this lemma later in Section 2.3):

LEMMA 2.1.7. *Let $a, x_0, b \in \mathbb{R}$ with $a < x_0 < b$ and let $f : (a,b)\setminus\{x_0\} \to \mathbb{R}$ be a function.*

(a) *If $\lim_{x\to x_0} f(x) = L \in \mathbb{R}$ exists, then $\limsup_{x\to x_0} |f(x) - L| = 0$.*
(b) *If $\limsup_{x\to x_0} |f(x) - L| = 0$ for some $L \in \mathbb{R}$, then $\lim_{x\to x_0} f(x)$ exists and $\lim_{x\to x_0} f(x) = L$.*
(c) *If $f \geq 0$ for all $x \in (a,b)\setminus\{x_0\}$ and $\lim_{x\to x_0} f(x) = L \in \mathbb{R}$ exists, then $\limsup_{x\to x_0} f(x) = L$.*

Similar results also hold true for right limit/limit superior from right as well as left limit/limit superior from left.

EXAMPLE 2.1.8. We now consider the function given in (2.1.7). If $L \leq \frac{1}{2}$, then for each $r > 0$, one can choose $x' \in B_r(x_0)\setminus\{x_0\}$ such that $f(x') = 1$, and hence

$$\sup_{x\in B_r(x_0)\setminus\{x_0\}} |f(x) - L| \geq |f(x') - L| = 1 - L \geq \frac{1}{2},$$

and thus $\limsup_{x\to x_0} |f(x) - L| \geq \frac{1}{2}$. Otherwise, if $L > \frac{1}{2}$, then for each $r > 0$, one can choose $x'' \in B_r(x_0)\setminus\{x_0\}$ such that $f(x'') = 0$, and hence

$$\sup_{x\in B_r(x_0)\setminus\{x_0\}} |f(x) - L| \geq |f(x'') - L| = L \geq \frac{1}{2},$$

and thus $\limsup_{x\to x_0} |f(x) - L| \geq \frac{1}{2}$. This means that, given any $L \in \mathbb{R}$, one always has

$$\limsup_{x\to x_0} |f(x) - L| \neq 0,$$

which concludes that $\lim_{x\to x_0} f(x)$ does not exist according to Lemma 2.1.7. We will later give a simpler proof in Example 2.3.12 below after expanding the definition of limits (Definition 2.1.1 and Definition 2.1.2).

EXERCISE 2.1.9. Let $a, x_0, b \in \mathbb{R}$ with $a < x_0 < b$ and let $g_1 : (a,b)\setminus\{x_0\} \to \mathbb{R}_{\geq 0}$ as well as $g_2 : (a,b)\setminus\{x_0\} \to \mathbb{R}_{\geq 0}$.

(a) Show that

$$\limsup_{x\to x_0} (g_1(x) + g_2(x)) \leq \limsup_{x\to x_0} g_1(x) + \limsup_{x\to x_0} g_2(x) \quad \text{(subadditivity)} \tag{2.1.9}$$

(b) Show that

$$\limsup_{x\to x_0} (g_1(x)g_2(x)) \leq \left(\limsup_{x\to x_0} g_1(x)\right)\left(\limsup_{x\to x_0} g_2(x)\right). \tag{2.1.10}$$

(c) If $g_1(x) \leq g_2(x)$ for all $x \in (a,b) \setminus \{x_0\}$, show that

$$\limsup_{x \to x_0} g_1(x) \leq \limsup_{x \to x_0} g_2(x) \quad \text{(monotonicity)}. \tag{2.1.11}$$

Similar results also hold true for limit suprerior from right and left.

REMARK 2.1.10. In general, unlike the linearity property in Lemma 2.1.4(a), we only have inequality in (2.1.9), see Remark 2.3.11 below for more details. Other than this, basically *limit superior is a replacement for the usual limit to avoid the difficulty to prove the existence of limits*.

REMARK 2.1.11 (Standard way to use Exercise 2.1.9). By using the *subadditivity property of Euclidean norm* (also known as the *triangle inequality*), we have

$$|f_1(x) - L| = |f_1(x) - f_2(x) + f_2(x) - L| \leq |f_1(x) - f_2(x)| + |f_2(x) - L|,$$

then the monotonicity property (2.1.11), and then consequently by the subadditivity property (2.1.9), imply that

$$\begin{aligned}\limsup_{x \to x_0} |f_1(x) - L| &\leq \limsup_{x \to x_0} (|f_1(x) - f_2(x)| + |f_2(x) - L|) \\ &\leq \limsup_{x \to x_0} |f_1(x) - f_2(x)| + \limsup_{x \to x_0} |f_2(x) - L|.\end{aligned}$$

Since the computations only involving inequality, rather than the equality, this gives possibility to simplify some computations.

EXAMPLE 2.1.12. We consider the function f given in (2.1.7), and define $g(x) := |x|f(x)$ for all $x \in \mathbb{R}$. It is not difficult to see that

$$0 \leq g(x) \leq |x| \quad \text{for all } x \in \mathbb{R}.$$

At the moment, since we do not know whether $\lim_{x \to 0} g(x)$ exists or not, one cannot directly use the monotonicity of limit in Lemma 2.1.4(b) to reach

$$0 \leq \lim_{x \to 0} g(x) \leq \lim_{x \to 0} |x|.$$

The proper argument should goes in the following way:

- **Method 1: via squeeze theorem (Lemma 2.1.5).** Since $\lim_{x \to 0} |x| = 0$ and $\lim_{x \to 0} 0 = 0$, then by using the squeeze theorem (Lemma 2.1.5) we conclude that $\lim_{x \to 0} g(x)$ exists and $\lim_{x \to 0} g(x) = 0$.
- **Method 2: via limit superior criteria (Lemma 2.1.7).** Since

  $$|g(x) - 0| = g(x) \leq |x| \quad \text{for all } x \in \mathbb{R},$$

  then by monotonicity property (2.1.11) and by part (c) the limit superior criteria (Lemma 2.1.7) we see that

  $$\limsup_{x \to 0} |g(x) - 0| \leq \limsup_{x \to 0} |x| = \lim_{x \to 0} |x| = 0,$$

  and finally we conclude that $\lim_{x \to 0} g(x)$ exists and $\lim_{x \to 0} g(x) = 0$ using part (a) the limit superior criteria (Lemma 2.1.7). This is just a demonstration of the standard way how to use the monotonicity property (2.1.11) and the limit superior criteria (Lemma 2.1.7), we will not going to exhibit all details after this example. We will give a simpler proof in Example 2.3.13 below after expanding the definition of limits (Definition 2.1.1 and Definition 2.1.2).

2.2. Continuous function

We now consider the very first example in Section 2.1: It is not difficult to see that the functions $f : \mathbb{R} \setminus \{1\} \to \mathbb{R}$ given by $f(x) = \frac{x^2-x}{x-1}$ for all $x \in \mathbb{R} \setminus \{1\}$ has the limit

$$\lim_{x\to 1} f(x) = 1.$$

We see that the function $g : \mathbb{R} \to \mathbb{R}$ given by $g(x) = x$ for all $x \in \mathbb{R}$ can be expressed as

$$g(x) = \begin{cases} f(x) & , x \in \mathbb{R} \setminus \{1\}, \\ \lim_{x\to 1} f(x) & , x = 1. \end{cases}$$

This observations suggest the following definition:

Definition 2.2.1. Let $a, x_0, b \in \mathbb{R}$ with $a < x_0 < b$ and let $f : (a, b) \to \mathbb{R}$ be a function. If the limit $\lim_{x\to x_0} f(x)$ exists in $\mathbb{R}$ and $\lim_{x\to x_0} f(x) = f(x_0)$, then we say that f is *continuous* at x_0. If $f : I \to \mathbb{R}$, where $I = (a, b)$, $I = (a, +\infty)$ or $I = (-\infty, b)$ or $I = \mathbb{R}$, and f is continuous at all points in I, then we say that f is continuous on I.

Definition 2.2.2. Let $x_0, b \in \mathbb{R}$ with $x_0 < b$ and let $f : [x_0, b) \to \mathbb{R}$ be a function. If the right limit $\lim_{x\to x_0+} f(x)$ exists in $\mathbb{R}$ and $\lim_{x\to x_0+} f(x) = f(x_0)$, then we say that f is *right continuous* at x_0.

Definition 2.2.3. Let $a, x_0 \in \mathbb{R}$ with $a < x_0$ and let $f : (a, x_0] \to \mathbb{R}$ be a function. If the left limit $\lim_{x\to x_0-} f(x)$ exists in $\mathbb{R}$ and $\lim_{x\to x_0-} f(x) = f(x_0)$, then we say that f is *left continuous* at x_0.

Remark 2.2.4. Let $f : I_1 \to I_2$ and $g : I_3 \to \mathbb{R}$ be functions, where I_1, I_2 and I_3 are intervals such that $I_2 \subset I_3$. If $g : I_3 \to \mathbb{R}$ is continuous, and $\lim_{x\to x_0} f(x)$ exists in $\mathbb{R}$ for some $x_0 \in I_1$, then

$$\lim_{x\to x_0} g(f(x)) = g\left(\lim_{x\to x_0} f(x)\right).$$

Similar properties also right limits/right continuity as well as left limits/left continuity.

Example 2.2.5. Let $a, b \in \mathbb{R}$ with $a < b$ and let $f : (a, b) \to \mathbb{R}$ be a function. If $\lim_{x\to x_0} f(x)$ exists in $\mathbb{R}$, then the continuity of Euclidean norm (absolute value) implies

$$\lim_{x\to x_0} |f(x)| = \left|\lim_{x\to x_0} f(x)\right|.$$

On the other hand, the continuity of exponential function also implies

$$\lim_{x\to x_0} e^{f(x)} = \exp\left(\lim_{x\to x_0} f(x)\right).$$

The composition of continuous functions is also continuous:

Lemma 2.2.6. *Let I_1, I_2, I_3, I_4 are open intervals in $\mathbb{R}$, which may unbounded in the sense of Example 1.2.5. If the functions $f : I_1 \to I_2$ and $g : I_3 \to I_4$ are continuous functions, with $I_2 \subset I_3$, then the composition $g \circ f : I_1 \to I_4$ is also continuous.*

It is remarkable to mention the following property of continuous functions.

THEOREM 2.2.7. *Let $a, b \in \mathbb{R}$ with $a < b$ and let $f : [a, b] \to \mathbb{R}$ be a continuous function. Then there exists $x_{\max} \in [a, b]$ and $x_{\min} \in [a, b]$ (not necessarily unique) such that*

$$f(x_{\max}) = \sup_{[a,b]} f \quad \textit{and} \quad f(x_{\min}) = \inf_{[a,b]} f,$$

that is, $f(x_{\min}) \leq f(x) \leq f(x_{\max})$ for all $x \in [a, b]$.

REMARK. The above theorem may not holds true of we remove a point from $[a, b]$. For example, if we consider the continuous function

$$f : [-1, 1] \setminus \{0\} \to \mathbb{R}, \quad f(x) := \frac{1}{x}, \tag{2.2.1}$$

we see that both $x_{\max}$ and $x_{\min}$ do not exist.

2.3. Limits at infinity, limit superior and limit inferior

We now consider the function given in (2.2.1), it is not difficult to see that the right limit $\lim_{x\to 0+} f(x)$ does not exist in $\mathbb{R}$. Intuitively, we observe the trend $f(x) \to +\infty$ as $x \to 0+$, but the problem is $+\infty \notin \mathbb{R}$ and $-\infty \notin \mathbb{R}$, therefore we cannot directly use Definition 2.1.1.

DEFINITION 2.3.1. Let $a, x_0, b \in \mathbb{R}$ with $a < x_0 < b$ and let $f : (a, b) \setminus \{x_0\} \to \mathbb{R}$ be a function. We say that $\lim_{x\to x_0} f(x) = +\infty$ if: Given any $M > 0$, there exists $\delta = \delta(\epsilon) > 0$ such that

$$0 < |x - x_0| < \delta \text{ implies } f(x) > M.$$

Similarly, we say that $\lim_{x\to x_0} f(x) = -\infty$ if: Given any $M > 0$, there exists $\delta = \delta(\epsilon) > 0$ such that

$$0 < |x - x_0| < \delta \text{ implies } f(x) < -M.$$

Similar to Definition 2.1.2, we also consider the following definition.

DEFINITION 2.3.2. Let $a, x_0, b \in \mathbb{R}$ with $a < x_0 < b$ and let $f : (x_0, b) \to \mathbb{R}$ and $g : (a, x_0) \to \mathbb{R}$ be functions. We say that the *right limit* $\lim_{x\to x_0+} f(x) = +\infty$ (resp. $\lim_{x\to x_0+} f(x) = -\infty$) if: Given any $M > 0$, there exists $\delta = \delta(\epsilon) > 0$ such that

$$0 < |x - x_0| < \delta \text{ and } x > x_0 \text{ together imply } f(x) > M \quad (\text{resp. } f(x) < -M).$$

Similarly, we say that the left limit $\lim_{x\to x_0-} g(x) = +\infty$ (resp. $\lim_{x\to x_0-} g(x) = -\infty$) if: Given any $M > 0$, there exists $\delta = \delta(\epsilon) > 0$ such that

$$0 < |x - x_0| < \delta \text{ and } x < x_0 \text{ together imply } g(x) > M \quad (\text{resp. } g(x) < -M).$$

In order to unify the notions, we summarize Definition 2.1.1, Definition 2.1.2, Definition 2.3.1 and Definition 2.3.2 together in the following definition.

DEFINITION 2.3.3. We unify the notion of limits as the followings:

- If either $\lim_{x\to x_0} f(x)$ exists in $\mathbb{R}$ or $\lim_{x\to x_0} f(x) = +\infty$ or $\lim_{x\to x_0} f(x) = -\infty$, we simply say that the limit $\lim_{x\to x_0} f(x)$ exists.
- If either $\lim_{x\to x_0+} f(x)$ exists in $\mathbb{R}$ or $\lim_{x\to x_0+} f(x) = +\infty$ or $\lim_{x\to x_0+} f(x) = -\infty$, we simply say that the limit $\lim_{x\to x_0+} f(x)$ exists.
- If either $\lim_{x\to x_0-} f(x)$ exists in $\mathbb{R}$ or $\lim_{x\to x_0-} f(x) = +\infty$ or $\lim_{x\to x_0-} f(x) = -\infty$, we simply say that the limit $\lim_{x\to x_0-} f(x)$ exists.

We now extend the limit superior and also introduce the limit inferior in the following definition, which are *always exist* in the sense of Definition 2.3.3 above.

DEFINITION 2.3.4. Let $a, x_0, b \in \mathbb{R}$ with $a < x_0 < b$ and let $g : (a,b) \setminus \{x_0\} \to \mathbb{R}$ be *any function*. We define the *limit superior/upper limit* and the *limit inferior/lower limit* by

$$\limsup_{x\to x_0} g(x) := \lim_{r\to 0+}\left(\sup_{B_r(x_0)\setminus\{x_0\}} g\right), \quad \liminf_{x\to x_0} g(x) := \lim_{r\to 0+}\left(\inf_{B_r(x_0)\setminus\{x_0\}} g\right).$$

The *limit superior from right/upper limit from right* and *limit inferior from right/lower limit from right* are defined by

$$\limsup_{x\to x_0+} g(x) := \lim_{r\to 0+}\left(\sup_{(x_0,x_0+r)} g\right), \quad \liminf_{x\to x_0+} g(x) := \lim_{r\to 0+}\left(\inf_{(x_0,x_0+r)} g\right)$$

and similarly the *limit superior from left* or *upper limit from left* and *limit inferior from left/lower limit from left* are defined by

$$\limsup_{x\to x_0-} g(x) := \lim_{r\to 0+}\left(\sup_{(x_0-r,x_0)} g\right), \quad \liminf_{x\to x_0-} g(x) := \lim_{r\to 0+}\left(\inf_{(x_0-r,x_0)} g\right).$$

REMARK 2.3.5. By definition, it is easy to see that

$$\liminf_{x\to x_0\pm} g(x) \le \limsup_{x\to x_0\pm} g(x)$$

for arbitrary functions $g : (a,b) \setminus \{x_0\} \to \mathbb{R}$. Therefore,

$$\begin{aligned} \liminf_{x\to x_0\pm} g(x) = +\infty \quad &\text{implies} \quad \limsup_{x\to x_0\pm} g(x) = +\infty, \\ \limsup_{x\to x_0\pm} g(x) = -\infty \quad &\text{implies} \quad \liminf_{x\to x_0\pm} g(x) = -\infty. \end{aligned}$$

We now can state the following theorem, which is extreme powerful to check whether the limit exists or not.

THEOREM 2.3.6. *Let $a, x_0, b \in \mathbb{R}$ with $a < x_0 < b$ and let $f : (a,b) \setminus \{x_0\} \to \mathbb{R}$ be a function.*

(a) *If $\lim_{x\to x_0} f(x)$ exists, then*

$$\limsup_{x\to x_0} f(x) = \liminf_{x\to x_0} f(x) = \lim_{x\to x_0} f(x). \tag{2.3.1}$$

In other words, if $\limsup_{x\to x_0} f(x) \neq \liminf_{x\to x_0} f(x)$, then the limit $\lim_{x\to x_0} f(x)$ does not exist.

(b) *If $\limsup_{x\to x_0} f(x) = \liminf_{x\to x_0} f(x)$, then $\lim_{x\to x_0} f(x)$ exists and (2.3.1) holds.*

Similar results also hold true for one-side limits/limit superior/limit inferior.

As an immediate consequence, one also can check the continuity of function easily:

COROLLARY 2.3.7. *Let $a, x_0, b \in \mathbb{R}$ with $a < x_0 < b$ and let $f : (a,b) \to \mathbb{R}$ be a function.*

(a) *If f is continuous at x_0, then*

$$\limsup_{x\to x_0} f(x) = \liminf_{x\to x_0} f(x) = \lim_{x\to x_0} f(x) = f(x_0).$$

In other words, if either one of the following holds:

- $\limsup_{x\to x_0} f(x) \neq \liminf_{x\to x_0} f(x)$ *or*
- $\limsup_{x\to x_0} f(x) \neq f(x_0)$ *or*
- $\liminf_{x\to x_0} f(x) \neq f(x_0)$,

then f is not continuous at x_0.

(b) *If $\limsup_{x\to x_0} f(x) = \liminf_{x\to x_0} f(x) = f(x_0)$, then f is continuous at x_0.*

Similar results also hold true for one-side limits/limit superior/limit inferior.

Remark 2.3.8. The limit superior criteria (Lemma 2.1.7) and the squeeze theorem (Lemma 2.1.5) are special cases of Theorem 2.3.6.

Using the same arguments as in Exercise 2.1.9, one can show the following proposition.

Proposition 2.3.9. *Let $a, x_0, b \in \mathbb{R}$ with $a < x_0 < b$ and let $g_1 : (a,b) \setminus \{x_0\} \to \mathbb{R}$ as well as $g_2 : (a,b) \setminus \{x_0\} \to \mathbb{R}$.*

(a) *The limit superior satisfies the subadditivity property:*

$$\limsup_{x\to x_0} (g_1(x) + g_2(x)) \le \limsup_{x\to x_0} g_1(x) + \limsup_{x\to x_0} g_2(x), \tag{2.3.2}$$

(b) *The limit inferior satisfies the superadditivity property:*

$$\liminf_{x\to x_0} (g_1(x) + g_2(x)) \ge \liminf_{x\to x_0} g_1(x) + \liminf_{x\to x_0} g_2(x), \tag{2.3.3}$$

(c) *Both limit superior and limit inferior satisfy the monotonicity property: If $g_1(x) \le g_2(x)$ for all $x \in (a,b) \setminus \{x_0\}$, then*

$$\limsup_{x\to x_0} g_1(x) \le \limsup_{x\to x_0} g_2(x), \quad \liminf_{x\to x_0} g_1(x) \le \liminf_{x\to x_0} g_2(x).$$

Similar results also hold true for one-side limits/limit superior/limit inferior.

Remark 2.3.10. Here we remark that the property (2.1.10), that is,

$$\limsup_{x\to x_0} (g_1(x) g_2(x)) \le \left(\limsup_{x\to x_0} g_1(x)\right)\left(\limsup_{x\to x_0} g_2(x)\right). \tag{2.3.4}$$

only holds true for non-negative functions g_1 and g_2.

Remark 2.3.11. As we mentioned above, we only have subadditivity/superadditivity property rather than the additivity. We now show that the linearity holds under extra assumptions. Suppose that all assumptions in Proposition 2.3.9 hold.

(a) If we *additionally assume that $\lim_{x\to x_0} g_2(x)$ exists in $\mathbb{R}$*, then applying the subadditivity/superadditivity property on the function $g_1(x) = (g_1(x) + g_2(x)) - g_2(x)$, one has

$$\begin{cases} \limsup_{x\to x_0} g_1(x) \le \limsup_{x\to x_0} (g_1(x) + g_2(x)) - \lim_{x\to x_0} g_2(x), \\ \liminf_{x\to x_0} g_1(x) \ge \liminf_{x\to x_0} (g_1(x) + g_2(x)) - \lim_{x\to x_0} g_2(x), \end{cases}$$

which implies

$$\begin{cases} \limsup_{x\to x_0} g_1(x) + \lim_{x\to x_0} g_2(x) \le \limsup_{x\to x_0} (g_1(x) + g_2(x)), \\ \liminf_{x\to x_0} g_1(x) + \lim_{x\to x_0} g_2(x) \ge \liminf_{x\to x_0} (g_1(x) + g_2(x)). \end{cases}$$

Now combine this with Proposition 2.3.9(a)(b) to conclude the additivity:

$$\begin{cases} \limsup_{x\to x_0} (g_1(x) + g_2(x)) = \limsup_{x\to x_0} g_1(x) + \lim_{x\to x_0} g_2(x), \\ \liminf_{x\to x_0} (g_1(x) + g_2(x)) = \liminf_{x\to x_0} g_1(x) + \lim_{x\to x_0} g_2(x). \end{cases} \tag{2.3.5}$$

(b) If we *additionally assume that $|g_1(x)| \leq M$ for all $x \in B_r(x_0)$ for some $r > 0$ and $\lim_{x\to x_0} g_2(x)$ exists in $\mathbb{R}_{\geq 0}$*, by writing

$$g_1(x)g_2(x) = g_1(x)\left(\lim_{t\to x_0} g_2(t)\right) + g_1(x)\left(g_2(x) - \lim_{t\to x_0} g_2(t)\right)$$

from (2.3.5) we see that

$$\begin{cases} \limsup\limits_{x\to x_0} (g_1(x)g_2(x)) = \left(\limsup\limits_{x\to x_0} g_1(x)\right)\left(\lim\limits_{t\to x_0} g_2(t)\right), \\ \liminf\limits_{x\to x_0} (g_1(x)g_2(x)) = \left(\liminf\limits_{x\to x_0} g_1(x)\right)\left(\lim\limits_{t\to x_0} g_2(t)\right). \end{cases} \tag{2.3.6}$$

In the particular case when $g_2(x) = c \geq 0$ for all $x \in B_r(x_0)$, we see that (2.3.6) reads

$$\begin{cases} \limsup\limits_{x\to x_0} (cg_1(x)) = c\limsup\limits_{x\to x_0} g_1(x), \\ \liminf\limits_{x\to x_0} (cg_1(x)) = c\liminf\limits_{x\to x_0} g_1(x). \end{cases}$$

One should be aware that, for constant $b \leq 0$, one sees that $b = -|b|$ and see that

$$\begin{aligned} \limsup_{x\to x_0} (bg_1(x)) &= \limsup_{x\to x_0} (-|b|g_1(x)) = -\liminf_{x\to x_0} (|b|g_1(x)) \\ &= -|b|\liminf_{x\to x_0} g_1(x) = b\liminf_{x\to x_0} g_1(x), \end{aligned}$$

and

$$\begin{aligned} \liminf_{x\to x_0} (bg_1(x)) &= \liminf_{x\to x_0} (-|b|g_1(x)) = -\limsup_{x\to x_0} (|b|g_1(x)) \\ &= -|b|\limsup_{x\to x_0} g_1(x) = b\limsup_{x\to x_0} g_1(x). \end{aligned}$$

This means that in general, the linearity does not hold true for general coefficients, which only holds true for positive coefficients. Similar result also holds true for one-side limit superior and limit inferior.

EXAMPLE 2.3.12 (Revisit of Example 2.1.8). We now consider the function given in (2.1.7), that is,

$$f : \mathbb{R} \to \mathbb{R}, \quad f(x) = \begin{cases} 1 & , x \in \mathbb{Q}, \\ 0 & , x \in \mathbb{R} \setminus \mathbb{Q}, \end{cases} \tag{2.3.7}$$

consider in Example 1.3.19. For each $x_0 \in \mathbb{R}$, it is easy to see that

$$\liminf_{x\to x_0} f(x) = 0 \neq 1 = \limsup_{x\to x_0} f(x),$$

which concludes that $\lim_{x\to x_0} f(x)$ does not exist.

EXAMPLE 2.3.13 (Revisit of Example 2.1.12). We now consider the function $g(x) := |x|f(x)$, where f is the function given in (2.3.7). It is easy to see that

$$0 \leq g(x) \leq |x| \quad \text{for all } x \in \mathbb{R}.$$

We now see that

$$0 \leq \liminf_{x\to 0} g(x) \leq \limsup_{x\to 0} g(x) \leq \limsup_{x\to 0} |x| = \lim_{x\to 0} |x| = 0,$$

which concludes that

$$\lim_{x\to 0} g(x) = \liminf_{x\to 0} g(x) = \limsup_{x\to 0} g(x) = 0.$$

This example demonstrates another alternative way to use squeeze theorem.

EXAMPLE 2.3.14. We now consider a difficult example, exhibited in Example 1.3.19(d): We consider the function

$$f : (0,1) \to \mathbb{R}, f(x) = \begin{cases} \frac{1}{q} & , \text{if } x = \frac{p}{q} \in (0,1)\cap\mathbb{Q}, q > 0, \gcd(p,q) = 1, \\ 0 & , \text{if } x \in (0,1)\setminus\mathbb{Q}. \end{cases} \tag{2.3.8}$$

In view of Corollary 2.3.7, it is easy to show that f is not continuous at all $x_1 \in (0,1)\cap\mathbb{Q}$, since

$$\liminf_{x\to x_1} f(x) = 0 < f(x_1).$$

We now show that f is continuous at all $x_0 \in (0,1)\setminus\mathbb{Q}$. Since $f(x) \geq 0$ for all $x \in (0,1)$ and $f(x_0) = 0$, it is suffice to show $\limsup_{x\to x_0} f(x) = 0$. For each integer $q \in \mathbb{N}$, we define the set of rational number with denominator at most q, that is,

$$\mathbb{Q}_q := \mathbb{Z} \cup \frac{1}{2}\mathbb{Z} \cup \frac{1}{3}\mathbb{Z} \cup \cdots \cup \frac{1}{q}\mathbb{Z}.$$

One sees that $(0,1)\cap\mathbb{Q}_q$ is a finite set, i.e. there are only finitely many points in that set. Since $x_0 \in (0,1)\setminus\mathbb{Q}$, then

$$\mathrm{dist}(x_0, (0,1)\cap\mathbb{Q}_q) = \min_{x\in(0,1)\cap\mathbb{Q}_q} |x - x_0| > 0.$$

This means that, if we define

$$r_q := \frac{1}{2}\mathrm{dist}(x_0, (0,1)\cap\mathbb{Q}_q),$$

we see that the set $B_{r_q}(x_0)\setminus\{x_0\}$ only consists of rational number with denominator $\geq q+1$, therefore

$$\sup_{B_{r_q}(x_0)\setminus\{x_0\}} f \leq \frac{1}{q+1}.$$

Hence, we see that

$$\limsup_{x\to x_0} f(x) = \lim_{r\to 0+}\left(\sup_{B_r(x_0)\setminus\{x_0\}} f\right) \leq \sup_{B_{r_q}(x_0)\setminus\{x_0\}} f \leq \frac{1}{q+1} \quad \text{for all } q \in \mathbb{N}.$$

Since the left hand side is independent of q, by arbitrariness of $q \in \mathbb{N}$, we now conclude that $\limsup_{x\to x_0} f(x) = 0$, and hence f is continuous at $x_0 \in (0,1)\setminus\mathbb{Q}$.
Conclusion. The function f given in (2.3.8) is continuous at each irrational point, but discontinuous at each rational point.

It is also possible to define the limit for $x \to \pm\infty$:

DEFINITION 2.3.15. Let $a \in \mathbb{R}$ and let $f : (a,+\infty) \to \mathbb{R}$ be a function. We say that $\lim_{x\to+\infty} f(x) = L \in \mathbb{R}$ exists if: Given any $\epsilon > 0$, there exists $N = N(\epsilon) > 0$ such that

$$x > M \text{ implies } |f(x) - L| < \epsilon.$$

We say that $\lim_{x\to+\infty} f(x) = +\infty$ if: Given any $M > 0$, there exists $N = N(M) > 0$ such that

$$x > M \text{ implies } f(x) > N,$$

and the limit $\lim_{x\to+\infty} f(x) = -\infty$ can be defined in an analogous way. If either $\lim_{x\to+\infty} f(x) = L \in \mathbb{R}$ exists or $\lim_{x\to+\infty} f(x) = +\infty$ or $\lim_{x\to+\infty} f(x) = -\infty$, then we say that $\lim_{x\to+\infty} f(x)$ exists, or we slightly abuse the notation by saying that $\lim_{x\to+\infty} f(x)$ exists in $[-\infty, +\infty]$. The notions $\lim_{x\to-\infty} g(x)$ for functions $g : (-\infty, a) \to \mathbb{R}$ can be defined using similar manner, here we omit the details.

We also can define the limit superior and limit inferior in a similar manner.

Definition 2.3.16. Let $a \in \mathbb{R}$ and let $f : (a, +\infty) \to \mathbb{R}$ be a function. We define

$$\limsup_{x\to+\infty} f(x) := \lim_{N\to+\infty} \left(\sup_{x>N} f(x)\right) \quad \text{and} \quad \liminf_{x\to+\infty} f(x) := \lim_{N\to+\infty} \left(\inf_{x>N} f(x)\right).$$

Similarly, for functions $g : (-\infty, a) \to \mathbb{R}$, we define

$$\limsup_{x\to-\infty} g(x) := \lim_{N\to-\infty} \left(\sup_{x<N} g(x)\right) \quad \text{and} \quad \liminf_{x\to-\infty} f(x) := \lim_{N\to-\infty} \left(\inf_{x<N} g(x)\right).$$

We also have the following theorem.

Theorem 2.3.17. *Let $a \in \mathbb{R}$ and let $f : (a, +\infty) \to \mathbb{R}$ be a function.*

(a) *If $\lim_{x\to+\infty} f(x)$ exists, then*

$$\lim_{x\to+\infty} f(x) = \limsup_{x\to+\infty} f(x) = \liminf_{x\to+\infty} f(x). \tag{2.3.9}$$

In other words, if $\limsup_{x\to+\infty} f(x) \neq \liminf_{x\to+\infty} f(x)$, then $\lim_{x\to+\infty} f(x)$ does not exist.

(b) *If $\limsup_{x\to+\infty} f(x) = \liminf_{x\to+\infty} f(x)$, then $\lim_{x\to+\infty} f(x)$ exists and (2.3.9) holds.*

Similar result also holds for the limit as $x \to -\infty$.

Using the same arguments as in Exercise 2.1.9, one can show the following proposition.

Proposition 2.3.18. *Let $a \in \mathbb{R}$ and let $g_1 : (a, +\infty) \to \mathbb{R}$ and $g_2 : (a, +\infty) \to \mathbb{R}$ be functions.*

(a) *The limit superior satisfies the subadditivity property:*

$$\limsup_{x\to+\infty} (g_1(x) + g_2(x)) \leq \limsup_{x\to+\infty} g_1(x) + \limsup_{x\to+\infty} g_2(x), \tag{2.3.10}$$

(b) *The limit inferior satisfies the superadditivity property:*

$$\liminf_{x\to+\infty} (g_1(x) + g_2(x)) \geq \liminf_{x\to+\infty} g_1(x) + \liminf_{x\to+\infty} g_2(x), \tag{2.3.11}$$

(b) *Both limit superior and limit inferior satisfy the monotonicity property: If $g_1(x) \leq g_2(x)$ for all $x \in (a, b) \setminus \{x_0\}$, then*

$$\limsup_{x\to+\infty} g_1(x) \leq \limsup_{x\to+\infty} g_2(x), \quad \liminf_{x\to+\infty} g_1(x) \leq \liminf_{x\to+\infty} g_2(x).$$

Similar result also holds for the limit as $x \to -\infty$.

Remark 2.3.19. Similar results in Remark 2.3.11 also holds true for the limit superior and limit inferior at infinity. Here we omit the details.

EXERCISE 2.3.20. Using Theorem 2.2.7 to show the followings:

(a) If $f : [a, +\infty) \to \mathbb{R}$ is continuous and $\lim_{x\to+\infty} f(x)$ exists in $\mathbb{R}$, then there exists a constant $C > 0$ such that $|f(x)| \leq C$ for all $x \geq a$.
(b) If $f : (-\infty, a] \to \mathbb{R}$ is continuous and $\lim_{x\to-\infty} f(x)$ exists in $\mathbb{R}$, then there exists a constant $C > 0$ such that $|f(x)| \leq C$ for all $x \leq a$.
(c) If $f : \mathbb{R} \to \mathbb{R}$ is continuous and both $\lim_{x\to+\infty} f(x)$ and $\lim_{x\to-\infty} f(x)$ exist in $\mathbb{R}$, then there exists a constant $C > 0$ such that $|f(x)| \leq C$ for all $x \in \mathbb{R}$.

REMARK 2.3.21. Unlike Theorem 2.2.7, the maximum and minima may not attained, for example the inverse tangent

$$\tan^{-1} : \mathbb{R} \to (-\pi/2, \pi/2)$$

satisfies

$$\lim_{x\to+\infty} \tan^{-1} x = \frac{\pi}{2}, \quad \lim_{x\to-\infty} \tan^{-1} x = -\frac{\pi}{2}.$$

EXAMPLE 2.3.22. Finally, by using the unified notations above, we also can do "change of variables" for limits, which simplify some computations. For example,

$$\lim_{x\to 0+} \frac{e^{-1/x}}{x} = \lim_{y\to+\infty} \frac{e^{-y}}{1/y} = \lim_{y\to+\infty} ye^{-y},$$

here we considered the change of variable $y = 1/x$, and see that $x \to 0+$ if and only if $y \to +\infty$. The above limit is not easy to compute using only the above definitions, since

$$\lim_{y\to+\infty} y = +\infty \quad \text{and} \quad \lim_{y\to+\infty} e^{-y} = 0.$$

We will resolve this difficulty in Example 3.2.6 of Section 3.2 below.

CHAPTER 3

Differentiation

3.1. Definition of Differentiation

We now use the notion of limit above to study the *infinitesimal rate of change of functions*, simply speaking, the *slope of the tangent line* at each point. Let $f : (a, b) \to \mathbb{R}$ be a function. We now pick any two points $x_0 \neq x_1 \in (a, b)$. The rate of change is given by

$$\frac{f(x_1) - f(x_0)}{x_1 - x_0}.$$

It is more convenient to denote $x_1 = x_0 + h$ for some $h \neq 0$, and write

$$\frac{f(x_0 + h) - f(x_0)}{h}. \tag{3.1.1}$$

Now it is natural to consider the following definition.

