CALCULUS

Siyuan Tang

Beijing International Center for Mathematical Research,

Peking University, Beijing 100871

***E-mail address*: siyuantang@pku.edu.cn**

Last updated: January 3, 2025

目录

Preface

This text is an expanded series of lecture notes based on the undergraduate calculus course that I taught at Peking University. It covers the various notions of limits and convergence, as well as differential calculus at the undergraduate level. Its main purpose is to help students understand how to approximate and solve general nonlinear problems using linear algebraic results. In particular, the fundamental theorem of calculus, the Taylor's formula, and the implicit function theorem are introduced. Also, it covers more advanced topics in analysis, including the completeness of real numbers, and the power series on the real line (Taylor series) and the circle (Fourier series). In addition, a number of auxiliary topics are discussed in Chapter 7. In my own course, I was able to cover the material in Chapters 1-6, using the exercises provided in Chapter 8 to assess the students. I also used the text [1] as a secondary source.

I would like to thank Chengfeng Shen and Yuxiang Jiao for their helpful conversation during the course. Finally, I would like to thank all who gave feedback on early drafts of this text including Zhiyuan Jiang, Kaicheng Que, Chunzhi Yao, and other students.

Chapter 1

Functions and Limits

1.1 Real numbers

1.1.1 Rational and irrational numbers

Recall the natural numbers $\mathbb{N} = \{0, 1, 2, \ldots\}$, the integers $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$, and the rational numbers $\mathbb{Q}$:

$$\mathbb{Q} = \left\{ \frac{m}{n} : m, n \in \mathbb{Z},\ n > 1 \right\}.$$

- $\mathbb{Z}$ is closed under addition and subtraction.
- $\mathbb{Q}$ is closed under addition, subtraction, multiplication and division.

Thus, $\mathbb{Q}$ is a *field* (数域), i.e. it is closed under addition, subtraction, multiplication and division, and it has commutativity, associativity, and distributivity (Definition 7.2.3). The motivation for $\mathbb{Q}$ is purely *algebraic* (代数的). However, $\mathbb{Q}$ is **NOT** *complete* (完备), i.e. there are still an infinite number of "gaps" or "holes" between the rationals.

Proposition 1.1.1 ([1, §1.1 命题 1])**.** $\sqrt{2}$ *is not rational.*

证明. Otherwise, $\sqrt{2} = m/n$ with $m, n \in \mathbb{N}$ and $(m, n) = 1$ (*greatest common divisor* (最大公约数)). Then $2n^2 = m^2$ and so m is even. But then $m = 2k$ and so $n^2 = 2k^2$. Then n is even too. A contradiction. □

Proposition 1.1.2 (Incompleteness of $\mathbb{Q}$)**.** *Any two rationals $a < b$, there is a rational $c \in (a,b) \cap \mathbb{Q}$, and an irrational $d \in (a,b) \cap \mathbb{Q}^c$.*

证明. First, consider the interval $[0,1]$. Clearly, $\frac{1}{2} \in (0,1)$ is rational, and $(\sqrt{2}-1) \in (0,1)$ is irrational by Proposition 1.1.1.

Next, for $k \in \mathbb{N}$, $[0, 2^{-k}] = 2^{-k} \cdot [0,1]$, we see that $2^{-k} \cdot \frac{1}{2} \in (0, 2^{-k})$ is rational, and $2^{-k} \cdot (\sqrt{2}-1) \in (0, 2^{-k})$ is irrational.

Moreover, for $n \in \mathbb{Z}$, $k \in \mathbb{N}$, $[n2^{-k}, (n+1)2^{-k}] = n2^{-k} + 2^{-k} \cdot [0,1]$, we see that $n2^{-k} + 2^{-k} \cdot \frac{1}{2} \in (n2^{-k}, (n+1)2^{-k})$ is rational, and $n2^{-k} + 2^{-k} \cdot (\sqrt{2}-1) \in (n2^{-k}, (n+1)2^{-k})$ is irrational.

Finally, for any $a < b$, there exists $n \in \mathbb{Z}$ and $k \in \mathbb{N}$ such that

$$[n2^{-k}, (n+1)2^{-k}] = n2^{-k} + 2^{-k} \cdot [0,1] \subset [a,b].$$

Hence, we establish the proposition. $\square$

The real numbers can be considered as any possible limit of the rational numbers:

$$\mathbb{R} = \left\{ \lim_{n\to\infty} a_n : a_n \in \mathbb{Q},\ \{a_n\} \text{ converges} \right\}.$$

The definition is not **rigorous**. In fact, it is a *circular definition* (循环定义), since the limit $\lim\limits_{n\to\infty} a_n$, as we shall see, is defined to be a real number. Actually, in the rigorous definition, $\lim\limits_{n\to\infty} a_n$ is explained as a *Cauchy sequence* of rational numbers (see Definitions 7.1.1 and 7.1.3).

$\mathbb{R}$ is also a field in the algebraic sense. But more importantly, $\mathbb{R}$ is *complete* in topological sense. In other words, there are no "gaps" or "holes" between the reals. (See also Appendix 7.1.)

Definition 1.1.3 (Monotonicity (单调))**.** A sequence $\{a_n\}$ is called *monotonic increasing* (单调递增) if $a_n \leq a_{n+1}$ for any $n \in \mathbb{N}$; called *monotonic decreasing* (单调递减) if $a_n \geq a_{n+1}$ for any $n \in \mathbb{N}$.

The completeness of $\mathbb{R}$ can be described as follows:

Proposition 1.1.4 (Monotone bounded sequences converge)**.** *Any monotonic and bounded (有界) sequence $\{b_n\} \subset \mathbb{R}$ has a limit.*

Example 1.1.5. $\{1+n^{-1}\}$ is bounded, and decreasing. $\{n\}$ is increasing, but not bounded.

Note that $\mathbb{Q}$ is not complete in the sense of Proposition 1.1.4. For instance, $1, 1.4, 1.41, 1.414, \ldots$ is a monotone and bounded sequence, but the limit $\sqrt{2}$ is not a rational number.

Example 1.1.6 (2022 Calculus A, Midterm)**.** Let $a_1 = \sqrt{2}$, $a_{n+1} = \sqrt{2}^{a_n}$, for $n \in \mathbb{N}^*$. Is the sequence $\{a_n\}$ convergent?

证明. If $0 < a_n < 2$, then $a_{n+1} = \sqrt{2}^{a_n} \in (0,2)$. Moreover, if $a_n > a_{n-1}$, then we have

$$\frac{a_{n+1}}{a_n} = \frac{\sqrt{2}^{a_n}}{\sqrt{2}^{a_{n-1}}} = 2^{\frac{a_n - a_{n-1}}{2}} > 1.$$

Then by induction, we conclude that $\{a_n\}$ is increasing. Then we calculate

$$\lim_{n\to\infty} a_n = 2.$$

□

1.1.2 Interval

The set of real numbers can be identified with real axis $\mathbb{R} = (-\infty, \infty)$. Intervals are denoted by e.g.

$$(a,b] = \{x \in \mathbb{R} : a < x \le b\}.$$

For an interval $I = [a,b]$, the number $b-a$ is the *length* (长度) of I, the number $c = \frac{a+b}{2}$ is the *center* (中心) of I, and the number $\frac{b-a}{2}$ is the *radius* (半径) of I.

1.1.3 Absolute values

For $x \in \mathbb{R}$, we write

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

Proposition 1.1.7 ([1, §1.1 命题 2])**.** *For any $x, y \in \mathbb{R}$, we have*

(1) $|x| \geq 0$*, and the equality holds if and only if* $x = 0$*;*

(2) $|x| = |-x|$*;*

(3) (Triangle inequality) $|x+y| \leq |x| + |y|$.

For $a, b \in \mathbb{R}$, $|a-b|$ indicates the distance between a and b.
By Proposition 1.1.7, we have

Proposition 1.1.8 ([1, §1.1 命题 3])**.** *For* $a, b, c \in \mathbb{R}$*, we have*

(1) $|a-b| \geq 0$*, and the equality holds if and only if* $a = b$*;*

(2) $|a-b| = |b-a|$*;*

(3) (Triangle inequality) $|a-c| \leq |a-b| + |b-c|$.

The interval $U_r(a) \coloneqq (a-r, a+r) = \{x \in \mathbb{R} : |x-a| < r\}$ is called the r*-neighborhood* (邻域) of a. Further

$$\mathring{U}_r(a) \coloneqq U_r(a) \setminus \{a\} = (a-r, a) \cup (a, a+r) = \{x \in \mathbb{R} : 0 < |x-a| < r\}$$

is called a *deleted* r*-neighbourhood* (空心邻域) of a. We may also ignore the radius and simply write $U(a)$ and $\mathring{U}(a)$ respectively.

1.2 Variable and Function

1.2.1 Function

Definition 1.2.1 ([1, §1.2 定义 1])**.** For sets $X, Y \subset \mathbb{R}$, a *function* (函数) $f : X \to Y$ is an assignment of one element $y \in Y$ to each element $x \in X$, i.e. f maps x to y,

$$f : x \mapsto y.$$

The set X is called the *domain* (定义域) of f, and set Y is called the *codomain* of f. $y = f(x)$ is called the *value of the function* (函数值) f at x, or the image of x under the function. The set $f(X) \subset Y$ is called the *image* (值域) of f.

In mathematics, a map or mapping is a function in its general sense.

Definition 1.2.2. For sets X, Y, a *map* or *mapping* (映射) $f : X \to Y$ is an assignment of one element $y \in Y$ to each element $x \in X$, i.e. f maps x to y,

$$f : x \mapsto y.$$

One can define notions related to maps in a way similar to functions, such as domain, codomain, and so on.

Example 1.2.3. A sequence $\{a_n\}$ can be considered as a function $a : \mathbb{N}^* \to \mathbb{R}$ defined by

$$a : n \mapsto a_n.$$

1.2.2 Basic elementary function

Basic elementary functions (基本初等函数):

(1) Constant functions (常数函数): $f(x) = c$ for some $c \in \mathbb{R}$. Its domain is $\mathbb{R}$.

(2) Powers of x (幂函数): $f(x) = x^a$ for some $0 \neq a \in \mathbb{R}$. Its domain always includes $(0, \infty)$.

(3) Exponential functions (指数函数): $f(x) = a^x$ for some $a > 0$ and $a \neq 1$. Its domain is $\mathbb{R}$.

(4) Logarithms (对数函数): $f(x) = \log_a x$ for some $a > 0$ and $a \neq 1$. Its domain is $(0, \infty)$. We write $\ln x = \log_e x$, where e is the base of the natural logarithm.

(5) Trigonometric functions (三角函数): $\sin x$, $\cos x$ on $\mathbb{R}$, $\tan x$ on $\mathbb{R} \setminus \{(n + \frac{1}{2})\pi : n \in \mathbb{Z}\}$, etc.

(6) Inverse trigonometric functions (反三角函数): $\arcsin x$, $\arccos x$ on $[-1, 1]$, $\arctan x$ on $\mathbb{R}$, etc.

Definition 1.2.4 ([1, §1.2 定义 2])**.** Suppose that we have two functions $f : X \to Y$ and $g : Y^* \to Z$ with $f(X) \subset Y^*$. The *composition* (复合函数) $g \circ f : X \to Z$ of f and g is defined by

$$g \circ f : x \mapsto g(f(x)).$$

Remark 1.2.5. When the domain or codomain of a function is not necessarily a subset of $\mathbb{R}$, then the function is called a *map* or *mapping* (映射).

Example 1.2.6 ([1, §1.2 例 1])**.** $f(x) = \sin x$, $g(y) = e^y$. Then $g \circ f(x) = e^{\sin x}$.

Example 1.2.7 ([1, §1.2 例 2])**.** $f(x) = x^2$, $g(y) = \sqrt{y}$. Then $g \circ f(x) = \sqrt{x^2} = |x|$.

Example 1.2.8 ([1, §1.2 例 3])**.** $f(x) = \sin x$, $g(x) = \arcsin x$. Then find the compositions $f \circ g$ and $g \circ f$, and sketch the graphs of the functions.

证明. Note that $f : \mathbb{R} \to [-1, 1]$, and $g : [-1, 1] \to [-\frac{\pi}{2}, \frac{\pi}{2}]$. Set $y = \arcsin x$. Then $x = \sin y$, $y \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. Then $f \circ g : [-1, 1] \to [-1, 1]$ is given by

$$f \circ g(x) = f(y) = \sin y = x.$$

On the other hand, note that $g \circ f : \mathbb{R} \to [-\frac{\pi}{2}, \frac{\pi}{2}]$. For $x \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, write $y = \sin x$. Then $x = \arcsin y$, and

$$g \circ f(x) = g(y) = \arcsin y = x.$$

In addition, for $x \in [\frac{\pi}{2}, \frac{3\pi}{2}]$, we have $\pi - x \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, and $\sin x = \sin(\pi - x)$. Write $y = \sin x = \sin(\pi - x)$. Then $\arcsin y = \pi - x$. Then we have

$$g \circ f(x) := \begin{cases} x & , \text{ if } x \in [-\frac{\pi}{2}, \frac{\pi}{2}] \\ \pi - x & , \text{ if } x \in [\frac{\pi}{2}, \frac{3\pi}{2}] \end{cases} .$$

□

Example 1.2.9 ([1, §1.2 例 4])**.** Proof that $\sin(\arccos x) = \sqrt{1 - x^2}$.

证明. For any $x \in [-1, 1]$, write $y = \arccos x$. Then $x = \cos y$ for $y \in [0, \pi]$. Then

$$\sin(\arccos x) = \sin y = \sqrt{1 - \cos^2 y} = \sqrt{1 - x^2}.$$

□

Definition 1.2.10. An *elementary function* (初等函数) is defined as taking sums, products, roots and compositions of finitely many basic elementary functions.

Example 1.2.11 ([1, §1.2 例 5])**.** $f(x) = \sum_{k=1}^{n} \frac{\sin kx}{k}$ is an elementary function.

Example 1.2.12. The *sign function* （符号函数）

$$f(x) = \operatorname{sgn} x \coloneqq \begin{cases} 1 & \text{, if } x > 0 \\ 0 & \text{, if } x = 0 \\ -1 & \text{, if } x < 0 \end{cases}$$

is not an elementary function.

Example 1.2.13 ([1, §1.2 例 6])**.** Hyperbolic functions:

- $\mathrm{sh}(x) = \sinh(x) = \frac{e^x - e^{-x}}{2}$, $\mathrm{ch}(x) = \cosh(x) = \frac{e^x + e^{-x}}{2}$,
- $\mathrm{th}(x) = \tanh(x) = \frac{\mathrm{sh}(x)}{\mathrm{ch}(x)}$, $\coth(x) = \frac{\mathrm{ch}(x)}{\mathrm{sh}(x)}$,
- $\mathrm{sech}(x) = \frac{1}{\mathrm{ch}(x)}$, $\mathrm{csch}(x) = \frac{1}{\mathrm{sh}(x)}$.

Then

- $\mathrm{ch}^2 x - \mathrm{sh}^2 x = 1$,
- $\mathrm{sh}(x \pm y) = \mathrm{sh}(x)\,\mathrm{ch}(y) \pm \mathrm{ch}(x)\,\mathrm{sh}(y)$,
- $\mathrm{ch}(x \pm y) = \mathrm{ch}(x)\,\mathrm{ch}(y) \pm \mathrm{sh}(x)\,\mathrm{sh}(y)$.