Definition 3.1.1. Let $a, b \in \mathbb{R}$ with $a < b$ and let $f : (a, b) \to \mathbb{R}$ be a function, and let $x_0 \in (a, b)$. We say that f is *differentiable* at x_0 if

$$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h} \text{ exists in } \mathbb{R}. \tag{3.1.2}$$

In this case, the number (3.1.2) is called the derivative of f at x_0, which we usually denoted as

$$f'(x_0) \text{ or } f'(x)|_{x=x_0} \text{ (Lagrange notation)},$$

which is more convenient in differentiation, or

$$\frac{\mathrm{d}f}{\mathrm{d}x}(x_0) \text{ or } \frac{\mathrm{d}f}{\mathrm{d}x}(x)\bigg|_{x=x_0} \text{ or } \frac{\mathrm{d}}{\mathrm{d}x}f(x)\bigg|_{x=x_0} \text{ (Leibniz notation)},$$

which is more convenient in Riemann/Lebesgue integration. Here I suggest the monograph [**WZ15**] for those interested in Lebesgue integral.

Remark 3.1.2. The notations $f'(x)|_{x=x_0}$ and $\frac{\mathrm{d}}{\mathrm{d}x}f(x)\big|_{x=x_0}$ both emphasize that "first differentiate and then evaluate the point $x = x_0$".We see that the quotient (3.1.1) is not well-defined at $h = 0$. According to the definition of limit (Definition 2.1.1), we remind the readers that the limit (3.1.2) does not require the pointwise evaluation at $h = 0$.

One sees that (3.1.2) is equivalent to

$$\limsup_{h \to 0} \frac{|f(x_0 + h) - f(x_0) - f'(x_0)h|}{|h|} = 0.$$

We also can rephrase Definition 3.1.1 in the followings implicit way:

DEFINITION 3.1.3. Let $a, b \in \mathbb{R}$ with $f : (a, b) \to \mathbb{R}$ be a function, and let $x_0 \in (a, b)$. We say that f is *differentiable* at x_0 if there exists $L \in \mathbb{R}$ such that

$$\limsup_{h \to 0} \frac{|f(x_0 + h) - f(x_0) - Lh|}{|h|} = 0. \tag{3.1.3}$$

In this case, the derivative $f'(x_0)$ of f at x_0 is defined by $f'(x_0) := L$.

REMARK 3.1.4. Unlike Definition 3.1.1, it is not so obvious that whether the number L in Definition 3.1.3 is unique or not. Suppose that (3.1.3) holds true for $L = L_1$ and $L = L_2$. We see that

$$\begin{aligned} |L_1 - L_2| &= \frac{|(f(x_0 + h) - f(x_0) - L_1 h) - (f(x_0 + h) - f(x_0) - L_2 h)|}{|h|} \\ &\le \frac{|(f(x_0 + h) - f(x_0) - L_1 h)|}{|h|} + \frac{|f(x_0 + h) - f(x_0) - L_2 h)|}{|h|}, \end{aligned}$$

and take limit superior to see that

$$\begin{aligned} |L_1 - L_2| &\le \limsup_{h \to 0} \left(\frac{|(f(x_0 + h) - f(x_0) - L_1 h)|}{|h|} + \frac{|f(x_0 + h) - f(x_0) - L_2 h)|}{|h|} \right) \\ &\le \limsup_{h \to 0} \frac{|(f(x_0 + h) - f(x_0) - L_1 h)|}{|h|} + \limsup_{h \to 0} \frac{|f(x_0 + h) - f(x_0) - L_2 h)|}{|h|} = 0, \end{aligned}$$

which concludes that $L_1 = L_2$. We again remind the readers that the limit superior only subaddivity property rather than the additivity.

The following lemma is an easy consequence of the definitions of continuity and differentiability of functions.

LEMMA 3.1.5. *Let $a, x_0, b \in \mathbb{R}$ with $a < x_0 < b$ and let $f : (a, b) \to \mathbb{R}$ be a function. If f is differentiable at x_0, then f is continuous at x_0.*

DEFINITION 3.1.6. Let $a, b \in \mathbb{R}$ with $a < b$, and let $f : (a, b) \to \mathbb{R}$ be a differentiable function. We say that f is *twice-differentiable* at $x_0 \in (a, b)$ if the function $f' : (a, b) \to \mathbb{R}$ is differentiable at x_0. In this case,

$$f''(x_0) := \left(\frac{\mathrm{d}}{\mathrm{d}x} f'(x) \right) \bigg|_{x = x_0} = \lim_{h \to 0} \frac{f'(x + h) - f'(x)}{h}.$$

In terms of Lagrange notation, we call $f'(x_0)$ the *first-order derivative* of f at x_0 and call $f''(x_0)$ the *second-order derivative* of f at x_0.

DEFINITION 3.1.7. Let $a, x_0, b \in \mathbb{R}$ with $a < x_0 < b$. A function $f : (a, b) \to \mathbb{R}$ is said to be *twice-differentiable* if it is differentiable and $f' : (a, b) \to \mathbb{R}$ is also differentiable. In this case, the function $f^{(0)} := f$ is called the *zeroth-order derivative* of f, $f^{(1)} := f'$ is called the *first-order derivative* of f, and say that $f^{(2)} := f''$ the *second-order derivative* of f. In this case, we also say that *the function $f : (a, b) \to \mathbb{R}$ is differentiable 2-times*. Inductively, for each $n \in \mathbb{N}$, we say that *a function $f : (a, b) \to \mathbb{R}$ is differentiable n-times* if the derivatives $f^{(j)} : (a, b) \to \mathbb{R}$ are differentiable for all $j = 0, \cdots, n - 1$, and we define $f^{(n)} := (f^{(n-1)})'$. In terms of Leibniz notation, we write

$$\frac{\mathrm{d}^n f}{\mathrm{d}x^n} \equiv \left(\left(\frac{\mathrm{d}}{\mathrm{d}x} \right)^n f(x) \right) := f^{(n)}(x),$$

or emphasizing the evaluation of points:

$$\frac{\mathrm{d}^n f}{\mathrm{d}x^n}(x_0) \equiv \left(\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^n f(x)\right)\Bigg|_{x=x_0} := f^{(n)}(x_0),$$

REMARK 3.1.8. Suppose that $f : (a,b) \to \mathbb{R}$ is differentiable n times, according to Lemma 3.1.5, the derivatives $f^{(j)} : (a,b) \to \mathbb{R}$ are continuous for all $j = 0, \cdots, n-1$. However, in general, the highest order derivative $f^{(n)} : (a,b) \to \mathbb{R}$ may not continuous. One should be careful that the set $\mathcal{C}^n(I)$, for any interval I, means that

$$\mathcal{C}^n(I) := \left\{ f : I \to \mathbb{R} \text{ is differentiable } n\text{-times} : f^{(n)} : I \to \mathbb{R} \text{ is continuous} \right\}.$$

EXERCISE 3.1.9. Let $a, b \in \mathbb{R}$ with $a < b$ and let $f : (a,b) \to \mathbb{R}$ be a function, and let $x_0 \in (a,b)$. Show that if f is differentiable at x_0 with derivative $f'(x_0)$, then

$$\lim_{h \to 0} \frac{f(x_0+h) - f(x_0-h)}{2h} \text{ exists in } \mathbb{R} \text{ and it is equal to } f'(x_0).$$

REMARK. The converse of Exercise 3.1.9 may not true. For example, we consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = |x|$, and we consider $x_0 = 0$. One sees that

$$\frac{f(0+h) - f(0-h)}{2h} = \frac{|h| - |h|}{2h} = 0 \quad \text{for all } h \neq 0,$$

and hence

$$\lim_{h \to 0} \frac{f(0+h) - f(0-h)}{2h} = 0.$$

However, one sees that

$$\lim_{h \to 0+} \frac{f(0+h) - f(0)}{h} = \lim_{h \to 0+} \frac{h}{h} = 1 \neq -1 = \lim_{h \to 0-} \frac{-h}{h} = \lim_{h \to 0-} \frac{f(0+h) - f(0)}{h},$$

hence the limit $\lim_{h \to 0} \frac{f(0+h)-f(0)}{h}$ does not exist, and thus f is not differentiable at $x_0 = 0$.

EXAMPLE 3.1.10. The definition of the natural exponential e, which is approximated by $2.71828\cdots$, means that

$$\frac{\mathrm{d}}{\mathrm{d}x} e^x \bigg|_{x=0} = 1.$$

From this, one sees that

$$\frac{\mathrm{d}}{\mathrm{d}x} e^x = \lim_{h \to 0} \frac{e^{x+h} - e^x}{h} = \lim_{h \to 0} \frac{e^x e^h - e^x}{h} = e^x \lim_{h \to 0} \frac{e^h - 1}{h} = e^x \lim_{h \to 0} \frac{\mathrm{d}}{\mathrm{d}x} e^x \bigg|_{x=0} = e^x \quad \text{for all } x \in \mathbb{R}.$$

EXAMPLE 3.1.11. Let $n \in \mathbb{N}$, and we consider the power function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^n$. By using the binomial theorem, one sees that

$$(x+h)^n = \sum_{j=0}^{n} \binom{n}{j} x^{n-j} h^j,$$

where $\binom{n}{j}$ is the number of ways to choose j elements from a set with n elements. For each $h \neq 0$, one sees that

$$\frac{(x+h)^n - x^n}{h} = \frac{1}{h} \sum_{j=1}^{n} \binom{n}{j} x^{n-j} h^j = \sum_{j=1}^{n} \binom{n}{j} x^{n-j} h^{j-1} = n x^{n-1} + \sum_{j=2}^{n} \binom{n}{j} x^{n-j} h^{j-1}.$$

By the linearity of the limit (Lemma 2.1.4), one sees that

$$f'(x) = \lim_{h\to 0} \frac{(x+h)^n - x^n}{h} = nx^{n-1} + \sum_{j=2}^{n} \binom{n}{j} x^{n-j} \lim_{h\to 0} h^{j-1} = nx^{n-1} \quad \text{for all } x \in \mathbb{R}.$$

EXAMPLE 3.1.12. We now consider the function $f : \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = \begin{cases} x^3 & , x < 0, \\ x^2 & , x \geq 0. \end{cases}$$

By using Example 3.1.11, one has $f'(x) = 3x^2$ for all $x < 0$ and $f'(x) = 2x$ for all $x > 0$. We are now asking whether f is differentiable at $x = 0$ or not. We only can check this directly from the definition:

$$\frac{f(0+h)-f(0)}{h} = \frac{h^2}{h} = h \text{ for all } h > 0, \quad \text{which gives} \quad \lim_{h\to 0+} \frac{f(0+h)-f(0)}{h} = 0,$$
$$\frac{f(0+h)-f(0)}{h} = \frac{h^3}{h} = h^2 \text{ for all } h < 0, \quad \text{which gives} \quad \lim_{h\to 0-} \frac{f(0+h)-f(0)}{h} = 0.$$

Since the left and right limits exist and coincide, we conclude that the limit $\lim_{h\to 0} \frac{f(0+h)-f(0)}{h}$ exists and equal to 0, and in fact it means that $f'(0) = 0$.

We will give more interesting examples after exhibit some differentiation rules in Section 3.3 below.

3.2. L' Hôpital's rule

Before continue, let us exhibit some important facts which further motivate the study of differentiation. We see that the limits (3.1.2) and (3.1.3) are both special case of the limit

$$\lim_{x\to 0} \frac{f(x)}{g(x)} \text{ with } \lim_{x\to 0} f(x) = 0 \text{ and } \lim_{x\to 0} g(x) = 0,$$

which is not so easy to compute. Despite some authors say that this it is the intermediate form $\frac{0}{0}$, however personally *I strongly suggested not to use this terminology*, since it is not rigorous and may cause ambiguity. Let $f, g : (a, b) \to \mathbb{R}$ be functions which are differentiable on (a, b) and their derivatives $f' : (a, b) \to \mathbb{R}$ and $g' : (a, b) \to \mathbb{R}$ are continuous. In this case. If there is a point $x_0 \in (a, b)$ such that

$$\lim_{x\to x_0} f(x) = 0 = f(x_0) \text{ and } \lim_{x\to x_0} g(x) = 0 = g(x_0) \text{ as well as } g'(x_0) \neq 0,$$

one sees that

$$\begin{aligned} \lim_{x\to x_0} \frac{f(x)}{g(x)} &= \lim_{x\to x_0} \frac{f(x)-0}{g(x)-0} = \lim_{x\to x_0} \frac{f(x)-f(x_0)}{g(x)-g(x_0)} \\ &= \lim_{x\to x_0} \frac{\frac{f(x)-f(x_0)}{x-x_0}}{\frac{g(x)-g(x_0)}{x-x_0}} = \frac{\lim_{x\to x_0} \frac{f(x)-f(x_0)}{x-x_0}}{\lim_{x\to x_0} \frac{g(x)-g(x_0)}{x-x_0}} = \frac{f'(x_0)}{g'(x_0)} = \lim_{x\to x_0} \frac{f'(x)}{g'(x)}, \end{aligned}$$

this proves the simplest version of *L' Hôpital's rule*. Although the rule is often attributed to de L' Hôpital, the theorem was first introduced to him in 1694 by the Swiss mathematician Johann Bernuolli. We first exhibit a version of L' Hôpital's rule in terms of limit superior and limit inferior.

THEOREM 3.2.1 ([**Tay52**]). *Suppose that $f : (a,b) \to \mathbb{R}$ and $g : (a,b) \to \mathbb{R}$ are differentiable functions for some $-\infty \leq a < b \leq +\infty$ in the sense of Example 1.2.5. Suppose that $g(x) \neq 0$ and $g'(x) \neq 0$ for all $x \in (a,b)$. If either one of the following holds for some $x_0 \in (a,b)$:*

(a) $\lim_{x\to x_0} f(x) = 0$ *and* $\lim_{x\to x_0} g(x) = 0$;
(b) $\lim_{x\to x_0} f(x) = \pm\infty$ *and* $\lim_{x\to x_0} g(x) = \pm\infty$;

then

$$\liminf_{x\to x_0} \frac{f'(x)}{g'(x)} \leq \liminf_{x\to x_0} \frac{f(x)}{g(x)} \leq \limsup_{x\to x_0} \frac{f(x)}{g(x)} \leq \limsup_{x\to x_0} \frac{f'(x)}{g'(x)}.$$

Combining this with Theorem 2.3.6, we immediately reach the following version L' Hôpital's rule:

COROLLARY 3.2.2. *Suppose that all assumptions in Theorem 3.2.1 hold. If we further assume that*

$$\text{the limit } \lim_{x\to x_0} \frac{f'(x)}{g'(x)} \text{ exists,}$$

then

$$\lim_{x\to x_0} \frac{f(x)}{g(x)} = \lim_{x\to x_0} \frac{f'(x)}{g'(x)}.$$

Here we also exhibit another version in terms of left and right limits. Despite its proof involving the Cauchy mean value theorem (Theorem 3.4.6) below, here we still present here in order to motivate the study, and we will not going to give the proof in this lecture note.

THEOREM 3.2.3 ([**Rud87**, Theorem 5.13]). *Suppose that $f : (a,b) \to \mathbb{R}$ and $g : (a,b) \to \mathbb{R}$ are differentiable functions for some $-\infty \leq a < b \leq +\infty$ in the sense of Example 1.2.5. Suppose that*

$$\text{the right limit } \lim_{x\to a+} \frac{f'(x)}{g'(x)} \text{ exists,}$$

where we interpret $\lim_{x\to a+}$ as $\lim_{x\to -\infty}$ if $a = -\infty$. If either one of the following holds:

(a) $\lim_{x\to a+} f(x) = 0$ *and* $\lim_{x\to a+} g(x) = 0$;
(b) $\lim_{x\to a+} f(x) = \pm\infty$ *and* $\lim_{x\to a+} g(x) = \pm\infty$;

then

$$\lim_{x\to a+} \frac{f(x)}{g(x)} = \lim_{x\to a+} \frac{f'(x)}{g'(x)}.$$

Similar results also hold true for the left limit $\lim_{x\to b-}$, which is interpreted as $\lim_{x\to +\infty}$ when $b = +\infty$.

REMARK 3.2.4. The limits in both Corollary 3.2.2 and Theorem 3.2.3 may take values $\pm\infty$, more precisely, they are understood in the sense of Definition 2.3.3 and Definition 2.3.15.

REMARK 3.2.5. In this calculus course, we only involving functions are quite smooth, therefore most of the assumptions in the L' Hôpital's rule (Corollary 3.2.2 and Theorem 3.2.3) can be satisfies easily. However, *one always need to emphasize the sufficient condition (a) or (b)* in the L' Hôpital's rule (Corollary 3.2.2 and Theorem 3.2.3) before using the theorem, *otherwise your marks will be deducted significantly*.

EXAMPLE 3.2.6. Now we can compute the limit in Example 2.3.22. We see that

$$\lim_{x\to 0+} \frac{e^{-1/x}}{x} = \lim_{y\to+\infty} \frac{e^{-y}}{1/y} = \lim_{y\to+\infty} \frac{y}{e^y}.$$

Since $\lim_{y\to+\infty} y = +\infty$ and $\lim_{y\to+\infty} e^y = +\infty$ (as mentioned above, *I strongly recommend not to say that this is an intermediate form $\frac{+\infty}{+\infty}$*), then by L' Hôpital's rule, one has

$$\lim_{y\to+\infty} \frac{y}{e^y} = \lim_{y\to+\infty} \frac{\frac{\mathrm{d}}{\mathrm{d}y}(y)}{\frac{\mathrm{d}}{\mathrm{d}y}(e^y)} = \lim_{y\to+\infty} \frac{1}{e^y} = 0,$$

which concludes that $\lim_{x\to 0+} \frac{e^{-1/x}}{x} = 0$.

3.3. Differentiation rules

The main theme of this section is to introduce the product rule, chain rule, implicit differentiation and how to take limit implicitly. We first exhibit the main properties of differentiations.

LEMMA 3.3.1 ([**Rud87**, Theorem 5.3]). *Let $a, b \in \mathbb{R}$ with $a < b$ and let $f_1 : (a,b) \to \mathbb{R}$ and $f_2 : (a,b) \to \mathbb{R}$ be functions.*

(a) ***Linearity.*** *If both f_1 and f_2 are differentiable at $x_0 \in (a,b)$, then for each $c_1, c_2 \in \mathbb{R}$, the function*

$$c_1 f_1 + c_2 f_2 : (a,b) \to \mathbb{R}, \quad (c_1 f_1 + c_2 f_2)(x) := c_1 f_1(x) + c_2 f_2(x) \text{ for all } x \in (a,b)$$

is also differentiable at such point x_0, and satisfying

$$(c_1 f_1 + c_2 f_2)'(x_0) = c_1 f_1'(x_0) + c_2 f_2'(x_0).$$

(b) ***Product rule.*** *If both f_1 and f_2 are differentiable at $x_0 \in (a,b)$, then the function (not to be confused with the composition of functions in Definition 1.3.6)*

$$f_1 f_2 : (a,b) \to \mathbb{R}, \quad (f_1 f_2)(x) := f_1(x) f_2(x) \text{ for all } x \in (a,b)$$

is also differentiable at x_0, and satisfying

$$(f_1 f_2)'(x_0) = f_1'(x_0) f_2(x_0) + f_1(x_0) f_2'(x_0).$$

LEMMA 3.3.2 (Chain rule [**Rud87**, Theorem 5.5]). *Let $a, b \in \mathbb{R}$ with $a < b$ and let $f : (a,b) \to I$ be a continuous function for some open interval I (may unbounded as in Example 1.2.5). Suppose that f is differentiable at $x \in (a,b)$ and suppose that $g : I \to \mathbb{R}$ is differentiable at $f(x)$, then the composition $g \circ f : (a,b) \to \mathbb{R}$ (as in Definition 1.3.6) is differentiable at x and satisfying*

$$(g \circ f)'(x) = g'(y)\big|_{y=f(x)} f'(x). \tag{3.3.1}$$

REMARK 3.3.3. In terms of composition (Definition 1.3.6), one also sees that

$$g'(y)\big|_{y=f(x)} = (g' \circ f)(x) = g'(f(x)).$$

One should be careful about the notations: The term $(g \circ f)'(x)$ means that we first composite the functions, and then differentiate the resulting function, while the term $g'(f(x)) = g'(y)\big|_{y=f(x)}$ means that we we first differentiate g, and then evaluate $y = f(x)$ after that. In general, $(g \circ f)'(x)$ and $g'(y)\big|_{y=f(x)}$ are different. Roughly speaking, chain rule says that *if one interchanging the order of "differentiation" and "evaluation", then the "price"* *of doing so is multiplying $f'(x)$*. Personally, I would suggest the notation $g'(y)|_{y=f(x)}$ rather than $g'(f(x))$ (*of course this is not mandatory, it is up to you to take the risk or not*). The proper way to write (3.3.1) in terms of Leibniz notation should be

$$\frac{\mathrm{d}}{dx}(g \circ f) = \left.\frac{\mathrm{d}g}{\mathrm{d}y}\right|_{y=f(x)} \frac{\mathrm{d}f}{\mathrm{d}x}. \tag{3.3.2}$$

Some authors abuse the notation by ignoring the evaluation $y = f(x)$ to write

$$\frac{\mathrm{d}g}{dx} = \frac{\mathrm{d}g}{\mathrm{d}f} \cdot \frac{\mathrm{d}f}{\mathrm{d}x}, \tag{3.3.3}$$

to formally canceled out the notation $\mathrm{d}f$.

REMARK 3.3.4. Suppose that both $f : (a,b) \to I$ and $g : I \to \mathbb{R}$ are both twice differentiable. By using the Lagrange notations, one sees that

$$\begin{aligned}(g \circ f)''(x) &= ((g' \circ f)(x)f'(x))' \quad \text{(chain rule)}\\ &= (g' \circ f)'(x)f'(x) + (g' \circ f)(x)f''(x) \quad \text{(product rule)}\\ &= g''(y)|_{y=f(x)} (f'(x))^2 + g'(y)|_{y=f(x)} f''(x) \quad \text{(chain rule)}.\end{aligned}$$

In terms of Leibniz notation, the above equality reads

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}(g \circ f) = \left.\frac{\mathrm{d}^2 g}{\mathrm{d}y^2}\right|_{y=f(x)} \left(\frac{\mathrm{d}f}{\mathrm{d}x}\right)^2 + \left.\frac{\mathrm{d}g}{\mathrm{d}y}\right|_{y=f(x)} \frac{\mathrm{d}^2 f}{\mathrm{d}x^2}. \tag{3.3.4}$$

If we abuse the notation by ignoring the evaluation $y = f(x)$ (like (3.3.3)) to write

$$\frac{\mathrm{d}^2 g}{\mathrm{d}x^2} = \frac{\mathrm{d}^2 g}{\mathrm{d}f^2} \left(\frac{\mathrm{d}f}{\mathrm{d}x}\right)^2 + \frac{\mathrm{d}g}{\mathrm{d}f} \frac{\mathrm{d}^2 f}{\mathrm{d}x^2},$$

which looks strange. Due to this reason, *I strongly not recommend to abuse the notation like (3.3.3)*.

REMARK 3.3.5 (Suggested notation). I would suggest a combination of Lagrange notation and Leibniz notation, with a lot of parentheses/brackets. For example, I like to write the chain rule in Lemma 3.3.2 as

$$(g(f(x)))' = \left.\frac{\mathrm{d}}{\mathrm{d}y} g(y)\right|_{y=f(x)} f'(x),$$

and the second order chain rule (3.3.4) as

$$(g(f(x)))'' = \left.\frac{\mathrm{d}^2 g}{\mathrm{d}y^2}\right|_{y=f(x)} (f'(x))^2 + \left.\frac{\mathrm{d}g}{\mathrm{d}y}\right|_{y=f(x)} f''(x)$$

to remind myself the "evaluation of points".

The term "implicit differentiation" is not really a theorem, which is more like an idea. We introduce this idea using the below example.

EXAMPLE 3.3.6 (Quotient rule). Let $a, b \in \mathbb{R}$ with $a < b$ and let $f : (a,b) \to \mathbb{R}$ and $g : (a,b) \to \mathbb{R}$ be functions, such that both f and g are differentiable at x_0 with $g(x) \neq 0$ for all $x \in (a,b)$. By using product rule and chain rule, one sees that

$$\frac{f}{g} : (a,b) \to \mathbb{R}, \quad \frac{f}{g}(x) := \frac{f(x)}{g(x)} = f(x)(g(x))^{-1} \text{ for all } x \in (a,b)$$

is differentiable at x_0, and an explicit formula can be computed accordingly. Here we exhibit another simple way to compute it. Write $h = \frac{f}{g}$, and we see that

$$f(x) = h(x)g(x). \tag{3.3.5}$$

Differentiate both sides of (3.3.5) at $x = x_0$, one sees that

$$f'(x_0) = (h(x)g(x))'|_{x=x_0} = h'(x_0)g(x_0) + h(x_0)g'(x_0),$$

that is,

$$h'(x_0) = \frac{f'(x_0) - h(x_0)g'(x_0)}{g(x_0)}. \tag{3.3.6}$$

We now combine (3.3.5) and (3.3.6) to reach

$$\left(\frac{f}{g}\right)'(x_0) = \frac{f'(x_0) - \frac{f(x_0)}{g(x_0)}g'(x_0)}{g(x_0)} = \frac{f'(x_0)g(x_0) - f(x_0)g'(x_0)}{(g(x_0))^2},$$

which is exactly the well-known *quotient rule*.

Main ideas of implicit differentiation. We first simplify the equation, then differentiate both sides, and in many case, one can compute it using product rule (Lemma 3.3.1) and chain rule (Lemma 3.3.2) above.

EXAMPLE 3.3.7 (Differentiation of logarithmic and exponential functions). By the definition of natural logarithm, one has

$$x = e^{\ln x} \quad \text{for all } x > 0.$$

Differentiate the above equation, and using chain rule, we see that

$$1 = \frac{\mathrm{d}}{\mathrm{d}x}x = \frac{\mathrm{d}}{\mathrm{d}x}(e^{\ln x}) = \frac{\mathrm{d}}{\mathrm{d}y}e^y\bigg|_{y=\ln x}(\ln x)' = e^y|_{y=\ln x}\frac{\mathrm{d}}{\mathrm{d}x}\ln x = x(\ln x)',$$

which concludes that

$$(\ln x)' = \frac{1}{x} \quad \text{for all } x \in \mathbb{R}.$$

From (1.3.5) and the linearity of differentiation, it is easy to see that

$$(\log_a x)' = \frac{(\ln x)'}{\ln a} = \frac{1}{\ln a}\frac{1}{x} \quad \text{for all } x > 0.$$

By using (1.3.6), we already know that $\log_a : \mathbb{R}_{>0} \to \mathbb{R}$ is the inverse function of the function $f : \mathbb{R} \to \mathbb{R}_{>0}$ given by $f(x) = a^x$ for all $x \in \mathbb{R}$. We now differentiate on both sides of

$$\log_a(a^x) = x \quad \text{for all } x \in \mathbb{R}$$

to see that

$$1 = \frac{\mathrm{d}}{\mathrm{d}x}x = \frac{\mathrm{d}}{\mathrm{d}x}(\log_a(a^x)) = \frac{\mathrm{d}}{\mathrm{d}y}\log_a y\bigg|_{y=a^x}(a^x)' = \frac{1}{\ln a}\frac{1}{x}\bigg|_{y=a^x}(a^x)' = \frac{1}{\ln a}a^{-x}(a^x)',$$

which conclude

$$(a^x)' = a^x \ln a \quad \text{for all } x > 0. \tag{3.3.7}$$

We also give another proof of (3.3.7), which is more direct: Since $a^x = e^{\ln(a^x)} = e^{x\ln a}$, then using chain rule one sees that

$$(a^x)' = \frac{\mathrm{d}}{\mathrm{d}y}e^y\bigg|_{y=x\ln a}\frac{\mathrm{d}}{\mathrm{d}x}(x\ln a) = e^y|_{y=x\ln a}\ln a = e^{x\ln a}\ln a = a^x\ln a.$$

Example 3.3.8 (Power function). Now one can easily extend Example 3.1.11 using implicit differentiation. Let $p \in \mathbb{R}$ and we consider the function $f : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ given by $f(x) = x^p$. According to the spirit of implicit differentiation, we first simplify the equation as

$$\ln(f(x)) = \ln(x^p) = p\ln x.$$

Now differentiate the above equation, and using product rule and chain rule to see that

$$p\frac{1}{x} = \frac{\mathrm{d}}{\mathrm{d}x}(p\ln x) = \frac{\mathrm{d}}{\mathrm{d}x}(\ln(f(x))) = \frac{\mathrm{d}}{\mathrm{d}y}\ln y\bigg|_{y=f(x)} f'(x) = \frac{1}{y}\bigg|_{y=f(x)} f'(x) = \frac{f'(x)}{f(x)},$$

which implies

$$f'(x) = pf(x)x^{-1} = px^{p-1} \quad \text{for all } x > 0.$$

Example 3.3.9. We define the function $f : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ by $f(x) = x^x$ for all $x > 0$, which is differentiable. According to the spirit of implicit differentiation, we first simplify the equation as

$$\ln(f(x)) = \ln(x^x) = x\ln x. \tag{3.3.8}$$

Now differentiate the above equation, and using product rule and chain rule to see that

$$\ln x + 1 = \frac{\mathrm{d}}{\mathrm{d}x}(x\ln x) = \frac{\mathrm{d}}{\mathrm{d}x}(\ln(f(x))) = \frac{\mathrm{d}}{\mathrm{d}y}\ln y\bigg|_{y=f(x)} f'(x) = \frac{1}{y}\bigg|_{y=f(x)} f'(x) = \frac{f'(x)}{f(x)},$$

which implies

$$f'(x) = f(x)(\ln x + 1) \quad \text{for all } x > 0. \tag{3.3.9}$$

One also can write (3.3.9) as $f'(x) = x^x(\ln x + 1)$ for all $x > 0$. It is much convenient to compute second derivative from (3.3.9):

$$f''(x) = (f(x)(\ln x+1))' = f'(x)(\ln x+1) + f(x)(\ln x)' = f'(x)(\ln x+1) + f(x)x^{-1} \tag{3.3.10}$$

for all $x > 0$, which also can be further simplify as

$$\begin{aligned} f''(x) &= f(x)(\ln x + 1)(\ln x + 1) + f(x)\frac{1}{x} \\ &= f(x)((\ln x + 1)^2 + x^{-1}) = x^x((\ln x + 1)^2 + x^{-1}). \end{aligned}$$

Again, it is more convenient to compute third derivative from (3.3.10), and the procedure can be done for arbitrary order of derivative.

Exercise 3.3.10. We define the function $g : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ by $g(x) = x^{x^x}$ for all $x > 0$. Compute its derivative $g' : \mathbb{R}_{>0} \to \mathbb{R}$.

Similar ideas also works for taking limit. which is also an efficient way to proof the existence of limits:

Taking limit implicitly. We first simplify the equation using continuous function, and then taking limit (L' Hôpital's rule is helpful). Finally solve the resulting equation.

EXAMPLE 3.3.11. Let f be the function given in Example 3.3.9. We want to show that $\lim_{x\to 0+} f(x)$ exists and compute its value by writing (3.3.8). Since $\lim_{x\to 0+} \ln x = -\infty$ and $\lim_{x\to 0+} x^{-1} = +\infty$, by using L' Hôpital's rule, we take limit on (3.3.8) to see that

$$\lim_{x\to 0+} \ln(f(x)) = \lim_{x\to 0+} x\ln x = \lim_{x\to 0+} \frac{\ln x}{x^{-1}} = \lim_{x\to 0+} \frac{(\ln x)'}{(x^{-1})'} = \lim_{x\to 0+} \frac{x^{-1}}{-x^{-2}} = -\lim_{x\to 0+} x = 0.$$

Note that we cannot directly use the continuity of $\ln : \mathbb{R}_{>0} \to \mathbb{R}$ to obtain $\ln(\lim_{x\to 0+} f(x))$ since we do not know whether $\lim_{x\to 0+} f(x)$ exists or not at the moment. The proper way to argue this is to use the continuity of $\exp : \mathbb{R} \to \mathbb{R}_{>0}$ to see that

$$1 = \exp(0) = \exp\left(\lim_{x\to 0+} \ln(f(x))\right) = \lim_{x\to 0+} \exp\left(\ln(f(x))\right) = \lim_{x\to 0+} f(x),$$

which conclude our result.

EXERCISE 3.3.12. Show that

$$\lim_{x\to 0}(1+x)^{1/x} = e.$$

We now consider the derivative of trigonometric functions. We begin the our discussion from the following lemma.

LEMMA 3.3.13. *One has*

$$\left.\frac{\mathrm{d}}{\mathrm{d}\theta}\sin\theta\right|_{\theta=0} = \lim_{h\to 0}\frac{\sin h}{h} = 1 \quad \textit{and} \quad \left.\frac{\mathrm{d}}{\mathrm{d}\theta}\cos\theta\right|_{\theta=0} = \lim_{h\to 0}\frac{\cos h - 1}{h} = 0.$$

PROOF. Regarding the first result, since

$$\frac{\sin h}{h} = \frac{\sin(-h)}{-h} \quad \text{for all } h \neq 0,$$

it is suffice to show the right limit

$$\lim_{h\to 0+}\frac{\sin h}{h} = 1. \tag{3.3.11}$$

By using the definition of angle (in radian), one observes that

$$\sin h \leq h \leq \tan h = \frac{\sin h}{\cos h} \quad \text{for all } 0 < h < \frac{\pi}{2},$$

see Figure 1.3.3, which implies that

$$\cos h \leq \frac{\sin h}{h} \leq 1 \quad \text{for all } 0 < h < \frac{\pi}{2},$$

hence

$$1 = \lim_{h\to 0+}\cos h = \liminf_{h\to 0+}\cos h \leq \liminf_{h\to 0+}\frac{\sin h}{h} \leq \limsup_{h\to 0+}\frac{\sin h}{h} = 1,$$

which conclude (3.3.11) by Theorem 2.3.6.

We now prove the second result from the first result. One sees that

$$\begin{aligned}\lim_{h\to 0}\frac{\cos h - 1}{h} &= \lim_{h\to 0}\left(\frac{\cos h - 1}{h}\frac{\cos h + 1}{\cos h + 1}\right) = \lim_{h\to 0}\frac{(\cos h)^2 - 1}{h(\cos h + 1)} \\ &= \lim_{h\to 0}\frac{-(\sin h)^2}{h(\cos h + 1)} = -\left(\lim_{h\to 0}\frac{\sin h}{h}\right)\left(\lim_{h\to 0}\frac{\sin h}{\cos h + 1}\right) = 0,\end{aligned}$$

which conclude our result. ☐

We now ready to compute the derivatives of trigonometric functions.

LEMMA 3.3.14. *One has*

$$(\sin\theta)' = \cos\theta \quad \textit{and} \quad (\cos\theta)' = \sin\theta.$$

PROOF. We recall the sum-to-product rule (can be easily proved using de Moivre theorem):

$$\begin{aligned}\cos(\theta_1+\theta_2) &= \cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2,\\ \sin(\theta_1+\theta_2) &= \cos\theta_1\sin\theta_2 + \sin\theta_1\cos\theta_2,\end{aligned}$$

which holds true for all $\theta_1, \theta_2 \in \mathbb{R}$. By using the linearity of limits, one computes that

$$\begin{aligned}(\sin\theta)' &= \lim_{h\to 0}\frac{\sin(\theta+h)-\sin\theta}{h} = \lim_{h\to 0}\frac{\cos\theta\sin h + \sin\theta\cos h - \sin\theta}{h}\\ &= \lim_{h\to 0}\left(\sin\theta\left(\frac{\cos h - 1}{h}\right) + \cos\theta\left(\frac{\sin h}{h}\right)\right)\\ &= \sin\theta\lim_{h\to 0}\left(\frac{\cos h - 1}{h}\right) + \cos\theta\lim_{h\to 0}\left(\frac{\sin h}{h}\right)\\ &= \cos\theta\end{aligned}$$

and

$$\begin{aligned}(\cos\theta)' &= \lim_{h\to 0}\frac{\cos(\theta+h)-\cos\theta}{h} = \lim_{h\to 0}\frac{\cos\theta\cos h - \sin\theta\sin h - \cos\theta}{h}\\ &= \lim_{h\to 0}\left(\cos\theta\left(\frac{\cos h - 1}{h}\right) - \sin\theta\left(\frac{\sin h}{h}\right)\right)\\ &= \cos\theta\lim_{h\to 0}\left(\frac{\cos h - 1}{h}\right) - \sin\theta\lim_{h\to 0}\left(\frac{\sin h}{h}\right)\\ &= -\sin\theta,\end{aligned}$$

which concludes the lemma. □

The derivative of the trigonometric functions $\tan\theta$, $\cot\theta$, $\sec\theta$ and $\csc\theta$ can be easily proved using product rule, chain rule as well as implicit differentiation, here we omit the details. Now lets summarize the ideas before using the following examples.

EXAMPLE 3.3.15. We define the function $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} x^2\sin(1/x) & , x \neq 0,\\ 0 & , x = 0.\end{cases}$$

We compute

$$\begin{aligned}f'(x) &= 2x\sin(1/x) + x^2(\sin(1/x))' \quad \text{(product rule)}\\ &= 2x\sin(1/x) + x^2\cos(1/x)(-x^{-2}) \quad \text{(chain rule)}\\ &= 2x\sin(1/x) - \cos(1/x) \quad \text{for all } x \in \mathbb{R}\setminus\{0\}.\end{aligned}$$

By using the additivity property (Remark 2.3.11), one sees that

$$\limsup_{x\to 0} f'(x) = 1 \neq -1 = \liminf_{x\to 0} f'(x),$$

which shows that $\lim_{x\to 0} f'(x)$ does not exist. We now show that f is differentiable at $x = 0$. We also compute that

$$\limsup_{h\to 0}\left|\frac{h^2\sin(1/h) - 0}{h}\right| = \limsup_{h\to 0}|h\sin(1/h)| \leq \limsup_{h\to 0}|h| = 0,$$

which conclude that

$$f'(0) = \lim_{h\to 0}\frac{f(h) - f(0)}{h} = \lim_{h\to 0}\frac{h^2\sin(1/h) - 0}{h} = 0.$$

Thus $f' : \mathbb{R} \to \mathbb{R}$ is a well-defined function, which is not continuous at $x = 0$.

The ideas in Example 3.3.7 can be further generalized:

Example 3.3.16. Let $f : I_1 \to I_2$ be a bijective function, which is differentiable. Suppose that its inverse function $f^{-1} : I_2 \to I_1$ is also differentiable. The definition of inverse function gives

$$x = f(f^{-1}(x)) \quad \text{for all } x \in I_2.$$

Differentiate the above equation, and using chain rule, we see that

$$1 = \frac{\mathrm{d}}{\mathrm{d}x}x = \frac{\mathrm{d}}{\mathrm{d}x}(f(f^{-1}(x))) = \left.\frac{\mathrm{d}f}{\mathrm{d}y}\right|_{y=f^{-1}(x)}(f^{-1}(x))',$$

that is,

$$(f^{-1}(x))' = \frac{1}{\left.\frac{\mathrm{d}f}{\mathrm{d}y}\right|_{y=f^{-1}(x)}}. \tag{3.3.12}$$

Indeed, Example 3.3.7 is nothing but just a special case $f = \exp : \mathbb{R} \to \mathbb{R}_{>0}$ with $f^{-1} = \ln : \mathbb{R}_{>0} \to \mathbb{R}$:

$$(\ln x)' = \frac{1}{e^y|_{y=\ln(x)}} = \frac{1}{x} \quad \text{for all } x > 0.$$

Remark. Some authors may abuse the notation by writing (3.3.12) as

$$\frac{\mathrm{d}x}{\mathrm{d}y} = \frac{1}{\frac{\mathrm{d}y}{\mathrm{d}x}},$$

especially while performing the change of variables in Riemann/Lebesgue integral. *I suggest not to abuse the notation like this.*

Example 3.3.17 (Differentiation of inverse trigonometric functions). Here we only exhibit the differentiation of $\sin^{-1} : (-1,1) \to (-\pi/2, \pi/2)$ based on the principal in Example 3.3.16. Differentiate the equation

$$x = \sin(\sin^{-1}x) \quad \text{for all } x \in (-1,1),$$

one sees that

$$1 = \frac{\mathrm{d}}{\mathrm{d}x}x = \frac{\mathrm{d}}{\mathrm{d}x}(\sin(\sin^{-1}x)) = \cos(\sin^{-1}x)(\sin^{-1}x)'.$$

Since $-\pi/2 < \sin^{-1}x < \pi/2$ for all $x \in (-1,1)$, then $\cos(\sin^{-1}x) > 0$ for all $x \in (-1,1)$, therefore dividing the above equation by $\cos(\sin^{-1}x)$ implies (*one has to make sure not to divide by 0*)

$$(\sin^{-1}x)' = \frac{1}{\cos(\sin^{-1}x)} \quad \text{for all } x \in (-1,1). \tag{3.3.13}$$

In fact, one can further simplify (not necessary) the formula: By choosing $\theta = \sin^{-1} x$ in the formula $(\cos\theta)^2 + (\sin\theta)^2 = 1$, one sees that

$$(\cos(\sin^{-1} x))^2 = 1 - x^2 \quad \text{for all } x \in (-1,1).$$

Since both $1 - x^2 > 0$ and $\cos(\sin^{-1} x) > 0$ for all $x \in (-1,1)$, then

$$\cos(\sin^{-1} x) = \sqrt{1 - x^2} \quad \text{for all } x \in (-1,1). \tag{3.3.14}$$

Combining (3.3.13) and (3.3.14) we reach

$$(\sin^{-1} x)' = \frac{1}{\sqrt{1 - x^2}} \quad \text{for all } x \in (-1,1).$$

EXERCISE 3.3.18. Compute the derivative of $\cos^{-1} : (-1,1) \to (0,\pi)$.

3.4. Mean value theorem

DEFINITION 3.4.1 (Local extrema in interior). Let $a, b \in \mathbb{R}$ with $a < b$. We say that $x_0 \in (a,b)$ is a *local maximum* (resp. *local minimum*) of $f : (a,b) \to \mathbb{R}$ if there exists $\delta > 0$ such that $f(x_0) \geq f(x)$ (resp. $f(x_0) \leq f(x)$) for all $x \in B_\delta(x_0)$.

In order to unify the notations, here we also introduce the following definition.

DEFINITION 3.4.2 (Local extrema at boundary). Let $a, b \in \mathbb{R}$ with $a < b$. We say that $x_0 = a$ is a *local maximum* (resp. *local minimum*) of $f : [a,b] \to \mathbb{R}$ if there exists $\delta > 0$ such that $f(x_0) \geq f(x)$ (resp. $f(x_0) \leq f(x)$) for all $a \leq x < a + \delta$. Similarly, we say that $x_0 = b$ is a *local maximum* (resp. *local minimum*) of $f : [a,b] \to \mathbb{R}$ if there exists $\delta > 0$ such that $f(x_0) \geq f(x)$ (resp. $f(x_0) \leq f(x)$) for all $b - \delta < x \leq b$.

In particular, the above definitions are just special case of the following general notion.

DEFINITION 3.4.3. Let E be any set in $\mathbb{R}$. We say that $x_0 \in E$ is a *local maximum* (resp. *local minimum*) of $f : E \to \mathbb{R}$ if there exists $\delta > 0$ such that $f(x_0) \geq f(x)$ (resp. $f(x_0) \leq f(x)$) for all $x \in B_\delta(x_0) \cap E$. In contrast, we say that f has a *global maximum* (resp. *local minimum*) at $x_0 \in E$ if $f(x_0) \geq f(x)$ (resp. $f(x_0) \leq f(x)$) for all $x \in E$.

REMARK. Obviously, if x_0 is a global maximum/minimum, then it is also a local maximum/minimum.

EXAMPLE 3.4.4. We define the function $f : [-1,1] \to \mathbb{R}$ by

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

One sees that $x = 0$ is a global maximum of $f : [-1,1] \to \mathbb{R}$ with value $f(0) = 1$. According to the above unify notations, we see that $x = 1$ is a local maximum of $f : [-1,1] \to \mathbb{R}$ with value $f(1) = 0$. However, $x = 1$ is *not* a local maximum of $f : [-1,1) \to \mathbb{R}$ because $1 \notin [-1,1)$, this reminds the readers that *one always need to write down the domain of functions* carefully. One also sees that $x = -1$ is a local minimum of $f : [-1,1] \to \mathbb{R}$, but it is not global since

$$\lim_{x \to 0-} f(x) = -2 < 0 = f(-1).$$

The following lemma suggests that one can find some candidate of *local* maximum/minimum using differentiation.

LEMMA 3.4.5. *Let $a, b \in \mathbb{R}$ with $a < b$ and let $f : (a, b) \to \mathbb{R}$. If f has a local maximum or local minimum at $x_0 \in (a, b)$ and if f is differentiable at x_0, then $f'(x_0) = 0$.*

PROOF. Suppose that f has a local maximum at $x_0 \in (a, b)$. By definition, there exists $\delta > 0$ such that $f(x_0) \geq f(x)$ for all $x \in B_\delta(x_0)$. We now see that

$$\frac{f(x) - f(x_0)}{x - x_0} \leq 0 \text{ for all } x_0 < x < x_0 + \delta,$$
$$\frac{f(x) - f(x_0)}{x - x_0} \geq 0 \text{ for all } x_0 - \delta < x < x_0.$$

Since $f'(x_0) = \lim_{x\to x_0} \frac{f(x)-f(x_0)}{x-x_0}$ exists, then our lemma immediately follows. □

THEOREM 3.4.6 (Cauchy mean value theorem). *Let $a, b \in \mathbb{R}$ with $a < b$. Suppose that $f_1 : [a, b] \to \mathbb{R}$ and $f_2 : [a, b] \to \mathbb{R}$ are continuous functions, where the continuity at boundary points are understood as $f_j(a) = \lim_{x\to a+} f_j(x)$ and $f_j(b) = \lim_{x\to b-} f_j(x)$ for each $j = 1, 2$, such that it is differentiable on (a, b). Then there exists a point $x_0 \in (a, b)$ such that*

$$(f_1(b) - f_1(a))f_2'(x_0) = (f_2(b) - f_2(a))f_1'(x_0).$$

PROOF. We define the function $h : [a, b] \to \mathbb{R}$ by

$$h(t) := (f_1(b) - f_1(a))f_2(t) - (f_2(b) - f_2(a))f_1(t) \quad \text{for all } t \in [a, b]$$

which is also differentiable on (a, b), and one can check that $h(a) = h(b)$. It is remains to show that $h'(x_0) = 0$.

Case 1: Suppose that there exists $t \in (a, b)$ such that $h(t) > h(a)$. By using Theorem 2.2.7, there exists $x_0 \in [a, b]$ such that

$$h(x_0) \geq h(x) \quad \text{for all } x \in [a, b].$$

In this case, one has $x_0 \neq a$ and $x_0 \neq b$, therefore Lemma 3.4.5 gives $h'(x_0) = 0$.

Case 2: Suppose that there exists $t \in (a, b)$ such that $h(t) < h(a)$. By using Theorem 2.2.7, there exists $x_0 \in [a, b]$ such that

$$h(x_0) \leq h(x) \quad \text{for all } x \in [a, b].$$

In this case, one has $x_0 \neq a$ and $x_0 \neq b$, therefore Lemma 3.4.5 gives $h'(x_0) = 0$.

Case 3: Suppose that both Case 1 and Case 2 do not hold. By definition of differentiation, one has $h(t) = h(a)$ for all $t \in [a, b]$, thus $h'(x) = 0$ for all $x \in (a, b)$. □

The following corollary corresponding to the special case $f_1(x) = f(x)$ and $f_2(x) = x$ in Theorem 3.4.6.

COROLLARY 3.4.7. *Let $a, b \in \mathbb{R}$ with $a < b$. Suppose that $f : [a, b] \to \mathbb{R}$ is a continuous functions such that it is differentiable on (a, b). Then there exists a point $x_0 \in (a, b)$ such that*

$$f(b) - f(a) = (b - a)f'(x_0).$$

The following corollary corresponding to the special case $f_1(x) = f(x)$ and $f_2(x) = x$ in Theorem 3.4.6 with $f(a) = f(b)$.

COROLLARY 3.4.8 (Rolle's theorem). *Let $a, b \in \mathbb{R}$ with $a < b$. Suppose that $f : [a, b] \to \mathbb{R}$ is a continuous functions such that $f(a) = f(b)$ and it is differentiable on (a, b). Then there exists a point $x_0 \in (a, b)$ such that $f'(x_0)$.*

By using Corollary 3.4.7, one immediately sees that:

COROLLARY 3.4.9. *Let $a, b \in \mathbb{R}$ with $a < b$ and let $f : (a, b) \to \mathbb{R}$ be a differentiable function.*

(a) *If $f'(x) \geq 0$ for all $x \in (a, b)$, then f is nondecreasing. If one has the strict inequality $f(x) > 0$ for all $x \in (a, b)$, then f is strictly increasing.*
(b) *If $f'(x) \leq 0$ for all $x \in (a, b)$, then f is nonincreasing. If one has the strict inequality $f(x) < 0$ for all $x \in (a, b)$, then f is strictly decreasing.*
(c) *If $f'(x) = 0$ for all $x \in (a, b)$, then f is a constant function.*

PROOF. In order to proof (a), we want to show $f(t_1) \leq f(t_2)$ for all $a < t_1 < t_2 < b$. By using the mean value theorem (Corollary 3.4.7) on $[t_1, t_2]$, there exists $t_3 \in (t_1, t_2)$ such that

$$f(t_2) - f(t_1) = (t_2 - t_1) f'(t_3) \geq 0,$$

which conclude the first result in (a). One sees that the second result in (a) can be easily proof as well. The proof of (b) can be done similarly. Combining (a) and (b) we obtain (c). ☐

3.5. Extreme values

We now consider the problem of finding maximums as well as minimums. We will see that this problem is actually extremely difficult for general function. Rather than solving the problem directly, we first find some "candidates". You can think about the election of president: We first nominate candidates first, and then vote for present among these candidates. Suggested by Lemma 3.4.5, we restricted ourselves for differentiable functions in this course.

DEFINITION 3.5.1. Let $a, b \in \mathbb{R}$ with $a < b$ and let $f : (a, b) \to \mathbb{R}$. If f is differentiable at x_0 and $f'(x_0) = 0$, then x_0 is called a *critical point* of f.

EXAMPLE 3.5.2. Let $f(x) = (x - x_0)^3$ for all $x \in \mathbb{R}$, which satisfies $f'(x_0) = 0$. Since $f : \mathbb{R} \to \mathbb{R}$ is non-decreasing, therefore x_0 is neither local maximum nor minimum of f.

DEFINITION 3.5.3. Let E be a set in $\mathbb{R}$. We define the (topological) boundary ∂E of E by

$$\partial E := \{x \in \mathbb{R} : B_r(x) \cap E \neq \emptyset \text{ and } B_r(x) \cap (\mathbb{R} \setminus E) \neq \emptyset \text{ for all } r > 0\}.$$

EXAMPLE 3.5.4. Let $a, b \in \mathbb{R}$ with $a < b$. We see that a is a boundary point of (a, b). We see that a is also a boundary point of $[a, b]$. This notion also works for unbounded set, for example, a is a boundary point of $(a, +\infty)$, as well as a boundary point of $[a, +\infty)$.

ALGORITHM 3.5.5 (See also Algorithm 3.5.11 below for a refinement). *Let E be a set in $\mathbb{R}$ and let $f : E \to \mathbb{R}$ be a function. Suppose that $f : E_0 \to \mathbb{R}$ is differentiable for some $E_0 \subset E$. All candidates must be either one of the followings:*

(a) *critical points in E_0 (i.e. among all points which are differentiable), which is based on Lemma 3.4.5.*
(b) *those points in $E \setminus E_0$, that is, those points which are not differentiable. (Note: the boundary points which are in E are element in $E \setminus E_0$)*

The boundary points which is not in E are not candidates, but its left limit/right limit/limit superior/limit inferior is helpful to decide whether the local maximum/minimum is global or not.

We now explain how to use Algorithm 3.5.5 in the following example.

EXAMPLE 3.5.6. We define the function $f : [-1,2) \to \mathbb{R}$ given by

$$f(x) = \begin{cases} (x-1/2)^2 & , 0 \leq x < 2, \\ \frac{1}{2}x + 1 & , -1 \leq x < 0. \end{cases}$$

We now find *all* candidates as in Algorithm 3.5.5:

(1) **Critical point.** We first observe that f is differentiable on $(-1,0) \cup (0,1)$, and note that

$$f'(x) = \begin{cases} 2(x-1/2) & , 0 < x < 2, \\ 1/2 & , -1 < x < 0. \end{cases}$$

We see that the only critical described in Algorithm 3.5.5 is $x_0 = \frac{1}{2}$, and one sees that

$$x_0 = \frac{1}{2} \text{ is a local minimum of } f : [-1,1) \to \mathbb{R} \text{ with } f(x_0) = 0.$$

(2) **Nondifferentiable points.** We see that f is not differentiable at $x_1 = -1$ (boundary point) and $x_3 = 0$. One sees that

$$x_1 = -1 \text{ is a local minimum of } f : [-1,1) \to \mathbb{R} \text{ with } f(x_1) = \frac{1}{2}.$$

On the other hand, since

$$f(x_3) = \lim_{x \to x_3+} f(x) = \frac{1}{2}, \quad \lim_{x \to x_3-} f(x) = 1,$$

and this is helpful to see that $x_3 = 0$ is neither local maximum nor local minimum. Note that the boundary point $x_2 = 2 \notin [-1,2)$, therefore it is not a candidate, but it is helpful to decide whether other candidate is local maximum/minimum or not. We will use the following fact later: $\lim_{x \to x_2-} f(x) = 9/4$.

Other than the above three types of points are all not candidate, and they are not possible to be local maximum/minimum at all. We now conclude that all local extrema of $f : [-1,1) \to \mathbb{R}$ is:

(i) Local minimum: $x_0 = \frac{1}{2}$ and $x_1 = -1$.
(ii) Local maximum: none.

In order to decide whether the local extrema are global, we list all values of candidates as well as interesting points:

$$f(x_0) = 0, \quad f(x_1) = \frac{1}{2}, \quad \lim_{x \to x_2-} f(x) = \frac{9}{4}, \quad f(x_3) = \lim_{x \to x_3+} f(x) = \frac{1}{2}, \quad \lim_{x \to x_3-} f(x) = 1.$$

We see that $f(x_0) = 0$ takes the smallest value, and thus we know that $x_0 = \frac{1}{2}$ is indeed a global minimum. But however, we see that $\lim_{x \to x_2-} f(x) = \frac{9}{4}$ takes the largest value, and since $x_2 \notin [-1,2)$, thus we conclude that there is *no global maximum* (another way to see this is there is no local maximum). In terms of election, the candidates are

$$x_0 = \frac{1}{2}, \quad x_1 = -1, \quad x_3 = 0.$$

After the "voting", $x_0 = \frac{1}{2}$ becomes local minimum, $x_1 = -1$ becomes global minimum, and x_3 is failed to be chosen (due to x_2, which is not a candidate). Even though x_2 is not a "candidate", but it can affect the result of the "election".

As promised at the beginning, we now exhibit an example to explain the difficulties of the problem consider in this section.

EXAMPLE 3.5.7. Let $a, b \in \mathbb{R}$ with $a < b$ and let $f : (a, b) \to \mathbb{R}$. If $x_0 \in (a, b)$ satisfies

$$f'(x) \le 0 \text{ for all } x \in (x_0 - \delta, x_0] \quad \text{and} \quad f'(x) \ge 0 \text{ for all } x \in [x_0 x_0 + \delta) \tag{3.5.1}$$

for some $\delta > 0$, then it is easy to see that x_0 is a local minimum of $f : (a, b) \to \mathbb{R}$. However, the converse may false. For example, we consider the function

$$f : \mathbb{R} \to \mathbb{R}, \quad f(x) = \begin{cases} x^2(2 + \sin(1/x)) & , x \in \mathbb{R} \setminus \{0\}, \\ 0 & , x = 0, \end{cases}$$

since $2 + \sin(1/x) \ge 1$for all $x \in \mathbb{R} \setminus \{0\}$, then $f(x) \ge x^2$ for all $x \in \mathbb{R} \setminus \{0\}$, which implies that

$$f(x) > 0 = f(x) \quad \text{for all } x \in \mathbb{R} \setminus \{0\},$$

which implies that $x_0 = 0$ is the unique global minimum of f. By using product rule and chain rule, one can compute

$$f'(x) = 4x + 2x\sin(1/x)) - \cos(1/x) \quad \text{for all } x \in \mathbb{R} \setminus \{0\}.$$

As mentioned in Example 3.3.15 before, one can prove that f is differentiable at $x = 0$ with $f'(0) = 0$, which is not possible to prove using product rule and chain rule. We now see that the derivative reads

$$f'(x) = \begin{cases} 4x + 2x\sin(1/x)) - \cos(1/x) & , x \in \mathbb{R} \setminus \{0\} \\ 0 & , x = 0, \end{cases}$$

which is not continuous since $\limsup_{x \to 0} f'(x) = 1 \neq 0 = f'(0)$ and it does not satisfy (3.5.1).

Despite Example (3.5.7) says that it is not possible to determine all local minimum by using only the intuitively criteria (3.5.1), but however the condition (3.5.1) is already good enough in many practical case. One way to guarantee (3.5.1) is that $f' : (x_0 - \delta, x_0 + \delta) \to \mathbb{R}$ is non-decreasing and $f'(x_0) = 0$. If $f : (x_0 - \delta, x_0 + \delta) \to \mathbb{R}$ is twice differentiable, this means that $f''(x) \ge 0$ for all $x \in (x_0 - \delta, x_0 + \delta)$. In fact, one have the following theorem (see also Lemma 4.5.1 below).

THEOREM 3.5.8 (Second derivative test). *Let $a, b \in \mathbb{R}$ with $a < b$. Suppose that $f \in C^2((a, b))$, which means that $f : (a, b) \to \mathbb{R}$ is twice differentiable and $f'' : (a, b) \to \mathbb{R}$ is continuous, then the following holds:*

(a) *If $f'(x_0) = 0$ and the* strict *inequality $f''(x_0) > 0$ hold for some $x_0 \in (a, b)$, then x_0 is a local minimum of $f : (a, b) \to \mathbb{R}$.*
(b) *If $f'(x_0) = 0$ and the* strict *inequality $f''(x_0) < 0$ hold for some $x_0 \in (a, b)$, then x_0 is a local maximum of $f : (a, b) \to \mathbb{R}$.*

EXAMPLE 3.5.9. The strict inequality in Theorem 3.5.8 is necessary. For example, if we consider the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^3$ for all $x \in \mathbb{R}$, we see that $f'(0) = 0$ and $f''(0) = 0$, but $x = 0$ is neither local maximum nor local minimum.

DEFINITION 3.5.10. Let $a, x_0, b \in \mathbb{R}$with $a < x_0 < b$ and let $f : (a, b) \to \mathbb{R}$ be a function. We say that f is C^2 *near* x_0 if there exists $\delta > 0$ such that $B_\delta(x_0) \subset (a, b)$ and $f \in C^2(B_\delta(x_0))$.

In view of Theorem 3.5.8, we may slightly enhance Algorithm 3.5.5 for C^2-functions (we highlight the refinements in blue text).

ALGORITHM 3.5.11. *Let E be a set in $\mathbb{R}$ and let $f : E \to \mathbb{R}$ be a function. Suppose that $f : E_0 \to \mathbb{R}$ is differentiable for some $E_0 \subset E$. All candidates must be either one of the followings:*

(a) *critical points in E_0 (i.e. among all points which are differentiable). If f is C^2 near a critical point, says x_0, then we can use Theorem 3.5.8 to check whether it is a local maximum/minimum.*
(b) *those points in $E \setminus E_0$, that is, those points which are not differentiable. (Note: the boundary points which are in E are element in $E \setminus E_0$)*

EXERCISE 3.5.12. Let $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} -x^2 & , x < 0, \\ x^3 & , x \geq 0. \end{cases}$$

Show that $f'(0)$ exists, but $f''(0)$ does not exist. [Note: One cannot assume the continuity of f' in a valid argument]

CHAPTER 4

Riemann integrals

4.1. Definition of Riemann integrals and fundamental theorems of calculus

Let $a, b \in \mathbb{R}$ with $a < b$ and let $f : [a,b] \to \mathbb{R}$ be a function. Suppose that

$$f : [a,b] \to \mathbb{R} \text{ is continuous and } f(x) \geq 0 \text{ for all } x \in [a,b], \tag{4.1.1}$$

we want to compute the area of the region under the graph f and above the interval $[a,b]$, more precisely, the area of the set

$$S = \left\{(x,y) \in \mathbb{R}^2 : a \leq x \leq b, 0 \leq y \leq f(x)\right\},$$

see also Figure 4.1.1 below:

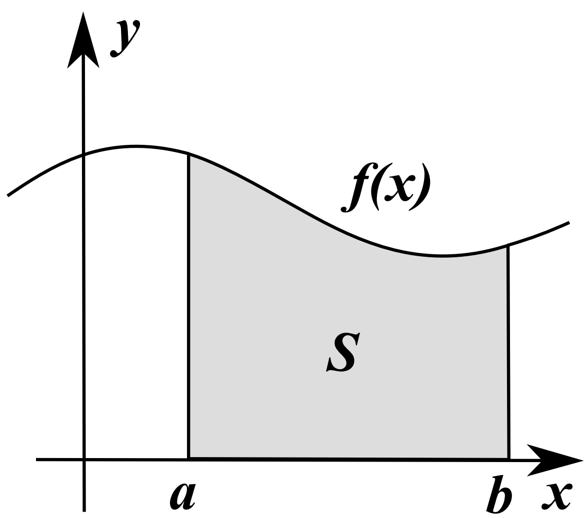

FIGURE 4.1.1. Motivation of Riemann integral: 4C, CC BY-SA 3.0, via Wikimedia Commons

Before giving a rigorous definition, we first approximate the area S intuitively.

DEFINITION 4.1.1. Let $[a,b]$ be a given interval. By a partition Γ of $[a,b]$ we mean a *finite* set of points $x_0, x_1, \cdots, x_n$, where

$$a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b.$$

In order to shorten the notations, we abuse the notation (throughout this course) by denoting the partition as $\Gamma = \{a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b\}$. We also define the *partition norm* as

$$|\Gamma| := \max_{i=1,\cdots,n} (x_i - x_{i-1}),$$

which is the length of the largest (closed) interval $[x_{i-1}, x_i]$.

For each $i \in \{1, \cdots, n\}$, we randomly choose a point $x_i^* \in [x_i, x_{i-1}]$, and we approximate the area S by the Riemann sum

$$R(f, \Gamma, \{y_i\}_{i=1}^n) = \sum_{i=1}^n f(x_i^*)(x_i - x_{i-1}) \tag{4.1.2}$$
$$= f(x_1^*)(x_1 - x_0) + f(x_2^*)(x_2 - x_1) + \cdots + f(x_n^*)(x_n - x_{n-1}),$$

where we observe that see $f(x_i^*)(x_i - x_{i-1})$ is exactly the area of the rectangle with base $[x_{i-1}, x_i]$ and height $f(x_i^*)$, see Figure 4.1.2 below for $i = 3$:

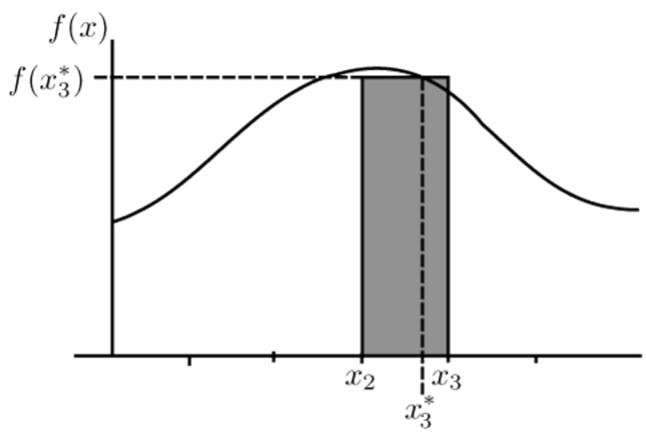

FIGURE 4.1.2. The area of the rectangle with base $[x_2, x_3]$ and height $f(x_3^*)$: Juliusross~commonswiki, CC BY 2.5, via Wikimedia Commons

Under assumptions (4.1.1),

$$\begin{gathered} R(f, \Gamma, \{y_i\}_{i=1}^n) \text{ will give a "fairly good" approximate} \\ \text{for the area of } S \text{ when } |\Gamma| \text{ is "small",} \end{gathered} \tag{4.1.3}$$

see Figure 4.1.3 below:

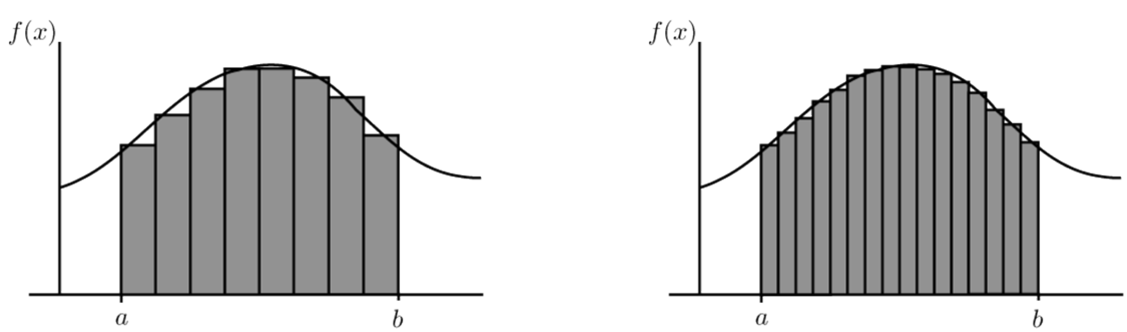

FIGURE 4.1.3. Approximate the area using Riemann sum: Juliusross~commonswiki, CC BY 2.5, via Wikimedia Commons

We now see that the figures in Figure 4.1.1, Figure 4.1.2 and Figure 4.1.3 are all drawn for the case when f satisfies (4.1.1). However, we see that the Riemann sum (4.1.2) is actually well-defined without assumption "$f(x) \geq 0$ for all $x \in [a,b]$", which means that we actually can define the "signed area" by using Riemann sum (4.1.2). For example, if we consider $f(x) = \sin x$ for all $x \in [-\pi/2, \pi/2]$, the Riemann sum will suggests that the area between f and the interval $[-\pi/2.\pi/2]$ will be 0.

In order to give a precise statement, the terms "fairly good" and "small" in the idea (4.1.3) need to be clarify. This can be done by using similar ideas for define limits (without assuming the redundant assumptions in (4.1.1)).

Definition 4.1.2 (Riemann integral via Riemann sum). Let $a,b \in \mathbb{R}$ with $a < b$. We say that $f : [a,b] \to \mathbb{R}$ is *Riemann integrable on* $[a,b]$ if there exists a number $L \in \mathbb{R}$ such that the following holds: Given any $\epsilon > 0$, there exists $\delta = \delta(\epsilon) > 0$ such that

$$|R(f,\Gamma,\{x_i^*\}_{i=1}^n) - L| < \epsilon$$

for all partition $\Gamma = \{a = x_0 < x_1 < \cdots < x_n = b\}$ with $|\Gamma| < \delta$ and for all $x_i^* \in [x_{i-1}, x_i]$. In this case, we denote $L = \int_a^b f(x)\,\mathrm{d}x$. Here the variable "$x$" can be replaced by other variables, for example, $\int_a^b f(x)\,\mathrm{d}x = \int_a^b f(t)\,\mathrm{d}t$.

Remark 4.1.3. In order to unify our notations, we use the convention

$$\int_b^a f(x)\,\mathrm{d}x := -\int_a^b f(x)\,\mathrm{d}x,$$

so that $\int_b^a f(x)\,\mathrm{d}x = 0$ if $a = b$.

In fact, one can simplify Definition 4.1.2 which looks rather complicated. Rather than consider arbitrary Riemann sum, we now always overestimate/underestimate the area: For any partition $\Gamma = \{a = x_0 < x_1 < \cdots < x_n = b\}$, we define

$$\begin{aligned} U(f,\Gamma) &:= \sum_{i=1}^n \left(\sup_{y \in [x_{i-1},x_i]} f(y) \right)(x_i - x_{i-1}) \quad \text{(upper sum)}, \\ L(f,\Gamma) &:= \sum_{i=1}^n \left(\inf_{y \in [x_{i-1},x_i]} f(y) \right)(x_i - x_{i-1}) \quad \text{(lower sum)}. \end{aligned}$$

It is clear that

$$L(f,\Gamma) \leq R(f,\Gamma,\{x_i^*\}_{i=1}^n) \leq U(f,\Gamma) \text{ for any partition } \Gamma \text{ of } [a,b].$$

In fact, by using [**Rud87**, Theorem 6.6], Definition 4.1.2 is equivalent to the following definition.

Definition 4.1.4 (An equivalent definition of Riemann integral: Darboux definition). Let $a,b \in \mathbb{R}$ with $a < b$. We say that $f : [a,b] \to \mathbb{R}$ is *(Riemann) integrable on* $[a,b]$ if: Given any $\epsilon > 0$, there exists a partition Γ_ϵ) of $[a,b]$ such that

$$U(f,\Gamma_\epsilon) - L(f,\Gamma_\epsilon) < \epsilon.$$

Remark. The partition norm $|\Gamma_\epsilon|$ of the partition Γ_ϵ in Definition 4.1.4 need not to be small. Therefore it is much more convenient to use the equivalent formulation in Definition 4.1.4 in mathematical proof.