Example 1.2.14 ([1, §1.2 例 7])**.** Let $[x]$ be the greatest integer less than or equal to x. It is called *floor function* (取整函数). Then the *fractional part* (小数部分) $f(x) = x - [x]$ is not an elementary function.

Example 1.2.15 ([1, §1.2 例 8])**.** Let

$$D(x) \coloneqq \begin{cases} 1 & \text{, if } x \in \mathbb{Q} \\ 0 & \text{, if } x \in \mathbb{Q}^c = \mathbb{R} \setminus \mathbb{Q} \end{cases}$$

It is called the *Dirichlet function* (狄利克雷函数). It is not even *Riemann integrable* (Definition 2.8.1)!

A map $f: E \to F$ is called *surjective* (满射), if $f(E) = F$.

A map $f: E \to F$ is called *injective* (单射) or *one-to-one* (一一映射), if $f(x_1) = f(x_2)$ implies $x_1 = x_2$.

A map is called *bijective* (双射) if it is both surjective and injective. A bijective map $f: E \to F$ naturally has an *inverse map* (逆映射 or 反函数).

Example 1.2.16 ([1, §1.2 例 9])**.** The function $\sin x$ is surjective, but not injective.

1.2.3 Bounded function

We say that $f: X \to Y$ has an *upper bound* (上界) if there is $M \in \mathbb{R}$ such that

$$f(x) \leq M$$

for any $x \in X$. One can similarly define the *lower bound* (下界). A function $f: X \to Y$ is called *bounded* (有界函数) if it has both upper bound and lower bound.

Example 1.2.17 ([1, §1.2 例 10])**.** Let

$$f(x) := \begin{cases} \frac{1}{x} & , \text{ if } x \in (0, a] \\ 0 & , \text{ if } x = 0 \end{cases} .$$

f is well-defined on $[0, a]$, but f is not bounded (unbounded).

Proposition 1.2.18. *A function $f: X \to Y$ is bounded if and only if there is $C > 0$ such that*

$$|f(x)| \leq C$$

for all $x \in X$.

1.3 Limit of a sequence

1.3.1 Definition

Definition 1.3.1 ([1, §1.3 定义])**.** Let $\{a_n\} \subset \mathbb{R}$ be a sequence. if there is a constant $l \in \mathbb{R}$ such that for any $\epsilon > 0$, there is a positive integer $N = N(\epsilon) > 0$ such that

$$|a_n - l| < \epsilon$$

for any $n > N$, then we say l is the *limit* of $\{a_n\}$. We write

$$\lim_{n\to\infty} a_n = l.$$

We also say that as n tends to infinity, a_n tends to (or converges to) l. We also write

$$a_n \to l \quad (\text{as } n \to \infty).$$

For a given $\{a_n\}$, if there is l so that $a_n \to l$, then we say $\{a_n\}$ has a limit (is convergent).

It is called the *epsilon-N definition* ($\epsilon - N$ 说法).

Example 1.3.2.

- $\lim_{n\to\infty} \frac{1}{n} = 0$.
- $\{n\}$ is unbounded.
- $\{(-1)^n\}$ is bounded but bouncing.

Example 1.3.3 ([1, §1.3 例 2])**.** Prove that $\lim_{n\to\infty} \frac{1}{n^\alpha} = 0$ for $\alpha > 0$.

证明. For given $\epsilon > 0$, in order to

$$\left|\frac{1}{n^\alpha} - 0\right| < \epsilon \quad \Longleftrightarrow \quad n^{-\alpha} < \epsilon$$

we need $n^\alpha > \epsilon^{-1}$, i.e. $n > \epsilon^{-\frac{1}{\alpha}}$. Thus, we choose $N = [\epsilon^{-\frac{1}{\alpha}}] + 1$. Then for any $n \geq N$, we have

$$\left|\frac{1}{n^\alpha} - 0\right| < \epsilon$$

as required. □

Example 1.3.4 ([1, §1.3 例 3])**.** For any $a > 1$, show that

$$\lim_{n\to\infty} \sqrt[n]{a} = 1.$$

证明. For any $\epsilon > 0$, let $N = [(\log_a(1+\epsilon))^{-1}] + 1$. Then for any $n \geq N$, we have

$$n \geq N > (\log_a(1+\epsilon))^{-1}.$$

Then $\log_a(1+\epsilon) > n^{-1}$, and so $1+\epsilon > a^{\frac{1}{n}} = \sqrt[n]{a}$. Thus,

$$\left|\sqrt[n]{a} - 1\right| = \sqrt[n]{a} - 1 < \epsilon,$$

as required. □

Remark 1.3.5. One may also show

$$\lim_{n\to\infty} \sqrt[n]{a} = 1$$

for $0 < a \leq 1$.

Example 1.3.6 ([1, §1.3 例 4])**.** For any $q \in \mathbb{R}$ with $|q| < 1$, show that

$$\lim_{n\to\infty} q^n = 0.$$

证明. If $q = 0$, then $q^n = 0$, and so $\lim_{n\to\infty} q^n = 0$.

Now assume that $0 < |q| < 1$. For any $\epsilon \in (0,1)$, let

$$N = \left[\log_{|q|} \epsilon\right] + 1.$$

Then for any $n \geq N$, we have

$$n \geq N > \log_{|q|} \epsilon.$$

Then $|q|^n < \epsilon$, and so

$$|q^n - 0| < \epsilon,$$

as required. □

Example 1.3.7 ([1, §1.3 例 6])**.** Prove that

$$\lim_{n\to\infty} \frac{n^2+3n+10}{3n^2-n+2} = \frac{1}{3}.$$

证明. Note that when $n > 3$, we have

$$\begin{aligned}\left|\frac{n^2+3n+10}{3n^2-n+2}-\frac{1}{3}\right| &= \left|\frac{10n+28}{3(3n^2-n+2)}\right| \\ &= \frac{10n+28}{3(2n^2+n^2-n+2)} \\ &\le \frac{10n+10n}{3\cdot 2n^2} = \frac{10}{3n}.\end{aligned}$$

Thus, for any $\epsilon \in (0,1)$, let $N = [\frac{10}{3\epsilon}] + 1 \ge 3$. Then for any $n \ge N$, we have

$$n \ge N > \frac{10}{3\epsilon} \quad \Longleftrightarrow \quad \frac{10}{3n} < \epsilon.$$

Then

$$\left|\frac{n^2+3n+10}{3n^2-n+2}-\frac{1}{3}\right| < \epsilon,$$

as required. □

1.3.2 Squeeze theorem

Theorem 1.3.8 (Squeeze theorem, 夹逼定理, [1, §1.3 定理 1])**.** *Let $\{a_n\}$, $\{b_n\}$, $\{c_n\}$ be sequences. Suppose that there is an integer $N_0 > 0$ such that*

$$c_n \le a_n \le b_n$$

for $n \ge N_0$. If both $\{b_n\}$ and $\{c_n\}$ have the same limit l, then l is also the limit of $\{a_n\}$.

证明. For given $\epsilon > 0$, there are N_1 and N_2 such that

$$|b_n - l| < \epsilon \quad \Longrightarrow \quad b_n < l + \epsilon$$

for $n \ge N_1$, and

$$|c_n - l| < \epsilon \quad \Longrightarrow \quad c_n > l - \epsilon$$

for $n \ge N_2$. Now let $N = \max\{N_0, N_1, N_2\}$. Then we have

$$l - \epsilon < c_n \le a_n \le b_n < l + \epsilon \quad \Longrightarrow \quad |a_n - l| < \epsilon$$

for $n > N$, as required. □

Example 1.3.9 ([1, §1.3 例 11])**.** For $a > 1$, integer $k > 1$, prove that

$$\lim_{n\to\infty} \frac{n}{a^n} = 0.$$

证明. Let $h = a - 1$. Then $h > 0$, and

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

Thus, for $n > k$, we have

$$a^n > \frac{n(n-1)}{2}h^2.$$

Then

$$0 \le \frac{n}{a^n} \le \frac{n}{\frac{n(n-1)}{2}h^2} \to 0.$$

The consequence follows from the squeeze theorem (Theorem 1.3.8). □

Remark 1.3.10. Similar idea shows that $\lim\limits_{n\to\infty} \frac{n^k}{a^n} = 0$, for $k \in \mathbb{N}$.

Example 1.3.11. Let x_n be the unique solution of the equation $x+x^2+\cdots+x^n = 1$ on $(0,1)$. Prove that

$$\lim_{n\to\infty} x_n = \frac{1}{2}.$$

证明. Let $f_n(x) = x + x^2 + \cdots + x^n$. Then f is increasing on $(0,1)$. Note that $f_n(x_n) = 1$. Also,

$$f_n\left(\frac{1}{2}\right) = \frac{1}{2} + \left(\frac{1}{2}\right)^2 + \cdots + \left(\frac{1}{2}\right)^n = 1 - \left(\frac{1}{2}\right)^n,$$

$$f_n\left(\frac{1}{2} + \left(\frac{1}{2}\right)^n\right) = \frac{1}{2} + \left(\frac{1}{2}\right)^n + \cdots + \left(\frac{1}{2} + \left(\frac{1}{2}\right)^n\right)^n > 1.$$

We have $\frac{1}{2} \le x_n \le \frac{1}{2} + \left(\frac{1}{2}\right)^n$. By Theorem 1.3.8, we conclude that

$$\lim_{n\to\infty} x_n = \frac{1}{2}.$$

□

1.3.3 Inequalities

Theorem 1.3.12 ([1, §1.3 定理 2])**.** *Let $\{a_n\}$ and $\{b_n\}$ be sequences that have limits l_1 and l_2 respectively. Also, $l_1 > l_2$. Then there is an integer $N > 0$ such that*

$$a_n > b_n$$

for $n > N$.

证明. For given $\epsilon > 0$, there are $N_1, N_2 > 0$ such that

$$|a_n - l_1| < \epsilon \quad \Longrightarrow \quad a_n > l_1 - \epsilon$$

for $n \geq N_1$, and

$$|c_n - l| < \epsilon \quad \Longrightarrow \quad b_n < l_2 + \epsilon$$

for $n \geq N_2$. Then for $2\epsilon < l_1 - l_2$, let $N = \{N_1, N_2\}$, and we have

$$b_n < l_2 + \epsilon < l_1 - \epsilon < a_n$$

for all $n \in N$. □

Theorem 1.3.13 ([1, §1.3 定理 3])**.** *Let $\{a_n\}$ and $\{b_n\}$ be sequences that have limits l_1 and l_2 respectively. Suppose there is $N_0 > 0$ such that*

$$a_n \geq b_n \tag{1.3.1}$$

for $n > N_0$. Then $l_1 \geq l_2$.

证明. Assume that $l_1 < l_2$. By Theorem 1.3.12, we have $N > 0$ such that for $n > N$, $a_n < b_n$, a contradiction. □

Remark 1.3.14. Note that we can only deduce $l_1 \geq l_2$, even if (1.3.1) is replaced by $a_n > b_n$. For example, $a_n = n^{-1}$, $b_n = 0$.

1.3.4 Algebraic limit theorem

Theorem 1.3.15 (Algebraic limit theorem, [1, §1.3 定理 4])**.** *Let* $\{a_n\}$ *and* $\{b_n\}$ *be sequences that have limits* l_1 *and* l_2 *respectively. Then*

$$\lim_{n\to\infty}(a_n \pm b_n) = l_1 \pm l_2, \quad \lim_{n\to\infty} a_n b_n = l_1 l_2,$$

and for $l_2 \neq 0$*, we have*

$$\lim_{n\to\infty}\frac{a_n}{b_n} = \frac{l_1}{l_2}.$$

证明. Only prove for multiplication.

$$\begin{aligned}
&|a_n b_n - l_1 l_2| \\
=&|a_n b_n - l_1 b_n + l_1 b_n - l_1 l_2| \\
\leq&|a_n - l_1||b_n| + |l_1||b_n - l_2| \\
=&|a_n - l_1||b_n - l_2 + l_2| + |l_1||b_n - l_2| \\
\leq&|a_n - l_1||b_n - l_2| + |a_n - l_1||l_2| + |l_1||b_n - l_2|.
\end{aligned} \tag{1.3.2}$$

Then for any $\epsilon' > 0$, there are $N_1, N_2 > 0$ such that

$$|a_n - l_1| < \epsilon'$$

for $n \geq N_1$, and

$$|b_n - l_2| < \epsilon'$$

for $n \geq N_2$. Then for $n \geq N = N(\epsilon') = \max\{N_1, N_2\}$, we have by (1.3.2) that

$$|a_n b_n - l_1 l_2| \leq (\epsilon')^2 + \epsilon'|l_2| + \epsilon'|l_1| = \epsilon'(\epsilon' + |l_2| + |l_1|).$$

Then for any $\epsilon \in (0,1)$, let $\epsilon' = \epsilon'(\epsilon) = \frac{\epsilon}{1+|l_2|+|l_1|}$, and we have

$$|a_n b_n - l_1 l_2| \leq \frac{\epsilon}{1 + |l_2| + |l_1|} \cdot (\epsilon' + |l_2| + |l_1|) < \epsilon$$

for $n \geq N = N(\epsilon'(\epsilon))$. □

Example 1.3.16 ([1, §1.3 例 12])**.** Find

$$\lim_{n\to\infty}\frac{n^3 + 5n + 1}{4n^3 + 8}.$$

证明. One calculates

$$\lim_{n\to\infty}\frac{n^3+5n+1}{4n^3+8}=\lim_{n\to\infty}\frac{1+\frac{5}{n^2}+\frac{1}{n^3}}{4+\frac{8}{n^3}}=\frac{\lim_{n\to\infty}1+\frac{5}{n^2}+\frac{1}{n^3}}{\lim_{n\to\infty}4+\frac{8}{n^3}}=\frac{1}{4}.$$

□

Given a sequence $\{a_n\}$, a *subsequence* (子序列) of $\{a_n\}$ is an infinite subset $\{a_{n_k}\}\subset\{a_n\}$.

Theorem 1.3.17 ([1, §1.3 定理 5])**.** *Suppose that $\{a_n\}$ has a limit l. Then any subsequence $\{a_{n_k}\}\subset\{a_n\}$ also tends to l.*

证明. For any $\epsilon>0$, there is $N>0$ such that

$$|a_n-l|<\epsilon$$

for $n\geq N$. In particular, for $k>N$, we have $n_k\geq N$, and so

$$|a_{n_k}-l|<\epsilon$$

for all $k>N$. □

1.3.5 An important limit

Example 1.3.18 (Euler's number e, [1, §1.3 例 14])**.** The sequence

$$\left\{\left(1+\frac{1}{n}\right)^n : n\in\mathbb{N}^*\right\}$$

converges.