EXAMPLE 4.1.5 (Areas between Curves). Let $a, b \in \mathbb{R}$ with $a < b$. Let $f : [a, b] \to \mathbb{R}$ and $g : [a, b] \to \mathbb{R}$ are integrable functions such that

$$f(x) \leq g(x) \quad \text{for all } x \in [a, b].$$

Then the area between curves f and g, more precisely, the area of the set

$$\{(x, y) \in [a, b] \times \mathbb{R} : f(x) \leq y \leq g(x)\},$$

is simply given by

$$\int_a^b (g(x) - f(x))\,\mathrm{d}x.$$

EXAMPLE 4.1.6 (Volume via slicing method). Let Ω be a bounded domain in $\mathbb{R}^3$ with smooth boundary[1]. For each $x \in \mathbb{R}$, we define the *slice* (or *cross-section*) by

$$\Omega_x := \{(y, z) \in \mathbb{R}^2 : (x, y, z) \in \Omega\},$$

which forms a bounded smooth domain in $\mathbb{R}^2$, so that its area can be computed. Since Ω is bounded, then there exists $a, b \in \mathbb{R}$ with $a < b$ such that

$$\Omega = \bigcup_{x \in [a,b]} \Omega_x.$$

We define $A : [a, b] \to \mathbb{R}$ by

$$A(x) := \mathrm{area}\,(\Omega_x) \quad \text{for all } x \in [a, b].$$

Then the volume of Ω is given by

$$\mathrm{volume}\,(\Omega) = \int_a^b A(x)\,\mathrm{d}x.$$

In practical, we need multivariable calculus (and even more advance calculus as well as Riemannian geometry) to compute the volume in $\mathbb{R}^3$, or Lebesgue measure for higher dimensional case as well as in Riemannian manifold. This is a special case of *coarea formula*, see e.g. the advance monograph [**Cha06**].

Even though the continuity of function is not necessary to ensure the integrability of functions, but it serves as a simple sufficient condition.

THEOREM 4.1.7. *Let $a, b \in \mathbb{R}$ with $a < b$ and let $f : [a, b] \to \mathbb{R}$ be a function. If*

(a) *there exists a number $M > 0$ such that $|f(x)| \leq M$ for all $x \in [a, b]$, and*
(b) *f is continuous except on at most finitely many points on $[a, b]$,*

then $f : [a, b] \to \mathbb{R}$ is Riemann integrable.

By using Theorem 2.2.7, we immediately obtain the following corollary.

COROLLARY 4.1.8. *Let $a, b \in \mathbb{R}$ with $a < b$. If $f : [a, b] \to \mathbb{R}$ is continuous, then it is also Riemann integrable.*

[1]Its rigorous definition requires quite advance calculus, here roughly understood that the domain is "regular". Think about the *space-filling curve* (or known as *Peano curve*).

EXAMPLE 4.1.9. It is important to have condition (a) in Theorem 4.1.7. For example, we now consider the function $f : [0,1] \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} x^{-1/2} & , 0 < x \leq 1, \\ 0 & , x = 0. \end{cases}$$

Given any partition $\Gamma = \{0 = x_0 < x_1 < \cdots < x_n = 1\}$, one sees that

$$U(f,\Gamma) \geq \left(\sup_{y \in [x_0,x_1]} f(y) \right) (x_1 - x_0) = +\infty$$

and

$$L(f,\Gamma) \leq \sum_{i=1}^{n} f(x_1)(x_i - x_{i-1}) = f(x_1) < +\infty,$$

which shows that

$$U(f,\Gamma) - L(f,\Gamma) = +\infty \text{ for any partition } \Gamma \text{ on } [0,1],$$

therefore $f : [0,1] \to \mathbb{R}$ is not Riemann integrable. Despite it is not Riemann integrable, but it may be integrate in the sense of improper integral as in Section 4.4 below.

EXAMPLE 4.1.10. The boundedness (i.e. condition (a) in Theorem 4.1.7) itself is not enough to ensure the Riemann integrability. For example, we now consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by

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

We now restrict the function on $[0,1]$. It is easy to see that

$$U(f,\Gamma) - L(f,\Gamma) = 1 \text{ for any partition } \Gamma \text{ on } [0,1],$$

therefore $f : [0,1] \to \mathbb{R}$ is not Riemann integrable. In fact, $\mathbb{Q}$ has Lebesgue measure zero, and $f : [0,1] \to \mathbb{R}$ is Lebesgue integrable with area 0. This example shows that Riemann integral is actually not good enough for practical application (including statistics), but however, we will not going to introduce this during this course.

From now on, the term "integrable" will refers the "Riemann integrable".

LEMMA 4.1.11 (Basic properties of Riemann integral). *Let $a,b \in \mathbb{R}$ with $a < b$.*

(a) ***Linearity.*** *If $f_1 : [a,b] \to \mathbb{R}$ and $f_2 : [a,b] \to \mathbb{R}$ are integrable, then for any constants $c_1, c_2 \in \mathbb{R}$ the function $c_1 f_1 + c_2 f_2 : [a,b] \to \mathbb{R}$ is also integrable and*

$$\int_a^b (c_1 f_1(x) + c_2 f_2(x))\,\mathrm{d}x = c_1 \int_a^b f_1(x)\,\mathrm{d}x + c_2 \int_a^b f_2(x)\,\mathrm{d}x.$$

(b) ***Monotonicity.*** *If $f_1 : [a,b] \to \mathbb{R}$ and $f_2 : [a,b] \to \mathbb{R}$ are integrable such that $f_1(x) \leq f_2(x)$ for all $x \in [a,b]$, then*

$$\int_a^b f_1(x)\,\mathrm{d}x \leq \int_a^b f_2(x)\,\mathrm{d}x.$$

(c) *If* $f : [a, x_0] \to \mathbb{R}$ *and* $f : [x_0, b] \to \mathbb{R}$ *are integrable for some* $x_0 \in [a, b]$*, then* $f : [a, b] \to \mathbb{R}$ *is also integrable and*

$$\int_a^b f(x)\,\mathrm{d}x = \int_a^{x_0} f(x)\,\mathrm{d}x + \int_{x_0}^b f(x)\,\mathrm{d}x.$$

(d) *If* $f : [a, b] \to \mathbb{R}$ *is integrable, then* $|f| : [a, b] \to \mathbb{R}$ *is integrable and*

$$\left| \int_a^b f(x)\,\mathrm{d}x \right| \leq \int_a^b |f(x)|\,\mathrm{d}x.$$

(e) *If* $f_1 : [a, b] \to \mathbb{R}$ *and* $f_2 : [a, b] \to \mathbb{R}$ *are integrable, then* $fg : [a, b] \to \mathbb{R}$ *is also integrable.*

Example 4.1.12. The integrablity of $|f| : [a, b] \to \mathbb{R}$ does not guarantee the integrability of $f : [a, b] \to \mathbb{R}$. For example, we consider the function

$$f : [0, 1] \to \mathbb{R}, \quad f(x) = \begin{cases} 1 & , x \in [0, 1] \cap \mathbb{Q}, \\ -1 & , x \in [0, 1] \setminus \mathbb{Q}, \end{cases}$$

which is not integrable (see Example 4.1.10), but one sees that $|f(x)| = 1$ for all $x \in [0, 1]$, which shows that $|f| : [0, 1] \to \mathbb{R}$ is integrable. This example also explains the "defectness" of Riemann integrable. As a comparison,

$f : [a, b] \to \mathbb{R}$ is Lebesgue integrable if and only if $|f| : [a, b] \to \mathbb{R}$ is Lebesgue integrable.

After explaining the mathematical aspect of integration, we now asking how to compute it. It is impractical to compute the integral by directly partition the intervals. Instead, we compute it via differentiation.

Theorem 4.1.13 (Fundamental theorem of calculus, part I [**Rud87**, Theorem 6.20]). *Let* $a, b \in \mathbb{R}$ *with* $a < b$ *and let* $f : [a, b] \to \mathbb{R}$ *be a* continuous *function. We define a continuous function*

$$F : [a, b] \to \mathbb{R}, \quad F(x) := \int_a^x f(t)\,\mathrm{d}t.$$

Then $F : (a, b) \to \mathbb{R}$ *is differentiable and satisfies* $F'(x) = f(x)$ *for all* $x \in (a, b)$*.*

Example 4.1.14. Theorem 4.1.13 only holds true for continuous functions. For example, we consider the Heaviside function

$$f : \mathbb{R} \to \mathbb{R}, \quad f(x) = \begin{cases} 1 & , x \geq 0, \\ 0 & , x < 0, \end{cases}$$

which is integrable on $[-1, 1]$. We see that the function

$$F : [-1, 1] \to \mathbb{R}, \quad F(x) := \int_{-1}^x f(t)\,\mathrm{d}t = \begin{cases} x & , x \geq 0 \\ 0 & , x < 0 \end{cases}$$

is continuous, but it is not differentiable at $x = 0$.

Theorem 4.1.15 (Fundamental theorem of calculus, part II [**Rud87**, Theorem 6.21]). *Let* $a, b \in \mathbb{R}$ *with* $a < b$ *and let* $f : [a, b] \to \mathbb{R}$ *be an (Riemann) integrable function. If there* *exists a* $\delta > 0$ *and a function* $F : (a-\delta, b+\delta) \to \mathbb{R}$ *such that it is differentiable on* $[a,b]$ *with* $F' = f$*, then*

$$\int_a^b f(x)\,\mathrm{d}x = F(b) - F(a). \tag{4.1.4}$$

Remark. We write

$$F(x)\Big|_{x=a}^{x=b} := F(b) - F(a). \tag{4.1.5}$$

From (4.1.4) we have

$$\int_a^b F'(x)\,\mathrm{d}x = F(x)\Big|_{x=a}^{x=b}. \tag{4.1.6}$$

This means that the total rate of change in $[a,b]$ is equal to the flux across the boundary of $[a,b]$, which is exactly the two-points set $\{a,b\}$. For example, if we want to count the number of people in a building, we only need to count the people across all doors. We see that, at the boundary point $x = b$ of $[a,b]$, the outward direction is right (i.e. positive), while at the boundary point $x = a$ of $[a,b]$, the outward direction is left (i.e. negative), therefore there is a negative sign in (4.1.5). In fact, (4.1.6) is exactly the *divergence theorem* for 1-dimensional case. The divergence theorem is a fundamental result in partial differential equations, one also can refer to my other lecture note [**Kow24**] for more details.

Let I be a *connected* open interval (may or may not bounded, i.e. either (a,b) or $(-\infty,b)$ or $(a,+\infty)$ or $\mathbb{R}$). We now consider the *functor*

$$\mathscr{D} : C^1(I) \to C^0(I), \quad \mathscr{D}F := F' \text{ for all } F \in C^1(I). \tag{4.1.7}$$

Here we recall that $C^0(I)$ is the collection of continuous functions $I \to \mathbb{R}$, while $C^1(I)$ is the collection of differentiable functions with continuous derivative $I \to \mathbb{R}$. Similar to sets, we can interpret functions as a "level-0" objects, while functors as "level-1" objects. We now fix any $x_0 \in I$. For each $f \in C^0(I)$, we define (with the convention 4.1.3)

$$F(x) := \int_{x_0}^x f(t)\,\mathrm{d}t \quad \text{for all } x \in I.$$

By using Theorem 4.1.13, we see that $F \in C^1(I)$ with $F' = f$, which shows that the functor (4.1.7) is surjective. However, since

$$F + C \in C^1(I) \text{ and } (F+C)' = F' \quad \text{for all } C \in \mathbb{R}, \tag{4.1.8}$$

one sees that the functor (4.1.7) is not injective. By using Corollary 3.4.9(c), it is easy to see the following lemma.

Lemma 4.1.16. *Let* I *be a connected open set. Let* $F_1 : I \to \mathbb{R}$ *and* $F_2 : I \to \mathbb{R}$ *are differentiable functions. If* $F_1'(x) = F_2'(x)$ *for all* $x \in I$*, then there exists a constant* $C \in \mathbb{R}$ *such that* $F_1(x) = F_2(x) + C$ *for all* $x \in I$*.*

In view of (4.1.8), for each $F \in C^1(I)$, we now consider the equivalence class

$$\{F + C\}_{C\in\mathbb{R}} := \{F + C : C \in \mathbb{R}\}.$$

We now define a mapping, which is similar to (4.1.7), by

$$\begin{aligned}\tilde{\mathscr{D}} &: \big\{\{F+C\}_{C\in\mathbb{R}} : F\in C^1(I)\big\} \to C^0(I), \\ \tilde{\mathscr{D}}&\left(\{F+C\}_{C\in\mathbb{R}}\right) := F' \text{ for any } F\in\{F+C\}_{C\in\mathbb{R}}.\end{aligned} \tag{4.1.9}$$

By using (4.1.8), one sees that (4.1.9) is a well-defined function (as mentioned in Section 1.3). Since (4.1.7) is surjective, then so is (4.1.9). By using Lemma 4.1.16, one sees that (4.1.9) is injective. Now we conclude that the mapping (4.1.9) is bijective, therefore the inverse mapping

$$\tilde{\mathscr{D}}^{-1} : C^0(I) \to \big\{\{F+C\}_{C\in\mathbb{R}} : F\in C^1(I)\big\}$$

is well-defined. This suggests the following definition.

DEFINITION 4.1.17. Let I be a connected open interval. The *antiderivative* of a continuous function $f : I\to\mathbb{R}$ is the equivalence class $\{F+C\}_{C\in\mathbb{R}}$ such that $F'(x) = f(x)$ for all $x\in I$.

REMARK. We also slightly abuse the notation by referring an element in $\{F+C\}_{C\in\mathbb{R}}$ the antiderivative. If we abuse the notation in this way, the antiderivative does not unique, and one should use "an" rather than "the". For example, we know that the antiderivative of $\cos:\mathbb{R}\to\mathbb{R}$ is the $\{\sin + C : \mathbb{R}\to\mathbb{R}\}_{C\in\mathbb{R}}$, where a

$$(\sin + C)(x) = \sin x + C \quad \text{for all } x\in\mathbb{R}.$$

If we abuse the notation in this way, we also say that $\sin x$ is "an" antiderivative of $\cos x$. In view of fundamental theorem of calculus, this kind of abuse of notation is *acceptable*, and we will abuse the notation for "antiderivative" in this way.

REMARK. Since

$$(F(x)+C)\Big|_{x=a}^{x=b} = (F(b)+C) - (F(a)+C) = F(b) - F(a) = F(x)\Big|_{x=a}^{x=b},$$

we also can write (4.1.4) as

$$\int_a^b f(x)\,\mathrm{d}x = (F(x)+C)\Big|_{x=a}^{x=b}.$$

This suggests many authors to *abuse the notation* by writing

$$\int f(x)\,\mathrm{d}x = F(x) + C, \tag{4.1.10}$$

but this may cause some ambiguity (see Section 4.2), therefore *personally I strongly suggests not to abuse the notation like* (4.1.10).

EXERCISE 4.1.18. Let $a,b\in\mathbb{R}$ with $a<b$. Let f be a continuous function on $[a,b]$, and let $\alpha,\beta\in C^1(\mathbb{R})$ be such that

$$a < \alpha(x) < b, \quad a < \beta(x) < b \quad \text{for all } x\in\mathbb{R}.$$

We define $g(x) := \int_{\alpha(x)}^{\beta(x)} f(t)\,\mathrm{d}t$ for all $x\in\mathbb{R}$. Show that

$$g'(x) = f(\beta(x))\beta'(x) - f(\alpha(x))\alpha'(x) \quad \text{for all } x\in\mathbb{R}.$$

4.2. Integration by parts and substitution rule

The main difficulty to compute $\int_a^b f(x)\,\mathrm{d}x$ is to find an antiderivative $F(x)$ of $f(x)$. For example, even though we know an antiderivatives $F(x)$ for $f(x)$, and an antiderivatives $G(x)$ for $g(x)$, but it is not easy to guess the antiderivative of $f(x)g(x)$. By using the product rule, one sees that

$$(F(x)G(x))' = F'(x)G(x) + F(x)G'(x) = f(x)G(x) + F(x)g(x) \quad \text{for all } x \in (a,b).$$

Under some suitable assumptions on F and G, one may use the fundamental theorem of calculus (Theorem 4.1.15) to see that

$$F(x)G(x)\Big|_{x=a}^{x=b} = \int_a^b (F(x)G(x))'\,\mathrm{d}x = \int_a^b f(x)G(x)\,\mathrm{d}x + \int_a^b F(x)g(x)\,\mathrm{d}x,$$

that is,

$$\int_a^b F(x)g(x)\,\mathrm{d}x = F(x)G(x)\Big|_{x=a}^{x=b} - \int_a^b f(x)G(x)\,\mathrm{d}x.$$

We now summarize the above observations in the following theorem.

THEOREM 4.2.1 (Integration by parts [**Rud87**, Theorem 6.22]). *Let $a,b \in \mathbb{R}$ with $a < b$ and let I be an open interval such that $[a,b] \subset I$. We consider functions $F : I \to \mathbb{R}$ and $G : I \to \mathbb{R}$. If both F and G are differentiable on $[a,b]$, such that $F' : [a,b] \to \mathbb{R}$ and $G' : [a,b] \to \mathbb{R}$ are Riemann integrable, then*

$$\int_a^b F(x)G'(x)\,\mathrm{d}x = F(x)G(x)\Big|_{x=a}^{x=b} - \int_a^b F'(x)G(x)\,\mathrm{d}x.$$

We see that integration by part is nothing but just an integral version of product rule.

EXAMPLE 4.2.2. We now compute $\int_0^1 xe^x\,\mathrm{d}x$. In view of the fundamental theorem of calculus (Theorem 4.1.15), the most difficult part here is to find a function F such that $F'(x) = xe^x$. Integration by parts (Theorem 4.2.1) suggests us that one can simplify the problem by guessing the antiderivative of the function x of the antiderivative of the function e^x. Since $(e^x)' = e^x$, then one sees that

$$\begin{aligned}
\int_0^1 xe^x\,\mathrm{d}x &= \int_0^1 x(e^x)'\,\mathrm{d}x \\
&= xe^x\Big|_{x=0}^{x=1} - \int_0^1 (x)'e^x\,\mathrm{d}x \quad \text{(integration by parts)} \\
&= e - \int_0^1 e^x\,\mathrm{d}x = e - e^x\Big|_{x=0}^{x=1} = 1,
\end{aligned}$$

which means that we compute the integral $\int_0^1 xe^x \,\mathrm{d}x$ without directly guessing an antiderivative of xe^x. Suggested by $(\frac{1}{2}x^2)' = x$, one may also consider the following attempt:

$$\begin{aligned}\int_0^1 xe^x \,\mathrm{d}x &= \frac{1}{2}\int_0^1 (x^2)'e^x \,\mathrm{d}x \\ &= \frac{1}{2}x^2e^x\bigg|_{x=0}^{x=1} - \frac{1}{2}\int_0^1 x^2(e^x)' \,\mathrm{d}x \quad \text{(integration by parts)} \\ &= \frac{1}{2}e - \frac{1}{2}\int_0^1 x^2e^x \,\mathrm{d}x,\end{aligned}$$

which somehow make the problem even worse. This failed attempt demonstrates that there is no standard way to use integration by parts, and this is *highly depends on personal experience*.

EXAMPLE 4.2.3. One can compute $\int_0^{\pi/2}(\sin x)^3 \,\mathrm{d}x$ using trigonometric identities. We now give another alternative way by using integration by parts (Theorem 4.2.1). We write

$$\begin{aligned}&\int_0^{\pi/2}(\sin x)^3 \,\mathrm{d}x = \int_0^{\pi/2}(\sin x)^2 \sin x \,\mathrm{d}x = -\int_0^{\pi/2}(\sin x)^2(\cos x)' \,\mathrm{d}x \\ &= -\overbrace{(\sin x)^2 \cos x\bigg|_{x=0}^{x=\pi/2}}^{=0} + \int_0^{\pi./2}((\sin x)^2)' \cos x \,\mathrm{d}x \\ &= \int_0^{\pi./2}(2\sin x \cos x)\cos x \,\mathrm{d}x \\ &= 2\int_0^{\pi/2}\sin x(\cos x)^2 \,\mathrm{d}x = 2\int_0^{\pi/2}\sin x(1-(\sin x)^2) \,\mathrm{d}x \\ &= 2\int_0^{\pi/2}\sin x \,\mathrm{d}x - 2\int_0^{\pi/2}(\sin x)^3 \,\mathrm{d}x,\end{aligned}$$

hence

$$3\int_0^{\pi/2}(\sin x)^3 \,\mathrm{d}x = 2\int_0^{\pi/2}\sin x \,\mathrm{d}x, \tag{4.2.1}$$

which implies

$$\int_0^{\pi/2}(\sin x)^3 \,\mathrm{d}x = \frac{2}{3}\int_0^{\pi/2}\sin x \,\mathrm{d}x = -\frac{2}{3}\cos x\bigg|_{x=0}^{x=\pi/2} = \frac{2}{3}.$$

EXAMPLE 4.2.4. The idea in (4.2.1) can be extend to a more general settings. For any real numbers $\alpha \geq 2$, one sees that

$$\begin{aligned}
\int_0^{\pi/2} (\sin x)^\alpha \,\mathrm{d}x &= \int_0^{\pi/2} (\sin x)^{\alpha-1} \sin x \,\mathrm{d}x = -\int_0^{\pi/2} (\sin x)^{\alpha-1} (\cos x)' \,\mathrm{d}x \\
&= \overbrace{-(\sin x)^{\alpha-1} \cos x \Big|_{x=0}^{x=\pi/2}}^{=0} + \int_0^{\pi./2} ((\sin x)^{\alpha-1})' \cos x \,\mathrm{d}x \\
&= \int_0^{\pi./2} ((\alpha-1)(\sin x)^{\alpha-2} \cos x) \cos x \,\mathrm{d}x = (\alpha-1) \int_0^{\pi./2} (\sin x)^{\alpha-2} (\cos x)^2 \,\mathrm{d}x \\
&= (\alpha-1) \int_0^{\pi./2} (\sin x)^{\alpha-2} (1-(\sin x)^2) \,\mathrm{d}x \\
&= (\alpha-1) \int_0^{\pi./2} (\sin x)^{\alpha-2} \,\mathrm{d}x - (\alpha-1) \int_0^{\pi./2} (\sin x)^\alpha \,\mathrm{d}x,
\end{aligned}$$

and hence

$$\alpha \int_0^{\pi./2} (\sin x)^\alpha \,\mathrm{d}x = (\alpha-1) \int_0^{\pi./2} (\sin x)^{\alpha-2} \,\mathrm{d}x.$$

One can, at least, compute the precise formula of

$$\int_0^{\pi./2} (\sin x)^n \,\mathrm{d}x$$

for all integer $n \geq 1$ by using the idea.

We now recall the chain rule:

$$(f(\varphi(x)))' = f'(y)|_{y=\varphi(x)} \varphi'(x) = f'(\varphi(x))\varphi'(x).$$

for some suitable differentiable function f and φ. Under some assumptions, by using the fundamental theorem of calculus (Theorem 4.1.15), one sees that

$$\begin{aligned}
\int_A^B f'(\varphi(x))\varphi'(x) \,\mathrm{d}x &= \int_A^B (f(\varphi(x)))' \,\mathrm{d}x = f(\varphi(B)) - f(\varphi(A)) \\
&= f(x)\Big|_{x=\varphi(A)}^{x=\varphi(B)} = \int_{\varphi(A)}^{\varphi(B)} f'(x) \,\mathrm{d}x.
\end{aligned}$$

If we write $F = f'$, then

$$\int_A^B F(\varphi(x))\varphi'(x) \,\mathrm{d}x = \int_{\varphi(A)}^{\varphi(B)} F(x) \,\mathrm{d}x.$$

We now write $a = \varphi(A)$ and $b = \varphi(B)$. If $\varphi : [A,B] \to [a,b]$ is *bijective* (note: since φ is differentiable, this implies that either φ is strictly increasing or strictly decreasing) then

$$\begin{aligned}
\int_{\varphi^{-1}(a)}^{\varphi^{-1}(b)} F(\varphi(t))\varphi'(t) \,\mathrm{d}t &= \int_{\varphi^{-1}(a)}^{\varphi^{-1}(b)} F(\varphi(x))\varphi'(x) \,\mathrm{d}x \\
&= \int_A^B F(\varphi(x))\varphi'(x) \,\mathrm{d}x = \int_{\varphi(A)}^{\varphi(B)} F(x) \,\mathrm{d}x = \int_a^b F(x) \,\mathrm{d}x.
\end{aligned}$$

In fact, the following theorem holds true:

THEOREM 4.2.5 (Integration by parts [**Rud87**, Theorems 6.17 and 6.19]). *Let $\varphi : [a,b] \to \mathbb{R}$ be a strictly increasing (or strictly decreasing) function, which is differentiable on (a,b). Suppose that φ' can be extend to function $[a,b] \to \mathbb{R}$ such that it is Riemann integrable. If a function $F : [a,b] \to \mathbb{R}$ is Riemann integrable, then $t \mapsto F(\varphi(t))\varphi'(t)$ is integrable on $[\varphi^{-1}(a), \varphi^{-1}(b)]$ and*

$$\int_a^b F(x)\,\mathrm{d}x = \int_{\varphi^{-1}(a)}^{\varphi^{-1}(b)} F(\varphi(t))\varphi'(t)\,\mathrm{d}t.$$

REMARK. It is convenient to write $\varphi(t) = x(t)$, so that

$$\int_a^b F(x)\,\mathrm{d}x = \int_{\varphi^{-1}(a)}^{\varphi^{-1}(b)} F(x(t))x'(t)\,\mathrm{d}t = \int_{\varphi^{-1}(a)}^{\varphi^{-1}(b)} F(x(t))\frac{\mathrm{d}x}{\mathrm{d}t}\,\mathrm{d}t, \tag{4.2.2}$$

which is quite convenient to memorize.

EXAMPLE 4.2.6 (Revisit of Example 4.2.3). One can compute $\int_0^{\pi/2}(\sin x)^3\,\mathrm{d}x$ using trigonometric identities or using the integration by parts formula in Example 4.2.3. We again write

$$\int_0^{\pi/2}(\sin x)^3\,\mathrm{d}x = \int_0^{\pi/2}(\sin x)^2 \sin x\,\mathrm{d}x = \int_0^{\pi/2}(1-(\cos x)^2)\sin x\,\mathrm{d}x.$$

We now consider the change of variable $\cos x(t) = t$. One sees that

$$\begin{aligned} x = 0 &\leftrightarrow t = 1, \\ x = \frac{\pi}{2} &\leftrightarrow t = 0, \end{aligned}$$

and acting $\frac{\mathrm{d}}{\mathrm{d}t}$ on the equation $\cos x(t) = t$ to see that

$$-(\sin x(t))\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}(\cos x(t)) = \frac{\mathrm{d}}{\mathrm{d}t}t = 1.$$

Now from (4.2.2) we see that

$$\begin{aligned} \int_0^{\pi/2}(\sin x)^3\,\mathrm{d}x &= \int_{x=0}^{x=\pi/2}(1-(\cos x)^2)\sin x\,\mathrm{d}x \\ &= \int_{t=1}^{t=0}(1-t^2)\overbrace{(\sin x(t))\frac{\mathrm{d}x}{\mathrm{d}t}}^{=-1}\,\mathrm{d}t = -\int_{t=1}^{t=0}(1-t^2)\,\mathrm{d}t \\ &= -\left(t - \frac{1}{3}t^3\right)\Bigg|_{t=1}^{t=0} = \frac{2}{3}. \end{aligned}$$

EXAMPLE 4.2.7. Let's do a generalization similar to Example 4.2.4. For any real numbers $\alpha > 1$, one sees that

$$\int_0^{\pi/2}(\sin x)^\alpha\,\mathrm{d}x = \int_0^{\pi/2}(\sin x)^{\alpha-1}\sin x\,\mathrm{d}x = \int_0^{\pi/2}(1-(\cos x)^2)^{\frac{\alpha-1}{2}}\sin x\,\mathrm{d}x,$$

since $\sin x > 0$ for all $x \in (0, \pi.2)$ and that $\sin x = (1-(\cos x)^2)^{1/2}$. By considering the same change of variables in Example 4.2.6, one sees that

$$\begin{aligned}\int_0^{\pi/2} (\sin x)^\alpha \,\mathrm{d}x &= \int_{x=0}^{x=\pi/2} (1-(\cos x)^2)^{\frac{\alpha-1}{2}} \sin x \,\mathrm{d}x \\ &= \int_{t=1}^{t=0} (1-t^2)^{\frac{\alpha-1}{2}} \overbrace{(\sin x(t))\frac{\mathrm{d}x}{\mathrm{d}t}}^{=-1} \,\mathrm{d}t = -\int_{t=1}^{t=0} (1-t^2)^{\frac{\alpha-1}{2}} \,\mathrm{d}t \\ &= \int_0^1 (1-t^2)^{\frac{\alpha-1}{2}} \,\mathrm{d}t.\end{aligned}$$

This method allow use to compute the formula of $\int_0^{\pi/2} (\sin x)^n \,\mathrm{d}x$ at least for odd integer $n \in \mathbb{N}$.

REMARK 4.2.8. It is important to check whether the mapping $\varphi : [a,b] \to \mathbb{R}$ is strict increasing/decreasing. We illustrate this precaution by the following simple integral

$$\int_{-\pi/2}^{\pi/2} (\sin x)^2 \,\mathrm{d}x = \int_{-\pi/2}^{\pi/2} (\sin x)(\sin x) \,\mathrm{d}x,$$

which is obviously > 0. If we consider the "change of variable" $\cos x(t) = t$, and we see that

$$\begin{aligned} x = -\pi/2 &\leftrightarrow t = 0, \\ x = \frac{\pi}{2} &\leftrightarrow t = 0, \end{aligned}$$

and acting $\frac{\mathrm{d}}{\mathrm{d}t}$ on the equation $\cos x(t) = t$ to see that

$$-(\sin x(t))\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}(\cos x(t)) = \frac{\mathrm{d}}{\mathrm{d}t} t = 1.$$

Hence

$$\int_{-\pi/2}^{\pi/2} (\sin x)^2 \,\mathrm{d}x \overset{(\times)}{=} -\int_{t=0}^{t=0} (\sin x(t)) \,\mathrm{d}t = 0,$$

which is obviously not true. The main reason in the above "argument" is the mapping $\cos : [-\pi/2, \pi/2] \to [0,1]$ is not bijective. We now show the correct way to do this: by writing

$$\int_{-\pi/2}^{\pi/2} (\sin x)^2 \,\mathrm{d}x = \int_{-\pi/2}^{0} (\sin x)^2 \,\mathrm{d}x + \int_0^{\pi/2} (\sin x)^2 \,\mathrm{d}x,$$

we consider the change of variable $x(t) = -y$, we see that

$$\int_{x=-\pi/2}^{x=0} (\sin x)^2 \,\mathrm{d}x = -\int_{y=\pi/2}^{y=0} (\sin(-y))^2 \,\mathrm{d}y = \int_0^{\pi/2} (\sin y)^2 \,\mathrm{d}y = \int_0^{\pi/2} (\sin x)^2 \,\mathrm{d}x,$$

and hence

$$\int_{-\pi/2}^{\pi/2} (\sin x)^2 \,\mathrm{d}x = 2\int_0^{\pi/2} (\sin x)^2 \,\mathrm{d}x.$$

Then the rest can be argued as in Example 4.2.4.

4.3. Integration by Partial Fractions

The main theme of this section is to introduce the rational form of the form

$$\frac{f(x)}{g(x)} \quad \text{for all } x \in \mathbb{R} \setminus \mathscr{Z}_g, \tag{4.3.1}$$

where $\mathscr{Z}_g := \{x \in \mathbb{R} : g(x) = 0\}$. First of all, we need to ensure that $\mathscr{Z}_g$ is not "too big" in order to make sure that the rational form (4.3.1) is meaningful.

DEFINITION 4.3.1. Let I be an open interval in $\mathbb{R}$ (may or may not bounded) and we consider a function $g : I \to \mathbb{R}$. If $x_0 \in I$ satisfies $g(x_0) = 0$, then we say that such x_0 is a *zero of g in I*.

Let $g(x) = \sum_{k=0}^{n} c_k x^k$ be a polynomial. If the *leading coefficient* $c_n \neq 0$, then we say that the degree of g is n, and we denote by $\deg(g) = n$. By using the fundamental theorem of algebra (see e.g. [**Kow23**, Theorem 4.3.6] and some further computations, one has the following corollary.

COROLLARY 4.3.2. *Let $g : \mathbb{R} \to \mathbb{R}$ be a nontrivial polynomial of degree n with leading coefficient $c_n \neq 0$, then one can factorize*

$$g(x) = c_n(x - x_1) \cdots (x - x_\ell)(x^2 + y_1 x + z_1) \cdots (x^2 + y_p x + z_p) \tag{4.3.2}$$

for some $\ell \in \mathbb{N}$, for some $p \in \mathbb{Z}_{\geq 0}$, for some $x_1, \cdots, x_\ell \in \mathbb{R}$ and for some $y_1, z_1, \cdots, y_p, z_p \in \mathbb{R}$ with $2p + \ell = n$. Here, the coefficients x_j, y_j and z_j are not necessarily distinct.

REMARK 4.3.3 (Some advance remarks). For those who familiar with complex analysis [**Kow23**], the fundamental theorem of algebra says that the analytic polynomial $g(z) = \sum_{k=0}^{n} c_k z^k$ with $c_n \neq 0$ always can be written as $g(z) = c_n(z - z_1) \cdots (z - z_n)$ for some $z_1, \cdots, z_n \in \mathbb{C}$. We now assume that the coefficients $c_0, \cdots, c_n \in \mathbb{R}$. In this case, if $g(w_0) = 0$, then one sees that

$$0 = \overline{g(w_0)} = \sum_{k=0}^{n} \overline{c_k w_0^k} = \sum_{k=0}^{n} c_k \overline{w_0}^k = g(\overline{w_0}).$$

This shows that, for the case when $c_0, \cdots, c_n \in \mathbb{R}$, the analytic polynomial $g(z) = \sum_{k=0}^{n} c_k z^k$ with $c_n \neq 0$ always can be written as

$$g(z) = c_n(z - w_1)(z - \overline{w_1}) \cdots (z - w_p)(z - \overline{w_p})(z - x_1) \cdots (z - x_{n-2p})$$

for some $w_1, \cdots, w_p \in \mathbb{C} \setminus \mathbb{R}$ and $x_1, \cdots, x_{n-2p} \in \mathbb{R}$. If $p = \frac{n}{2}$, this means that $f(z) = c_n(z - w_1)(z - \overline{w_1}) \cdots (z - w_p)(z - \overline{w_p})$. We also see that

$$(z - w_j)(z - \overline{w_j}) = z^2 - (w_j + \overline{w_j})z + w_j \overline{w_j} = z^2 - 2\Re w_j z + |w_j|.$$

Corollary 4.3.2 is simply a restriction on $\mathbb{R}$.

EXAMPLE 4.3.4. The polynomial $g(x) = x^2 + x = x(x+1)$ has zeros 0 and -1. The polynomial $g(x) = x^2 - 2x + 1 = (x-1)^2$ has one repeated zero $1 \in \mathbb{R}$. The polynomial $g(x) = x^2 + 1$ has no zero in $\mathbb{R}$.