证明. First, $\left\{\left(1+\frac{1}{n}\right)^n\right\}$ is monotonic increasing. To see this, using *AM-GM inequality* (均值不等式, cf. Example 4.5.3), we have

$$1\cdot\left(1\pm\frac{1}{n}\right)^n\leq\left(\frac{1+(1\pm\frac{1}{n})+\cdots+(1\pm\frac{1}{n})}{n+1}\right)^{n+1}=\left(\frac{n+1\pm 1}{n+1}\right)^{n+1}=\left(1\pm\frac{1}{n+1}\right)^{n+1}.$$

Note that $(\frac{n-1}{n})^n\leq(\frac{n}{n+1})^{n+1}$ implies that

$$\frac{1}{4}=\left(\frac{2-1}{2}\right)^2\leq\left(\frac{n}{n+1}\right)^{n+1}\quad\Longrightarrow\quad\left(\frac{n+1}{n}\right)^{n+1}\leq 4.$$

Thus,

$$1 \leq \left(1+\frac{1}{n}\right)^n \leq \left(1+\frac{1}{n}\right)^{n+1} \leq 4.$$

Thus, $\{(1+\frac{1}{n})^n\}$ is bounded and so is convergent. $\square$

Then we define

$$e \coloneqq \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n.$$

Example 1.3.19 ([1, §1.3 例 17])**.** Find $\lim\limits_{n\to\infty}\left(1-\frac{1}{n}\right)^n$.

证明. Note that

$$\left(1-\frac{1}{n}\right)^n = \left(\frac{n-1}{n}\right)^n = \frac{1}{\left(\frac{n}{n-1}\right)^n} = \frac{1}{\left(1+\frac{1}{n-1}\right)^{n-1}\left(1+\frac{1}{n-1}\right)}.$$

Thus, we have

$$\lim_{n\to\infty}\left(1-\frac{1}{n}\right)^n = \frac{1}{e\cdot 1} = e^{-1}.$$

$\square$

Example 1.3.20 ([1, P.46 习题 9])**.** Prove that

$$e = \lim_{n\to\infty}\left(1+1+\frac{1}{2}+\cdots+\frac{1}{n!}\right).$$

证明. First, note that

$$\begin{aligned}\left(1+\frac{1}{n}\right)^n =& 1+\frac{1}{1!}\cdot\frac{n!}{n(n-1)!}+\frac{1}{2!}\cdot\frac{n!}{n^2(n-2)!}+\cdots+\frac{1}{n!}\cdot\frac{n!}{n^n 1!}\\ \leq& 1+1+\frac{1}{2}+\cdots+\frac{1}{n!}.\end{aligned}$$

Taking the limit, we have

$$e \leq \lim_{n\to\infty}\left(1+1+\frac{1}{2}+\cdots+\frac{1}{n!}\right).$$

On the other hand, for any $k \in \mathbb{N}$, $n > k$, we have

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

Letting $n \to \infty$, we have

$$e \geq 1 + 1 + \frac{1}{2} + \cdots + \frac{1}{k!}$$

for any $k \in \mathbb{N}$. Thus, e is an upper bound of the sequence $\{1+1+\frac{1}{2}+\cdots+\frac{1}{k!} : k \in \mathbb{N}\}$. Letting $k \to \infty$, we have

$$e \geq \lim_{k\to\infty}\left(1 + 1 + \frac{1}{2} + \cdots + \frac{1}{k!}\right).$$

□

1.3.6 Stolz-Cesàro theorem

Proposition 1.3.21 (Cesàro summation, 切萨罗求和). *Let $\{a_n\}$ be a sequence so that $\lim_{n\to\infty} a_n = l$. Then*

$$\lim_{n\to\infty} \frac{1}{n}(a_1 + \cdots + a_n) = l.$$

证明. Fix $\epsilon > 0$. Then there exists $N_1 = N_1(\epsilon) > 0$ such that

$$|a_n - l| < \frac{\epsilon}{2}$$

for any $n > N_1$. It follows that

$$\frac{1}{n}\sum_{k=N_1+1}^{n} |a_k - l| < \frac{\epsilon}{2}$$

for any $n > N_1$.

On the other hand, for sufficiently large $N_2 = N_2(N_1(\epsilon)) > 0$, we have

$$\frac{1}{n}\sum_{k=1}^{N_1} |a_k - l| \leq \frac{\epsilon}{2}$$

for any $n > N_2$.

Thus, for any $n > \max\{N_1, N_2\}$, we have

$$\left|\frac{1}{n}\sum_{k=1}^{n} a_k - l\right| \leq \frac{1}{n}\sum_{k=1}^{N_1} |a_k - l| + \frac{1}{n}\sum_{k=N_1+1}^{n} |a_k - l| < \epsilon.$$

Hence, we establish the proposition. □

In general, we have the following *Stolz-Cesàro theorem* which, can be viewed as a *L'Hôpital's rule* (Theorem 4.2.1) for sequences.

Theorem 1.3.22 (Stolz–Cesàro theorem)**.** *Let $\{a_n\}$ and $\{b_n\}$ be two sequences of $\mathbb{R}$ such that the following limit exists:*

$$\lim_{n\to\infty}\frac{a_{n+1}-a_n}{b_{n+1}-b_n}.$$

Suppose further that one of the following holds:

(1) $\{b_n\}$ is strictly monotone, $a_n\to 0$, $b_n\to 0$,

(2) $\{b_n\}$ is strictly monotone and divergent.

Then we have

$$\lim_{n\to\infty}\frac{a_n}{b_n}=\lim_{n\to\infty}\frac{a_{n+1}-a_n}{b_{n+1}-b_n}. \tag{1.3.3}$$

证明. Assume that $b_{n+1}-b_n>0$ for all n. Suppose that $\lim_{n\to\infty}\frac{a_{n+1}-a_n}{b_{n+1}-b_n}=\ell$. Then for any $\epsilon>0$, there is $N_1=N_1(\epsilon)>0$ such that

$$\frac{a_{n+1}-a_n}{b_{n+1}-b_n}\leq \ell+\epsilon$$

for $n\geq N_1$. Then for $n\geq m\geq N_1$, one calculates

$$\begin{aligned}a_{n+1}=a_m+\sum_{k=m}^{n}(a_{k+1}-a_k)&\leq a_m+(\ell+\epsilon)\sum_{k=m}^{n}(b_{k+1}-b_k)\\&=a_m-(\ell+\epsilon)b_m+(\ell+\epsilon)b_{n+1}.\end{aligned} \tag{1.3.4}$$

Then we discuss the two different situations:

(i) In the case (1), since b_n is increasing, $b_n<0$. Then we divide (1.3.4) by b_m:

$$\frac{a_m}{b_m}\leq(\ell+\epsilon)+\frac{a_{n+1}-(\ell+\epsilon)b_{n+1}}{b_m}$$

for any $n\geq m\geq N_1$. Letting $n\to\infty$, we see that

$$\frac{a_m}{b_m}\leq \ell+\epsilon$$

for $m\geq N_1$.

(ii) In the case (2), note that $b_n > 0$ for all n. Then we divide (1.3.4) by b_{n+1}:

$$\frac{a_{n+1}}{b_{n+1}} \leq (\ell + \epsilon) + \frac{a_m - (\ell + \epsilon)b_m}{b_{n+1}}$$

for any $n \geq m \geq N_1$. Choose sufficiently large $N_2 = N_2(N_1(\epsilon)) > 0$ so that

$$\frac{a_{N_1} - (\ell + \epsilon)b_{N_1}}{b_{n+1}} < \epsilon$$

for $n \geq N_2$. Then we have

$$\frac{a_{n+1}}{b_{n+1}} \leq \ell + 2\epsilon$$

for $n \geq \max\{N_1, N_2\}$.

Therefore, in both cases, there exists some $N = N(\epsilon) > 0$ such that

$$\frac{a_{n+1}}{b_{n+1}} \leq \ell + 2\epsilon$$

for $n \in N$. A similar argument shows that there exists $N' = N'(\epsilon) > 0$ such that

$$\frac{a_{n+1}}{b_{n+1}} \geq \ell - 2\epsilon$$

for $n \geq N'$. This implies (1.3.3). $\square$

Also, Proposition 1.3.21 is related to the *Dirichlet's test.*

Theorem 1.3.23 (Dirichlet's test, [2, §10.3 定理 4])**.** *Suppose that* $\{a_n\}$ *is a monotonic sequence of* $\mathbb{R}$ *with* $\lim_{n\to\infty} a_n = 0$*, and* $\{b_n\}$ *is a sequence of* $\mathbb{R}$ *so that* $\{\sum_{n=1}^m b_n\}_{m\geq 1}$ *is bounded. Then the series*

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

converges.

Example 1.3.24 (2016 Calculus B, Midterm)**.** Find

$$\lim_{n\to\infty} \frac{1^9 + \cdots + n^9}{n^{10}}.$$

证明. Using Theorem 1.3.22, we have

$$\lim_{n\to\infty}\frac{1^9+\cdots+n^9}{n^{10}}=\lim_{n\to\infty}\frac{n^9}{n^{10}-(n-1)^{10}}=\lim_{n\to\infty}\frac{n^9}{10n^9-45n^8+\cdots}=\frac{1}{10}.$$

Or consider the integral (see Section 2.8),

$$\lim_{n\to\infty}\frac{1^9+\cdots+n^9}{n^{10}}=\lim_{n\to\infty}\sum_{k=1}^{n}\left(\frac{k}{n}\right)^9\cdot\frac{1}{n}=\int_0^1 x^9dx=\frac{1}{10}.$$

□

Example 1.3.25. Find the limit

$$\lim_{n\to\infty}\frac{\sqrt[n]{n!}}{n}.$$

证明. Note that

$$\lim_{n\to\infty}\ln\frac{\sqrt[n]{n!}}{n}=\lim_{n\to\infty}\frac{1}{n}\ln n!-\ln n=\lim_{n\to\infty}\frac{1}{n}(\ln n!-n\ln n).$$

By letting $a_n=(\ln n!-n\ln n)-(\ln(n-1)!-(n-1)\ln(n-1))$, using Proposition 1.3.21, we have

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}(\ln n!-n\ln n)&=\lim_{n\to\infty}a_n\\&=\lim_{n\to\infty}(\ln n-n\ln n+(n-1)\ln(n-1))\\&=\lim_{n\to\infty}-(n-1)\ln\left(1+\frac{1}{n-1}\right)=-1.\end{aligned}$$

Thus, by the continuity of e^x, we have

$$\lim_{n\to\infty}\frac{\sqrt[n]{n!}}{n}=e^{-1}.$$

□

1.4 Limit of a function

1.4.1 One-sided limits

Definition 1.4.1 ([1, §1.4 定义 1])**.** Let $f(x)$ be a function on (a,b). Suppose there is l such that for any $\epsilon>0$, there is $\delta>0$ such that

$$|f(x)-l|<\epsilon$$

for $0 < x - a < \delta$. Then we say that $f(x)$ has the *right-sided limit* (右极限) l as $x \to a + 0$, and write

$$\lim_{x \to a+0} f(x) = l.$$

Or we say that $f(x)$ tends to l, as $x \to a + 0$, and write

$$f(x) \to l \quad (x \to a + 0).$$

It is called epsilon-delta definition ($\epsilon - \delta$ 说法). One can similarly define the *left-sided limit* (左极限).

1.4.2 Two-sided limits

Definition 1.4.2 ([1, §1.4 定义 2]). Let $f(x)$ be a function on a deleted neighborhood

$$\mathring{U}_r(a) = (a - r, a) \cup (a, a + r).$$

Suppose there is l such that for any $\epsilon > 0$, there is $\delta > 0$ such that

$$|f(x) - l| < \epsilon$$

for $0 < |x - a| < \delta$. Then we say that $f(x)$ has the *limit* l as $x \to a$, and write

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

Or we say that $f(x)$ tends to l, as $x \to a$, and write

$$f(x) \to l \quad (x \to a).$$

For a given f, if there is l so that $\lim_{x \to a} f(x) = l$, then we say f has a limit (is convergent) as $x \to a$ (at a).

The limit of a function exists only if the left- and right-sided limit exist and equal.

Example 1.4.3 ([1, §1.4 例 1]). $f(x) = \operatorname{sgn} x$ does not have limits at 0. In fact,

$$\lim_{x \to 0+0} \operatorname{sgn} x = 1, \quad \lim_{x \to 0-0} \operatorname{sgn} x = -1.$$

Remark 1.4.4. Note that the limit of f at a is not related to $f(a)$. For example, $\lim_{x\to 0} |\operatorname{sgn} x| = 1$, but $\operatorname{sgn} 0 = 0$.

Example 1.4.5 ([1, §1.4 例 3])**.** Prove that

$$\lim_{x\to 2} \frac{x^2 - x - 2}{x^2 - 4} = \frac{3}{4}.$$

证明. Observe that

$$\frac{x^2 - x - 2}{x^2 - 4} - \frac{3}{4} = \frac{(x+1)(x-2)}{(x+2)(x-2)} - \frac{3}{4} = \frac{x+1}{x+2} - \frac{3}{4} = \frac{x-2}{4(x+2)}$$

for $x \neq \pm 2$.

For $0 < |x-2| < 1$, we have

$$\left|\frac{x-2}{4(x+2)}\right| < \frac{|x-2|}{12}.$$

Thus, for any $\epsilon > 0$, choose $\delta = \min\{1, 12\epsilon\}$. Then

$$\left|\frac{x^2 - x - 2}{x^2 - 4} - \frac{3}{4}\right| < \frac{|x-2|}{12} < \frac{\delta}{12} \leq \epsilon$$

for any $0 < |x-2| < \delta$, as required. □

1.4.3 Theorems of limits of functions

Theorem 1.4.6 ([1, §1.4 定理 1])**.** *Let $f(x)$, $g(x)$, $h(x)$ be functions on some $\mathring{U}_r(a)$ that satisfy*

$$h(x) \leq f(x) \leq g(x).$$

If $\lim\limits_{x\to a} g(x) = \lim\limits_{x\to a} h(x) = l$, then $\lim\limits_{x\to a} f(x) = l$.

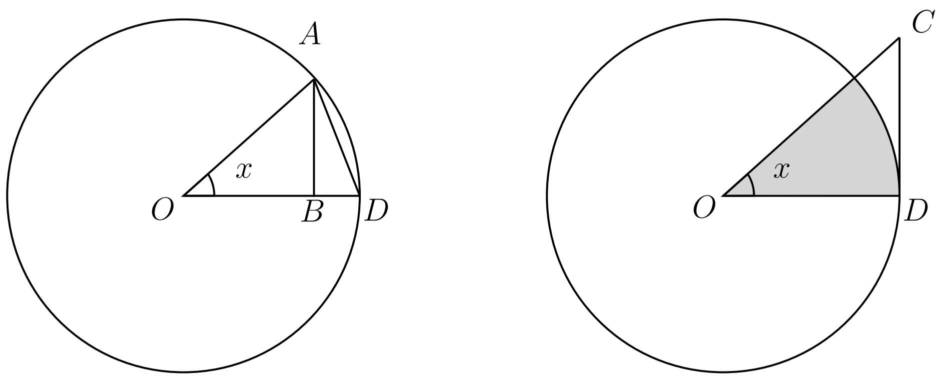
Example 1.4.7 (An important limit, [1, §1.4 例 5])**.** Prove that

$$\lim_{x\to 0} \frac{\sin x}{x} = 1.$$

证明. Considering the arcs, we have

$$\sin x = |AB| < |AD| < |\overset{\frown}{AD}| = x$$

for $x \in (0, \frac{\pi}{2})$.



Considering the areas, we have

$$\frac{1}{2}x = \text{Area(Shade Sector)} < \text{Area}(\Delta ODC) = \frac{1}{2}\tan x$$

for $x \in (0, \frac{\pi}{2})$. Then we have

$$\sin x < x < \frac{\sin x}{\cos x}$$

for $x \in (0, \frac{\pi}{2})$, i.e.

$$\cos x < \frac{\sin x}{x} < 1$$

for $x \in (0, \frac{\pi}{2})$.

This is also true for $x \in (-\frac{\pi}{2}, 0)$, since $\cos x$ and $\frac{\sin x}{x}$ are even. Thus, the consequence follows from $\lim_{x\to 0}\cos x = 1$ and Theorem 1.4.6. $\square$

Theorem 1.4.8 (Algebraic limit theorem, [1, §1.4 定理 2])**.** *Let $f(x)$ and $g(x)$ be functions on some $\mathring{U}_r(a)$. If*

$$\lim_{x\to a} f(x) = l_1, \quad \lim_{x\to a} g(x) = l_2,$$

then

$$\lim_{x\to a}(f(x) \pm g(x)) = l_1 \pm l_2, \quad \lim_{x\to a} f(x)g(x) = l_1 l_2,$$

and for $l_2 \neq 0$, we have

$$\lim_{x\to a}\frac{f(x)}{g(x)} = \frac{l_1}{l_2}.$$

Example 1.4.9 ([1, §1.4 例 7])**.** Find

$$\lim_{x\to 0}\frac{\sqrt{1+x}-1}{x+\sin x}.$$

证明. Observe that

$$\frac{\sqrt{1+x}-1}{x+\sin x}=\frac{\frac{\sqrt{1+x}-1}{x}}{\frac{x+\sin x}{x}}=\frac{\frac{\sqrt{1+x}-1}{x}}{1+\frac{\sin x}{x}}.$$

Note that

$$\lim_{x\to 0} 1+\frac{\sin x}{x}=2,$$

and

$$\frac{\sqrt{1+x}-1}{x}=\frac{(\sqrt{1+x}-1)(\sqrt{1+x}+1)}{x(\sqrt{1+x}+1)}=\frac{1}{\sqrt{1+x}+1}\to\frac{1}{2}$$

as $x\to 0$. Then by Theorem 1.4.8, we get

$$\lim_{x\to 0}\frac{\sqrt{1+x}-1}{x+\sin x}=\frac{\frac{1}{2}}{2}=\frac{1}{4}.$$

□

Theorem 1.4.10 ([1, §1.4 定理 3]). *Let $f(x)$ and $g(x)$ be functions on a deleted neighborhood $\mathring{U}_r(a)$, and*

$$\lim_{x\to a} f(x)=l_1,\quad \lim_{x\to a} g(x)=l_2.$$

Also, $l_1>l_2$. Then there is $\delta>0$ such that

$$f(x)>g(x)$$

for $0<|x-a|<\delta$.

Theorem 1.4.11 ([1, §1.4 定理 4]). *Let $f(x)$ and $g(x)$ be functions on a deleted neighborhood $\mathring{U}_r(a)$ that satisfy $f(x)\geq g(x)$. Suppose that as $x\to a$, $f(x)$ and $g(x)$ both have limits. Then*

$$\lim_{x\to a} f(x)\geq\lim_{x\to a} g(x).$$

Theorem 1.4.12 ([1, §1.4 定理 5]). *Let $f(x)$ be a function on some $\mathring{U}_r(a)$ such that*

$$\lim_{x\to a} f(x)=l. \tag{1.4.1}$$

If $\{x_n\}\subset\mathring{U}_r(a)$, and

$$\lim_{n\to\infty} x_n=a, \tag{1.4.2}$$

then we have

$$\lim_{n\to\infty} f(x_n)=l.$$

证明. By (1.4.1), for any $\epsilon > 0$, there is $\delta > 0$ such that

$$|f(x) - l| < \epsilon$$

for $0 < |x - a| < \delta$. In particular, we have

$$|f(x_n) - l| < \epsilon$$

for $0 < |x_n - a| < \delta$.

By (1.4.2), for $\delta > 0$, there is $N > 0$ such that

$$|x_n - a| < \delta$$

for $n > N$. Since $x_n \neq a$, we further have $0 < |x_n - a| < \delta$. Then we have

$$|f(x_n) - l| < \epsilon$$

for $n > N$, as required. □

Example 1.4.13 ([1, §1.4 例 9])**.** Show that $\sin \frac{1}{x}$ does not have limits as $x \to 0$.

证明. Let $x'_n = \frac{1}{2n\pi}$. Then $\sin \frac{1}{x'_n} = 0$ and so

$$\lim_{n\to\infty} \sin \frac{1}{x'_n} = 0.$$

Let $x''_n = \frac{1}{2n\pi + \frac{\pi}{2}}$. Then $\sin \frac{1}{x''_n} = 1$ and so

$$\lim_{n\to\infty} \sin \frac{1}{x''_n} = 1.$$

Then by (1.4.12), we conclude that $\lim_{x\to 0} \sin \frac{1}{x}$ does not exist. □

1.4.4 Variable goes to infinity

There are three possible situations as x goes to infinity:

- $x \to +\infty$,
- $x \to -\infty$,

- $x \to \infty$, it means both $x \to +\infty$ and $x \to -\infty$.

Remark 1.4.14. In the theory of order and topology,

- ∞ is the *one-point compactification* 单点紧化 (or *Alexandroff compactification*) of $\mathbb{R}$. $\mathbb{R} \cup \{\infty\}$ is called the *projectively real number system*. (Ideally, it is just a circle!) However, it breaks the algebraic structure (e.g. $0 \times \infty = ?$), and the order structure ($1 < \infty < -1$?) of $\mathbb{R}$. Only topologists are satisfied.

- $\pm\infty$ is another compactification of $\mathbb{R}$. $\mathbb{R} \cup \{\pm\infty\}$ is called the *extended real number system* (扩张实数系). It is the *Dedekind-MacNeille completion* of the real numbers. $\mathbb{R} \cup \{\pm\infty\}$ preserves the order structure of $\mathbb{R}$, but it still breaks the algebraic structure ($-\infty + (+\infty) = ?$) of $\mathbb{R}$. Algebraists are not happy.

Definition 1.4.15 ([1, §1.4 定义 3])**.** Let $f(x)$ be a function on $(a, +\infty)$. If there is a constant $l \in \mathbb{R}$ such that for any $\epsilon > 0$, there is $A > a$, such that

$$|f(x) - l| < \epsilon,$$

for $x > A$, then we say l is the *limit* of $f(x)$ as $x \to +\infty$, and write

$$\lim_{x \to +\infty} f(x) = l.$$

Remark 1.4.16. One can similarly define $\lim\limits_{x \to -\infty} f(x)$.

Definition 1.4.17 ([1, §1.4 定义 4])**.** Let $f(x)$ be functions on $(-\infty, -a) \cup (a, \infty)$. If there is a constant $l \in \mathbb{R}$ such that for any $\epsilon > 0$, there is $A > a$, such that

$$|f(x) - l| < \epsilon,$$

for $|x| > A$, then we say l is the *limit* of $f(x)$ as $x \to \infty$, and write

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

Remark 1.4.18. $\lim\limits_{x \to \infty} f(x)$ exists if and only if $\lim\limits_{x \to +\infty} f(x)$ and $\lim\limits_{x \to -\infty} f(x)$ exist and equal.

Example 1.4.19.

- $\lim\limits_{x\to+\infty} e^x$ does not exist, and $\lim\limits_{x\to-\infty} e^x = 0$. Thus, $\lim\limits_{x\to\infty} e^x$ does not exist.
- $\lim\limits_{x\to+\infty} \operatorname{sgn} x = 1$, $\lim\limits_{x\to-\infty} \operatorname{sgn} x = -1$, and so $\lim\limits_{x\to\infty} \operatorname{sgn} x$ does not exist.

Remark 1.4.20. If $\lim\limits_{x\to\infty} f(x)$ exists, then $\lim\limits_{y\to 0} f\left(\frac{1}{y}\right)$ exists, and

$$\lim_{x\to\infty} f(x) = \lim_{y\to 0} f\left(\frac{1}{y}\right);$$

vice versa.

The theorems of limits also work for x going to infinity.

Example 1.4.21 ([1, §1.4 例 13])**.** Show that $\lim\limits_{x\to\infty}\left(1+\frac{1}{x}\right)^x = e$.

证明. First, note that for $x > 1$, we have

$$\left(1+\frac{1}{[x]+1}\right)^{[x]} \le \left(1+\frac{1}{x}\right)^x \le \left(1+\frac{1}{[x]}\right)^{[x]+1}.$$

By the squeeze theorem (Theorem 1.3.8), we have

$$\lim_{x\to+\infty}\left(1+\frac{1}{x}\right)^x = e.$$

Now consider the case when $x\to-\infty$. Let $y = -x$, then $y\to+\infty$. Note that

$$\left(1+\frac{1}{x}\right)^x = \left(1-\frac{1}{y}\right)^{-y} = \left(\frac{y-1}{y}\right)^{-y} = \left(1+\frac{1}{y-1}\right)^y.$$

We conclude that

$$\lim_{x\to-\infty}\left(1+\frac{1}{x}\right)^x = \lim_{y\to+\infty}\left(1+\frac{1}{y-1}\right)^y = e.$$

Thus, since $\lim\limits_{x\to+\infty}\left(1+\frac{1}{x}\right)^x = \lim\limits_{x\to-\infty}\left(1+\frac{1}{x}\right)^x = e$, we conclude that $\lim\limits_{x\to\infty}\left(1+\frac{1}{x}\right)^x$ exists and equals e. $\square$

1.4.5 Infinitely large function

Definition 1.4.22 ([1, §1.4 定义 5])**.** Let $f(x)$ be a function on $\mathring{U}_r(x_0)$. If for any $M > 0$, there is $\delta > 0$, such that

$$|f(x)| > M,$$

for $0 < |x - x_0| < \delta$, then we say $f(x)$ is *infinitely large* as $x \to x_0$, and write

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

Remark 1.4.23. Note that if $f(x)$ is infinitely large as $x \to x_0$, then the limit $\lim\limits_{x\to\infty} f(x)$ does not exist.

1.5 Continuous function

1.5.1 Definitions

Definition 1.5.1 ([1, §1.5 定义 1])**.** Let $f(x)$ be a function on (a, b). Suppose for $x_0 \in (a, b)$, there is $l \in \mathbb{R}$ such that $\lim\limits_{x\to x_0} f(x) = l$, and $f(x_0) = l$. Then we say that f is *continuous* at x_0. If f is continuous at any point in (a, b), then we say f is *continuous* on (a, b), or f is a *continuous function* on (a, b).

Remark 1.5.2. In terms of the epsilon-delta definition, a function f is continuous at x_0 if and only if for any $\epsilon > 0$, there is $\delta > 0$, such that

$$|f(x) - f(x_0)| < \epsilon$$

for $|x - x_0| < \delta$.

Example 1.5.3 (Lipschitz condition, [1, §1.5 例 2])**.** Let $L > 0$, and let $f : (a, b) \to \mathbb{R}$ satisfy

$$|f(x_1) - f(x_2)| \leq L|x_1 - x_2|$$

for any $x_1, x_2 \in (a, b)$. Then f is continuous on (a, b).

证明. For any $x_0 \in (a,b)$, $\epsilon > 0$, let $\delta = \frac{\epsilon}{L}$. Then for $|x - x_0| < \delta$, we have

$$|f(x) - f(x_0)| \leq L|x - x_0| < \epsilon,$$

as desired. □

Note also that we can define the function f is *right continuous* at x_0 if

$$\lim_{x \to x_0+0} f(x) = f(x_0);$$

is *left continuous* at x_0 if

$$\lim_{x \to x_0-0} f(x) = f(x_0).$$

Definition 1.5.4 ([1, §1.5 定义 2])**.** A function $f : [a,b] \to \mathbb{R}$ is *continuous* on $[a,b]$ if it is continuous (a,b), and right continuous at a, and left continuous at b.

Example 1.5.5. The function $f(x) = x - [x]$ is continuous on $[0, \frac{1}{2}]$, but not continuous on $[0,1]$.

Theorem 1.5.6 ([1, §1.5 定理 1])**.** *Let f and g be functions on $U_r(x_0)$ that are continuous at x_0. Then $f(x) \pm g(x)$, $f(x)g(x)$, and $\frac{f(x)}{g(x)}$ $(g(x_0) \neq 0)$ are continuous at x_0.*

1.5.2 Composition

Theorem 1.5.7 ([1, §1.5 定理 2])**.** *Let $f : (a,b) \to (c,d)$ be continuous at x_0, and $g : (c,d) \to \mathbb{R}$ be continuous at $f(x_0)$. Then the composition $g \circ f$ is continuous at x_0.*

证明. For any $\epsilon > 0$, there is $\delta_1 > 0$ such that

$$|g(y) - g(f(x_0))| < \epsilon$$

for $|y - f(x_0)| < \delta_1$. Next, for $\delta_1 > 0$, there is $\delta(\delta_1) > 0$ such that

$$|f(x) - f(x_0)| < \delta_1$$

for $|x - x_0| < \delta$. It follows that

$$|g(f(x)) - g(f(x_0))| < \epsilon$$

for $|x - x_0| < \delta$. □

Theorem 1.5.8 ([1, §1.5 定理 3])**.** *Let $g : (c, d) \to \mathbb{R}$ be continuous at y_0. Suppose that*

$$\lim_{x \to a} f(x) = y_0.$$

Then

$$\lim_{x \to a} g(f(x)) = g(y_0) = g(\lim_{x \to a} f(x)).$$

Here a can be replaced by $+\infty$, $-\infty$, or ∞.

Example 1.5.9 ([1, §1.5 例 5])**.** Show that $\lim_{x \to +\infty} \sin(\sqrt{x+1} - \sqrt{x}) = 0$.

证明. Note that

$$\lim_{x \to +\infty} \sqrt{x+1} - \sqrt{x} = \lim_{x \to +\infty} \frac{1}{\sqrt{x+1} + \sqrt{x}} = 0.$$

Then by the continuity of sine functions:

$$\lim_{x \to +\infty} \sin(\sqrt{x+1} - \sqrt{x}) = \sin(\lim_{x \to +\infty} \sqrt{x+1} - \sqrt{x}) = \sin(0) = 0.$$

□

1.5.3 Inverse functions

Definition 1.5.10 ([1, §1.5 定义 3])**.** If $f : X \to Y$ satisfies

$$f(x_1) \leq f(x_2)$$

whenever $x_1 < x_2$, then we say f is *monotonically increasing*. If f satisfies

$$f(x_1) < f(x_2)$$

whenever $x_1 < x_2$, then we say f is *strictly increasing*. One can similarly define *decreasing* and *strictly decreasing* function. We say a function is (strictly) *monotone* if it is (strictly) increasing or decreasing.

Example 1.5.11. $\sin x$ and $\cos x$ are strictly monotone on $(-\pi/2, \pi/2)$ and $(0, \pi)$, respectively.