Since $\mathscr{Z}_g$ is a finite set, then we now see the rational form (4.3.1) is meaningful. First of all, we consider the case when $\deg(f) \geq \deg(g) \geq 2$, and we want to write (4.3.1) as

$$\frac{f(x)}{g(x)} = P(x) + \frac{\tilde{f}(x)}{g(x)} \quad \text{for all } x \in \mathbb{R} \setminus \mathscr{Z}_g, \tag{4.3.3}$$

for some polynomials P and $\tilde{f}$ such that $0 \leq \deg(\tilde{f}) < \deg(g)$. This can be done by the long division described by the following algorithm:

Algorithm 1 Long division for polynomials

if $\deg(f) < \deg(g)$ **then**
    Set $\tilde{f}(x) = f(x)$ and $P(x) = 0$ for all $x \in \mathbb{R}$.
else
    Set $j = 0$, $x_0 \neq 0$ be the leading coefficient of g, and set $f_0(x) := f(x)$ for all $x \in \mathbb{R}$.
    while $\deg(f_j) \geq \deg(g)$ **do**
        Set $A_j \neq 0$ be the leading coefficient of f_j.
        Define $f_{j+1}(x) := f_j(x) - \frac{A_j}{c_0} x^{\deg(f_j)-\deg(g)} g(x)$ for all $x \in \mathbb{R}$.
        Set $j = j + 1$.
    end while
    Set $\tilde{f}(x) = f_j(x)$ and $P(x) = \sum_{k=0}^{j-1} \frac{A_k}{c_0} x^{\deg(f_k)-\deg(g)}$ for all $x \in \mathbb{R}$.
end if

LEMMA 4.3.5. *Algorithm 1 must terminate within finite steps and it solves (4.3.3).*

PROOF. The case when $\deg(f) < \deg(g)$ is trivial. We now consider the case when $\deg(f) \geq \deg(g)$. Since $\deg(f_{j+1}) \leq \deg(f_j) - 1$, then we see that Algorithm 1 must terminate within finite steps, says at $j = j_0$, then we compute that

$$\begin{aligned}
\tilde{f}(x) = f_{j_0+1}(x) &= f_{j_0}(x) - \frac{A_{j_0}}{c_0} x^{\deg(f_{j_0})-\deg(g)} g(x) \\
&= f_{j_0-1}(x) - \sum_{j=j_0-1}^{j_0} \frac{A_j}{c_0} x^{\deg(f_j)-\deg(g)} g(x) \\
&\vdots \\
&= f_0(x) - \sum_{j=0}^{j_0} \frac{A_j}{c_0} x^{\deg(f_j)-\deg(g)} g(x) \\
&= f(x) - \sum_{j=0}^{j_0} \frac{A_j}{c_0} x^{\deg(f_j)-\deg(g)} g(x) \quad \text{for all } x \in \mathbb{R},
\end{aligned}$$

and then divide the above equation by $g(x)$ we reach

$$\frac{\tilde{f}(x)}{g(x)} = \frac{f(x)}{g(x)} - \overbrace{\sum_{j=0}^{j_0} \frac{A_j}{c_0} x^{\deg(f_j)-\deg(g)}}^{=P(x)} \quad \text{for all } x \in \mathbb{R} \setminus \mathscr{Z}_g,$$

which conclude our lemma. $\square$

We now give an example.

EXAMPLE 4.3.6. We now divide $f(x) = 4x^4 + 1$ by using $g(x) = 2x^2 + 1$ (with $c_0 = 2$). Set $f_0(x) = f(x) = 4x^4 + 1$ (with $A_0 = 4$), then we reach

$$\begin{aligned} f_1(x) &= f_0(x) - \frac{A_0}{c_0} x^{\deg(f_0)-\deg(g)} g(x) \\ &= 4x^4 + 1 - 2x^2(2x^2+1) = -2x^2 + 1 \quad (\text{with } A_1 = -2). \end{aligned}$$

Now we check that $\deg(f_1) = 2 \geq 2 = \deg(g)$, then the algorithm still continue:

$$\begin{aligned} f_2(x) &= f_1(x) - \frac{A_1}{c_0} x^{\deg(f_1)-\deg(g)} g(x) \\ &= -2x^2 + 1 + (2x^2+1) = 2. \end{aligned}$$

Now we see that $\deg(f_2) = 0 < 2 = \deg(g)$, thus the while loop is terminate at $j = 2$. According to Algorithm 1, we now output

$$\tilde{f}(x) := f_2(x) = 2$$

and

$$\begin{aligned} P(x) &= \sum_{k=0}^{1} \frac{A_k}{c_0} x^{\deg(f_k)-\deg(g)} = \frac{A_0}{c_0} x^{\deg(f_0)-\deg(g)} + \frac{A_1}{c_0} x^{\deg(f_1)-\deg(g)} \\ &= \frac{4}{2}x^2 + \frac{-2}{2} = 2x^2 - 1. \end{aligned}$$

As a demonstration, we verify that

$$P(x) + \frac{\tilde{f}(x)}{g(x)} = 2x^2 - 1 + \frac{2}{2x^2+1} = \frac{(2x^2-1)(2x^2+1)+2}{2x^2+1} = \frac{4x^4-1+2}{2x^2+1} = \frac{f(x)}{g(x)}.$$

We now focus on the rational form

$$\frac{\tilde{f}(x)}{g(x)} \quad \text{for all } x \in \mathbb{R} \setminus \mathscr{Z}_g, \tag{4.3.4}$$

with $\deg(\tilde{f}) < \deg(g)$. In fact, the above rational form can be further decomposed as follows:

THEOREM 4.3.7 (A special case of [**Kow23**, Theorem 5.2.7]). *We now consider the rational form* (4.3.4) *with* $\deg(\tilde{f}) < \deg(g)$. *If the polynomial g can be decomposed by Corollary 4.3.2 as*

$$g(x) = c_n(x-x_1)^{\mathfrak{m}(x_1)} \cdots (x-x_\ell)^{\mathfrak{m}(x_\ell)} (x^2+y_1x+z_1)^{\mathfrak{m}(y_1,z_1)} \cdots (x^2+y_px+z_p)^{\mathfrak{m}(y_p,z_p)}$$

for some $\ell \in \mathbb{N}$, for some $p \in \mathbb{Z}_{\geq 0}$, for some $x_1, \cdots, x_\ell \in \mathbb{R}$ and for some $y_1, z_1, \cdots, y_p, z_p \in \mathbb{R}$ so that $\{x_1, \cdots, x_\ell\}$ are all distinct and all the pairs $\{(y_1,z_1), \cdots, (y_p,z_p)\}$ are all distinct. Then the rational form (4.3.4) *can be uniquely decomposed as*

$$\frac{\tilde{f}(x)}{g(x)} = \sum_{j=1}^{\ell} \sum_{r=1}^{\mathfrak{m}(x_j)} \frac{A_{jr}}{(x-x_j)^r} + \sum_{j=1}^{p} \sum_{r=1}^{\mathfrak{m}(y_j,z_j)} \frac{B_{jr}x + C_{jr}}{(x^2+y_1x+z_1)^r} \quad \textit{for all } x \in \mathbb{R} \setminus \mathscr{Z}_g,$$

for some $A_{jr}, B_{jr}, C_{jr} \in \mathbb{R}$.

REMARK. Now one can easily compute the antiderivative of $\frac{A_{jr}}{(x-x_j)^r}$. It is difficult to compute the antiderivative of $\frac{B_{jr}x+C_{jr}}{(x^2+y_1x+z_1)^r}$ by using real numbers itself: In fact, this can be easily handle by using complex analysis.

We now close this section by the following example.

EXAMPLE 4.3.8. By using Theorem 4.3.7, one has a unique decomposition

$$\frac{1}{(x+1)^2(x^2+1)} = \frac{A}{x+1} + \frac{B}{(x+1)^2} + \frac{Cx+D}{x^2+1} \quad \text{for all } x \in \mathbb{R} \setminus \{-1\}. \tag{4.3.5}$$

It is recommend to begin with terms for those easy to handle. We first multiply (4.3.5) by $(x+1)^2$ to see that

$$\frac{1}{x^2+1} = A(x+1) + B + (x+1)^2 \frac{Cx+D}{x^2+1} \quad \text{for all } x \in \mathbb{R} \setminus \{-1\},$$

and consequently by taking the limit $x \to -1$ (**note.** we cannot directly take $x = -1$ since the above equation is not well-defined at it) we see that

$$B = \frac{1}{2}.$$

Now from (4.3.5) we have

$$\begin{aligned} \frac{A}{x+1} + \frac{Cx+D}{x^2+1} &= \frac{1}{(x+1)^2(x^2+1)} - \frac{1}{2(x+1)^2} \\ &= \frac{2-(x^2+1)}{2(x+1)^2(x^2+1)} = \frac{-x^2+1}{2(x+1)^2(x^2+1)} \\ &= \frac{-(x+1)(x-1)}{2(x+1)^2(x^2+1)} = \frac{1-x}{2(x+1)(x^2+1)} \quad \text{for all } x \in \mathbb{R} \setminus \{-1\}. \end{aligned} \tag{4.3.6}$$

Now we multiply the above equation by $(x+1)$ to see that

$$A + (x+1)\frac{Cx+D}{x^2+1} = \frac{1-x}{2(x^2+1)} \quad \text{for all } x \in \mathbb{R} \setminus \{-1\},$$

and consequently by taking the limit $x \to -1$ we see that

$$A = \frac{1}{2}.$$

Now from (4.3.6) we see that

$$\begin{aligned} \frac{Cx+D}{x^2+1} &= \frac{1-x}{2(x+1)(x^2+1)} - \frac{1}{2(x+1)} \\ &= \frac{1-x-(x^2+1)}{2(x+1)(x^2+1)} = \frac{-x-x^2}{2(x+1)(x^2+1)} \\ &= \frac{-x(x+1)}{2(x+1)(x^2+1)} = \frac{-x}{2(x^2+1)} \quad \text{for all } x \in \mathbb{R} \setminus \{-1\}. \end{aligned}$$

Multiplying the above equation by (x^2+1), we now see that

$$Cx + D = -\frac{1}{2}x.$$

By taking $x = 0$ and $x = 1$ (or a less rigorous statement "comparing the coefficients") we conclude that

$$C = -\frac{1}{2}, \quad D = 0.$$

Now we put everything into (4.3.5) to conclude that

$$\frac{1}{(x+1)^2(x^2+1)} = \frac{1}{2(x+1)} + \frac{1}{2(x+1)^2} - \frac{x}{2(x^2+1)} \quad \text{for all } x \in \mathbb{R} \setminus \{-1\}.$$

4.4. Improper Integrals

As motivated by Example 4.1.9 above, we see that the function $f : [0,1] \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} x^{-1/2} & , 0 < x \leq 1, \\ 0 & , x = 0, \end{cases}$$

is not Riemann integrable on $[0,1]$. We see that f is continuous on $(0,1]$. This suggests us to approximate the area of the unbounded area by

$$\int_\epsilon^1 f(x)\,\mathrm{d}x = \int_\epsilon^1 x^{-1/2}\,\mathrm{d}x = 2x^{1/2}\Big|_{x=\epsilon}^{x=1} = 2 - 2\sqrt{\epsilon}$$

for a "small" parameter $\epsilon > 0$. This suggests us to consider the limit $\epsilon \to 0+$ to obtain

$$\lim_{\epsilon \to 0+} \int_\epsilon^1 f(x)\,\mathrm{d}x = 2.$$

This is called the improper integral of $f : (0,1] \to \mathbb{R}$. We see that the above improper integral does nothing with the value $f(0)$. We now summarize the above observation by the following definition.

Definition 4.4.1. Let $a, b \in \mathbb{R}$ with $a < b$.

(a) Suppose that $f : (a,b] \to \mathbb{R}$ is a continuous function. If $\lim_{\epsilon \to 0+} \int_{a+\epsilon}^b f(x)\,\mathrm{d}x$ exists, then we define the improper integral by

$$\int_a^b f(x)\,\mathrm{d}x := \lim_{\epsilon \to 0+} \int_{a+\epsilon}^b f(x)\,\mathrm{d}x.$$

(b) Suppose that $f : [a,b) \to \mathbb{R}$ is a continuous function. If $\lim_{\epsilon \to 0+} \int_a^{b-\epsilon} f(x)\,\mathrm{d}x$ exists, then we define the improper integral by

$$\int_a^b f(x)\,\mathrm{d}x := \lim_{\epsilon \to 0+} \int_a^{b-\epsilon} f(x)\,\mathrm{d}x.$$

(c) Let $a < c < b$. Suppose that $f : [a,b] \setminus \{c\} \to \mathbb{R}$ is a continuous function. If both $\lim_{\epsilon_2 \to 0+} \int_a^{c-\epsilon_2} f(x)\,\mathrm{d}x$ and $\lim_{\epsilon_1 \to 0+} \int_{c+\epsilon_1}^b f(x)\,\mathrm{d}x$ exist and the following two situations do not happen:
- $\lim_{\epsilon_2 \to 0+} \int_a^{c-\epsilon_2} f(x)\,\mathrm{d}x = +\infty$ and $\lim_{\epsilon_1 \to 0+} \int_{c+\epsilon_1}^b f(x)\,\mathrm{d}x = -\infty$,
- $\lim_{\epsilon_2 \to 0+} \int_a^{c-\epsilon_2} f(x)\,\mathrm{d}x = -\infty$ and $\lim_{\epsilon_1 \to 0+} \int_{c+\epsilon_1}^b f(x)\,\mathrm{d}x = +\infty$,

then we define the improper integral by

$$\int_a^b f(x)\,\mathrm{d}x := \lim_{\epsilon_2 \to 0+} \int_a^{c-\epsilon_2} f(x)\,\mathrm{d}x + \lim_{\epsilon_1 \to 0+} \int_{c+\epsilon_1}^b f(x)\,\mathrm{d}x. \tag{4.4.1}$$

Remark 4.4.2. If $f : [a,b] \to \mathbb{R}$ is continuous (hence integrable), the its Riemann integral is identical to the improper integrals above.

EXAMPLE 4.4.3. A closely related notion is called the *principal value integration*:

$$\mathrm{pv} \int_a^b g(x)\,\mathrm{d}x := \lim_{\epsilon \to 0+} \left(\int_a^{c-\epsilon} g(x)\,\mathrm{d}x + \int_{c+\epsilon}^b g(x)\,\mathrm{d}x \right). \tag{4.4.2}$$

This is different to the improper integral in Definition 4.4.1: the improper integral in (4.4.1) means we first take limit on each term, and then summing the resulting numbers, while the principal value integration in (4.4.2) means that we first sum the truncated integral with *same truncation level*, and then take the limit. If the improper integral of $g : [a,b] \setminus \{c\} \to \mathbb{R}$ exists in the sense of (4.4.1), then

$$\int_a^b g(x)\,\mathrm{d}x = \mathrm{pv} \int_a^b g(x)\,\mathrm{d}x.$$

However, it is possible that $\mathrm{pv} \int_a^b f(x)\,\mathrm{d}x$ exists but its improper integral does not exist. For example, we consider the function

$$f : \mathbb{R} \setminus \{0\} \to \mathbb{R}, \quad f(x) = \frac{1}{x} \text{ for all } x \in \mathbb{R} \setminus \{0\}.$$

For each $0 < \epsilon < 1$, we compute that

$$\int_\epsilon^1 \frac{1}{x}\,\mathrm{d}x = \ln x \Big|_{x=\epsilon}^{x=1} = -\ln \epsilon,$$

and by the change of variable formula one sees that

$$\int_{-1}^{-\epsilon} \frac{1}{x}\,\mathrm{d}x = -\int_\epsilon^1 \frac{1}{y}\,\mathrm{d}y = \ln \epsilon.$$

Hence one sees that

$$\int_{-1}^{-\epsilon} \frac{1}{x}\,\mathrm{d}x + \int_\epsilon^1 \frac{1}{x}\,\mathrm{d}x = 0 \quad \text{for all } 0 < \epsilon < 1,$$

which gives

$$\mathrm{pv} \int_{-1}^1 f(x)\,\mathrm{d}x = \lim_{\epsilon \to 0+} \left(\int_{-1}^{-\epsilon} \frac{1}{x}\,\mathrm{d}x + \int_\epsilon^1 \frac{1}{x}\,\mathrm{d}x \right) = 0.$$

However, since

$$\lim_{\epsilon \to 0+} \int_\epsilon^1 \frac{1}{x}\,\mathrm{d}x = +\infty, \quad \lim_{\epsilon \to 0} \int_{-1}^{-\epsilon} \frac{1}{x}\,\mathrm{d}x = -\infty,$$

thus the improper integral of f on $[-1,1] \setminus \{0\}$ does not exist. In fact, for each $c_1 > 0$ and $c_2 > 0$, one sees that

$$\lim_{\epsilon \to 0+} \left(\int_{-1}^{-c_1\epsilon} \frac{1}{x}\,\mathrm{d}x + \int_{c_2\epsilon}^1 \frac{1}{x}\,\mathrm{d}x \right) = \lim_{\epsilon \to 0+} (\ln(c_1\epsilon) - \ln(c_2\epsilon)) = \lim_{\epsilon \to 0+} \ln \frac{c_1\epsilon}{c_2\epsilon} = \ln \frac{c_1}{c_2},$$

this means that the limit is even depends on the "speed of convergence", which shows that *the area of unbounded regions may not well-defined without any restrictions*. If the limit is independent of the "speed of convergence", then the area of unbounded region is well-defined as described in Definition 4.4.1 above. We use the notations ϵ_1 and ϵ_2 there to emphasize the convergence rate of two limits may arbitrary.

We now generalize Example 4.1.9 in the following example.

EXAMPLE 4.4.4. Let $p > 0$ and we consider the function $f : (0, \infty) \to \mathbb{R}$ by $f(x) = x^{-p}$. For each $\epsilon > 0$, one computes that

$$\int_{\epsilon}^{1} f(x)\,\mathrm{d}x = \int_{\epsilon}^{1} x^{-p}\,\mathrm{d}x = \begin{cases} \frac{1}{1-p}(1-\epsilon^{1-p}) & \text{for all } p > 0 \text{ with } p \neq 1, \\ -\ln \epsilon & \text{when } p = 1. \end{cases}$$

Hence we compute the improper integral by

$$\lim_{\epsilon \to 0+} \int_{\epsilon}^{1} f(x)\,\mathrm{d}x = \begin{cases} \frac{1}{1-p} & \text{for all } p < 1, \\ +\infty & \text{for all } p \geq 1. \end{cases}$$

We see that $p = 1$ is a critical value.

EXERCISE 4.4.5. Let $p > 0$ and we consider the continuous function

$$f : (0, 1/e] \to \mathbb{R}, \quad f(x) = \frac{1}{x|\ln x|^p} = \frac{1}{x(-\ln x)^p} \text{ for all } x \in (0, 1/e].$$

Compute the improper integral

$$\lim_{\epsilon \to 0+} \int_{\epsilon}^{1/e} f(x)\,\mathrm{d}x$$

for each $p > 0$.

Another similar notion of improper integrals can be defined as follows:

DEFINITION 4.4.6. Given a real number $a \in \mathbb{R}$.

(a) Suppose that $f : [a, +\infty) \to \mathbb{R}$ is a continuous function. If $\lim_{M \to +\infty} \int_a^M f(x)\,\mathrm{d}x$ exists, then we define the improper integral by

$$\int_a^{+\infty} f(x)\,\mathrm{d}x := \lim_{M \to +\infty} \int_a^M f(x)\,\mathrm{d}x.$$

(b) Suppose that $f : (-\infty, a] \to \mathbb{R}$ is a continuous function. If $\lim_{M \to +\infty} \int_{-M}^a f(x)\,\mathrm{d}x$ exists, then we define the improper integral by

$$\int_{-\infty}^{a} f(x)\,\mathrm{d}x := \lim_{M \to +\infty} \int_{-M}^{a} f(x)\,\mathrm{d}x.$$

(c) Suppose that $f : \mathbb{R} \to \mathbb{R}$ is continuous function. If there exists $b \in \mathbb{R}$ such that both $\lim_{M_1 \to +\infty} \int_{-M_1}^{b} f(x)\,\mathrm{d}x$ and $\lim_{M_2 \to +\infty} \int_{b}^{M_2} f(x)\,\mathrm{d}x$ exist and the following two situations do not happen:
   - $\lim_{M \to +\infty} \int_{-M_1}^{b} f(x)\,\mathrm{d}x = +\infty$ and $\lim_{M_2 \to +\infty} \int_{b}^{M_2} f(x)\,\mathrm{d}x = -\infty$,
   - $\lim_{M \to +\infty} \int_{-M_1}^{b} f(x)\,\mathrm{d}x = -\infty$ and $\lim_{M_2 \to +\infty} \int_{b}^{M_2} f(x)\,\mathrm{d}x = +\infty$,

   then we define the improper integral by

$$\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x := \lim_{M_1 \to +\infty} \int_{-M_1}^{b} f(x)\,\mathrm{d}x + \lim_{M_2 \to +\infty} \int_{b}^{M_2} f(x)\,\mathrm{d}x. \tag{4.4.3}$$

REMARK 4.4.7. The integral (4.4.3) is independent of the choice of b, therefore it is well-defined. We also can similar define the *principal value integration centered at* b_0 by

$$\mathrm{pv} \int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x := \lim_{M \to +\infty} \int_{b_0 - M}^{b_0 + M} f(x)\,\mathrm{d}x.$$

In general, the principal value integration may depends on the choice b. It is important to mention that: If the improper integral (4.4.3) of $f : \mathbb{R} \to \mathbb{R}$ is well-defined, then

$$\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = \mathrm{pv}\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x$$

and its value is independent of the center b_0. Similar to above, the converse may false without any restriction.

In view of Definition 4.4.6(c), one should expect *the area under non-negative function is always well-defined in* $[0,+\infty]$. In fact, this is confirmed by the monotone convergence theorem (for Lebesgue integral). Here we formulate a more precise statement in terms of Riemann integral.

LEMMA 4.4.8 (Monotone convergence theorem for improper integrals). *Let I be an unbounded interval. If $f : I \to \mathbb{R}$ is a continuous function such that $f(x) \geq 0$ for all $x \in \mathbb{R}$. Then the improper integral*

$$\int_I f(x)\,\mathrm{d}x$$

is always well-defined (in the sense of (4.4.3)) with value in $[0,+\infty]$.

EXERCISE 4.4.9. Let $p > 0$. Show that

$$\int_1^{+\infty} x^{-p}\,\mathrm{d}x < +\infty \quad \text{if and only if} \quad p \geq 1.$$

Determine a necessary and sufficient condition for which

$$\int_1^{+\infty} \frac{1}{x|\ln x|^p}\,\mathrm{d}x < +\infty.$$

EXAMPLE 4.4.10 (Normal distribution). Given constants $\mu > 0$ and $\sigma > 0$, and we define the continuous function

$$p_{\mu,\sigma} : \mathbb{R} \to \mathbb{R}_{>0}, \quad p_{\mu,\sigma}(x) := \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \text{ for all } x \in \mathbb{R}.$$

By using Lemma 4.4.8, the improper integral $\int_{-\infty}^{+\infty} p_{\mu,\sigma}(x)\,\mathrm{d}x$ exists. In fact, one has (unfortunately, it is not possible to be computed using only the knowledge until at this point)

$$\int_{-\infty}^{+\infty} p_{\mu,\sigma}(x)\,\mathrm{d}x = 1.$$

This shows that $p_{\mu,\sigma}$ is a density of a probability distribution.

EXERCISE 4.4.11. For each $m \in \mathbb{N}$, show that the improper integral $\int_{-\infty}^{+\infty} x^m p_{\mu,\sigma}(x)\,\mathrm{d}x$ exists. In addition, compute

(a) the mean $\mathbb{E}(p_{\mu,\sigma}) := \int_{-\infty}^{+\infty} x p_{\mu,\sigma}(x)\,\mathrm{d}x$; and
(b) the variance $\mathrm{var}(p_{\mu,\sigma}) := \int_{-\infty}^{+\infty} (x - \mathbb{E}(p_{\mu,\sigma}))^2\, p_{\mu,\sigma}(x)\,\mathrm{d}x$,

of the normal distribution mentioned above.

Finally, we close this section by giving a remark that, unfortunately there is no unified definition for Definition 4.4.1 and Definition 4.4.6 in terms of Riemann integral. In fact, a consistent and unify framework can be given in terms of Lebesgue integral [**WZ15**], which we will not introduce in this course.

4.5. Some fundamental inequalities

We begin this section by the following fact, which can be found in [**BV04**, Section 3.1.4] (note that the second derivative test in Theorem 3.5.8 is an immediate consequence of this fact).

LEMMA 4.5.1. *Let $a, b \in \mathbb{R}$ with $a < b$ and let $f : (a, b) \to \mathbb{R}$ which is twice differentiable. The following are equivalent:*

(1) *f is* concave*, that is, $f(\alpha t_1 + \beta t_2) \geq \alpha f(t_1) + \beta f(t_2)$ for all $t_1, t_2 \in (a, b)$ and for all $\alpha \geq 0$, $\beta \geq 0$ with $\alpha + \beta = 1$.*
(2) *$f''(x) \leq 0$ for all $x \in (a, b)$.*

We now consider the logarithmic function $\ln : (0, \infty) \to \mathbb{R}$. One sees that

$$(\ln x)'' = (x^{-1})' = -x^{-2} < 0 \quad \text{for all } x \in (0, \infty),$$

then Lemma 4.5.1 says that $\ln : (0, \infty) \to \mathbb{R}$ is convex, that is,

$$\ln(\alpha t_1 + \beta t_2) \geq \alpha \ln(t_1) + \beta \ln(t_2) \quad \text{for all } t_1, t_2 > 0 \text{ and for all } \alpha \geq 0, \beta \geq 0 \text{ with } \alpha + \beta = 1.$$

Let $p > 1$, $q > 1$ be such that $\frac{1}{p} + \frac{1}{q} = 1$, and we choose $\alpha = \frac{1}{p}$ and $\beta = \frac{1}{q}$ in the above inequality to reach

$$\ln\left(\frac{1}{p} t_1 + \frac{1}{q} t_2\right) \geq \frac{1}{p} \ln(t_1) + \frac{1}{q} \ln(t_2) = \ln(t_1^{1/p} t_2^{1/q}) \quad \text{for all } t_1, t_2 > 0.$$

We now write $s_1 = t_1^{1/p} > 0$ and $s_2 = t_2^{1/q} > 0$ to see that $\ln\left(\frac{1}{p} s_1^p + \frac{1}{q} s_2^p\right) \geq \ln(s_1 s_2)$ for all $s_1 > 0$ and $s_2 > 0$. Since $\exp : \mathbb{R} \to \mathbb{R}_{>0}$ is strictly increasing, then we see that

$$\frac{1}{p} s_1^p + \frac{1}{q} s_2^p \geq s_1 s_2 \quad \text{for all } s_1 > 0 \text{ and } s_2 > 0.$$

The above inequality obviously holds true for either $s_1 = 0$ or $s_2 = 0$ as well, and we now conclude the following lemma.

LEMMA 4.5.2 (Young's inequality). *For each $p > 1$ and $q > 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$, we have the inequality*

$$s_1 s_2 \leq \frac{1}{p} s_1^p + \frac{1}{q} s_2^p \quad \text{for all } a \geq s_1 \text{ and } s_2 \geq 0.$$

For simplicity, let I be any connected interval in $\mathbb{R}$, and we consider any continuous functions $f : I \to \mathbb{R}$ and $g : I \to \mathbb{R}$. By Lemma 4.4.8, the improper integral

$$\|f\|_{L^p(I)} := \left(\int_I |f(x)|^p \,\mathrm{d}x\right)^{1/p} \text{ and } \|g\|_{L^q(I)} := \left(\int_I |g(x)|^q \,\mathrm{d}x\right)^{1/q} \text{ both exist.}$$

Suppose that $\|f\|_{L^p(I)} \neq 0$ and $\|g\|_{L^q(I)} \neq 0$. For each $x \in I$ we choose

$$s_1 := \frac{|f(x)|}{\|f\|_{L^p(I)}} \quad \text{and} \quad s_2 := \frac{|g(x)|}{\|g\|_{L^q(I)}}$$

in the Young's inequality (Lemma 4.5.2) to see that

$$\frac{|f(x)g(x)|}{\|f\|_{L^p(I)} \|g\|_{L^q(I)}} \leq \frac{1}{p} \frac{|f(x)|^p}{\|f\|_{L^p(I)}^p} + \frac{1}{q} \frac{|g(x)|^q}{\|g\|_{L^q(I)}^q} \quad \text{for all } x \in I.$$

Integrate both sides on $I \cap [-M, M]$, we see that

$$\begin{aligned}\frac{\int_{I\cap[-M,M]} |f(x)g(x)|\,\mathrm{d}x}{\|f\|_{L^p(I)}\|g\|_{L^q(I)}} &\leq \frac{1}{p}\frac{\int_{I\cap[-M,M]} |f(x)|^p\,\mathrm{d}x}{\|f\|_{L^p(I)}^p} + \frac{1}{q}\frac{\int_{I\cap[-M,M]} |g(x)|^q\,\mathrm{d}x}{\|g\|_{L^q(I)}^q} \\ &\leq \frac{1}{p}\frac{\int_I |f(x)|^p\,\mathrm{d}x}{\|f\|_{L^p(I)}^p} + \frac{1}{q}\frac{\int_I |g(x)|^q\,\mathrm{d}x}{\|g\|_{L^q(I)}^q} = \frac{1}{p} + \frac{1}{q} = 1.\end{aligned}$$

Again, by Lemma 4.4.8 the improper integral $\int_I |f(x)g(x)|\,\mathrm{d}x$ exists, then taking limit $M \to +\infty$ and we reach

$$\frac{\int_I |f(x)g(x)|\,\mathrm{d}x}{\|f\|_{L^p(I)}\|g\|_{L^q(I)}} \leq 1,$$

that is,

$$\|fg\|_{L^1(I)} = \int_I |f(x)g(x)|\,\mathrm{d}x \leq \|f\|_{L^p(I)}\|g\|_{L^q(I)}.$$

One sees that $\|f\|_{L^p(I)} = 0$ if and only if $f(x) = 0$ for all $x \in I$, and similarly, $\|g\|_{L^q(I)} = 0$ if and only if $g(x) = 0$ for all $x \in I$. Therefore the above inequality also holds true if either $\|f\|_{L^p(I)} = 0$ or $\|g\|_{L^q(I)} = 0$. It is easy to see that

$$\|fg\|_{L^1(I)} \leq \|f\|_{L^\infty(I)}\|g\|_{L^1(I)} \quad \text{if we write } \|f\|_{L^\infty(I)} = \sup_{y\in I} |f(y)|.$$

We now summarize the above computations in the following theorem.

THEOREM 4.5.3 (Hölder's inequality). *Let I be a connected interval in $\mathbb{R}$, then*

$$\|fg\|_{L^1(I)} \leq \|f\|_{L^p(I)}\|g\|_{L^q(I)} \quad \textit{for all } p \geq 1, q \geq 1 \textit{ with } \frac{1}{p} + \frac{1}{q} = 1$$

for all continuous functions $f : I \to \mathbb{R}$ and $g : I \to \mathbb{R}$.

REMARK 4.5.4 (Optimality). When $p > 1$ (if and only if $q > 1$), we choose $g = |f|^{\frac{p}{q}}$, then we see that

$$\|f\|_{L^p(I)}\|g\|_{L^q(I)} = \|f\|_{L^p(I)}\left(\int_I |f(x)|^p\,\mathrm{d}x\right)^{1/q} = \|f\|_{L^p(I)}\|f\|_{L^p(I)}^{\frac{p}{q}} = \|f\|_{L^p(I)}^{1+\frac{p}{q}} = \|f\|_{L^p(I)}^p$$

and

$$\|fg\|_{L^1(I)} = \int_I |f(x)|^{1+\frac{p}{q}}\,\mathrm{d}x = \int_I |f(x)|^p\,\mathrm{d}x = \|f\|_{L^p(I)}^p.$$

Combining the above two equations, we reach $\|f\|_{L^p(I)}\|g\|_{L^q(I)} = \|fg\|_{L^1(I)}$ when $g = |f|^{\frac{p}{q}}$. This shows that the exponents in Hölder's inequality (Theorem 4.5.3) are optimal, but however, the regularity of functions are not optimal: the continuity of functions and the connectness of domain are actually not required.

By using triangle inequality of the absolute value (1-dimensional Euclidean norm), one can easily check that

$$\|f + g\|_{L^1(I)} \leq \|f\|_{L^1(I)} + \|g\|_{L^1(I)},$$

and

$$\|f + g\|_{L^\infty(I)} \leq \|f\|_{L^\infty(I)} + \|g\|_{L^\infty(I)},$$

for all continuous functions $f : I \to \mathbb{R}$ and $g : I \to \mathbb{R}$. For $1 < p < +\infty$, we see that

$$
\begin{aligned}
|f(x)+g(x)|^p &= |f(x)+g(x)|^{p-1}|f(x)+g(x)| \\
&\le |f(x)+g(x)|^{p-1}|f(x)| + |f(x)+g(x)|^{p-1}|g(x)| \quad \text{for all } x \in I.
\end{aligned} \tag{4.5.1}
$$

By using Hölder's inequality (Theorem 4.5.3), for each $M > 0$, we compute that

$$
\begin{aligned}
&\int_{I\cap[-M,M]} |f(x)+g(x)|^{p-1}|f(x)|\,\mathrm{d}x \le \left(\int_{I\cap[-M,M]} |f(x)+g(x)|^{q(p-1)}\,\mathrm{d}x\right)^{1/q} \|f\|_{L^p(I)} \\
&= \left(\int_{I\cap[-M,M]} |f(x)+g(x)|^{p}\,\mathrm{d}x\right)^{1/q} \|f\|_{L^p(I)} = \|f+g\|_{L^p(I\cap[-M,M])}^{\frac{p}{q}} \|f\|_{L^p(I)}
\end{aligned}
$$

and similarly,

$$
\int_{I\cap[-M,M]} |f(x)+g(x)|^{p-1}|g(x)|\,\mathrm{d}x \le \|f+g\|_{L^p(I\cap[-M,M])}^{\frac{p}{q}} \|f\|_{L^p(I)}.
$$

Integrating (4.5.1) on $I \cap [-M, M]$ and then combining the resulting inequality with these two inequalities, we now see that

$$
\begin{aligned}
\|f+g\|_{L^p(I\cap[-M,M])}^p &= \int_{I\cap[-M,M]} |f(x)+g(x)|^p\,\mathrm{d}x \\
&\le \|f+g\|_{L^p(I\cap[-M,M])}^{\frac{p}{q}} \|f\|_{L^p(I)} + \|f+g\|_{L^p(I\cap[-M,M])}^{\frac{p}{q}} \|g\|_{L^p(I)} \\
&= \|f+g\|_{L^p(I\cap[-M,M])}^{\frac{p}{q}} (\|f\|_{L^p(I)} + \|g\|_{L^p(I)}).
\end{aligned}
$$

Since $f : I \to \mathbb{R}$ and $g : I \to \mathbb{R}$ are continuous, then $\|f + g\|_{L^p(I\cap[-M,M])} < +\infty$. If $\|f + g\|_{L^p(I\cap[-M,M])} \neq 0$, then we divide both sides by $\|f + g\|_{L^p(I\cap[-M,M])}$ to see that

$$
\|f+g\|_{L^p(I\cap[-M,M])} = \|f+g\|_{L^p(I\cap[-M,M])}^{p-\frac{p}{q}} \le \|f\|_{L^p(I)} + \|g\|_{L^p(I)}.
$$

The above inequality obviously hold trues when $\|f + g\|_{L^p(I\cap[-M,M])} = 0$. By using Lemma 4.4.8, we can take the limit $M \to +\infty$ to obtain

$$
\|f+g\|_{L^p(I)} \le \|f\|_{L^p(I)} + \|g\|_{L^p(I)}.
$$

We now summaraize the above discussions in the following theorem.