Theorem 1.5.12 ([1, §1.5 定理 4]). *Let $f : (a, b) \to (c, d)$ be bijective, and strictly monotone. Then f is continuous on (a, b), and its inverse f^{-1} is continuous on (c, d).*

证明. For any $x_0 \in (a, b)$, write $y_0 = f(x_0)$. Fix any $\epsilon > 0$ with

$$c < y_0 - \epsilon < y_0 < y_0 + \epsilon < d.$$

Without loss of generality, we assume that f is strictly increasing. Let

$$x_1 = f^{-1}(y_0 - \epsilon), \quad x_2 = f^{-1}(y_0 + \epsilon).$$

Then $x_1 < x_0 < x_2$. Moreover, by the monotonicity of f, we have

$$y_0 - \epsilon < f(x) < y_0 + \epsilon$$

for $x_1 < x < x_2$. Choose $\delta = \min\{x_2 - x_0, x_0 - x_1\}$. Then $U_\delta(x_0) \subset (x_1, x_2)$, and so

$$|f(x) - f(x_0)| < \epsilon$$

for $|x - x_0| < \delta$. Thus, f is continuous at x_0. Since x_0 is chosen arbitrarily, f is continuous on (a, b).

Note that if f is strictly monontone, then f^{-1} is strictly monotone. Thus, f^{-1} is continuous on (c, d). □

Remark 1.5.13. The open intervals can be replaced by closed intervals etc.

Example 1.5.14 ([1, §1.5 例 7]). $\arcsin x$ is continuous.

Example 1.5.15 ([1, §1.5 例 8]). Since $f(x) = a^x$ $(a > 0,\ a \neq 1)$ is monotone, a^x and $\log_a x$ are continuous on $\mathbb{R}$ and $(0, +\infty)$.

Example 1.5.16 ([1, §1.5 例 9]). Find $\lim_{x\to+\infty}(1 + \frac{1}{x})^{x+\sqrt{x}}$.

证明. Observe that

$$\left(1 + \frac{1}{x}\right)^{x+\sqrt{x}} = e^{(x+\sqrt{x})\ln\left(1+\frac{1}{x}\right)} = e^{(1+\frac{1}{\sqrt{x}})x\ln\left(1+\frac{1}{x}\right)}.$$

Now by the continuity of logarithms,

$$\lim_{x\to+\infty} x\ln\left(1+\frac{1}{x}\right) = \lim_{x\to+\infty} \ln\left(1+\frac{1}{x}\right)^x = \ln\lim_{x\to+\infty}\left(1+\frac{1}{x}\right)^x = \ln e = 1.$$

Then we have

$$\lim_{x\to+\infty}(1+\frac{1}{\sqrt{x}})x\ln\left(1+\frac{1}{x}\right) = 1.$$

Then by the continuity of exponential functions,

$$\left(1+\frac{1}{x}\right)^{x+\sqrt{x}} = e^{(x+\sqrt{x})\ln\left(1+\frac{1}{x}\right)}$$

□

Powers of x: $x^a = e^{a\ln x}$ is a composition of continuous functions, and so is continuous.

Thus, we conclude that elementary functions are continuous on every interval.

Remark 1.5.17. The domain of an elementary function can have isolated points, e.g. $f(x) = \sqrt{x^2(x^2-1)}$ is defined on $(-\infty,-1]\cup\{0\}\cup[1,+\infty)$. Since we did not define the continuity at isolated points, we only say elementary functions are continuous on every interval.

1.5.4 Discontinuities

If $f(x)$ is not continuous at x_0, then we say x_0 is *discontinuity* of f.

(1) If $\lim_{x\to x_0+0} f(x)$ and $\lim_{x\to x_0-0} f(x)$ exist, but they are not equal, or not equal to $f(x_0)$ (even if they are equal), then x_0 is called a *discontinuity of the first kind* (第一类间断点). There are two possibilities:

- If they are not equal, then x_0 is called a *jump discontinuity*.
- If they are equal but not equal to $f(x_0)$, then x_0 is *removable discontinuity* (可去间断点).

(2) If one of $\lim_{x\to x_0+0} f(x)$ and $\lim_{x\to x_0-0} f(x)$ does not exist, then x_0 is called a *discontinuity of the second kind* (第二类间断点), or an *essential discontinuity*.

Example 1.5.18 ([1, §1.5 例 10])**.** Find the discontinuities of

$$f(x)=\begin{cases}\frac{\sin x}{x(1-x)} & ,\ \text{if } x\neq 0,1\\ -1 & ,\ \text{if } x=0\\ 1 & ,\ \text{if } x=1\end{cases}.$$

If there is a removable discontinuity, modify the function so that f is continuous at there.

证明. Note that

$$\lim_{x\to 0}\frac{\sin x}{x(1-x)}=\lim_{x\to 0}\frac{\sin x}{x}\cdot\frac{1}{1-x}=1.$$

But $f(0)=-1$. Thus, $x=0$ is a removable discontinuity. Modify $f(0)=1$. Then f is continuous at $x=0$.

On the other hand, since

$$\lim_{x\to 1+0}\frac{\sin x}{x(1-x)}=-\infty,$$

we conclude that $x=1$ is a discontinuity of the second kind. □

Example 1.5.19 (2014 Calculus B, Midterm)**.** Discuss the continuity of the function $f(x)=\dfrac{\ln(e^n+x^n)}{n}$.

证明. Let $m=\max\{e,x\}$. Then we have

$$\ln m=\frac{\ln m^n}{n}\leq\frac{\ln(e^n+x^n)}{n}\leq\frac{\ln 2m^n}{n}=\ln m+\frac{\ln 2}{n}\to\ln m\quad(n\to\infty).$$

By the squeeze theorem (Theorem 1.3.8), we have

$$f(x)=\ln\max\{e,x\}.$$

It is continuous on $[0,\infty)$. (See e.g. [1, P.75, Exercise 19].) □

Example 1.5.20 (2017 Calculus B, Midterm)**.** Let f be an increasing function on $[a,b]$ so that $f([a,b])=[f(a),f(b)]$. Discuss the continuity of f.

证明. Let $x_0 \in (a, b]$. Since f is increasing and bounded, the left limit exists

$$\lim_{x \to x_0 - 0} f(x) = A.$$

Thus, $f(a) \le f(x) \le A$ for $x \in [a, x_0)$. Now suppose that $f(x_0) > A$. Then pick $B = \frac{1}{2}(A + f(x_0)) \in [f(a), f(b)]$, and we have

$$f(x) \le A < B$$

for $x \in [a, x_0)$ and

$$f(x) \le f(x_0) > B$$

for $x \in [x_0, b]$. Thus, $B \notin f([a, b])$ and $B \in [f(a), f(b)]$, which leads to a contradiction. Thus, $f(x_0) = A$, and f is left continuous at $x = x_0$.

Similar idea shows that f is right continuous at $x = x_0$, and so f is continuous on $[a, b]$. □

1.6 Continuous functions on closed intervals

The theorems in this section are the consequences of the *completeness* of $\mathbb{R}$.

Theorem 1.6.1 (Intermediate value theorem, [1, §1.6 定理 1])**.** *Let $f : [a, b] \to \mathbb{R}$ be a continuous function with $f(a) \neq f(b)$. Then for any value η between $f(a)$ and $f(b)$, there is $\xi \in (a, b)$, such that*

$$f(\xi) = \eta.$$

证明. Suppose that $f(a) < f(b)$ and $\eta \in (f(a), f(b))$. Let

$$E = \{x \in [a, b] : f(x) < \eta\}.$$

Clearly, $a \in E$, we see that E is nonempty. By Theorem 7.1.7, $\sup(E)$ exists. Let

$$c = \sup(E).$$

Since b is an upper bound of E, $c \le b$. Since $a \in E$, $c \ge a$ and so $c \in [a, b]$.

Now for any $n \geq 1$, there exists $x_n \in E$ such that

$$c - \frac{1}{n} \leq x_n \leq c.$$

By Theorem 1.3.8, we have $x_n \to c$ as $n \to \infty$. Thus, by the continuity of f, we have

$$\lim_{n\to\infty} f(x_n) = f(c).$$

Since $x_n \in E$, by Theorem 1.4.11, we conclude that $f(c) \leq \eta$.

Now since c is the supremum of E, we have $c + \frac{1}{n} \in E^c \cap [a, b]$ for sufficiently large n. Then by the continuity of f, we have

$$\lim_{n\to\infty} f\left(c + \frac{1}{n}\right) = f(c).$$

Since $c + \frac{1}{n} \in E^c$, we have $f\left(c + \frac{1}{n}\right) \geq \eta$. By Theorem 1.4.11, we conclude that $f(c) \geq \eta$. Therefore, $f(c) = \eta$. □

Example 1.6.2 ([1, §1.6 例 1])**.** Prove that $x - \cos x = 0$ has a unique solution on $(0, \frac{\pi}{2})$.

证明. Let $f(x) = x - \cos x$. Then $f(x)$ is continuous on $[0, \frac{\pi}{2}]$, and

$$f(0) = -1, \quad f\left(\frac{\pi}{2}\right) = \frac{\pi}{2}.$$

Let $\eta = 0$. Then $f(0) < \eta < f\left(\frac{\pi}{2}\right)$. By the intermediate value theorem (Theorem 1.6.1), there is $\xi \in \left(0, \frac{\pi}{2}\right)$ such that

$$f(\xi) = \eta = 0.$$

That is ξ is the solution of $x - \cos x = 0$ on $\left(0, \frac{\pi}{2}\right)$.

Suppose there is another solution $\xi' \in \left(0, \frac{\pi}{2}\right)$. Then

$$\xi - \cos\xi = \xi' - \cos\xi' = 0 \quad \Longleftrightarrow \quad \xi - \xi' = \cos\xi - \cos\xi'.$$

Assume $\xi > \xi'$. Then $\xi - \xi' > 0$, but $\cos\xi - \cos\xi' < 0$, a contradiction. Thus, ξ is the unique solution. □

Theorem 1.6.3 (Extreme value theorem, [1, §1.6 定理 2]). *Let $f : [a, b] \to \mathbb{R}$ be a continuous function. Then f must attain a maximum, and a minimum, each at least once. That is, there are $x_1, x_2 \in [a, b]$ such that*

$$f(x_1) \leq f(x) \leq f(x_2)$$

for any $x \in [a, b]$.

Theorem 1.6.3 is more specific than the related boundedness theorem.

Theorem 1.6.4 (Boundedness theorem, [1, §1.6 定理 3]). *Let f be a continuous function on $[a, b]$. Then f is bounded.*

证明. Suppose that f does not have an upper bound. Then we can find a sequence $x_n \in [a, b]$, such that

$$f(x_n) > n.$$

Since $\{x_n\} \subset [a, b]$ is bounded, by Theorem 7.1.8, we can find a subsequence $\{x_{n_k}\}$ so that $x_{n_k} \to \ell \in [a, b]$ $(k \to \infty)$. Then by the continuity of f,

$$\lim_{k\to\infty} f(x_{n_k}) = f(\ell) \neq +\infty.$$

A contradiction. Thus, we see that f is bounded on $[a, b]$. □

Proof of Theorem 1.6.3. Consider the image

$$E = \{f(x) : x \in [a, b]\}.$$

By Theorem 1.6.4, we see that E is bounded. Thus, by Theorem 7.1.7, $\sup(E)$ exists. Then for any $n \in \mathbb{N}^*$, there exists $y_n \in E$ such that

$$\sup(E) - \frac{1}{n} \leq y_n \leq \sup(E). \tag{1.6.1}$$

Then by the definition of E, there exists $x_n \in [a, b]$ such that $f(x_n) = y_n$. By Theorem 7.1.8, we can find a subsequence $\{x_{n_k}\}$ so that $x_{n_k} \to \ell \in [a, b]$ $(k \to \infty)$. Then by the continuity of f, we conclude

$$\lim_{k\to\infty} f(x_{n_k}) = f(\ell).$$

By Theorem 1.3.8 and (6.3.1), we have $f(\ell) = \sup(E)$. One can similarly discuss the infima. □

Theorems 1.6.1 and 1.6.3 imply that the images of continuous functions on closed intervals are closed intervals.

Example 1.6.5 ([1, §1.6 例 2])**.** For $a_0, \ldots, a_5 \in \mathbb{R}$ with $a_0 > 0$, the polynomial

$$P(x) = a_0x^5 + a_1x^4 + \cdots + a_5$$

has a real root.

证明. Write

$$P(x) = x^5 \left(a_0 + \frac{a_1}{x} + \cdots + \frac{a_5}{x^5}\right)$$

for $|x| > 0$. For $x \to \infty$, we have

$$a_0 + \frac{a_1}{x} + \cdots + \frac{a_5}{x^5} \to a_0 > 0.$$

It follows that for sufficiently large $|x|$, $P(x)/x^5 = a_0 + \frac{a_1}{x} + \cdots + \frac{a_5}{x^5} > 0$. That is, $P(x)$ and x^5 have the same sign. Thus, we have

$$P(x) \to +\infty \ (x \to +\infty), \quad P(x) \to -\infty \ (x \to -\infty).$$

Thus, there is $b \in \mathbb{R}$ with $P(b) > 0$, and there is $a \in \mathbb{R}$ with $P(a) < 0$. Since $P(x)$ is continuous on $\mathbb{R}$, by the intermediate value theorem, there is c between a and b, such that $P(c) = 0$, i.e. $P(x)$ has a real root. $\square$

Example 1.6.6 ([1, §1.6 例 3])**.** Let f be continuous on (a, b). If f is injective, then f is strictly monotone.

证明. Assume that f is not strictly monotone. Then there are $x_1, x_2, x_3 \in [a, b]$ with $x_1 < x_2 < x_3$ such that

$$f(x_2) < \min\{f(x_1), f(x_3)\}, \qquad \text{or} \qquad f(x_2) > \max\{f(x_1), f(x_3)\}.$$

for any $x \in [a, b]$.

Then by the intermediate value theorem (Theorem 1.6.1), for any η between $f(x_2)$ and $\min\{f(x_1), f(x_3)\}$, or between $f(x_2)$ and $\max\{f(x_1), f(x_3)\}$, there are $\xi_1 \in (x_1, x_2)$ and $\xi_2 \in (x_2, x_3)$ such that

$$\eta = f(\xi_1), \quad \eta = f(\xi_2).$$

Since $\xi_1 \neq \xi_2$, f is not injective, which leads to a contradiction. $\square$

By Theorem 1.5.12 in the last section and Example 1.6.6, we obtain

Theorem 1.6.7 ([1, §1.6 定理 4])**.** *Let* $f : (a, b) \to (c, d)$ *be continuous. If* $f(x)$ *is bijective, then* f^{-1} *is continuous on* (c, d)*.*

Chapter 2

Basic Concepts of Calculus

2.1 Derivative

2.1.1 Definitions

Definition 2.1.1 ([1, §2.1 定义])**.** Let $f(x)$ be a function on (a, b). Suppose for $x_0 \in (a, b)$, there is $L \in \mathbb{R}$ such that

$$\lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} = L.$$

Then we say f is *differentiable* at x_0, and the limit L is called the *derivative* of f at x_0, and we write

$$L = f'(x_0) = \frac{df}{dx}(x_0) = f'|_{x=x_0}.$$

If f is differentiable for all $x_0 \in (a, b)$, then f' is a function on (a, b), called the *derivative function*, or the *derivative* of f.

Remark 2.1.2. Similarly, one can define the *left/right derivatives* of a function. Clearly, a function f is differentiable at x_0 if and only if f has left and right derivatives and they are equal.

Example 2.1.3. Consider the derivative of $f(x) = |x|$ at $x = 0$.

证明. The right derivative is

$$\lim_{\Delta x \to 0+0} \frac{|\Delta x| - 0}{\Delta x} = 1.$$

The left derivative is

$$\lim_{\Delta x \to 0-0} \frac{|\Delta x| - 0}{\Delta x} = -1.$$

Thus, f is not differentiable at $x = 0$. □

Remark 2.1.4. Let

$$\eta(\Delta x) = \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} - f'(x_0). \tag{2.1.1}$$

In terms of the epsilon-delta definition, a function f is differentiable at x_0 if and only if for any $\epsilon > 0$, there is $\delta > 0$, such that

$$|\eta(\Delta x)| < \epsilon$$

for $0 < |\Delta x| < \delta$.

By (2.1.1), it means that

$$f(x_0 + \Delta x) - f(x_0) = f'(x_0)\Delta x + \eta(\Delta x)\Delta x. \tag{2.1.2}$$

Thus, we see that the difference $f(x_0 + \Delta x) - f(x_0)$ is **locally approximated by a linear map** $f'(x_0)\Delta x$. Moreover, since $\eta(\Delta x)\Delta x$ tends to 0 faster than $f'(x_0)\Delta x$ as $\Delta x \to 0$, $f'(x_0)\Delta x$ is the **best linear approximation** to f at x_0.

Similarly, for $m, n \in \mathbb{N}$, $f : \mathbb{R}^m \to \mathbb{R}^n$, we can consider the best linear approximation: For $\overrightarrow{x_0} \in \mathbb{R}^m$, find some linear map $L : \mathbb{R}^m \to \mathbb{R}^n$ so that

$$f(\overrightarrow{x_0} + \overrightarrow{\Delta x}) - f(\overrightarrow{x_0}) \sim L(\overrightarrow{\Delta x})$$

for $\overrightarrow{\Delta x} \in \mathbb{R}^m$ with $\overrightarrow{\Delta x} \to 0$. It is the origin of *differential geometry*, *Riemannian geometry*, etc.

Proposition 2.1.5. *If f is differentiable at x_0, then f is continuous at x_0.*

证明. Let $\eta(\Delta x)$ be as in (2.1.1). Then since f is differentiable at x_0, $\eta(\Delta x) \to 0$ as $\Delta x \to 0$. Then by (2.1.2), we have

$$f(x_0 + \Delta x) - f(x_0) = f'(x_0)\Delta x + \eta(\Delta x)\Delta x \to 0$$

as $\Delta x \to 0$. Thus, f is continuous at x_0. □

Theorem 2.1.6 (Weierstrass function 魏尔斯特拉斯函数)**.** *There is a function on* $\mathbb{R}$ *that is continuous everywhere, but differentiable nowhere.*

Example 2.1.7 ([1, §2.1 例 3])**.** Let $C \in \mathbb{R}$ and $f(x) = C$. Then $f'(x) = 0$.

Conversely, if a function $f : (a, b) \to \mathbb{R}$ has $f'(x) = 0$ for all $x \in (a, b)$, then $f(x) = C$ for some $C \in \mathbb{R}$. The rigorous proof follows from the *Lagrange's mean value theorem* (see Corollary 4.1.7).

Example 2.1.8 ([1, §2.1 例 4])**.** $(\sin x)' = \cos x$.

证明. Note that

$$\sin(x + \Delta x) - \sin x = 2 \sin \frac{\Delta x}{2} \cos\left(x + \frac{\Delta x}{2}\right).$$

Then

$$\begin{aligned}
\lim_{\Delta x \to 0} \frac{\sin(x + \Delta x) - \sin x}{\Delta x} &= \lim_{\Delta x \to 0} \frac{2 \sin \frac{\Delta x}{2} \cos\left(x + \frac{\Delta x}{2}\right)}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{\sin \frac{\Delta x}{2}}{\frac{\Delta x}{2}} \lim_{\Delta x \to 0} \cos\left(x + \frac{\Delta x}{2}\right) \\
&= \cos x.
\end{aligned}$$

□

Example 2.1.9 ([1, §2.1 例 7])**.** $(e^x)' = e^x$.

证明. Note that

$$e^{x+\Delta x} - e^x = e^x(e^{\Delta x} - 1).$$

Let $\eta = e^{\Delta x} - 1$. Then $\eta \to 0$ as $\Delta x \to 0$, and $\Delta x = \ln(1 + \eta)$. Then

$$\begin{aligned}
\lim_{\Delta x \to 0} \frac{e^{x+\Delta x} - e^x}{\Delta x} &= e^x \lim_{\Delta x \to 0} \frac{e^{\Delta x} - 1}{\Delta x} \\
&= e^x \lim_{\eta \to 0} \frac{\eta}{\ln(1 + \eta)} = e^x \lim_{\eta \to 0} \frac{1}{\ln(1 + \eta)^{\frac{1}{\eta}}} = e^x.
\end{aligned}$$

□

Similarly, we have

Example 2.1.10 ([1, §2.1 例 8])**.** $(a^x)' = a^x \ln a$.

Derivatives are limits! e.g. [1, P.75, Exercises 14, 17(1)(2)].

Example 2.1.11 ([1, P.75, 习题 14])**.** $\lim_{n\to\infty} n(a^{\frac{1}{n}} - 1) = (a^x)'|_{x=0} = \ln a$.

2.1.2 Elementary arithmetic of derivatives

Theorem 2.1.12 ([1, §2.1 定理]). *Let f and g be differentiable functions on (a,b). Then for $x \in (a,b)$, we have*

$$(f \pm g)'(x) = f'(x) \pm g'(x),$$

$$(f \cdot g)'(x) = f'(x)g(x) + f(x)g'(x),$$

and for $g(x) \neq 0$, we have

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

证明. Only prove the quotient. For $g(x) \neq 0$, we have

$$\begin{aligned}
&\frac{f(x+\Delta x)}{g(x+\Delta x)} - \frac{f(x)}{g(x)} \\
=&\frac{f(x+\Delta x)g(x) - f(x)g(x+\Delta x)}{g(x+\Delta x)g(x)} \\
=&\frac{(f(x+\Delta x) - f(x))g(x) - f(x)(g(x+\Delta x) - g(x))}{g(x+\Delta x)g(x)}.
\end{aligned}$$

The consequence follows immediately. □

Example 2.1.13 ([1, §2.1 例 11]). Find $(\tan x)'$.

证明. One calculates

$$(\tan x)' = \left(\frac{\sin x}{\cos x}\right)' = \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = \frac{1}{\cos^2 x}.$$

□

2.2 Composition and inverse function

Theorem 2.2.1 ([1, §2.2 定理 1]). *Let $f : (a,b) \to (A,B)$ be differentiable at x_0, and $g : (A,B) \to \mathbb{R}$ be differentiable at $y_0 = f(x_0)$. Then the composition $g \circ f$ is differentiable at x_0, and*

$$(g \circ f)'(x_0) = g'(f(x_0))f'(x_0).$$

证明. For $\Delta y \neq 0$, let

$$\eta(\Delta y) = \frac{g(y_0 + \Delta y) - g(y_0)}{\Delta y} - g'(y_0).$$

Then $\lim_{\Delta y \to 0} \eta(\Delta y) = 0$. Let $\eta(0) = 0$. Then

$$g(y_0 + \Delta y) - g(y_0) = (g'(y_0) + \eta(\Delta y))\Delta y \tag{2.2.1}$$

for any sufficiently small Δy (including $\Delta y = 0$).

Let $\Delta y = f(x_0 + \Delta x) - f(x_0)$. Then (2.2.1) implies

$$\begin{aligned} \frac{g(f(x_0 + \Delta x)) - g(f(x_0))}{\Delta x} &= \frac{g(y_0 + \Delta y) - g(y_0)}{\Delta x} \\ &= (g'(y_0) + \eta(\Delta y))\frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}. \end{aligned} \tag{2.2.2}$$

By the continuity of f, we have $\Delta y \to 0$ as $\Delta x \to 0$. The consequence follows. □

Note by the equation (2.2.2) that if f is continuous at x_0 and $g'(f(x_0)) \neq 0$, then the limit

$$\lim_{\Delta x \to 0} \frac{g(f(x_0 + \Delta x)) - g(f(x_0))}{\Delta x}$$

exists if and only if the limit

$$\lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$

exists. Hence, we obtain the following variation:

Theorem 2.2.1′. Let $f : (a, b) \to (A, B)$ be continuous at x_0, and $g : (A, B) \to \mathbb{R}$ be differentiable at $f(x_0)$ with $g'(f(x_0)) \neq 0$. Then the composition $g \circ f$ is differentiable at x_0 if and only if f is differentiable at x_0, and

$$(g \circ f)'(x_0) = g'(f(x_0))f'(x_0).$$

Theorem 2.2.2 ([1, §2.2 定理 2])**.** *Let $f : (a, b) \to (A, B)$ be continuous, strictly monotone and surjective. Suppose that for $y_0 \in (A, B)$, we have $(f^{-1})'(y_0) \neq 0$. Then for $x_0 = f^{-1}(y_0)$, we have*

$$f'(x_0) = \frac{1}{(f^{-1})'(y_0)}, \quad \text{or} \quad f'(x_0) = \frac{1}{(f^{-1})'(f(x_0))}.$$

Remark 2.2.3. Geometrically speaking, Theorem 2.2.2 indicates that the slopes of the tangent line with respect to x-axis and y-axis are reciprocal.

证明 *1.* Note that $f^{-1} \circ f(x) = x$, f is continuous at x_0, and $(f^{-1})'(f(x_0)) \neq 0$. Then by Theorem 2.2.1′, we have

$$1 = (f^{-1})'(f(x_0)) \cdot f'(x_0).$$

□

证明 *2.* Let $\Delta x \neq 0$. Then since f is strictly monotone, $f(x_0 + \Delta x) \neq f(x_0)$. Also note that $\Delta x = f^{-1}(f(x_0 + \Delta x)) - f^{-1}(f(x_0))$. Then

$$\frac{f(x_0+\Delta x) - f(x_0)}{\Delta x} = \frac{1}{\frac{\Delta x}{f(x_0+\Delta x)-f(x_0)}} = \frac{1}{\frac{f^{-1}(f(x_0+\Delta x))-f^{-1}(f(x_0))}{f(x_0+\Delta x)-f(x_0)}} \to \frac{1}{(f^{-1})'(f(x_0))}$$

as $\Delta x \to 0$. □

Example 2.2.4 ([1, §2.2 例 6])**.** Find the derivative of $f(x) = \arcsin x$ on $(-1, 1)$.

证明. Write $y = \arcsin x$. Then $x = \sin y$ on $(-\frac{\pi}{2}, \frac{\pi}{2})$. Then

$$(\arcsin x)' = \frac{1}{\cos y} = \frac{1}{\sqrt{1-\sin^2 y}} = \frac{1}{\sqrt{1-x^2}}$$

for $x \in (-1, 1)$. □

Example 2.2.5 ([1, §2.2 例 9])**.** Prove that

$$(\ln|x|)' = \frac{1}{x}$$

for $x \neq 0$.

证明. For $x > 0$, if $y = \ln x$, then $x = e^y$. By Theorem 2.2.2, we have

$$(\ln x)' = \frac{1}{e^y} = \frac{1}{e^{\ln x}} = \frac{1}{x}.$$

For $x < 0$, let $t = -x$. Then $\ln|x| = \ln t$. Then

$$\frac{d}{dx}\ln|x| = \frac{d}{dt}\ln t \cdot \frac{dt}{dx} = \frac{1}{t} \cdot (-1) = \frac{1}{x}.$$

□

Example 2.2.6 ([1, §2.2 例 10])**.** Prove that for $a \in \mathbb{R}$, we have

$$(x^a)' = ax^{a-1}$$

for $x > 0$.

证明. For $a = 0$, it is constant. For $a \neq 0$, we can write

$$x^a = e^{a \ln x}.$$

Then

$$(x^a)' = e^{a \ln x} \cdot (a \ln x)' = x^a \cdot \frac{a}{x} = ax^{a-1}.$$

□

Example 2.2.7 ([1, §2.2 例 11])**.** For $y = x^x$ $(x > 0)$, find y'.

证明. Write

$$y = e^{x \ln x}.$$

Differentiate both sides, we get

$$y' = e^{x \ln x}(x \ln x)' = x^x(\ln x + 1).$$

□

Example 2.2.8 ([1, §2.2 例 12])**.** For $y = \sqrt[3]{\frac{(x+1)^2(2-x)}{(3-x)^2(x-4)}}$, find y'.

证明. Note that

$$\ln |y| = \frac{1}{3}(2 \ln |x+1| + \ln |2-x| - 2 \ln |3-x| - 2 \ln |x-4|).$$

Differentiate on both sides, we get

$$\frac{y'}{y} = \frac{1}{3}\left(\frac{2}{x+1} - \frac{1}{2-x} - \frac{2}{3-x} - \frac{2}{x-4}\right).$$

Thus, we have

$$y' = \frac{1}{3}\sqrt[3]{\frac{(x+1)^2(2-x)}{(3-x)^2(x-4)}}\left(\frac{2}{x+1} - \frac{1}{2-x} - \frac{2}{3-x} - \frac{2}{x-4}\right).$$

□

2.3 Differential

2.3.1 Infinitesimal

An *infinitesimal* is a variable that tends to 0. For instance, when $x \to 0$, x, $\cos x - 1$, and $\ln(1+x)$ are infinitesimals. An infinitesimal $f(x)$ $(x \to a)$ is equivalent to an infinitely large $1/f(x)$ $(x \to a)$.

Suppose that $\alpha(x), \beta(x)$ are infinitesimals as $x \to a$.

- If there is $l \neq 0$ such that
$$\lim_{x\to a} \frac{\alpha(x)}{\beta(x)} = l$$
then we say α and β are *infinitesimals of the same order*.

  In particular, if $l = 1$, then we say α and β are *equivalent infinitesimals*, and write $\alpha \sim \beta$ $(x \to a)$.

  If $\beta(x) = (x-a)^n$, then we say $\alpha(x)$ is an infinitesimal of order n as $x \to a$.

- If there is another infinitesimal $\eta(x)$ such that
$$\alpha(x) = \eta(x)\beta(x),$$
then we say $\alpha(x)$ is *of higher order* than $\beta(x)$, and write
$$\alpha(x) = o(\beta(x)) \quad (x \to a).$$

Example 2.3.1 ([1, §2.3 例 1])**.** $\sin x \sim x$ $(x \to 0)$.

Example 2.3.2 ([1, §2.3 例 5])**.** $x \sin x^2 = o(x^2)$ $(x \to 0)$.