THEOREM 4.5.5 (Minkowski's inequality). *Let I be a connected interval in $\mathbb{R}$, then for each $1 \le p \le +\infty$ one has*

$$
\|f+g\|_{L^p(I)} \le \|f\|_{L^p(I)} + \|g\|_{L^p(I)}
$$

for all continuous functions $f : I \to \mathbb{R}$ and $g : I \to \mathbb{R}$.

The results in Theorem 4.5.5 does not hold true for $0 < p < 1$. For this case, we need different treatments. For any numbers $a > 0$ and $b > 0$, one sees that

$$
1 = \frac{a}{a+b} + \frac{b}{a+b} \le \left(\frac{a}{a+b}\right)^p + \left(\frac{b}{a+b}\right)^p,
$$

then

$$
(a+b)^p \le a^p + b^p \quad \text{for all } a > 0 \text{ and } b > 0. \tag{4.5.2}
$$

The above inequality obviously holds true for the case when $a = 0$ or $b = 0$. Now we choose

$$a = |f(x)| \quad \text{and} \quad b = |g(x)|,$$

and we reach

$$|f(x) + g(x)|^p \leq (|f(x)| + |g(x)|)^p \leq |f(x)|^p + |g(x)|^p \quad \text{for all } x \in I.$$

By using Lemma 4.4.8, we can integrate both sides on I to conclude the following theorem.

THEOREM 4.5.6 (Minkowski's inequality). *Let I be a connected interval in $\mathbb{R}$, then for each $0 < p < 1$ one has*

$$\|f + g\|_{L^p(I)}^p \leq \|f\|_{L^p(I)}^p + \|g\|_{L^p(I)}^p$$

for all continuous functions $f : I \to \mathbb{R}$ and $g : I \to \mathbb{R}$.

Minkowski's inequalities are exactly the triangle inequality, which says that the notion of "length" also can be introduced to function spaces, that is, by viewing functions as points, we can define the "distance between them". Here we close this section by remarking that the results in Theorem 4.5.3, Theorem 4.5.5 and Theorem 4.5.6 are not optimal, and they are far away from optimal. The optimal version has to be formulated in terms of Lebesgue integral [**WZ15**]. Finally, we end this semester by giving a remark that the results in this section also can be proved for series, see Section 5.6 below.

Part 2

Spring 2025 (113-2, 000713012)

CHAPTER 5

Numerical sequences and series

5.1. Convergence of sequences

We begin this semester by the following definitions.

DEFINITION 5.1.1. A *sequence* in $\mathbb{R}$ is a function $a : \mathbb{N} \to \mathbb{R}$. For convenience, we usually denote

$$a(i) = a_i \quad \text{for all } i \in \mathbb{N},$$

and we slightly abuse the notation by writing $\{a_i\}_{i\in\mathbb{N}}$ or $\{a_i\}_{i=1}^{+\infty}$.

DEFINITION 5.1.2. Let $\{a_i\}_{i=1}^{+\infty}$ be a sequence in $\mathbb{R}$.

(a) We say that the sequence $\{a_i\}_{i=1}^{+\infty}$ converges to some $a \in \mathbb{R}$ if: Given any $\epsilon > 0$, there exists $N = N(\epsilon) > 0$ such that

$$i \geq N \text{ implies } |a_i - a| < \epsilon.$$

In this case, we also say that $\lim_{i\to+\infty} a_i$ exists in $\mathbb{R}$, or we simply write $\lim_{i\to+\infty} a_i = a$ for some $a \in \mathbb{R}$.

(b) We say that the sequence $\{a_i\}_{i=1}^{+\infty}$ convergence to $+\infty$ if: Given any $M > 0$, there exists $N = N(M) > 0$ such that

$$i \geq N \text{ implies } a_i \geq M.$$

In this case, we also write $\lim_{i\to+\infty} a_i = +\infty$.

(c) We say that the sequence $\{a_i\}_{i=1}^{+\infty}$ convergence to $-\infty$ if: Given any $M > 0$, there exists $N = N(M) > 0$ such that

$$i \geq N \text{ implies } a_i \leq -M.$$

In this case, we also write $\lim_{i\to+\infty} a_i = -\infty$.

(d) If either (a), (b) or (c) holds, we unify the notations by saying that $\lim_{i\to+\infty} a_i$ exists, or by slightly abuse the notation by saying that $\lim_{i\to+\infty} a_i$ exists in $[-\infty, +\infty]$.

(c) Otherwise, if (d) does not hold, then we say the sequence $\{a_i\}_{i=1}^{+\infty}$ diverges.

We see that Definition 5.1.2 is nothing by just a special case of the usual limit for functions (see Definition 2.3.15 above). Therefore, similar properties will holds as well:

LEMMA 5.1.3. *Let $\{a_i\}_{i=1}^{+\infty}$ and $\{b_i\}_{i=1}^{+\infty}$ be a sequences in $\mathbb{R}$.*

(a) *If both $\{a_i\}_{i=1}^{+\infty}$ and $\{b_i\}_{i=1}^{+\infty}$ converge and the following two situations do not happen:*
- $\lim_{i\to+\infty} a_i = +\infty$ *and* $\lim_{i\to+\infty} b_i = -\infty$*;*
- $\lim_{i\to+\infty} a_i = -\infty$ *and* $\lim_{i\to+\infty} b_i = +\infty$*;*

then $\lim_{i\to+\infty}(a_i + b_i)$ exists and

$$\lim_{i\to+\infty}(a_i + b_i) = \lim_{i\to+\infty} a_i + \lim_{i\to+\infty} b_i.$$

(b) *If* $\{a_i\}_{i=1}^{+\infty}$ *converges, then for each* $c \in \mathbb{R}$ *the sequence* $\{ca_i\}_{i=1}^{+\infty}$ *converges and*

$$\lim_{i\to+\infty}(ca_i) = c\lim_{i\to+\infty} a_i.$$

(c) *If both* $\{a_i\}_{i=1}^{+\infty}$ *and* $\{b_i\}_{i=1}^{+\infty}$ *converge and the following two situations do not happen:*
- $\lim_{i\to+\infty} a_i = \pm\infty$ *and* $\lim_{i\to+\infty} b_i = 0$*;*
- $\lim_{i\to+\infty} a_i = 0$ *and* $\lim_{i\to+\infty} b_i = \pm\infty$*;*

then $\lim_{i\to+\infty}(a_ib_i)$ *exists and*

$$\lim_{i\to+\infty}(a_ib_i) = \left(\lim_{i\to+\infty} a_i\right)\left(\lim_{i\to+\infty} b_i\right).$$

(d) *If* $a_i \neq 0$ *for all* $i \in \mathbb{N}$ *and* $\{a_i\}_{i=1}^{+\infty}$ *converges in* $[-\infty,+\infty]\setminus\{0\}$*, then*

$$\lim_{i\to+\infty}\frac{1}{a_n} = \frac{1}{\lim_{i\to+\infty} a}.$$

We first remark that one can use continuous function to proof the existence of limits and compute it, similar to Example 3.3.11 above. For reader's convenience, here we rewrite the ideas in the following example, in terms of sequences.

Example 5.1.4. We consider the sequence $\{i^{1/i}\}_{i=1}^{+\infty}$. By using L' Hôpital's rule in Section 3.2 above (remember to check sufficient conditions!), one sees that

$$\lim_{i\to+\infty}\ln(i^{1/i}) = \lim_{i\to+\infty}\frac{1}{i}\ln i = 0.$$

One cannot use the continuity of $\ln : (0,+\infty) \to \mathbb{R}$ to write $\ln(\lim_{i\to+\infty} i^{1/i})$, since we do not know whether $\lim_{i\to+\infty} i^{1/i}$ exists or not at the moment. The proper way to argue this is to use the continuity of $\exp : \mathbb{R} \to (0,+\infty)$ and see that

$$1 = e^0 = \exp\left(\lim_{i\to+\infty}\ln(i^{1/i})\right) = \lim_{i\to+\infty}\exp\left(\ln(i^{1/i})\right) = \lim_{i\to+\infty} i^{1/i} = \lim_{i\to+\infty} i^{1/i},$$

which completes the proof.

We first introduce a simple criteria to check whether the limit exists or not.

Lemma 5.1.5 (Monotone convergence theorem). *Let* $\{a_i\}_{i=1}^{+\infty}$ *be a sequence in* $\mathbb{R}$*.*

(a) *If* $\{a_i\}_{i=1}^{+\infty}$ *is non-decreasing and there exists a number* $b \in \mathbb{R}$ *such that* $a_i \leq b$ *for all* $i \in \mathbb{N}$*, then* $\lim_{i\to+\infty} a_i = a$ *for some* $a \in \mathbb{R}$ *with* $a \leq b$*.*
(b) *If* $\{a_i\}_{i=1}^{+\infty}$ *is non-increasing and there exists a number* $b \in \mathbb{R}$ *such that* $a_i \geq b$ *for all* $i \in \mathbb{N}$*, then* $\lim_{i\to+\infty} a_i = a$ *for some* $a \in \mathbb{R}$ *with* $a \geq b$*.*
(c) *If* $\{a_i\}_{i=1}^{+\infty}$ *is monotone (that is, either non-decreasing or non-increasing), then* $\lim_{i\to+\infty} a_i$ *converges in* $[-\infty,+\infty]$*.*

While taking limit, we always need to check whether it exists or not, which is very inconvenient. For future convenience, we similarly introduce the notion of *limit superior* and *limit inferior* for sequences. Here we follow [**Rud76**, Definition 3.16].

The above lemma only guarantees the existence, but the exact value of the limit is unknown. We demonstrate this in the following example.

Example 5.1.6. Let $\{a_i\}_{i=1}^{+\infty}$ be a sequence in $\mathbb{R}$, which defined by the recurrence relation

$$a_1 = \sqrt{2},\quad a_{i+1} = \sqrt{2+\sqrt{a_i}} \text{ for all } i = 2,3,4,\cdots. \tag{5.1.1}$$

We first check $0 < a_i < 2$ for all $i = 1, 2, \cdots$ by using mathematical induction. The base case is immediately verified by $0 < a_1 = \sqrt{2} < 2$. It remains to show the following:

$$\overbrace{0 < a_i < 2 \text{ for some } i \in \mathbb{N}}^{\text{induction hypothesis}} \implies 0 < a_{i+1} < 2.$$

Since $\sqrt{\cdot} : [0, +\infty) \to [0, +\infty)$ is strictly increasing, it is not difficult to see this, and the details are left to the readers as exercises. Now the mathematical induction implies that

$$0 < a_i < 2 \quad \text{for all } i \in \mathbb{N}. \tag{5.1.2}$$

We now want to show that $\{a_i\}_{i=1}^{+\infty}$ is strictly increasing. First of all, from $a_1 > 0$ we check that

$$a_2 = \sqrt{2 + \sqrt{a_1}} > \sqrt{2 + \sqrt{0}} = \sqrt{2} = a_1,$$

which confirmed the base case. It remains to show the following:

$$\overbrace{a_{i+1} > a_i \text{ for some } i \in \mathbb{N}}^{\text{induction hypothesis}} \implies a_{i+2} > a_{i+1}.$$

Since $\sqrt{\cdot} : [0, +\infty) \to [0, +\infty)$ is strictly increasing, From $a_{i+1} > a_i$, we see that $\sqrt{a_{i+1}} > \sqrt{a_i}$, and consequently one sees that $2 + \sqrt{a_{i+1}} > 2 + \sqrt{a_i}$. Now, again using the strictly increasing function $\sqrt{\cdot} : [0, +\infty) \to [0, +\infty)$, we now see that

$$a_{i+2} = \sqrt{2 + \sqrt{a_{i+1}}} > \sqrt{2 + \sqrt{a_i}} = a_{i+1}.$$

Now the mathematical induction implies that

$$a_i > a_i \quad \text{for all } i \in \mathbb{N}. \tag{5.1.3}$$

In view of (5.1.2) and (5.1.3), now the monotone convergence theorem (Lemma 5.1.5) guarantees that there exists a number $0 \leq a \leq 2$ such that

$$\lim_{i \to +\infty} a_i = a. \tag{5.1.4}$$

At this moment, we do not know the precise value of a. By taking the limit $i \to +\infty$ in the recurrence relation $a_{i+1} = \sqrt{2 + \sqrt{a_i}}$, from the continuity of $\sqrt{\cdot} : [0, +\infty) \to [0, +\infty)$ one sees that $a = \sqrt{2 + \sqrt{a}}$, and thus

$$(\sqrt{a})^4 = 2 + \sqrt{a}.$$

This means that $\sqrt{a}$ solves the polynomial $y^4 - y - 2 = 0$. In this case, by taking account to the condition $0 \leq a \leq 2$, by using advance algebra, one can compute that

$$a = \frac{1}{9}\left(1 - 2\sqrt[3]{\frac{2}{47 + 3\sqrt{249}}} + \sqrt[3]{\frac{1}{2}(47 + 3\sqrt{249})}\right)^2 \approx 1.831177\cdots.$$

We finally remark that the limit a is called the fixed point of the recurrence relation (5.1.1), and *it may not unique*! Even though the limit (5.1.4) follows by taking the limit $i \to +\infty$ in the recurrence relation $a_{i+1} = \sqrt{2 + \sqrt{a_i}}$, it actually *also depends on the initial condition* a_1, that is why *we always need to check base case in mathematical induction* (even though it may looks trivial).

DEFINITION 5.1.7. Given any sequence $\{a_i\}_{i=1}^{+\infty}$ in $\mathbb{R}$, we define

$$\limsup_{i\to+\infty} a_i := \lim_{i\to+\infty} \sup_{j\geq i} a_j \equiv \inf_{i\in\mathbb{N}} \sup_{j\geq i} a_j,$$
$$\liminf_{n\to+\infty} a_i := \lim_{i\to+\infty} \inf_{j\geq i} a_j \equiv \sup_{i\in\mathbb{N}} \inf_{j\geq i} a_j.$$

This is nothing by just a special case of Definition 2.3.16. Unlike limit, the monotone convergence theorem (Lemma 5.1.5) implies that both limit superior and limit inferior always exist (because $\sup_{j\geq i} a_j$ and $\inf_{j\geq i} a_j$ are monotone as the index i increasing) in $[-\infty, +\infty]$. It is clear that

$$\liminf_{i\to+\infty} a_i \leq \limsup_{i\to+\infty} a_i$$

and if $a_i \leq b_i$ for all $i \geq N$ for some $N > 0$ one has

$$\limsup_{i\to+\infty} a_i \leq \limsup_{i\to+\infty} b_i, \quad \liminf_{i\to+\infty} a_i \leq \liminf_{i\to+\infty} b_i.$$

In addition, one has

$$\lim_{i\to+\infty} a_i = a \in \mathbb{R} \iff \limsup_{i\to+\infty} a_i = \liminf_{i\to+\infty} a_i = a \in \mathbb{R} \iff \limsup_{i\to+\infty} |a_i - a| = 0,$$
$$\lim_{i\to+\infty} a_i = +\infty \iff \liminf_{i\to+\infty} a_i = +\infty,$$
$$\lim_{i\to+\infty} a_i = -\infty \iff \limsup_{i\to+\infty} a_i = -\infty.$$

EXAMPLE 5.1.8 (Oscillating sequences). We consider the sequence $\{(-1)^i\}_{i=1}^{+\infty}$. One sees that

$$\limsup_{i\to+\infty} (-1)^i = 1 \neq -1 = \liminf_{i\to-\infty} (-1)^i,$$

which shows that the sequence $\{(-1)^i\}_{i=1}^{+\infty}$ is divergent.

We also consider the sequence $\{i(-1)^i\}_{i=1}^{+\infty}$. One sees that

$$\limsup_{i\to+\infty} i(-1)^i = +\infty \neq -\infty = \liminf_{i\to-\infty} i(-1)^i,$$

which shows that the sequence $\{i(-1)^i\}_{i=1}^{+\infty}$ is divergent. This example shows that the sequence may oscillating with magnitude $+\infty$.

EXAMPLE 5.1.9. We consider the sequence $\{a_i\}_{i=1}^{+\infty}$ defined by

$$a_i = \begin{cases} i^{-1} & \text{for all odd } i \in \mathbb{N}, \\ i^{-2} & \text{for all even } i \in \mathbb{N}. \end{cases}$$

Since $\lim_{i\to+\infty} i^{-1} = 0$ and $\lim_{i\to+\infty} i^{-2} = 0$, then

$$\limsup_{i\to+\infty} |a_i - 0| = \limsup_{i\to+\infty} a_i = 0,$$

which concludes that $\lim_{i\to+\infty} a_i = 0$.

However, one has to be careful that, *we only have subadditivity (resp. superaddivity) property for limit supremum (resp. limit infimum)*:

$$\begin{cases} \limsup_{i\to\infty}(a_i + b_i) \leq \limsup_{i\to\infty} a_i + \limsup_{i\to\infty} b_i \\ \liminf_{i\to\infty}(a_i + b_i) \geq \liminf_{i\to\infty} a_i + \liminf_{i\to\infty} b_i \end{cases} \tag{5.1.5}$$

holds whenever the right hand side is not $\infty - \infty$ or $-\infty + \infty$. **For the case when $\lim_{i\to\infty} b_i$ exists and finite**, by writing $a_i = (a_i + b_i) + (-b_i)$, using (5.1.5) we obtain

$$\begin{cases} \limsup_{i\to\infty} a_i \le \limsup_{i\to\infty}(a_i+b_i) - \lim_{i\to\infty} b_i \\ \liminf_{i\to\infty} a_i \ge \liminf_{i\to\infty}(a_i+b_i) - \lim_{i\to\infty} b_i \end{cases}$$

which implies

$$\begin{cases} \limsup_{i\to\infty} a_i + \lim_{i\to\infty} b_i \le \limsup_{i\to\infty}(a_i+b_i), \\ \liminf_{i\to\infty} a_i + \lim_{i\to\infty} b_i \ge \liminf_{i\to\infty}(a_i+b_i). \end{cases}$$

Combining this with (5.1.5), we reach

$$\begin{cases} \limsup_{i\to\infty}(a_i+b_i) = \limsup_{i\to\infty} a_i + \lim_{i\to\infty} b_i \\ \liminf_{i\to\infty}(a_i+b_i) = \liminf_{i\to\infty} a_i + \lim_{i\to\infty} b_i \end{cases} \quad \text{when } \lim_{i\to\infty} b_i \text{ exists and finite.} \tag{5.1.6}$$

If $\{a_i\}$ is bounded and $\lim_{i\to\infty} b_i$ exists which converges to some $b \ge 0$, by writing $a_i b_i = a_i b + a_i(b_i - b)$ and using (5.1.6), one sees that

$$\begin{cases} \limsup_{i\to\infty}(a_i b_i) = \limsup_{i\to\infty}(a_i b) \overset{(\because b\ge 0)}{\equiv} \left(\limsup_{i\to\infty} a_i\right)\left(\lim_{i\to\infty} b_i\right), \\ \liminf_{i\to\infty}(a_i b_i) = \liminf_{i\to\infty}(a_i b) \overset{(\because b\ge 0)}{\equiv} \left(\liminf_{i\to\infty} a_i\right)\left(\lim_{i\to\infty} b_i\right). \end{cases} \tag{5.1.7}$$

If we choose trivial sequence $b_i = b \ge 0$ for all $i \in \mathbb{N}$, then we reach

$$\limsup_{i\to\infty}(b a_i) = b \limsup_{i\to\infty} a_i \quad \text{for } b \ge 0. \tag{5.1.8}$$

However, one should be aware that when $b \ge 0$, we have

$$\limsup_{i\to\infty}(b a_i) = -\liminf_{i\to\infty}(|b| a_i) = -|b| \liminf_{i\to\infty} a_i = b \liminf_{i\to\infty} a_i \quad \text{for } b \le 0.$$

EXERCISE 5.1.10. Compute $\limsup_{i\to\infty}(a_i b_i)$ and $\liminf_{i\to\infty}(a_i b_i)$ when $\{a_i\}$ is bounded and $\lim_{i\to\infty} b_i$ exists which converges to some $b \le 0$.

If both $\{a_i\}$ and $\{b_i\}$ are non-negative, one also has

$$\begin{cases} \limsup_{i\to\infty}(a_i b_i) \le \left(\limsup_{i\to\infty} a_i\right)\left(\limsup_{i\to\infty} b_i\right) \\ \liminf_{i\to\infty}(a_i b_i) \ge \left(\liminf_{i\to\infty} a_i\right)\left(\liminf_{i\to\infty} b_i\right) \end{cases} \quad \text{for } \textbf{non-negative } \{a_i\}_{i=1}^{+\infty}, \{b_i\}_{i=1}^{+\infty} \tag{5.1.9}$$

holds whenever the right hand side is not $0 \cdot \infty$ or $\infty \cdot 0$.

5.2. Absolute and conditional convergence of series

Let $\{a_i\}_{i=1}^{+\infty}$ be a sequence in $\mathbb{R}$. We now asking whether we can summing up all the elements in $\{a_i\}_{i=1}^{+\infty}$. If we define the function $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} a_i & x \in [i-1, i), \\ 0 & \text{otherwise,} \end{cases}$$

then for each $N \in \mathbb{N}$ we see that

$$\sum_{i=1}^{N} a_i := a_1 + \cdots + a_N = \int_0^N f(x)\,\mathrm{d}x.$$

We now introduce the following definition, which is a analogue to the improper integral (Definition 4.4.6).

Definition 5.2.1. Let $\{a_i\}_{i=1}^{+\infty}$ be a sequence in $\mathbb{R}$, for each $N \in \mathbb{N}$ we define the *partial sum* by

$$s_N := \sum_{i=1}^{N} a_i.$$

Note that $\{s_N\}_{N=1}^{+\infty}$ is also a sequence in $\mathbb{R}$. We say that the series $\sum_{i=1}^{+\infty} a_i$ converges in $[-\infty, +\infty]$ if $\{s_N\}_{N=1}^{+\infty}$ converges in $[-\infty, +\infty]$ and we write

$$\sum_{i=1}^{+\infty} a_i := \lim_{N\to+\infty} s_N = \lim_{N\to+\infty} \sum_{i=1}^{N} a_i.$$

Let us introduce the following technical lemma before begin our discussions.

Lemma 5.2.2 (Alternating series test). *If the sequence $\{a_i\}_{i=1}^{+\infty}$ in $\mathbb{R}$ is* nonincreasing *and satisfies $\lim_{i\to+\infty} a_i = 0$, then the alternating series $\sum_{i=1}^{+\infty}(-1)^i a_i$ converges to some $a \in \mathbb{R}$ in the sense of Definition 5.2.1.*

Example 5.2.3. We now consider the sequence $\{a_i\}_{i=1}^{+\infty}$ in $\mathbb{R}$ given by

$$a_i = (-1)^i \frac{1}{i} \quad \text{for all } i \in \mathbb{N}.$$

The alternating series test (Lemma 5.2.2) guarantees that $\sum_{i=1}^{+\infty} a_i$ converges to some $a \in \mathbb{R}$, more precisely,

$$\lim_{N\to+\infty} \sum_{i=1}^{N} (-1)^i \frac{1}{i} = a, \tag{5.2.1}$$

In fact $a = -\ln 2$, but the computation requires further advance tools, see Example 5.5.13 below. We begin with the series written in usual order:

$$\ln 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} - \frac{1}{8} + \frac{1}{9} - + \cdots .$$

We now collect the positive terms $\{b_1, b_2, b_3, \cdots\} := \{1, \frac{1}{3}, \frac{1}{5}, \cdots\}$ and negative terms $\{c_1, c_2, c_3, \cdots\} := \{-\frac{1}{2}, -\frac{1}{4}, -\frac{1}{6}, \cdots\}$. We now rearrange the terms as described by the following algorithm:

Algorithm 2 One way of rearranging alternating harmonic series

for $i = 1, 2, 3, \cdots$ **do**
    Define $\tilde{a}_{3i-2} = b_i$, $\tilde{a}_{3i-1} = c_{2i-1}$ and $\tilde{a}_{3i} = c_{2i}$
end for

The resulting sequence $\{\tilde{a}_i\}_{i=1}^{+\infty}$ is a rearrangement of $\{a_i\}_{i=1}^{+\infty}$, and the corresponding series is

$$
\begin{aligned}
&\overbrace{\left(1-\frac{1}{2}-\frac{1}{4}\right)}^{(i=1)}+\overbrace{\left(\frac{1}{3}-\frac{1}{6}-\frac{1}{8}\right)}^{(i=2)}+\cdots+\left(\frac{1}{2i-1}-\frac{1}{2(2i-1)}-\frac{1}{4i}\right)+\cdots \\
&=\lim_{N\to+\infty}\sum_{i=1}^{N}\left(\frac{1}{2i-1}-\frac{1}{2(2i-1)}-\frac{1}{4i}\right) \quad \text{(precise meaning)} \\
&=\lim_{N\to+\infty}\sum_{i=1}^{N}\left(\frac{1}{2(2i-1)}-\frac{1}{4i}\right) \\
&=\lim_{N\to+\infty}\frac{1}{2}\sum_{i=1}^{N}\left(\frac{1}{2i-1}-\frac{1}{2i}\right) \\
&=\frac{1}{2}\left(\lim_{N\to+\infty}\sum_{i=1}^{2N}(-1)^i\frac{1}{i}\right)=\frac{1}{2}a \quad \text{(using (5.2.1))}.
\end{aligned}
$$

This example shows that *the order of series cannot be change in general*. If we go back to the definition of series (Definition 5.1.1), the precise meaning of the sequences $\{a_i\}_{i=1}^{+\infty}$ and $\{\tilde{a}_i\}_{i=1}^{+\infty}$ are functions $a:\mathbb{N}\to\mathbb{R}$ and $\tilde{a}:\mathbb{N}\to\mathbb{R}$. These two functions are different, but there exists a bijection between their ranges, in other words $\{a_i\}_{i=1}^{+\infty}$ and $\{\tilde{a}_i\}_{i=1}^{+\infty}$ are identical as sets.

Even though the above example demonstrates some ambiguity may appear if we abuse the notation like Definition 5.1.1, but however this does not cause major problem in practical due to the following theorem.

THEOREM 5.2.4 (A sufficient condition). *Let $\{a_i\}_{i=1}^{+\infty}$ be a sequence in $\mathbb{R}$. If the series $\sum_{i=1}^{+\infty}|a_i|$ converges in $\mathbb{R}$, then $\sum_{i=1}^{+\infty}a_i$ converges in $\mathbb{R}$ and*

$$\sum_{i=1}^{+\infty}a_i=\sum_{i=1}^{+\infty}a_{\sigma(i)} \quad \textit{for all bijection } \sigma:\mathbb{N}\to\mathbb{N}.$$

REMARK 5.2.5. If $\sum_{i=1}^{+\infty}|a_i|$ converges in $\mathbb{R}$, then we also has

$$\left|\sum_{i=1}^{+\infty}a_i\right|\le\sum_{i=1}^{+\infty}|a_i| \tag{5.2.2}$$

which is valid as $\sum_{i=1}^{+\infty}a_i$ converges in $\mathbb{R}$.

REMARK 5.2.6. For non-negative sequence $\{b_i\}_{i=1}^{+\infty}$, one sees that its partial sum $\{S_N\}_{N=1}^{+\infty}$ given by

$$S_N=\sum_{i=1}^{N}b_i$$

is a nondecreasing sequence. By using the monotone convergence theorem (Lemma 5.1.5), one sees that $\lim_{N\to+\infty} S_N$ converges in $[0,+\infty]$, which means that there only two possibilities:

$$\sum_{i=1}^{+\infty} b_i = b \text{ for some } b \in \mathbb{R} \quad \text{or} \quad \sum_{i=1}^{+\infty} b_i = +\infty.$$

Since $\{S_N\}_{N=1}^{+\infty}$ is nondecreasing, then

$$\sum_{i=1}^{+\infty} b_i = b \text{ for some } b \in \mathbb{R} \iff \sum_{i=1}^{+\infty} b_i < +\infty. \tag{5.2.3}$$

The notation (5.2.3) is only valid for nonnegative sequence $\{b_i\}_{i=1}^{+\infty}$. Therefore, the assumption in Theorem 5.2.4 can be written as

$$\sum_{i=1}^{+\infty} |a_i| < +\infty.$$

Together with (5.2.2), we also see that

$$\left|\sum_{i=1}^{+\infty} a_i\right| \leq \sum_{i=1}^{+\infty} |a_i| < +\infty. \tag{5.2.4}$$

EXAMPLE 5.2.7. We now give an example to demonstrate the notation (5.2.3) may not valid for arbitrary series. For example, we consider

$$a_i = (-1)^i \quad \text{for all } i \in \mathbb{N}.$$

The partial sum $S_N = \sum_{i=1}^N a_i$ is given by

$$S_N = \begin{cases} -1 & \text{for all odd } i \in \mathbb{N}, \\ 0 & \text{for all even } i \in \mathbb{N}. \end{cases}$$

Since $\limsup_{N\to+\infty} S_N = 1 \neq 0 = \liminf_{N\to+\infty} S_N$, thus $\lim_{N\to+\infty} S_N$ does not exist, in other words, the series $\sum_{i=1}^{+\infty} a_i$ diverges. However, one sees that

$$\left|\sum_{i=1}^{N} a_i\right| \leq 1 \quad \text{for all } N \in \mathbb{N},$$

it is illegal to denote $\left|\sum_{i=1}^{+\infty} a_i\right| \leq 1$ for divergence series $\sum_{i=1}^{+\infty} a_i$. It is interesting to compare this example with (5.2.4).

In view of Theorem 5.2.4, we finally end this section by introducing some definitions.

DEFINITION 5.2.8. Let $\{a_i\}_{i=1}^{+\infty}$ be a sequence in $\mathbb{R}$.

(a) We say that the series $\sum_{i=1}^{+\infty} a_i$ is *absolutely convergent* if $\sum_{i=1}^{+\infty} |a_i| < +\infty$.
(b) We say that the series $\sum_{i=1}^{+\infty} a_i$ is *conditionally convergent* if $\sum_{i=1}^{+\infty} |a_i| = +\infty$ but $\sum_{i=1}^{+\infty} a_i$ converges in $\mathbb{R}$ in the sense of Definition 5.2.1.

EXAMPLE 5.2.9. The sequence in Example 5.2.3 is conditionally convergent. In fact, one can see this by the estimate

$$\begin{aligned}
&1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\frac{1}{6}+\frac{1}{7}+\frac{1}{8}+\cdots \\
&\quad = 1+\frac{1}{2}+\left(\frac{1}{3}+\frac{1}{4}\right)+\left(\frac{1}{5}+\frac{1}{6}+\frac{1}{7}+\frac{1}{8}\right)+\cdots \\
&\quad \geq 1+\frac{1}{2}+\left(\frac{1}{4}+\frac{1}{4}\right)+\left(\frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}\right)+\cdots \\
&\quad \geq 1+\frac{1}{2}+\frac{1}{2}+\frac{1}{2}+\cdots = +\infty.
\end{aligned}$$

We will later give a more systematic way in Example 5.3.2.

5.3. Convergence of nonnegative sequence

In many practical situation, we are interested in absolutely convergent series rather than the conditionally convergent one. As we discussed above, the convergence of series also can be understood as the well-defineness of improper integral. This suggests us to consider non-negative sequences $\{b_i\}_{i=1}^{+\infty}$, which is the main theme of this section. Throughout this section, we will use the notation in (5.2.3). We already introduce a criteria, called the monotone convergence theorem (Lemma 5.1.5), to check the convergence of the series $\sum_{i=1}^{+\infty} b_i$. We exhibit some other way to check the convergence of non-negative series.

5.3.1. Integral test. By using the monotone convergence theorems in Lemma 4.4.8 and Lemma 5.1.5, we see that the series $\sum_{i=1}^{+\infty} b_i$ for *non-negative* sequences $\{b_i\}_{i=1}^{+\infty}$ is exactly identical to the improper integral

$$\int_0^{+\infty} f(x)\,\mathrm{d}x \quad \text{where} \quad f(x) = \begin{cases} b_i & x \in [i-1,i) \text{ for all } i \in \mathbb{N}, \\ 0 & \text{otherwise.} \end{cases}$$

In view of the upper sum and lower sum with the special partition $\Gamma_N = \{0 < 1 < 2 < \cdots < N\}$ on $[0,N]$, we immediately reach the following useful test:

LEMMA 5.3.1 (Integral test for *nonnegative* series). *Suppose that* $\{a_i\}_{i=1}^{+\infty}$ *is a* nonnegative nonincreasing *sequence. For each nonnegative continuous function* $f : [1,+\infty) \to \mathbb{R}$ *with* $f(i) = a_i$ *for all* $i \in \mathbb{N}$, *one has*

$$\sum_{i=1}^{+\infty} a_i < +\infty \quad \textit{if and only if} \quad \int_1^{+\infty} f(x)\,\mathrm{d}x < +\infty.$$

EXAMPLE 5.3.2 (p-series). For each $p > 0$, we want to study the convergence of the positive series $\sum_{i=1}^{+\infty} i^{-p}$, which is the so-called p-series. The case when $p = 1$ corresponding to the harmonic series mentioned in Example 5.2.9 above. From Exercise 4.4.9 we know that

$$\int_1^{+\infty} x^{-p}\,\mathrm{d}x < +\infty \quad \text{if and only if} \quad p \geq 1.$$

Since $\{i^{-p}\}_{i=1}^{+\infty}$ is a *nonnegative nonincreasing* sequence, then the integral test is valid with continuous function $f : [1, +\infty) \to \mathbb{R}$ $f(x) = x^{-p}$, so that we conclude that

$$\sum_{i=1}^{+\infty} i^{-p} < +\infty \quad \text{if and only if} \quad p \geq 1.$$

EXAMPLE 5.3.3. We now give a examples to demonstrate that, even with the "positivity" condition, *the integral test may fails for both directions without "nonincreasing" assumption on* $f : [1, +\infty) \to \mathbb{R}$.

(1) We first consider the function

$$f(x) = (\sin(\pi x))^2 + \frac{1}{x^2} \quad \text{with} \quad a_i := f(i) = \frac{1}{i^2} \text{ for all } i \in \mathbb{N}.$$

One sees that

$$\sum_{i=1}^{+\infty} a_i < +\infty \quad \text{but} \quad \int_1^{+\infty} f(x)\,\mathrm{d}x \geq \int_1^{+\infty} (\sin(\pi x))^2\,\mathrm{d}x = +\infty.$$

(2) We now define the continuous function $\phi : \mathbb{R} \to \mathbb{R}$ given by

$$\phi(x) := \begin{cases} 1 + x & , -1 \leq x \leq 0, \\ 1 - x & , 0 < x \leq 1, \\ 0 & , \text{otherwise}. \end{cases}$$

It is easy to see that $\int_{-\infty}^{+\infty} \phi(x)\,\mathrm{d}x = \int_{-1}^{1} \phi(x)\,\mathrm{d}x = 1$ with $\phi(0) = 1$. We now define the function $f : (-\infty, +\infty) \to +\infty$ by

$$f(x) := \sum_{j=1}^{+\infty} j^{-1} \phi(2^j(x - j)),$$

which is non-negative and continuous. We now write $a_i := f(i) = i^{-1}$ for all $i \in \mathbb{N}$. We see that (this can be determined by solving the equations $2^j(x - j) = \pm 1$)

$$\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = \sum_{j=1}^{+\infty} j^{-1} \int_{j-2^{-j}}^{j+2^{-j}} \phi(2^j(x - j))\,\mathrm{d}x = \sum_{j=1}^{+\infty} j^{-1} 2^{-j} < +\infty$$

but however $\sum_{i=1}^{+\infty} a_i = +\infty$.