Note however that equivalent infinitesimals **CANNOT** exchange arbitrarily. In fact, the equivalent infinitesimals can only be used in the quotient.

Example 2.3.3 ([1, §2.3 例 6])**.** Find the limit $\lim_{x\to 0} \frac{\tan x - \sin x}{x^3}$.

证明. The problem cannot directly use the equivalent infinitesimals: $\tan x \sim x$, $\sin x \sim x$; Otherwise, the limit would be equal to 0, which is wrong.

To see the limit,

$$\frac{\tan x - \sin x}{x^3} = \frac{\sin x}{x^3} \cdot \frac{1 - \cos x}{\cos x} = \frac{\sin x}{x^3} \cdot \frac{2(\sin \frac{1}{2}x)^2}{\cos x} \to \frac{1}{2}$$

as $x \to 0$. □

Proposition 2.3.4 ([1, §2.3 性质])**.** *$\alpha(x) \sim \beta(x)$ $(x \to a)$ if and only if*

$$\alpha(x) = \beta(x) + o(\beta(x)) \quad (x \to a).$$

证明. $\alpha(x) \sim \beta(x)$ $(x \to a)$ if and only if

$$\lim_{x \to a} \frac{\alpha(x)}{\beta(x)} = 1.$$

$\alpha(x) = \beta(x) + o(\beta(x))$ $(x \to a)$ if and only if

$$\lim_{x \to a} \frac{\alpha(x) - \beta(x)}{\beta(x)} = 0.$$

□

Example 2.3.5 (2015 Calculus B, Midterm)**.** Find the limit

$$\lim_{n \to 0} \frac{\ln \cos x}{\sin x^2}.$$

证明. Using $\sin x \sim x$, we have

$$\lim_{n \to 0} \frac{\ln \cos x}{\sin x^2} = \lim_{n \to 0} \frac{\ln \left(1 - 2\sin^2 \frac{x}{2}\right)}{\sin x^2} = \lim_{n \to 0} \frac{-2\sin^2 \frac{x}{2}}{\sin x^2} = -\frac{1}{2}.$$

□

Example 2.3.6 (2018 Calculus B, Midterm)**.** For $0 < x < 1$, find the limit

$$\lim_{n \to \infty} ((n+1)^x - n^x).$$

证明. Using $(1+y)^x \sim xy$, we have

$$\begin{aligned}
&\lim_{n \to \infty} ((n+1)^x - n^x) \\
=& \lim_{n \to \infty} n^x \left(\left(1 + \frac{1}{n}\right)^x - 1 \right) \\
=& \lim_{n \to \infty} n^x \left(x \cdot \frac{1}{n} \right) = x \lim_{n \to \infty} n^{x-1} = 0.
\end{aligned}$$

□

2.3.2 Definitions of differentials

Definition 2.3.7 ([1, §2.3 定义]). Let $y = f(x)$ be a function on $U_r(x_0)$. Suppose that there is a linear map $L : \mathbb{R} \to \mathbb{R}$, such that

$$f(x_0 + \Delta x) - f(x_0) = L\Delta x + o(\Delta x) \quad (\Delta x \to 0).$$

Then we say f is *differentiable* at x_0. And $L\Delta x$ is called the *differential* of f at x_0, and is denoted by df or dy.

Clearly, if f is differentiable at x_0, then its differential is given by $df = f'(x)\Delta x$. As mentioned in Remark 2.1.4, the differential $df = f'(x)\Delta x$ is the **best linear approximation** to f at x. This is the key idea in the field of differential calculus: Use *linear algebra* (线性代数) to (locally) approximate and study nonlinear objects.

Example 2.3.8. For $f(x) = x$, $dx = \Delta x$.

Thus, for differentiable f, we can rewrite

$$df = f'(x)dx.$$

Remark 2.3.9. Unlike the independent variable dx, the variable $dy = F(x, dx)$ is always a dependent variable. It depends on both x and dx. Usually, dx is a fixed small value, and dy is considered as a function of x.

Remark 2.3.10. The difference between differentials and derivatives is subtle. For a fixed point x_0, the derivative $f'(x_0)$ is a **number**. This is also the case in high-dimensional, where we study the *partial/directional derivative* (偏导数/方向导数, Definitions 6.4.1, 6.5.1).

On the other hand, in most literature, the differential $df(x_0)$ is a **linear map** (线性映射, Appendix 7.2.2). In 1-dimensional, a linear map can be represented as a number; but in high-dimensional, a linear map is represented by a *matrix* (矩阵, Definition 7.2.24), i.e. a group of numbers, called the *entries* (元素) of the matrix. In fact, a differential is represented by a matrix where the entries are partial derivatives (see Theorem 6.4.16).

Example 2.3.11 ([1, §2.3 例 8]). $d(xe^x) = e^x(1 + x)dx$.

For differentiable functions f and g of x, the elementary arithmetic of differentials follows easily from the corresponding derivatives:

(1) $d(f \pm g) = df \pm dg$,

(2) $d(f \cdot g) = f dg + g df$,

(3) $d\left(\frac{f}{g}\right) = \frac{g df - f dg}{g^2}$.

Remark 2.3.12. The formula (2) is usually called the *product rule* or *Leibniz rule* (莱布尼兹公式).

Example 2.3.13 ([1, §2.3 例 10])**.** Find $d(e^x \sin x)$.

证明. By the product rule, we have

$$\begin{aligned} d(e^x \sin x) &= \sin x d(e^x) + e^x d(\sin x) \\ &= e^x \sin x dx + e^x \cos x dx \\ &= e^x (\sin x + \cos x) dx. \end{aligned}$$

□

2.4 Change of variables

Recall the derivative of the composition of functions follows the *chain rule*:

Theorem 2.4.1 (Chain rule)**.** *If $f_1, \ldots, f_n$ are differentiable functions, then the derivative of the composition is given by*

$$\frac{d}{dx}(f_1 \circ \cdots \circ f_n) = \frac{df_1}{df_2} \cdot \frac{df_2}{df_3} \cdots \frac{df_{n-1}}{df_n} \cdot \frac{df_n}{dx}.$$

In terms of differentials, the chain rule implies the following:

Proposition 2.4.2 (Change of variables)**.** *Suppose that y is a differentiable function of x, and z is a differentiable function of y. Consider the composition*

$$x \mapsto y(x) \mapsto z(y(x)).$$

Then the differentials satisfy

$$\frac{dz}{dy}dy = z'dy = dz = (z \circ y)'dx = \frac{dz}{dy}\frac{dy}{dx}dx = \frac{dz}{dx}dx. \tag{2.4.1}$$

This can simplify certain calculation.

2.4.1 Implicit function

Suppose that y is a differentiable function of x defined by the equation

$$F(x, y) = 0 \tag{2.4.2}$$

for some $F : \mathbb{R}^2 \to \mathbb{R}$. Then we can find the derivative y' without first finding the explicit formula of $y(x)$. In fact, by thinking $F = F(x, y(x))$ as a function of x, we differentiate both sides of (2.4.2) and obtain

$$dF(x, y) = 0. \tag{2.4.3}$$

Then try to dy/dx from (2.4.3) directly.

Example 2.4.3. The function y of x is determined by the equation

$$y - x - \epsilon \sin y = 0$$

for $\epsilon \in (0, 1)$. Find the derivative y'.

证明. Differentiate both sides of the equation:

$$dy - dx - \epsilon \cos y dy = 0.$$

Then we have

$$(1 - \epsilon \cos y)dy = dx$$

and so

$$\frac{dy}{dx} = \frac{1}{1 - \epsilon \cos y}.$$

□

Example 2.4.4 ([1, §2.4 例 2])**.** The function y of x is determined by the equation

$$e^{xy} + x^2y - 1 = 0.$$

Find the derivative y'.

证明. Differentiate both sides of the equation:

$$de^{xy} + dx^2y = 0.$$

By the chain rule, we get

$$e^{xy}dxy + dx^2y = 0.$$

By the Leibniz rule, we get

$$e^{xy}(xdy + ydx) + (x^2dy + 2xydx) = 0.$$

Then we obtain

$$\frac{dy}{dx} = -\frac{e^{xy}y + 2xy}{e^{xy}x + x^2}.$$

□

2.4.2 Parametric system

A function can also be determined by a *parametric system* (参数方程组):

$$\begin{cases} x = \varphi(t) \\ y = \psi(t) \end{cases}, \quad t \in [\alpha, \beta]. \tag{2.4.4}$$

Here t is called the *parameter* (参变量). Then again we can find the derivative y' without first finding the explicit formula of $y(x)$, if φ and ψ are differentiable. To see this, we consider $t = t(x)$ as a function of x. Then, we differentiate both sides of (2.4.4) and (by Proposition 2.4.2) obtain

$$\begin{cases} dx = \dfrac{d\varphi}{dt}dt \\ \\ dy = \dfrac{d\psi}{dt}dt \end{cases}. \tag{2.4.5}$$

Thus, if $\varphi'(t) \neq 0$, we obtain dy/dx by (2.4.5) directly:

$$\frac{dy}{dx} = \frac{\psi'(t)}{\varphi'(t)}. \tag{2.4.6}$$

Example 2.4.5 ([1, §2.4 例 3])**.** Consider the ellipse:

$$\begin{cases} x = a\cos\theta \\ y = b\sin\theta \end{cases}, \quad \theta \in [0, 2\pi].$$

Find the angle φ between the x-axis and the tangent line of the ellipse at $\theta = \frac{\pi}{4}$.

证明. By (2.4.6), we have

$$\frac{dy}{dx} = -\frac{b\cos\theta}{a\sin\theta}.$$

For $\theta = \frac{\pi}{4}$, we find the angle

$$\tan\varphi = -\frac{b}{a}.$$

□

2.5 Linearization

Linearization is finding the linear approximation to a function at a given point.

Definition 2.5.1. If $f(x)$ is differentiable at x_0, then we have

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + o(x - x_0).$$

Then approximation function

$$A(x) = f(x_0) + f'(x_0)(x - x_0)$$

is called the *linearization* of f at x_0.

Some important linearizations when $x \to 0$:

- $e^x \approx 1 + x$,
- $\ln(1 + x) \approx x$,
- $(1 + x)^\alpha \approx 1 + \alpha x$ for $\alpha > 0$.

The error will be studied in the future.

Example 2.5.2 ([1, §2.5 例 1])**.** Use the linearization to estimate $\sqrt{4.6}$.

证明. Since $\sqrt{4+x} \approx \sqrt{4} + \frac{1}{2\sqrt{4}}x$ for $x \to 0$, we have

$$\sqrt{4.6} \approx \sqrt{4} + \frac{1}{2\sqrt{4}} \times 0.6 = 2.15.$$

□

2.6 Higher-order derivatives and differentials

Definition 2.6.1. Suppose $f(x)$ has a derivative function $f'(x)$ that is also differentiable. Then the derivative of $f'(x)$:

$$(f')'(x) = f''(x) = \frac{d}{dx}\frac{df}{dx} = \frac{d^2 f}{dx^2}$$

is called the *second derivative* of f.

Similarly, by induction, we can define the *n-th derivative* of f for $n \geq 1$: For a differentiable $f^{(n-1)}(x)$, we define

$$(f^{(n-1)})'(x) = f^{(n)}(x) = \frac{d^n f}{dx^n}.$$

Example 2.6.2 ([1, §2.6 例 1])**.** Find $(\sin x)^{(n)}$ for $n \in \mathbb{N}^*$.

证明. One calculates

$$\begin{aligned}
(\sin x)' &= \cos x = \sin\left(x + \frac{\pi}{2}\right), \\
(\sin x)'' &= \cos\left(x + \frac{\pi}{2}\right) = \sin\left(x + \frac{2\pi}{2}\right), \\
(\sin x)''' &= \cos\left(x + \frac{2\pi}{2}\right) = \sin\left(x + \frac{3\pi}{2}\right).
\end{aligned}$$

Thus, by induction, we conclude

$$(\sin x)^{(n)} = \sin\left(x + \frac{n\pi}{2}\right).$$

□

Proposition 2.6.3 (General Leibniz rule 莱布尼茨公式, [1, §2.6 命题])**.** *Suppose that f and g are n-times differentiable functions. Then the product fg is also n-times differentiable functions, and its n-th derivative is given by*

$$(fg)^{(n)} = \sum_{k=0}^{n} \binom{n}{k} f^{(n-k)} g^{(k)},$$

where $\binom{n}{k} = \dfrac{n!}{k!(n-k)!}$ *is the binomial coefficient,* $0! = 1$*, and* $f^{(0)} = f$*.*

Example 2.6.4 ([1, §2.6 例 4])**.** Find the 5th derivative $y^{(5)}$ for the function $y = x^2 \sin x$.

证明. By the general Leibniz rule, we have

$$\begin{aligned} y^{(5)} &= x^2 (\sin x)^{(5)} + 5(x^2)'(\sin x)^{(4)} + 10(x^2)''(\sin x)^{(3)} \\ &= x^2 \sin\left(x + \frac{5\pi}{2}\right) + 10x \sin\left(x + \frac{4\pi}{2}\right) + 20 \sin\left(x + \frac{3\pi}{2}\right) \\ &= (x^2 - 20)\cos x + 10x \sin x. \end{aligned}$$

□

Note that there is **NO** general formulas for the higher-order derivatives of the **composition** of two functions. In particular, the chain rule does not hold for the higher-order derivatives of the composition of two functions.

Example 2.6.5 ([1, §2.6 例 7])**.** Let $y = y(x)$ be the function determined by the parametric system

$$\begin{cases} x = t - \sin t \\ y = 1 - \cos t \end{cases}.$$

Find the second derivative $\dfrac{d^2y}{dx^2}$.

证明. First, one calculates

$$\begin{cases} dx = 1 - \cos t \\ dy = \sin t \end{cases}.$$

Then we obtain the parametric system

$$\begin{cases} x = t - \sin t \\ \\ \dfrac{dy}{dx} = \dfrac{\sin t}{1 - \cos t} \end{cases}.$$

Then applying (2.4.6) again, we get

$$\frac{d^2y}{dx^2} = \frac{\dfrac{d}{dt}\left(\dfrac{dy}{dx}\right)}{\dfrac{dx}{dt}} = \frac{\dfrac{\cos t(1-\cos t) - \sin t \sin t}{(1-\cos t)^2}}{1 - \cos t} = \frac{\cos t - 1}{(1-\cos t)^3} = -\frac{1}{(1-\cos t)^2}.$$

□

We now discuss the *higher-order differentials*. Suppose that f is n-times differentiable. Recall from Remark 2.3.9 that a differential $df = F(x, dx) = f'(x)dx$ is considered as function depending on both x and dx. We fix dx. Then we can differentiate df again:

$$d^2 f = d(df) = d(f'(x)dx) = (f'(x)dx)'dx = f''(x)dx^2$$

where dx^2 denotes $(dx)^2$.

Similarly, we can define the higher-order differential:

$$d^n f = d(d^{n-1}f) = f^{(n)}(x)dx^n$$

for $n \in \mathbb{N}$.