5.3.2. Limit comparison test. Let $\{a_i\}_{i=1}^{+\infty}$ and $\{b_i\}_{i=1}^{+\infty}$ are non-negative sequences. If there exists positive constants C and N such that

$$a_i \leq C b_i \quad \text{for all } i \geq N, \tag{5.3.1}$$

then we immediately see that (which sometimes referred as *comparison test*):

$$\sum_{i=1}^{+\infty} b_i < +\infty \text{ implies } \sum_{i=1}^{+\infty} a_i < +\infty.$$

However, it is difficult to check (5.3.1) in many practical case. Instead, the following theorem is helpful:

THEOREM 5.3.4 (Limit comparison test). *Assume that* $\{a_i\}_{i=1}^{+\infty}$ *and* $\{b_i\}_{i=1}^{+\infty}$ *are* positive *sequences.*

(a) *If there exists* $c \in \mathbb{R} \setminus \{0\}$ *such that* $\lim_{i\to+\infty} \frac{a_i}{b_i} = c$, *then*

$$\sum_{i=1}^{+\infty} a_i < +\infty \quad \textit{if and only if} \quad \sum_{i=1}^{+\infty} b_i < +\infty.$$

(b) *If* $\lim_{i\to+\infty} \frac{a_i}{b_i} = 0$, *then*

$$\sum_{i=1}^{+\infty} b_i < +\infty \quad \textit{implies} \quad \sum_{i=1}^{+\infty} a_i < +\infty.$$

Roughly speaking, the condition in Theorem 5.3.4(a) means that a_i looks very similar to b_i up to a multiplicative constant, and the theorem confirms our intuition that the convergence behavior of the series $\sum_{i=1}^{+\infty} a_i < +\infty$ and $\sum_{i=1}^{+\infty} b_i < +\infty$ are the same. The condition in Theorem 5.3.4(b) means that the the convergence a_i to 0 is significantly faster than b_i, and the theorem confirms our intuition as well.

Example 5.3.5. We now consider the sequence $\{a_i\}_{i=1}^{+\infty}$ given by

$$a_i = \frac{e^i - 10}{e^i i^2 + 1} \quad \text{for all } i \in \mathbb{N}.$$

We see that $a_i > 0$ for all $i = 3, 4, 5, \cdots$, and we consider the sequence $\{a_i\}_{i=3}^{+\infty}$. By observing that

$$a_i = \frac{1 - 10e^{-i}}{i^2 + e^{-i}} \quad \text{for all } i \in \mathbb{N},$$

this suggests us to consider the positive sequence $\{b_i\}_{i=3}^{+\infty}$ given by

$$b_i = \frac{1}{i^2} \quad \text{for all } i \in \mathbb{N},$$

Now since

$$\lim_{i\to+\infty} \frac{a_i}{b_i} = \lim_{i\to+\infty} \frac{1 - 10e^{-i}}{1 + i^{-2}e^{-i}} = 1 \quad \text{and} \quad \sum_{i=3}^{+\infty} b_i < +\infty,$$

the limit comparison test guarantees that $\sum_{i=3}^{+\infty} a_i < +\infty$, and thus we conclude that

$$\sum_{i=1}^{+\infty} a_i \text{ converges.}$$

Intuitively, we can interpret the sequence $\{b_i\}_{i=3}^{+\infty}$ be the principal part of $\{a_i\}_{i=3}^{+\infty}$.

5.4. Root test

In many cases, especially for sequences that change sign for infinitely many times, the monotone convergence theorem (Lemma 5.1.5), the integral test (Section 5.3.1) and limit comparison test (Section 5.3.2) are not applicable. At this point, we only introduced alternating series test for changing-sign sequence, which is still quite restrictive in practical applications. There are some other tests for general sequences as well, for example, the Dirichlet test [**Apo74**, Theorem 8.28] and the Abel test [**Apo74**, Theorem 8.29]. We are not going to introduce them during this course. Instead, we will introduce a criteria which is quite fundamental, especially for those interested in complex analysis, see e.g. my lecture note [**Kow23**]. In order to motivate the criteria, lets begin with the following example.

EXAMPLE 5.4.1 (Geometric sequence). Let $r \in \mathbb{R}$, we now consider the sequence $\{a_i\}_{i=0}^{+\infty}$ given by $a_i = ar^i$, which is the well-known *geometric sequence with ratio* r. If $r = 1$, then $a_i = a$ for all $i \in \mathbb{N}$, in which obviously that $\sum_{i=1}^{+\infty} a_i = +\infty$. If $r = -1$, then $\sum_{i=1}^{+\infty} a_i$ diverges as demonstrated as in Example 5.2.7. We now consider the case when $-1 < r < 1$. In this case, for each $N \in \mathbb{N}$, one sees that

$$(1 - |r|) \sum_{i=0}^{N} |a_i| = |a| \sum_{i=0}^{N} (|r|^i - |r|^{i+1}) = |a|(1 - |r|^{N+1}).$$

Since $|r| < 1$, then one has

$$\sum_{i=0}^{N} |a_i| = \frac{|a|(1 - |r|^{N+1})}{1 - |r|},$$

which gives

$$\lim_{N \to +\infty} \sum_{i=0}^{N} |a_i| = \frac{|a|}{1 - |r|} < +\infty,$$

in other words, $\sum_{i=0}^{+\infty} a_i$ absolutely when $-1 < r < 1$. Similar arguments show that

$$(1 - r) \sum_{i=0}^{N} a_i = a \sum_{i=0}^{N} (r^i - r^{i+1}) = a(1 - r^{N+1}),$$

and thus

$$\sum_{i=0}^{N} a_i = \frac{a(1 - r^{N+1})}{1 - r}, \tag{5.4.1}$$

thus we know that the absolutely convergent series $\sum_{i=0}^{+\infty} a_i$ takes the value

$$\sum_{i=0}^{+\infty} a_i = \frac{a}{1 - r}.$$

We now consider the case when $r < -1$ or $r > 1$. In this case, from (5.4.1) it is not difficult to see that

$$\limsup_{N \to +\infty} \sum_{i=0}^{N} a_i = +\infty \neq -\infty = \liminf_{N \to +\infty} \sum_{i=0}^{N} a_i,$$

which demonstrates that $\sum_{i=0}^{+\infty} a_i$ diverges.

In the geometric sequence (Example 5.4.1), one sees that the convergence only depends on the *absolute value of the ratio*. We observe that

$$\lim_{i \to +\infty} |a_i|^{1/i} = \lim_{i \to +\infty} |a|^{1/i} |r| = |r|,$$

which suggests the absolute value of the "ratio" of a *general* sequence $\{a_i\}_{i=1}^{+\infty}$ should be given by $\lim_{i \to +\infty} |a_i|^{1/i}$. In fact, this idea works even the when the limit $\lim_{i \to +\infty} |a_i|^{1/i}$ does not exist:

THEOREM 5.4.2 (Root test, a special case of [**Apo74**, Theorem 8.26]). *Given a sequence* $\{a_i\}_{i=1}^{+\infty}$ *in* $\mathbb{R}$, *and let*

$$\rho := \limsup_{i \to +\infty} |a_i|^{1/i}. \tag{5.4.2}$$

(a) *If* $\rho < 1$, *then* $\sum_{i=1}^{+\infty} |a_i| < +\infty$.
(b) *If* $\rho > 1$, *then either* $\sum_{i=1}^{+\infty} a_i = \pm\infty$ *or* $\sum_{i=1}^{+\infty} a_i$ *diverges.*

REMARK 5.4.3. The case when $\rho = 1$ is inconclusive, which we will give examples in Section 5.5 later.

If we consider the geometric sequence in Example 5.4.1, one also see that

$$\frac{|a_{i+1}|}{|a_i|} = \frac{|ar^{i+1}|}{|a_i r^i|} = |r| \quad \text{for all } i \in \mathbb{N},$$

which suggests another characterization of the "ratio" of a *general* sequence $\{a_i\}_{i=1}^{+\infty}$ should be given by $\lim_{i\to+\infty} \frac{|a_{i+1}|}{|a_i|}$. In fact, one has the following fact:

LEMMA 5.4.4 ([**Apo74**, Exercise 8.4]). *Given a sequence* $\{a_i\}_{i=1}^{+\infty}$ *in* $\mathbb{R}$ *such that* $a_i \neq 0$ *for all* $i \in \mathbb{N}$, *one has*

$$\liminf_{i\to+\infty} \frac{|a_{i+1}|}{|a_i|} \leq \liminf_{i\to+\infty} |a_i|^{1/i} \leq \limsup_{i\to+\infty} |a_i|^{1/i} \leq \limsup_{i\to+\infty} \frac{|a_{i+1}|}{|a_i|}.$$

We now exhibit the following corollary, which gives a widely-used way to check the sufficient condition in the root test (Theorem 5.4.2).

COROLLARY 5.4.5. *Given a sequence* $\{a_i\}_{i=1}^{+\infty}$ *in* $\mathbb{R}$ *such that* $a_i \neq 0$ *for all* $i \in \mathbb{N}$. *If*

$$\lim_{i\to+\infty} \frac{|a_{i+1}|}{|a_i|} \text{ exists in } [0,+\infty], \tag{5.4.3}$$

then $\lim_{i\to+\infty} |a_i|^{1/i}$ *exists and the generalized radius* ρ *in* (5.4.2) *satisfies*

$$\rho = \lim_{i\to+\infty} |a_i|^{1/i} = \lim_{i\to+\infty} \frac{|a_{i+1}|}{|a_i|}.$$

We sometimes refer the "ratio test" if we check the criteria (5.4.3). Here we give an example to demonstrate that the root test is stronger than ratio test.

EXAMPLE 5.4.6. Define a sequence $\{a_i\}_{i=1}^{+\infty}$ by

$$a_i = \begin{cases} 2^{-(i+1)} & \text{if } i \text{ is odd,} \\ 2^{-i} & \text{if } i \text{ is even.} \end{cases}$$

One sees that

$$|a_i|^{1/i} = \begin{cases} 2^{-\frac{i+1}{i}} & \text{if } i \text{ is odd,} \\ 2^{-1} & \text{if } i \text{ is even,} \end{cases}$$

and thus

$$\lim_{i\to+\infty} |a_i|^{1/i} = \frac{1}{2} < 1,$$

and thus by root test (Theorem 5.4.2) one conclude that $\sum_{i=1}^{+\infty} |a_i| < +\infty$. Since for each odd i we have $a_i = a_{i+1} = 2^{-(i+1)}$ for all odd i, and for each even i we have $a_i = 2^{-i}$ and $a_{i+1} = 2^{-(i+2)}$, thus we have

$$\liminf_{i\to+\infty} \frac{|a_{i+1}|}{|a_i|} = \frac{1}{4} < 1 = \limsup_{i\to+\infty} \frac{|a_{i+1}|}{|a_i|},$$

which shows that (5.4.3) does not hold, in other words, the ratio test fails.

5.5. Power series: approximating functions by polynomials

We recall that a polynomial is a function $f : \mathbb{R} \to \mathbb{R}$ for which takes the form $f(x) = \sum_{i=0}^{N} c_i x^i$ for some $N \in \mathbb{N}$. If $c_N \neq 0$, then we say that the polynomial has degree N. For each fixed $x \in \mathbb{R}$, one sees that $\{c_i x^i\}_{i=0}^{+\infty}$ is actually a sequence in $\mathbb{R}$, therefore it is possible to discuss the "polynomial with degree $+\infty$" more precisely, the *power series*. This idea can be done by using root test (Theorem 5.4.2).

THEOREM 5.5.1 (A special case of [**Kow23**, Theorem 2.2.2], see also [**Rud87**]). *Given a sequence $\{c_i\}_{i=0}^{+\infty}$ in $\mathbb{R}$. We define the number*

$$\rho := \limsup_{i\to+\infty} |c_i|^{1/i}.$$

(a) *If $\rho = 0$, then $\sum_{i=0}^{+\infty} |c_i||x|^i < +\infty$ for each $x \in \mathbb{R}$.*
(b) *If $\rho = +\infty$, then $\sum_{i=1}^{+\infty} c_i x^i = \pm\infty$ or $\sum_{i=1}^{+\infty} c_i x^i$ diverges for each $x \in \mathbb{R} \setminus \{0\}$.*
(c) *If $0 < \rho < +\infty$, then $\sum_{i=0}^{+\infty} |c_i||x|^i < +\infty$ for each $x \in \mathbb{R}$ with $|x| < \rho^{-1}$ and $\sum_{i=0}^{+\infty} c_i x^i = \pm\infty$ or $\sum_{i=0}^{+\infty} c_i x^i$ diverges for each $x \in \mathbb{R}$ with $|x| > \rho^{-1}$.*

DEFINITION 5.5.2. If $0 < \rho < +\infty$, we define $R := \rho^{-1}$; If $\rho = 0$, we define $R := +\infty$; If $\rho = +\infty$, we define $R := 0$. In many cases, we simply abuse the notation by simply writing

$$R := \rho^{-1} \quad \text{for } \rho \in [0, +\infty],$$

and the number R is called the *radius of convergence*. If $R = +\infty$, we interpret $B_R = \mathbb{R}$.

We also have the following properties similar to polynomials.

THEOREM 5.5.3 (A special case of [**Kow23**, Theorem 2.2.9]). *Suppose that the power series $\sum_{i=0}^{+\infty} c_i x^i$ has the radius of convergence $R \in (0, +\infty]$, then it defines a differentiable function on $B_R \to \mathbb{R}$. In addition, the radius of convergence of the power series $\sum_{i=1}^{+\infty} c_i i x^{i-1}$ is $\geq R$ and the following identity holds:*

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\sum_{i=0}^{+\infty} c_i x^i\right) = \sum_{i=1}^{+\infty} c_i i x^{i-1} \quad \text{for all } x \in B_R.$$

COROLLARY 5.5.4. *Any power series is infinitely differentiable within its radius of convergence.*

If we look into Theorem 5.5.1(c), we see that the statement do not mention any result on $|x| = \rho^{-1}$. We now explain the situation here by using the following example.

EXAMPLE 5.5.5. We now give examples to demonstrate in the case Theorem 5.5.1(c) that it is inconclusive at $|x| = R$. This also serves as examples mentioned in Remark 5.4.3

(1) By using Example 5.1.4, we see that the power series $\sum_{i=1}^{+\infty} i x^i$ has the radius of convergence

$$R = \left(\lim_{i\to+\infty} i^{1/i}\right)^{-1} = 1.$$

We see that $\sum_{i=1}^{+\infty} i x^i = +\infty$ when $x = 1$ and $\sum_{i=1}^{+\infty} i x^i$ diverges when $x = -1$ (similar to Example 5.2.7).

(2) We also see that the power series $\sum_{i=1}^{+\infty} i^{-2}x^i$ has the radius of convergence

$$R = \left(\lim_{i\to+\infty} i^{1/i}\right)^2 = 1.$$

We see that $\sum_{i=1}^{+\infty} i^{-2}x^i$ converges absolutely at $x = \pm 1$.

(3) Similarly, we also see that the power series $\sum_{i=1}^{+\infty} i^{-1}x^i$ has the radius of convergence

$$R = \lim_{i\to+\infty} i^{1/i} = 1.$$

From Example 5.2.9 we see that $\sum_{i=1}^{+\infty} i^{-1}x^i = +\infty$ at $x = 1$, but from Example 5.2.3 we see that $\sum_{i=1}^{+\infty} i^{-1}x^i$ converges at $x = -1$.

We see that all above three three cases are not distinguishable by the criteria of root test (Theorem 5.4.2) as well as Theorem 5.5.1, but they have different convergence behavior at $|x| = 1$. In other words, the results in the root test (Theorem 5.4.2) as well as Theorem 5.5.1 are already somehow optimal.

However, if one can prove the convergence at $x = R$ or $x = -R$, then we have the following remarkable result.

Theorem 5.5.6 (Abel's limit theorem, [**Apo74**, Theorem 9.31]). *Suppose that the power series $f(x) = \sum_{i=0}^{+\infty} c_i x^i$ has the radius of convergence $R \in (0, +\infty]$. If $\sum_{i=0}^{+\infty} c_i R^i$ converges, then*

$$\lim_{x\to R-} f(x) = \sum_{i=0}^{+\infty} c_i R^i.$$

Similarly, if $\sum_{i=0}^{+\infty} c_i(-R)^i$ converges, then

$$\lim_{x\to -R+} f(x) = \sum_{i=0}^{+\infty} c_i(-R)^i.$$

Let $f : (x_0 - \epsilon, x_0 + \epsilon) \to \mathbb{R}$ be an infinitely differentiable function, with derivatives $f^{(n)} : (x_0 - \epsilon, x_0 + \epsilon) \to \mathbb{R}$ of order n. We now want to approximate f using the power series $\sum_{i=0}^{+\infty} c_i(x - x_0)^i$. If we consider the function $g : (-\epsilon, \epsilon) \to \mathbb{R}$ defined by $g(x) := f(x + x_0)$, then one sees that $g^{(n)}(x) = f^{(n)}(x + x_0)$. If we can approximate g using a power series $\sum_{i=0}^{+\infty} c_i x^i$, then $\sum_{i=0}^{+\infty} c_i(x - x_0)^i$ is an approximation of f.

Before we make things rigorous, let's do some formal computation first. If we formally write

$$f(x) = \sum_{i=0}^{+\infty} c_i(x - x_0)^i. \tag{5.5.1}$$

In view of Corollary 5.5.4, one immediately know that one necessary condition to do so is the function f must infinitely differentiable. We now take $x = x_0$ in (5.5.1) to see that $c_0 = f(x_0)$. In view of Theorem 5.5.3, we formally write

$$f'(x) = \sum_{i=1}^{+\infty} i c_i (x - x_0)^{i-1}.$$

Now taking $x = x_0$ we see that $c_1 = f'(x_0)$. In view of Theorem 5.5.3, we formally write

$$f''(x) = \sum_{i=2}^{+\infty} i(i-1)c_i(x-x_0)^{i-2}.$$

Now taking $x = x_0$ we see that $c_2 = f''(x_0)$. Repeating the above process, one can obtain

$$c_i = \frac{f^{(i)}(x_0)}{i!} \quad \text{for all } i = 0,1,2,\cdots .$$

In other words, if we formally approximate f like (5.5.1), it is necesarily to have the formula

$$f(x) = \sum_{i=0}^{+\infty} \frac{f^{(i)}(x_0)}{i!}(x-x_0)^i. \tag{5.5.2}$$

However, the above ideas fail in general, which can be demonstrated by the following example.

EXAMPLE 5.5.7. We define the continuous function $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} e^{-|x|^{-2}} & , x > 0, \\ 0 & , x \le 0. \end{cases}$$

In fact, by using the mathematical induction, one can verify that $f : \mathbb{R} \to \mathbb{R}$ is infinitely differentiable with i^{th}-order derivatives

$$f^{(i)}(x) = \begin{cases} P_i(x)e^{-|x|^{-2}} & , x > 0, \\ 0 & , x \le 0, \end{cases}$$

for some polynomial $P_i(x)$. This implies that $f^{(n)}(0) = 0$ for all $n = 0,1,2,\cdots$, so that the power series $\sum_{i=0}^{+\infty} \frac{f^{(i)}(0)}{i!}x^i$ has radius of convergent $R = +\infty$ with

$$\sum_{i=0}^{+\infty} \frac{f^{(i)}(0)}{i!}x^i = 0 \quad \text{for all } x \in \mathbb{R}.$$

This shows that the above ideas fail in general even within the radius of convergence.

The above idea is still holds true for some function. Before we state some condition, we need the following definition.

DEFINITION 5.5.8. Let $f : \mathbb{R} \to \mathbb{R}$ be an infinitely differentiable function. We say that f is *real analytic* near $x_0 \in \mathbb{R}$ if there exists $\epsilon > 0$ such that $\sum_{i=0}^{+\infty} \frac{f^{(i)}(x_0)}{i!}(x-x_0)^i$ converges for all $x \in B_\epsilon(x_0)$ and the representation (5.5.2) holds for all $x \in B_\epsilon(x_0)$. If f is real analytic at all points in $\mathbb{R}$, then we say that $f : \mathbb{R} \to \mathbb{R}$ is *real analytic* or *entire*.

We now described the following criteria for which the above ideas work properly:

THEOREM 5.5.9. *Let $f : \mathbb{R} \to \mathbb{R}$ be an infinitely differentiable function. The the following are equivalent:*

(a) *$f : \mathbb{R} \to \mathbb{R}$ is real analytic;*
(b) *For each $M > 0$, there exists a constant $C_M > 0$, which depends on M, such that*

$$\sup_{x \in [-M,M]} |f^{(i)}(x)| \le C_M^{i+1} i! \quad \textit{for all } i = 0,1,2,\cdots . \tag{5.5.3}$$

Here the constant C_M is independent of i.

In this case, for each $x_0 \in \mathbb{R}$*, let* $R = R(x_0) > 0$ *be the radius of convergence of the power series* $\sum_{i=0}^{+\infty} \frac{f^{(i)}(x_0)}{i!}(x-x_0)^i$*, then the power series representation* (5.5.2) *holds for all* $x \in B_R(x_0)$.

Roughly speaking, this means that the above ideas only works when the derivatives $f^{(n)}$ does not grow too fast. Combining Theorem 5.5.9 with Theorem 5.5.3, we immediately reach the following corollary:

COROLLARY 5.5.10. *If* $f : \mathbb{R} \to \mathbb{R}$ *is an analytic function, then its derivative* $f' : \mathbb{R} \to \mathbb{R}$ *is also analytic.*

There are quite a lot examples of analytic function, for example, Bessel functions, and their spherical versions. Here we will only exhibit some basic examples.

EXAMPLE 5.5.11. Let $\exp : \mathbb{R} \to \mathbb{R}$ be the exponential function. For each $M > 0$, it is easy to see that

$$\sup_{x\in[-M,M]} |\exp^{(i)}(x)| \leq \sup_{x\in[-M,M]} |e^x| \leq e^M,$$

which verifies (5.5.3) with $C_M = e^M$. For each $x_0 \in \mathbb{R}$, we see that the radius of convergence of $\sum_{i=0}^{+\infty} \frac{e^{x_0}(x-x_0)^i}{i!}$ is $+\infty$ (the details left as exercises), and therefore Theorem 5.5.2 guarantees that

$$e^x = \sum_{i=0}^{+\infty} \frac{e^{x_0}(x-x_0)^i}{i!} \quad \text{for all } x \in \mathbb{R}.$$

If write $y = x - x_0$, this is exactly equivalent to the well-known formula

$$e^y = \sum_{i=0}^{+\infty} \frac{y^i}{i!} \quad \text{for all } y \in \mathbb{R}.$$

If we take $y = 1$, the transcendental number $e \approx 2.718\cdots$ also can be written as the

$$e = \sum_{i=0}^{+\infty} \frac{1}{i!}.$$

An application of this function is the *moment generating function* in the branch of probability.

EXERCISE 5.5.12. Prove that $\sin : \mathbb{R} \to \mathbb{R}$ and $\cos : \mathbb{R} \to \mathbb{R}$ are real analytic functions, and prove that

$$\sin x = \sum_{k=0}^{+\infty} \frac{(-1)^k x^{1+2k}}{(1+2k)!} \quad \text{for all } x \in \mathbb{R}.$$

By using Corollary 5.5.10, we immediately see that $\cos : \mathbb{R} \to \mathbb{R}$ is real analytic. Use Theorem 5.5.3 to compute the power series of $\cos x$ centered at $x_0 = 0$.

We finally end this section by the following example.

EXAMPLE 5.5.13. We define $f : (-1,1) \to \mathbb{R}$ by $f(x) := \ln(1-x)$. One sees that

$$f^{(i)}(x) = (-1)^{i+1}(i-1)!(x-1)^{-i} \quad \text{for all } i = 1,2,\cdots,$$

and for each $0 < M < 1$ one has

$$\sup_{-M\leq x\leq M} |f^{(i)}(x)| = (1-M)^{-i}(i-1)! \quad \text{for all } i = 1,2,\cdots$$

and a generalized version of Theorem 5.5.9 shows that $f : (-1, 1) \to \mathbb{R}$ is real analytic, and one has the power series representation

$$\ln(1-x) = \sum_{i=0}^{+\infty} \frac{f^{(i)}(0)}{i!} x^i = -\sum_{i=1}^{+\infty} \frac{(i-1)!}{i!} x^i = -\sum_{i=1}^{+\infty} \frac{x^i}{i} \quad \text{for all } x \in (-1, 1),$$

since $f(0) = 0$ and $f^{(i)}(0) = (-1)^{i+1}(i-1)!(-1)^{-i} = -(i-1)!$ for all $i = 1, 2, 3, \cdots$. Since the power series $\sum_{i=1}^{+\infty} \frac{x^i}{i}$ converges at $x = -1$ (Example 5.2.3), then the Abel's limit theorem (Theorem 5.5.6) guarantees that

$$\ln 2 = \lim_{x \to -1+} \ln(1-x) = -\sum_{i=1}^{+\infty} \frac{(-1)^i}{i},$$

which concludes that

$$\sum_{i=1}^{+\infty} \frac{(-1)^i}{i} = -\ln 2.$$

5.6. Some fundamental inequalities

We now introduce some fundamental inequalities for series which are analogue to Section 4.5.

DEFINITION 5.6.1. For each sequence $\{a_i\}_{i=1}^{+\infty}$, we write

$$\|\{a_i\}_{i=1}^{+\infty}\|_{\ell^p} := \left(\sum_{i=1}^{+\infty} |a_i|^p \right)^{1/p} \qquad \text{for each } 0 < p < +\infty,$$
$$\|\{a_i\}_{i=1}^{+\infty}\|_{\ell^\infty} := \sup_{i \in \mathbb{N}} |a_i|.$$

We say that $\{a_i\}_{i=1}^{+\infty} \in \ell^p$ if $\|\{a_i\}_{i=1}^{+\infty}\|_{\ell^p} < +\infty$.

We now show the following result, which is analogue to the Hölder's inequality for integral (Theorem 4.5.3).

THEOREM 5.6.2 (Hölder's inequality for series). *For each $p \geq 1$ and $q \geq 1$ with $\frac{1}{p} + \frac{1}{q} = 1$, one has*

$$\|\{a_i b_i\}_{i=1}^{+\infty}\|_{\ell^1} \leq \|\{a_i\}_{i=1}^{+\infty}\|_{\ell^p} \|\{b_i\}_{i=1}^{+\infty}\|_{\ell^q}$$

for all $\{a_i\}_{i=1}^{+\infty} \in \ell^p$ and $\{b_i\}_{i=1}^{+\infty} \in \ell^q$.

PROOF. The case when $(p, q) = (1, +\infty)$ or $(p, q) = (+\infty, 1)$ are easy, we left the details for readers as an exercise. We now consider the case when $p > 1$ (if and only if $q > 1$). The result is trivial when either $\|\{a_i\}_{i=1}^{+\infty}\|_{\ell^p} = 0$ or $\|\{b_i\}_{i=1}^{+\infty}\|_{\ell^q} = 0$, we again left the details for readers as an exercise.

If $\|\{a_i\}_{i=1}^{+\infty}\|_{\ell^p} \neq 0$ and $\|\{b_i\}_{i=1}^{+\infty}\|_{\ell^q} \neq 0$, then for each i we can choose

$$s_1 = \frac{|a_i|}{\|\{a_i\}_{i=1}^{+\infty}\|_{\ell^p}} \quad \text{and} \quad s_2 = \frac{|b_i|}{\|\{b_i\}_{i=1}^{+\infty}\|_{\ell^q}}$$

in the Young's inequality (Lemma 4.5.2) to reach

$$\frac{|a_i b_i|}{\|\{a_i\}_{i=1}^{+\infty}\|_{\ell^p} \|\{b_i\}_{i=1}^{+\infty}\|_{\ell^q}} \leq \frac{1}{p} \frac{|a_i|^p}{\|\{a_i\}_{i=1}^{+\infty}\|_{\ell^p}^p} + \frac{1}{q} \frac{|b_i|^q}{\|\{b_i\}_{i=1}^{+\infty}\|_{\ell^q}}.$$

For each $N \in \mathbb{N}$, we sum the above inequality from $i = 1$ to $i = N$ to reach

$$\frac{\sum_{i=1}^N |a_i b_i|}{\|\{a_i\}_{i=1}^{+\infty}\|_{\ell^p} \|\{b_i\}_{i=1}^{+\infty}\|_{\ell^q}} \le \frac{1}{p} \frac{\sum_{i=1}^N |a_i|^p}{\|\{a_i\}_{i=1}^{+\infty}\|_{\ell^p}^p} + \frac{1}{q} \frac{\sum_{i=1}^N |b_i|^q}{\|\{b_i\}_{i=1}^{+\infty}\|_{\ell^q}} \le \frac{1}{p} + \frac{1}{q} = 1,$$

because

$$\sum_{i=1}^N |a_i|^p \le \|\{a_i\}_{i=1}^{+\infty}\|_{\ell^p}^p \quad \text{and} \quad \sum_{i=1}^N |b_i|^q \le \|\{b_i\}_{i=1}^{+\infty}\|_{\ell^q}.$$

Thus, we have

$$\sum_{i=1}^N |a_i b_i| \le \|\{a_i\}_{i=1}^{+\infty}\|_{\ell^p} \|\{b_i\}_{i=1}^{+\infty}\|_{\ell^q} \quad \text{for all } N \in \mathbb{N}.$$

Since $\{\sum_{i=1}^N |a_i b_i|\}_{N=1}^{+\infty}$ is a nondecreasing sequence in $\mathbb{R}$, by using the monotone convergence theorem (Lemma 5.1.5) one sees that the limit $\lim_{N\to+\infty} \sum_{i=1}^N |a_i b_i|$ exists, and hence we conclude our result by taking the limit $N \to +\infty$ in the above inequality. $\square$

REMARK 5.6.3. Similar to Remark 4.5.4, when $p > 1$ (if and only if $q > 1$), we can check the equality holds if one chooses $b_n = |a_n|^{\frac{p}{q}}$.

We also can obtain the following result, which is analogue to the Minkowski's inequaltiy for integral (Theorem 4.5.5).

THEOREM 5.6.4 (Minkowski's inequality for series). *If $1 \le p \le +\infty$, then one has*

$$\|\{a_i + b_i\}_{i=1}^{+\infty}\|_{\ell^p} \le \|\{a_i\}_{i=1}^{+\infty}\|_{\ell^p} + \|\{b_i\}_{i=1}^{+\infty}\|_{\ell^p}$$

for all $\{a_i\}_{i=1}^{+\infty} \in \ell^p$ and $\{b_i\}_{i=1}^{+\infty} \in \ell^p$.

PROOF. Then case $p = 1$ and the case $p = +\infty$ are easy, we left the details for readers for an exercise. We now consider the case when $1 < p < +\infty$. We write

$$|a_i + b_i|^p = |a_i + b_i|^{p-1} |a_i + b_i| \le |a_i + b_i|^{p-1} |a_i| + |a_i + b_i|^{p-1} |b_i|.$$

Let $q > 1$ with $\frac{1}{p} + \frac{1}{q} = 1$. By using the Hölder's inequality (Theorem 5.6.2), for each $N \in \mathbb{N}$ we see that

$$\begin{aligned}\sum_{i=1}^N |a_i + b_i|^{p-1} |a_i| &\le \left(\sum_{i=1}^N |a_i + b_i|^{q(p-1)} \right)^{1/q} \|\{a_i\}_{i=1}^{+\infty}\|_{\ell^p} \\ &= \left(\sum_{i=1}^N |a_i + b_i|^p \right)^{1/q} \|\{a_i\}_{i=1}^{+\infty}\|_{\ell^p},\end{aligned}$$

and similarly,

$$\sum_{i=1}^N |a_i + b_i|^{p-1} |a_i| \le \left(\sum_{i=1}^N |a_i + b_i|^p \right)^{\frac{1}{q}} \|\{b_i\}_{i=1}^{+\infty}\|_{\ell^p}.$$

Combining all the equations above, we now see that

$$\sum_{i=1}^N |a_i + b_i|^p \le \left(\sum_{i=1}^N |a_i + b_i|^p \right)^{\frac{1}{q}} \left(\|\{a_i\}_{i=1}^{+\infty}\|_{\ell^p} + \|\{b_i\}_{i=1}^{+\infty}\|_{\ell^p} \right),$$

which gives

$$\left(\sum_{i=1}^{N}|a_i+b_i|^p\right)^{\frac{1}{p}} = \left(\sum_{i=1}^{N}|a_i+b_i|^p\right)^{1-\frac{1}{q}} \leq \|\{a_i\}_{i=1}^{+\infty}\|_{\ell^p} + \|\{b_i\}_{i=1}^{+\infty}\|_{\ell^p}.$$

Since $\left\{\left(\sum_{i=1}^{N}|a_i+b_i|^p\right)^{\frac{1}{p}}\right\}_{N=1}^{+\infty}$ is a nondecreasing sequence in $\mathbb{R}$, by using the monotone convergence theorem (Lemma 5.1.5) one sees that the limit $\lim_{N\to+\infty}\left(\sum_{i=1}^{N}|a_i+b_i|^p\right)^{\frac{1}{p}}$ exists, and hence we conclude our result by taking the limit $N\to+\infty$ in the above inequality. □

The case when $0<p<1$ also can be discussed similar as in Theorem 4.5.6.

THEOREM 5.6.5 (Minkowski's inequality for series). *If $0<p<1$, then*

$$\|\{a_i+b_i\}_{i=1}^{+\infty}\|_{\ell^p}^p \leq \|\{a_i\}_{i=1}^{+\infty}\|_{\ell^p}^p + \|\{b_i\}_{i=1}^{+\infty}\|_{\ell^p}^p$$

for all $\{a_i\}_{i=1}^{+\infty}\in\ell^p$ and $\{b_i\}_{i=1}^{+\infty}\in\ell^p$.

PROOF. By choosing $a=|a_i|$ and $b=|b_i|$ in (4.5.2), we reach

$$|a_i+b_i|^p \leq (|a_i|+|b_i|)^p \leq |a_i|^p+|b_i|^p \quad \text{for all } i\in\mathbb{N}.$$

Since $\left\{\sum_{i=1}^{N}|a_i+b_i|^p\right\}_{N=1}^{+\infty}$ is a nondecreasing sequence in $\mathbb{R}$, by using the monotone convergence theorem (Lemma 5.1.5) one sees that the limit $\lim_{N\to+\infty}\sum_{i=1}^{N}|a_i+b_i|^p$ exists. Thus acting the operator $\sum_{i=1}^{+\infty}$ on the above inequality gives our desired theorem. □

5.7. A quick introduction of Fourier series

A Fourier series is an expansion of a periodic function into a sum of trigonometric functions. The Fourier series is an example of a trigonometric series, but not all trigonometric series are Fourier series. The Fourier series has many such applications in electrical engineering, vibration analysis, acoustics, optics, signal processing, image processing etc. Here we will only give a very rough introduction, one can see e.g. my lecture note [**Kow22**] for more details, via a modern approach, which requires Lebesgue integral. Here we will only exhibit some results in terms of Riemann integral.

We now define the normalized L^2-inner product on $(-\pi,\pi)$ by

$$(f,g)_{L^2(-\pi,\pi)} := \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)g(x)\,\mathrm{d}x, \tag{5.7.1}$$

and one we define the normalized L^2-norm on $(-\pi,\pi)$ by

$$\|f\|_{L^2(-\pi,\pi)} := (f,f)_{L^2(-\pi,\pi)}^{\frac{1}{2}} := \left(\frac{1}{\pi}\int_{-\pi}^{\pi}|f(x)|^2\,\mathrm{d}x\right)^{\frac{1}{2}}. \tag{5.7.2}$$

In view of the Minkowski's inequality for integral (Theorem 4.5.5), one can interpret $\|f\|_{L^2(-\pi,\pi)}$ be the "length" of the function f. Similar to the Euclidean space, we also can interpret

$$\cos^{-1}\left(\frac{f}{\|f\|_{L^2(-\pi,\pi)}},\frac{g}{\|g\|_{L^2(-\pi,\pi)}}\right)_{L^2(-\pi,\pi)}$$

be the "angle" between functions f and g. We say that f is perpendicular to g if $(f,g)_{L^2(-\pi,\pi)}=0$, or we simply denoted by $f\perp g$.

By using the product-to-sum formula (Example 1.3.19), for each $n \in \mathbb{N}$ and $m \in \mathbb{N}$, one sees that

$$\begin{aligned}
&(\sin(n\cdot), \sin(m\cdot))_{L^2(-\pi,\pi)} = \frac{1}{\pi}\int_{-\pi}^{\pi} \sin(nx)\sin(mx)\,\mathrm{d}x \\
&\quad = \frac{1}{2\pi}\int_{-\pi}^{\pi} (\cos((n-m)x) - \cos((n+m)x))\,\mathrm{d}x = \frac{1}{2\pi}\int_{-\pi}^{\pi} \cos((n-m)x)\,\mathrm{d}x \\
&\quad = \begin{cases} \frac{1}{2\pi}\int_{-\pi}^{\pi} 1\,\mathrm{d}x & \text{when } n = m \\ \frac{1}{2\pi}\frac{1}{n-m}\sin((n-m)x)\Big|_{x=-\pi}^{x=\pi} & \text{when } n \neq m \end{cases} \\
&= \begin{cases} 1 & \text{when } n = m, \\ 0 & \text{when } n \neq m. \end{cases}
\end{aligned}$$

Similarly, for each $n \in \mathbb{N} \cup \{0\}$ and $m \in \mathbb{N} \cup \{0\}$, one sees that

$$\begin{aligned}
&(\cos(n\cdot), \cos(m\cdot))_{L^2(-\pi,\pi)} = \frac{1}{\pi}\int_{-\pi}^{\pi} \cos(nx)\cos(mx)\,\mathrm{d}x \\
&\quad = \frac{1}{2\pi}\int_{-\pi}^{\pi} (\cos((n-m)x) + \cos((n+m)x))\,\mathrm{d}x = \frac{1}{2\pi}\int_{-\pi}^{\pi} \cos((n-m)x)\,\mathrm{d}x \\
&= \begin{cases} 1 & \text{when } n = m \\ 0 & \text{when } n \neq m. \end{cases}
\end{aligned}$$

For each $n \in \mathbb{N} \cup \{0\}$ and $m \in \mathbb{N}$, we also see that

$$\begin{aligned}
(\cos(n\cdot), \sin(m\cdot))_{L^2(-\pi,\pi)} &= \frac{1}{\pi}\int_{-\pi}^{\pi} \cos(nx)\sin(mx)\,\mathrm{d}x \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} (\sin((n+m)x) - \sin((n-m)x))\,\mathrm{d}x = 0.
\end{aligned}$$

We now can summarize the above in the following lemma.

LEMMA 5.7.1. *The set $\{\cos(n\cdot)\}_{n=0}^{+\infty} \cup \{\sin(m\cdot)\}_{m=1}^{+\infty}$ forms an orthonormal set with respect to the normalized inner product (5.7.1). In other words, the elements in $\{\cos(n\cdot)\}_{n=0}^{+\infty} \cup \{\sin(m\cdot)\}_{m=1}^{+\infty}$ are perpendicular to each other, and each of them has length 1 with respect to the norm (5.7.2).*

Let's do some formula computations in order to motivate the Fourier series. Suppose that a function $f : (-\pi, \pi) \to \mathbb{R}$ takes the form

$$f(x) = \frac{1}{2}b_0 + \sum_{n=1}^{+\infty} b_n \cos(nx) + \sum_{m=1}^{+\infty} a_m \sin(nx) \quad \text{for all } x \in (-\pi, \pi). \tag{5.7.3a}$$

In view of Lemma 5.7.1, one has

$$\begin{aligned}
\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,\mathrm{d}x &= (f,1)_{L^2(-\pi,\pi)} = \left(\frac{1}{2}b_0, 1\right)_{L^2(-\pi,\pi)} = b_0, \\
\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(nx)\,\mathrm{d}x &= (f,\cos(n\cdot))_{L^2(-\pi,\pi)} = (b_n\cos(n\cdot),\cos(n\cdot))_{L^2(-\pi,\pi)} = b_n, \\
\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(mx)\,\mathrm{d}x &= (f,\sin(m\cdot))_{L^2(-\pi,\pi)} = (a_n\sin(m\cdot),\sin(m\cdot))_{L^2(-\pi,\pi)} = a_n.
\end{aligned}$$

The above three equations can be summarized as

(5.7.3b) $$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(nx)\,\mathrm{d}x \quad \text{for all } n \in \mathbb{N}\cup\{0\},$$

(5.7.3c) $$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(mx)\,\mathrm{d}x \quad \text{for all } m \in \mathbb{N}.$$

DEFINITION 5.7.2. The series (5.7.3a) with coefficients given in (5.7.3b) and (5.7.3c) is called the *Fourier series* of f on $(-\pi,\pi)$.

In order to study its convergence, we again consider the partial sum

$$S_N(x) := \frac{1}{2}b_0 + \sum_{n=1}^{N} b_n\cos(nx) + \sum_{m=1}^{N} a_m\sin(nx) \quad \text{for all } x \in (-\pi,\pi).$$

By using Lemma 5.7.1, one sees that

$$(f,S_N)_{L^2(-\pi,\pi)} = (S_N,S_N)_{L^2(-\pi,\pi)} = \|S_N\|_{L^2(-\pi,\pi)}^2.$$

Now the Hölder's inequality for integral (Theorem 4.5.3) implies that

$$\|S_N\|_{L^2(-\pi,\pi)}^2 \le \|f\|_{L^2(-\pi,\pi)}\|S_N\|_{L^2(-\pi,\pi)} \le \frac{1}{2}\|f\|_{L^2(-\pi,\pi)}^2 + \frac{1}{2}\|S_N\|_{L^2(-\pi,\pi)}^2,$$

and hence $\|S_N\|_{L^2(-\pi,\pi)} \le \|f\|_{L^2(-\pi,\pi)}$ for all $N \in \mathbb{N}$. By using Lemma 5.7.1, one also sees that

$$\|S_N\|_{L^2(-\pi,\pi)}^2 = \frac{b_0^2}{2} + \sum_{n=1}^{N}|b_n|^2 + \sum_{n=1}^{N}|a_n|^2,$$

and we now reach

$$\frac{b_0^2}{2} + \sum_{n=1}^{N}|b_n|^2 + \sum_{n=1}^{N}|a_n|^2 \le \|f\|_{L^2(-\pi,\pi)}^2 \quad \text{for all } N \in \mathbb{N}.$$

Since both $\{\sum_{n=1}^{N}|b_n|^2\}_{N=1}^{+\infty}$ and $\{\sum_{n=1}^{N}|a_n|^2\}_{N=1}^{+\infty}$ are nondecreasing sequences, then by monotone convergence theorem (Theorem 5.1.5), the limits $\lim_{N\to+\infty}\sum_{n=1}^{N}|b_n|^2$ and $\lim_{N\to+\infty}\sum_{n=1}^{N}|a_n|^2$ both exists. Now taking the limit $N \to +\infty$, we reach the following lemma.

LEMMA 5.7.3 (Bessel's inequality). *Suppose that $f : [-\pi,\pi] \to \mathbb{R}$ is continuous except for finitely many points, then*

$$\frac{b_0^2}{2} + \|\{b_n\}_{n=1}^{+\infty}\|_{\ell^2}^2 + \|\{a_n\}_{n=1}^{+\infty}\|_{\ell^2}^2 \le \|f\|_{L^2(-\pi,\pi)}^2,$$

where a_n and b_n are Fourier coefficients given in (5.7.3b) and (5.7.3c).

We now state a sufficient condition to guarantee the pointwise convergence of Fourier series, see e.g. [**Kow22**, Theorem 1.4.1] or [**Str08**, Theorem 5.4.4∞].

THEOREM 5.7.4. *If $f : [-\pi, \pi] \to \mathbb{R}$ is continuously differentiable except for finitely many points, then one has*

$$\frac{1}{2}\left(\lim_{y\to x-} f(y) + \lim_{y\to x+} f(y)\right) = \frac{1}{2}b_0 + \sum_{n=1}^{+\infty} b_n \cos(nx) + \sum_{m=1}^{+\infty} a_m \sin(nx) \quad \textit{for all } x \in (-\pi, \pi)$$

with Fourier coefficients given in (5.7.3b) *and* (5.7.3c)*.*

REMARK 5.7.5. If f is continuous at x, then

$$f(x) = \frac{1}{2}\left(\lim_{y\to x-} f(y) + \lim_{y\to x+} f(y)\right).$$

Despite the mathematical theorems looks beautiful, but actually the convergence is not good. This is quite make sense, since we are attempting to use functions which are oscillating to fit arbitrary (nonoscillating) functions. In mathematical terms, despite the partial sum S_N converges to f pointwisely to the piecewise C^1 function f, the partial sum S_N produces large peaks around the jump of f, which overshoot and undershoot the function's actual values. This approximation error approaches a limit of about 9%. This phenomenon is called the Gibbs-Wilbraham phenomenon, see e.g. [**Kow22**, Theorem 1.4.2] or [**HH79**, Theorem F] for a precise statement. This explains why the signal is noisy if we do not do any further treatment.

EXERCISE 5.7.6. Compute the Fourier series of $f(x) = 1$ on $(-\pi, \pi)$.

EXERCISE 5.7.7. Compute the Fourier series of $f(x) = x$ on $(-\pi, \pi)$.

CHAPTER 6

Multivariable calculus

6.1. Euclidean space $\mathbb{R}^n$

DEFINITION 6.1.1. Let $n \geq 1$ be an integer. We define the n-dimensional Euclidean space by

$$\mathbb{R}^n := \{(x_1, \cdots, x_n) : x_i \in \mathbb{R} \text{ for all } i = 1, \cdots, n\}.$$

The elements in $\mathbb{R}^n$ usually denoted by a single bold-face letter, for example $\boldsymbol{x} = (x_1, \cdots, x_n) \in \mathbb{R}^n$ as well as $\boldsymbol{y} = (y_1, \cdots, y_n) \in \mathbb{R}^n$.

We usually identify the 1-dimensional Euclidean space $\mathbb{R}^1$ with $\mathbb{R}$. The 2-dimensional Euclidean space $\mathbb{R}^2$ can be understood as a "plane". We now introduce basic algebraic operators on n-dimensional points:

DEFINITION 6.1.2. Let $\boldsymbol{x} = (x_1, \cdots, x_n) \in \mathbb{R}^n$ as well as $\boldsymbol{y} = (y_1, \cdots, y_n) \in \mathbb{R}^n$.

(a) We denote $\boldsymbol{x} = \boldsymbol{y}$ when $x_i = y_i$ for all $i = 1, \cdots, n$.
(b) For each $a, b \in \mathbb{R}$, the *linear combination* $a\boldsymbol{x} + b\boldsymbol{y}$ is defined as the vector $(ax_1 + by_1, \cdots, ax_n + by_n)$. We simply denote $(-1)\boldsymbol{x} = -\boldsymbol{x}$.
(c) We denote $\mathbf{0} = (0, \cdots, 0)$ be the *origin* or *zero vector*.
(d) We denote the Euclidean norm

$$|\boldsymbol{x}| := \left(\sum_{i=1}^{n} |x_i|^2 \right)^{1/2}.$$

REMARK 6.1.3. When $n = 1$, we see that the Euclidean norm is exactly same as the absolute value function. If we consider the sequence $\{a_i\}_{i=1}^{+\infty}$ with

$$a_i = x_i \text{ for all } i = 1, \cdots, n, \quad a_i = x_i \text{ for all } i > n,$$

then one sees that the Euclidean norm can be expressed as

$$|\boldsymbol{x}| = \|\{a_i\}_{i=1}^{+\infty}\|_{\ell^2}.$$

Therefore the space ℓ^2 defined in Definition 5.6.1 can be regarded as "infinite-dimensional Euclidean space".

DEFINITION 6.1.4. Let $\boldsymbol{x} = (x_1, \cdots, x_n) \in \mathbb{R}^n$ as well as $\boldsymbol{y} = (y_1, \cdots, y_n) \in \mathbb{R}^n$.

(a) The *inner product* or *dot product* $\boldsymbol{x} \cdot \boldsymbol{y}$ is defined as a scalar $\sum_{i=1}^{n} x_i y_i$. The angle θ between $\boldsymbol{x}$ and $\boldsymbol{y}$ is defined by

$$\theta := \cos^{-1} \left(\frac{\boldsymbol{x}}{|\boldsymbol{x}|} \cdot \frac{\boldsymbol{y}}{|\boldsymbol{y}|} \right), \tag{6.1.1}$$

where $\cos^{-1} : [-1, 1] \to [0, \pi]$ is the usual inverse cosine function.
(b) The outer product or juxtaposition $\boldsymbol{x} \otimes \boldsymbol{y}$ is defined as a $n \times n$ matrix, with entries $(\boldsymbol{x} \otimes \boldsymbol{y})_{ij} = x_i y_j$ for all $i = 1, \cdots, n$ and $j = 1, \cdots, n$.

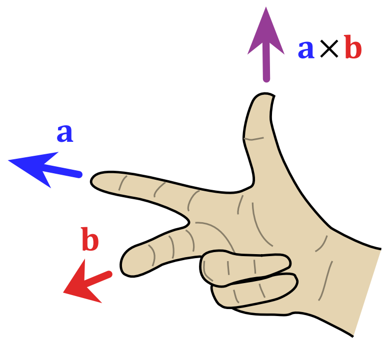

FIGURE 6.1.1. Right-hand rule for cross product: Acdx, CC BY-SA 3.0, via Wikimedia Commons

(c) When $n = 3$, the cross product $\boldsymbol{x} \times \boldsymbol{y}$ is defined by the formula

$$\boldsymbol{x} \times \boldsymbol{y} := (|\boldsymbol{x}||\boldsymbol{y}| \sin\theta)\boldsymbol{n},$$

where θ is the angle between $\boldsymbol{x}$ and $\boldsymbol{y}$ given by (6.1.1) and $\boldsymbol{n}$is a unit vector perpendicular to the plane containing $\boldsymbol{x}$ and $\boldsymbol{y}$, with direction as indicated in Figure 6.1.1. In fact, if we denote $\boldsymbol{i} = (1,0,0)$, $\boldsymbol{j} = (0,1,0)$ and $\boldsymbol{k} = (0,0,1)$, the cross product can be expressed as

$$\boldsymbol{x} \times \boldsymbol{y} = (x_2y_3 - x_3y_2)\boldsymbol{i} + (x_3y_1 - x_1y_3)\boldsymbol{j} + (x_1y_2 - x_2y_1)\boldsymbol{k},$$

or sometimes we abuse the notation by writing

$$\boldsymbol{x} \times \boldsymbol{y} = \det \begin{pmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{pmatrix}.$$

REMARK 6.1.5. The idea behind the Fourier series (Section 5.7) comes from the Euclidean space.

6.2. Limits and continuity

Some notions in Section 2.1 and Section 2.3 can be easily extended to higher dimensional case. Let's us walk through the details here. For later convenience, let's us introduce the following topological notion.

DEFINITION 6.2.1. A subset $\Omega \subset \mathbb{R}^n$is said to be open if for each $x \in \Omega$ there exists $\epsilon > 0$ such that $B_\epsilon(x) \subset \Omega$. Here and after, the open ball $B_R(x)$ is defined by

$$B_R(x) := \{y \in \mathbb{R}^n : |x - y| < \epsilon\}.$$

DEFINITION 6.2.2. Let Ω be an open set in $\mathbb{R}^n$ with $\boldsymbol{x}_0 \in \Omega$ and we consider a function $f : \Omega \setminus \{\boldsymbol{x}_0\} \to \mathbb{R}$. We say that the limit $\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} f(\boldsymbol{x}) = L \in \mathbb{R}$ exists if: Given any $\epsilon > 0$, there exists $\delta = \delta(\epsilon) > 0$ such that

$$0 < |\boldsymbol{x} - \boldsymbol{x}_0| < \delta \text{ implies } |f(x) - L| < \epsilon.$$

We say that the $\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} f(\boldsymbol{x}) = +\infty$ exists if: Given any $M > 0$, there exists $\delta = \delta(\epsilon) > 0$ such that

$$0 < |\boldsymbol{x} - \boldsymbol{x}_0| < \delta \text{ implies } f(x) \geq M.$$

Similarly, we say that the $\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} f(\boldsymbol{x}) = -\infty$ exists if: Given any $M > 0$, there exists $\delta = \delta(\epsilon) > 0$ such that

$$0 < |\boldsymbol{x} - \boldsymbol{x}_0| < \delta \text{ implies } f(x) \leq -M.$$

We also unify the above notions by saying that $\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} f(\boldsymbol{x})$ exists in $[-\infty, +\infty]$.

One also has similar properties as in Lemma 2.1.4:

LEMMA 6.2.3. *Let Ω be an open set in $\mathbb{R}^n$ with $\boldsymbol{x}_0 \in \Omega$ and we consider functions $g_1 : \Omega \setminus \{\boldsymbol{x}_0\} \to \mathbb{R}$ and $g_2 : \Omega \setminus \{\boldsymbol{x}_0\} \to \mathbb{R}$. If both limits $\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} g_1(\boldsymbol{x})$ and $\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} g_2(\boldsymbol{x})$ exist in $\mathbb{R}$, then the following holds true:*

(a) *for each $c_1 \in \mathbb{R}$ and $c_2 \in \mathbb{R}$the limit $\lim_{\boldsymbol{x}\to\boldsymbol{x}_0}(c_1 g_1(\boldsymbol{x}) + c_2 g_2(\boldsymbol{x}))$ exists in $\mathbb{R}$ and satisfies*
$$\lim_{\boldsymbol{x}\to\boldsymbol{x}_0}(c_1 g_1(\boldsymbol{x}) + c_2 g_2(\boldsymbol{x})) = c_1 \lim_{\boldsymbol{x}\to\boldsymbol{x}_0} g_1(\boldsymbol{x}) + c_2 \lim_{\boldsymbol{x}\to\boldsymbol{x}_0} g_2(\boldsymbol{x}) \quad \textit{(linearity)}.$$

(b) *if $g_1(\boldsymbol{x}) \leq g_2(\boldsymbol{x})$ for all $\boldsymbol{x} \in B_\epsilon(\boldsymbol{x}_0)$ for some $\epsilon > 0$, then*
$$\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} g_1(\boldsymbol{x}) \leq \lim_{\boldsymbol{x}\to\boldsymbol{x}_0} g_2(\boldsymbol{x}) \quad \textit{(monotonicity)}.$$

(c) *the limit $\lim_{\boldsymbol{x}\to\boldsymbol{x}_0}(g_1(\boldsymbol{x})g_2(\boldsymbol{x}))$ exists in $\mathbb{R}$ and satisfies*
$$\lim_{\boldsymbol{x}\to\boldsymbol{x}_0}(g_1(\boldsymbol{x})g_2(\boldsymbol{x})) = \left(\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} g_1(\boldsymbol{x})\right)\left(\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} g_2(\boldsymbol{x})\right).$$

(d) *if we additionally assume that $\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} g_2(\boldsymbol{x}) \neq 0$, then the limit $\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} \frac{g_1(\boldsymbol{x})}{g_2(\boldsymbol{x})}$ exists in $\mathbb{R}$ and satisfies*
$$\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} \frac{g_1(\boldsymbol{x})}{g_2(\boldsymbol{x})} = \frac{\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} g_1(\boldsymbol{x})}{\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} g_2(\boldsymbol{x})}.$$

There is no natural generalization for left/right limits. However, one still have natural generalization for limit superior and limit inferior.

DEFINITION 6.2.4. Let Ω be an open set in $\mathbb{R}^n$ with $\boldsymbol{x}_0 \in \Omega$ and let $f : \Omega \setminus \{\boldsymbol{x}_0\} \to \mathbb{R}$ be a function. We define the *limit superior/upper limit* and the *limit inferior/lower limit* by

$$\limsup_{\boldsymbol{x}\to\boldsymbol{x}_0} f(x) := \lim_{r\to 0+}\left(\sup_{B_r(\boldsymbol{x}_0)\setminus\{\boldsymbol{x}_0\}} f\right), \quad \liminf_{\boldsymbol{x}\to\boldsymbol{x}_0} f(x) := \lim_{r\to 0+}\left(\inf_{B_r(\boldsymbol{x}_0)\setminus\{\boldsymbol{x}_0\}} f\right).$$

One can check whether the limit exists or not by using the following theorem.

THEOREM 6.2.5. *Let Ω be an open set in $\mathbb{R}^n$ with $\boldsymbol{x}_0 \in \Omega$ and let $f : \Omega \setminus \{\boldsymbol{x}_0\} \to \mathbb{R}$ be a function.*

(a) *If $\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} f(\boldsymbol{x})$ exists in $[-\infty, +\infty]$, then*
$$\limsup_{\boldsymbol{x}\to\boldsymbol{x}_0} f(x) = \liminf_{\boldsymbol{x}\to\boldsymbol{x}_0} f(x) = \lim_{\boldsymbol{x}\to\boldsymbol{x}_0} f(\boldsymbol{x}). \tag{6.2.1}$$
In other words, if $\limsup_{\boldsymbol{x}\to\boldsymbol{x}_0} f(x) \neq \liminf_{\boldsymbol{x}\to\boldsymbol{x}_0} f(x)$, then the limit $\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} f(\boldsymbol{x})$ does not exist.

(b) *If $\limsup_{\boldsymbol{x}\to\boldsymbol{x}_0} f(x) \neq \liminf_{\boldsymbol{x}\to\boldsymbol{x}_0} f(x)$, then the limit $\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} f(\boldsymbol{x})$ exists in $[-\infty, +\infty]$ and (6.2.1).*

Using the same arguments as in Proposition 2.3.9, one can show the following proposition.

PROPOSITION 6.2.6. *Let Ω be an open set in $\mathbb{R}^n$ with $\boldsymbol{x}_0 \in \Omega$ and we consider functions $g_1 : \Omega \setminus \{\boldsymbol{x}_0\} \to \mathbb{R}$ and $g_2 : \Omega \setminus \{\boldsymbol{x}_0\} \to \mathbb{R}$.*

(a) *The limit superior satisfies the subadditivity property:*

$$\limsup_{\boldsymbol{x}\to\boldsymbol{x}_0} (g_1(\boldsymbol{x}) + g_2(\boldsymbol{x})) \leq \limsup_{\boldsymbol{x}\to\boldsymbol{x}_0} g_1(\boldsymbol{x}) + \limsup_{\boldsymbol{x}\to\boldsymbol{x}_0} g_2(\boldsymbol{x}), \tag{6.2.2}$$

(b) *The limit inferior satisfies the superadditivity property:*

$$\liminf_{\boldsymbol{x}\to\boldsymbol{x}_0} (g_1(\boldsymbol{x}) + g_2(\boldsymbol{x})) \geq \liminf_{\boldsymbol{x}\to\boldsymbol{x}_0} g_1(\boldsymbol{x}) + \liminf_{\boldsymbol{x}\to\boldsymbol{x}_0} g_2(\boldsymbol{x}), \tag{6.2.3}$$

(c) *Both limit superior and limit inferior satisfy the monotonicity property: If there exists $\epsilon > 0$ such that $g_1(\boldsymbol{x}) \leq g_2(\boldsymbol{x})$ for all $\boldsymbol{x} \in B_\epsilon(\boldsymbol{x}_0) \setminus \{\boldsymbol{x}_0\}$, then*

$$\limsup_{\boldsymbol{x}\to\boldsymbol{x}_0} g_1(\boldsymbol{x}) \leq \limsup_{\boldsymbol{x}\to\boldsymbol{x}_0} g_2(\boldsymbol{x}), \quad \liminf_{\boldsymbol{x}\to\boldsymbol{x}_0} g_1(\boldsymbol{x}) \leq \liminf_{\boldsymbol{x}\to\boldsymbol{x}_0} g_2(\boldsymbol{x}).$$

REMARK 6.2.7. Similar to (2.3.4), the inequality

$$\limsup_{\boldsymbol{x}\to\boldsymbol{x}_0} (g_1(\boldsymbol{x})g_2(\boldsymbol{x})) \leq \left(\limsup_{\boldsymbol{x}\to\boldsymbol{x}_0} g_1(\boldsymbol{x})\right)\left(\limsup_{\boldsymbol{x}\to\boldsymbol{x}_0} g_2(\boldsymbol{x})\right). \tag{6.2.4}$$

only holds true for non-negative functions g_1 and g_2.

The discussions in Remark 2.3.11 is also valid in higher dimensional setting. Rather than repeating all the details here, we only exhibit the results and the details are left to readers for an exercise.

REMARK 6.2.8. As we mentioned above, we only have subadditivity/superadditivity property rather than the additivity. We now show that the linearity holds under extra assumptions. Suppose that all assumptions in Proposition 6.2.6 hold.

(a) If $\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} g_2(x)$ *exists in* $\mathbb{R}$, then

$$\begin{cases} \limsup_{\boldsymbol{x}\to\boldsymbol{x}_0} (g_1(\boldsymbol{x}) + g_2(\boldsymbol{x})) = \limsup_{\boldsymbol{x}\to\boldsymbol{x}_0} g_1(\boldsymbol{x}) + \lim_{\boldsymbol{x}\to\boldsymbol{x}_0} g_2(\boldsymbol{x}), \\ \liminf_{\boldsymbol{x}\to\boldsymbol{x}_0} (g_1(\boldsymbol{x}) + g_2(\boldsymbol{x})) = \liminf_{\boldsymbol{x}\to\boldsymbol{x}_0} g_1(\boldsymbol{x}) + \lim_{\boldsymbol{x}\to\boldsymbol{x}_0} g_2(\boldsymbol{x}). \end{cases} \tag{6.2.5}$$

(b) If *there exists $\epsilon > 0$ such that $|g_1(\boldsymbol{x})| \leq M$ for all $\boldsymbol{x} \in B_\epsilon(\boldsymbol{x}_0)$ and $\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} g_2(\boldsymbol{x})$ exists in $\mathbb{R}_{\geq 0}$*, then

$$\begin{cases} \limsup_{\boldsymbol{x}\to\boldsymbol{x}_0} (g_1(\boldsymbol{x})g_2(\boldsymbol{x})) = \left(\limsup_{\boldsymbol{x}\to\boldsymbol{x}_0} g_1(\boldsymbol{x})\right)\left(\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} g_2(\boldsymbol{x})\right), \\ \liminf_{\boldsymbol{x}\to\boldsymbol{x}_0} (g_1(\boldsymbol{x})g_2(\boldsymbol{x})) = \left(\liminf_{\boldsymbol{x}\to\boldsymbol{x}_0} g_1(\boldsymbol{x})\right)\left(\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} g_2(\boldsymbol{x})\right). \end{cases} \tag{6.2.6}$$

In the particular case when $g_2(\boldsymbol{x}) = c \geq 0$ for all $x \in B_r(\boldsymbol{x}_0)$, we see that (6.2.6) reads

$$\begin{cases} \limsup_{\boldsymbol{x}\to\boldsymbol{x}_0} (cg_1(\boldsymbol{x})) = c \limsup_{\boldsymbol{x}\to\boldsymbol{x}_0} g_1(\boldsymbol{x}), \\ \liminf_{\boldsymbol{x}\to\boldsymbol{x}_0} (cg_1(\boldsymbol{x})) = c \liminf_{\boldsymbol{x}\to\boldsymbol{x}_0} g_1(\boldsymbol{x}). \end{cases}$$

One should be aware that, for constant $b \leq 0$, one sees that $b = -|b|$ and see that

$$\begin{aligned} \limsup_{\boldsymbol{x}\to\boldsymbol{x}_0} (bg_1(\boldsymbol{x})) &= \limsup_{\boldsymbol{x}\to\boldsymbol{x}_0} (-|b|g_1(\boldsymbol{x})) = -\liminf_{\boldsymbol{x}\to\boldsymbol{x}_0} (|b|g_1(\boldsymbol{x})) \\ &= -|b| \liminf_{\boldsymbol{x}\to\boldsymbol{x}_0} g_1(\boldsymbol{x}) = b \liminf_{\boldsymbol{x}\to\boldsymbol{x}_0} g_1(\boldsymbol{x}), \end{aligned}$$

and

$$\begin{aligned}\liminf_{\boldsymbol{x}\to\boldsymbol{x}_0}(bg_1(\boldsymbol{x})) &= \liminf_{\boldsymbol{x}\to\boldsymbol{x}_0}(-|b|g_1(\boldsymbol{x})) = -\limsup_{\boldsymbol{x}\to\boldsymbol{x}_0}(|b|g_1(\boldsymbol{x})) \\ &= -|b|\limsup_{\boldsymbol{x}\to\boldsymbol{x}_0} g_1(\boldsymbol{x}) = b\limsup_{\boldsymbol{x}\to\boldsymbol{x}_0} g_1(\boldsymbol{x}).\end{aligned}$$

This means that in general, the linearity does not hold true for general coefficients, which only holds true for positive coefficients.

EXAMPLE 6.2.9. One sees that the above definition of limit (Definition 6.2.2) is equivalent to $\lim_{\boldsymbol{x}\to\boldsymbol{x}_0}|f(\boldsymbol{x})-L|=0$, similar discussions also hold for limit superior and limit inferior. It is convenient to understood as the right limit:

$$\lim_{|\boldsymbol{x}-\boldsymbol{x}_0|\to 0+}|f(\boldsymbol{x})-L|=0.$$

For example, we consider the function $f:\mathbb{R}^2\setminus\{\mathbf{0}\}\to\mathbb{R}$ by

$$f(x_1,x_2)=\frac{x_1^2x_2}{x_1^2+x_2^2}\quad\text{for all } \boldsymbol{x}=(x_1,x_2)\neq(0,0).$$

Since $|x_1|\leq|\boldsymbol{x}|$ and $|x_2|\leq|\boldsymbol{x}|$ for all $\boldsymbol{x}=(x_1,x_2)\in\mathbb{R}^2$, one sees that

$$|f(\boldsymbol{x})|=\frac{|x_1|^2|x_2|}{|\boldsymbol{x}|^2}\leq\frac{|\boldsymbol{x}|^3}{|\boldsymbol{x}|^2}=|\boldsymbol{x}|\quad\text{for all } \boldsymbol{x}=(x_1,x_2)\neq(0,0).$$

Thus we have

$$\limsup_{|\boldsymbol{x}|\to 0}|f(\boldsymbol{x})|\leq\limsup_{|\boldsymbol{x}|\to 0}|\boldsymbol{x}|=\lim_{|\boldsymbol{x}|\to 0}|\boldsymbol{x}|=0,$$

which concludes that $\lim_{\boldsymbol{x}\to\mathbf{0}}f(\boldsymbol{x})=0$.

*** (to be appeared) ***

6.3. Definition of differentiation

We now extend the notion of "differentiation" for functions $f:\mathbb{R}^n\to\mathbb{R}$. Since one cannot divide by a vector, one cannot directly extend the definition of differentiation directly from the standard definition of 1-dimensional in Definition 3.1.1. In view of its equivalent definition in Definition 3.1.3, now the generalization for higher-dimensional case is much more natural.

*** (to be appeared) ***

Bibliography

[Apo74] T. M. Apostol. *Mathematical analysis*. Addison-Wesley Publishing Co., second edition, 1974. MR0344384, Zbl:0309.26002.

[BV04] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, Cambridge, 2004. MR2061575, Zbl:1058.90049, http://www.stanford.edu.

[Cha06] I. Chavel. *Riemannian geometry. A modern introduction*, volume 98 of *Cambridge Stud. Adv. Math.* Cambridge University Press, Cambridge, second edition, 2006. MR2229062, Zbl:1099.53001, doi:10.1017/CBO9780511616822.

[HH79] E. Hewitt and R.E. Hewitt. The Gibbs-Wilbraham phenomenon: an episode in Fourier analysis. *Arch. Hist. Exact Sci.*, 21(2):129–160, 1979. MR0555102, doi:10.1007/BF00330404.

[HB10] L. D. Hoffmann and G. L. Bradley. *Applied calculus: for business, economics, and the social and life sciences*. McGraw-Hill, New York, tenth edition, 2010. NCCU library.

[Kow22] P.-Z. Kow. *Fourier analysis and distribution theory*. University of Jyväskylä, 2022. https://puzhaokow1993.github.io/homepage.

[Kow23] P.-Z. Kow. *Complex Analysis*. National Chengchi University, Taipei, 2023. https://puzhaokow1993.github.io/homepage.

[Kow24] P.-Z. Kow. *An introduction to partial differential equations and functional analysis*. National Chengchi University, 2024. https://puzhaokow1993.github.io/homepage.

[Rud76] W. Rudin. *Principles of mathematical analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill Book Co., New York-Auckland-Düsseldorf, thid edition, 1976. MR0385023, Zbl:0346.26002.

[Rud87] W. Rudin. *Real and complex analysis*. McGraw-Hill Book Co., New York, third edition, 1987. MR0924157, Zbl:0925.00005.

[SCW21] J. Stewart, D. Clegg, and S. Watson. *Calculus: early transcendentals*. Cengage, Singapore, ninth edition, 2021. NCCU library.

[Str08] W. A. Strauss. *Partial differential equations: An introduction*. John Wiley & Sons, Ltd., Chichester, second edition, 2008. MR2398759, Zbl:1160.35002.

[Tay52] A. E. Taylor. L'Hospital's rule. *Amer. Math. Monthly*, 59:20–24, 1952. MR0044602, Zbl:0046.06202, doi:10.2307/2307183, JSTOR:2307183.

[WZ15] R. L. Wheeden and A. Zygmund. *Measure and integral. An introduction to real analysis*. Pure Appl. Math. CRC Press, Boca Raton, FL, second edition, 2015. MR3381284, Zbl:1326.26007.

[Win10] G. Winskel. *Set theory for computer science*. University of Cambridge, Cambridge, 2010. https://www.cl.cam.ac.uk.