Remark 2.6.6. Since the higher-order derivative of the composition of two functions does not have the chain rule, the (2.4.1) does not hold for the higher-order differentials. In other words, for a function

$$x \mapsto y(x) \mapsto z(y(x)),$$

we usually do **NOT** have

$$\frac{d^n z}{dy^n} dy^n = \frac{d^n z}{dx^n} dx^n \tag{2.6.1}$$

for $n \geq 2$. Actually, the equality (2.6.1) holds only if $y = cx$ for some $c \in \mathbb{R}$.

Example 2.6.7 ([1, §2.6 例 8])**.** Let $y = e^x \cos 2x$. Find d^2y.

证明. Differentiate both sides, we get

$$dy = (e^x \cos 2x - 2e^x \sin 2x)dx.$$

Differentiate both sides again, we get

$$\begin{aligned} d^2y &= (e^x \cos 2x - 2e^x \sin 2x - 2e^x \sin 2x - 4e^x \cos 2x)dx^2 \\ &= -(3\cos 2x + 4\sin 2x)e^x dx^2. \end{aligned}$$

□

2.7 Antiderivative

In the previous sections, we studied the *differentiation* - one of the two pillars of calculus. In the following, we shall study the other pillar, *integration*.

Definition 2.7.1 (Antiderivative, [1, §2.7 定义])**.** Let f be a function on (a, b). We say that a function $F : (a, b) \to \mathbb{R}$ is an *antiderivative* of f, if F is differentiable on (a, b) and

$$F'(x) = f(x)$$

for all $x \in (a, b)$. The process of taking an antiderivative is called *integration*. The collection of all antiderivatives of f is called the *indefinite integral*, and denoted by

$$\int f(x)dx.$$

The function $f(x)$ is the *integrand* of the integral.

Given a function f, if a function F is an antiderivative of f, then for any $C \in \mathbb{R}$, $F + C$ is also an antiderivative of f. On the other hand, if F and G are antiderivatives of f, then

$$(F - G)' = F' - G' = f - f = 0.$$

Thus, $F - G \equiv C$ for some $C \in \mathbb{R}$. Therefore, we conclude that if F is an antiderivative of f, then

$$\int f(x)dx = F(x) + C$$

for $C \in \mathbb{R}$.

Integration formulas:

(1) $\displaystyle\int x^\alpha dx = \frac{1}{\alpha+1}x^{\alpha+1} + C$ for $\alpha \neq -1$.

(2) $\displaystyle\int \cos x dx = \sin x + C$, $\displaystyle\int \sin x dx = -\cos x + C$,

$\displaystyle\int \frac{1}{\cos^2 x}dx = \tan x + C$, $\displaystyle\int \frac{1}{\sin^2 x}dx = -\frac{1}{\tan x} + C$.

(3) $\displaystyle\int \frac{dx}{\sqrt{1-x^2}} = \arcsin x + C = -\arccos x + C$.

$\displaystyle\int \frac{dx}{1+x^2} = \arctan x + C$.

(4) $\displaystyle\int a^x dx = \frac{1}{\ln a}a^x + C$ for $a > 0$, $a \neq 1$.

(5) $\displaystyle\int \frac{1}{x}dx = \ln|x| + C$.

Integration is linear:

(1) $\displaystyle\int (f(x) \pm g(x))dx = \int f(x)dx \pm \int g(x)dx$.

(2) $\displaystyle\int cf(x)dx = c\int f(x)dx$ for $c \neq 0$.

Example 2.7.2 ([1, §2.7 例 3])**.** Find $\displaystyle\int \left(e^x + \frac{3x^2}{1+x^2}\right)dx$.

证明.

$$\begin{aligned}
\int \left(e^x + \frac{3x^2}{1+x^2}\right)dx &= \int e^x dx + \int \frac{3x^2}{1+x^2}dx \\
&= e^x + 3\int \frac{x^2}{1+x^2}dx \\
&= e^x + 3\left(\int 1dx - \int \frac{1}{1+x^2}dx\right) \\
&= e^x + 3x - 3\arctan x + C.
\end{aligned}$$

□

Example 2.7.3 ([1, §2.7 例 5])**.** Solve the differential equation

$$\frac{d^2s}{dt^2} = -g$$

where $s(0) = h_0$, $s'(0) = v_0$, and g, h_0, v_0 are constant.

证明. Integrate both sides

$$\frac{ds}{dt} = \int -gdt = -gt + C_1. \tag{2.7.1}$$

Since $s'(0) = v_0$, we have $C_1 = v_0$. Integrate both sides of (2.7.1):

$$s = \int -gt + v_0dt = -\frac{1}{2}gt^2 + v_0t + C_2.$$

Since $s(0) = h_0$, we have $C_2 = h_0$. Thus, we conclude

$$s(t) = -\frac{1}{2}gt^2 + v_0t + h_0.$$

□

A differential equation of the form

$$\frac{dy}{dx} = f(x)g(y)$$

is called *separable*. To solve this, we first separate the variables:

$$\frac{1}{g(y)}dy = f(x)dx$$

and then integrate both sides.

Example 2.7.4 ([1, §2.7 例 6])**.** Solve the differential equation

$$-\frac{dm}{dt} = km(t), \quad m(0) = m_0$$

where $m_0, k > 0$ are constant.

证明. First, write it as differentials:

$$-dm = kmdt.$$

Divide both sides by m:

$$-\frac{1}{m}dm = kdt.$$

Integrate both sides:

$$\int -\frac{1}{m}dm = \int kdt.$$

Then we obtain

$$-\ln|m| = kt + C.$$

In other words, we obtain

$$m(t) = \pm e^C e^{-kt}.$$

Since $m(0) = m_0$, we have $\pm e^C = m_0$. Then we conclude

$$m(t) = m_0 e^{-kt}.$$

□

Now suppose we have a differential equation of the form

$$g(x, y)\frac{dy}{dx} = f(x, y).$$

To solve this, we first write it in terms of differentials:

$$g(x, y)dy - f(x, y)dx = 0.$$

Then try using the identities of differentials to move the functions f, g to the right side of the symbol "d". More precisely, try to find a (differentiable) function $u(x, y)$ so that

$$g(x, y)dy - f(x, y)dx = du(x, y).$$

Then the function $y(x)$ is determined by the equation

$$u(x, y) = C.$$

Some useful identities:

- (Chain rule) $z(y(x))y'(x)dx = z(y)dy$.
- (Product rule) $ydx + xdy = dxy$, $\dfrac{ydx - xdy}{y^2} = \dfrac{1}{y}dx + xd\left(\dfrac{1}{y}\right) = d\left(\dfrac{x}{y}\right)$.

Example 2.7.5. Solve the differential equation

$$f'(x) = -f(x) + e^{-x}\cos x.$$

证明. Write it in terms of differentials:

$$df = -f\,dx + e^{-x}\cos x dx.$$

Multiply by e^x:

$$e^x df + e^x f dx = \cos x dx.$$

By the chain rule, we have

$$e^x df + f de^x = d\sin x.$$

By the product rule, we have

$$dfe^x = d\sin x.$$

Integrate both sides:

$$fe^x = \sin x + C.$$

Thus, we conclude that

$$f(x) = e^{-x}\sin x + Ce^{-x}.$$

□

Examples 2.7.4 and 2.7.5 are the *first-order linear differential equations* (一阶线性微分方程). For the solution of the general first-order linear differential equation, see Appendix 7.3.

2.8 Riemann integral

2.8.1 Definitions

Definition 2.8.1 ([1, §2.8 定义])**.** Let $[a, b]$ be an interval. We say that a finite subset $T = \{x_0, \ldots, x_n\} \subset [a, b]$ is a *partition* of $[a, b]$ if

$$a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b.$$

Given a partition T of $[a, b]$, write

$$\Delta x_i = x_i - x_{i-1} \quad (i = 1, \ldots, n), \qquad \text{and} \qquad \lambda(T) \coloneqq \max\{\Delta x_i : i = 1, \ldots, n\}.$$

Let $f(x)$ be a function defined on a closed interval $[a, b]$. Suppose that for any partition T, any $\xi_i \in [x_{i-1}, x_i]$ $(i = 1, \ldots, n)$, the limit of the *Riemann sums*

$$\lim_{\lambda(T)\to 0} \sum_{i=1}^{n} f(\xi_i)\Delta x_i \tag{2.8.1}$$

exists. Then we say f is *Riemann integrable* (or *integrable* for short) on $[a, b]$. The limit (2.8.1) is called the *definite integral* (定积分) or *Riemann integral* (黎曼积分), and written as

$$\int_a^b f(x)dx.$$

Here $f(x)$ is called the integrand, a, b are called the *lower/upper limit of integration*, x is called the *variable of integration*.

Remark 2.8.2. In terms of the epsilon-delta definition, the existence of the limit (2.8.1) means the following:

Given any $\epsilon > 0$, there is a $\delta > 0$ such that for every partition $T = \{x_0, \ldots, x_n\}$ of $[a, b]$ with $|\lambda(T)| < \delta$, and any $\xi_i \in [x_{i-1}, x_i]$ $(i = 1, \ldots, n)$, we have

$$\left| \sum_{i=1}^{n} f(\xi_i)\Delta x_i - \int_a^b f(x)dx \right| < \epsilon.$$

Remark 2.8.3 (Riemann-Stieltjes integral)**.** Note that Δ can be interpreted as a function of intervals $[x_{i-1}, x_i]$:

$$\Delta x_i = \Delta[x_{i-1}, x_i] = x_i - x_{i-1}.$$

More generally, let $g : [a, b] \to \mathbb{R}$ be a function. Then we define the *g-length* of an interval $[c, d] \subset [a, b]$ (with $c \leq d$) by

$$g[c, d] = g(d) - g(c).$$

Replacing Δ by g in Definition 2.8.1, we obtain the *Riemann-Stieltjes integral* (黎曼-斯蒂尔杰斯积分), which is denoted by

$$\int_{x=a}^{b} f(x)dg(x).$$

Definition 2.8.4. For an integrable function $f(x)$, $a, b \in \mathbb{R}$, we define

- if $a = b$, then $\int_a^b f(x)dx = 0$;
- if $a > b$, then $\int_a^b f(x)dx = -\int_b^a f(x)dx$.

Remark 2.8.5 (Area under the graph)**.** If f is a continuous function f on $[a, b]$, then f is Riemann integrable.

- If $f(x) \geq 0$ for $x \in [a, b]$, then $\int_a^b f(x)dx$ is the area under the graph of f.
- If $f(x) \leq 0$ for $x \in [a, b]$, then $-\int_a^b f(x)dx$ is the area above the graph of f.

Example 2.8.6. Point out the integral that is approached by the limit

$$\lim_{n\to\infty} \sum_{i=1}^{n} \frac{1}{n} \sin \frac{i}{n}. \tag{2.8.2}$$

证明. Given $n \in \mathbb{N}^*$, consider the uniform partition $T = \{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{n}\}$ of $[0, 1]$. Then $\lambda(T) = \frac{1}{n}$. Let $\xi_i = \frac{i}{n} \in [\frac{i-1}{n}, \frac{i}{n}]$ $(i = 1, \ldots, n)$. Then the sum in (2.8.2) is the Riemann sum

$$\sum_{i=1}^{n} \frac{1}{n} \sin \frac{i}{n} = \sum_{i=1}^{n} \sin(\xi_i) \cdot \left(\frac{i}{n} - \frac{i-1}{n}\right).$$

Thus, the limit (2.8.2) is the Riemann integral

$$\lim_{n\to\infty} \sum_{i=1}^{n} \frac{1}{n} \sin \frac{i}{n} = \int_0^1 \sin x dx.$$

□

Note that Riemann integrability is a property of a function:

- A continuous function on $[a, b]$ is integrable. (See Appendix 7.5.)
- A bounded monotone function on $[a, b]$ is integrable.
- A bounded function on $[a, b]$ having finitely many discontinuities is integrable. (In fact, the discontinuities of a Riemann integrable function can be even countably infinite! See Appendix 7.4.)
- Dirichlet function is not integrable.

For more discussion about the theory of integration, see Appendix 7.4.

2.8.2 Properties of Riemann integrals

Theorem 2.8.7. *Let $[a, b]$ be a bounded interval, and let $f, g : [a, b] \to \mathbb{R}$ be Riemann integrable functions.*

(1) If $f(x) \geq 0$, then

$$\int_a^b f(x)dx \geq 0. \tag{2.8.3}$$

(2) $f(x) \pm g(x)$ are Riemann integrable on $[a, b]$, and

$$\int_a^b (f(x) \pm g(x))dx = \int_a^b f(x)dx \pm \int_a^b g(x)dx. \tag{2.8.4}$$

(3) If $f(x) \geq g(x)$ for all $x \in [a, b]$, then

$$\int_a^b f(x)dx \geq \int_a^b g(x)dx. \tag{2.8.5}$$

(4) For any $c \in \mathbb{R}$, $cf(x)$ is also Riemann integrable and

$$\int_a^b cf(x)dx = c\int_a^b f(x)dx.$$

(5) For any $c \in (a, b)$, $f(x)$ is Riemann integrable on $[a, c]$ and $[c, b]$, and

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx. \tag{2.8.6}$$

For general $c \in \mathbb{R}$, if f is Riemann integrable on $[a, c]$ and $[c, b]$, then (2.8.6) still holds.

(6) Suppose that $h(x)$ is a function on $[a, b]$ so that $\{x \in [a, b] : f(x) \neq h(x)\}$ is finite. Then $h(x)$ is Riemann integrable, and

$$\int_a^b f(x)dx = \int_a^b h(x)dx.$$

In particular, there is no difference in saying that a function is Riemann integrable on $[a, b]$ or (a, b).

(7) $|f(x)|$ is Riemann integrable and

$$\left|\int_a^b f(x)dx\right| \leq \int_a^b |f(x)|dx.$$

证明. One may establish (1) via the Riemann sum.

For (2), note that

$$\begin{aligned}
&\left|\sum_{i=1}^{n}(f(\xi_i) \pm g(\xi_i))\Delta x_i - \int_a^b f(x)dx \pm \int_a^b g(x)dx\right| \\
=&\left|\left(\sum_{i=1}^{n} f(\xi_i)\Delta x_i - \int_a^b f(x)dx\right) \pm \left(\sum_{i=1}^{n} g(\xi_i) - \int_a^b g(x)dx\right)\right| \\
\leq&\left|\sum_{i=1}^{n} f(\xi_i)\Delta x_i - \int_a^b f(x)dx\right| + \left|\sum_{i=1}^{n} g(\xi_i) - \int_a^b g(x)dx\right|.
\end{aligned}$$

Then we establish (2.8.4).

For (3), note that if $f(x) \geq g(x)$, then $f(x) - g(x) \geq 0$. Then combine the equations (2.8.3) and (2.8.4).

For (4), note that

$$\left|\sum_{i=1}^{n} cf(\xi_i)\Delta x_i - c\int_a^b f(x)dx\right| = \left|c\sum_{i=1}^{n} f(\xi_i)\Delta x_i - c\int_a^b f(x)dx\right|.$$

Then one may conclude the result via the Riemann sum directly.

(5) For $c \in (a,b)$, assume that $f(x)$ is not integrable on $[a,c]$. This means that there exists a $\epsilon_0 > 0$, such that for any $\delta > 0$ there are two partitions $T_1 = \{r_0, \ldots, r_m\}, T_2 = \{s_0, \ldots, s_n\}$ of $[a,b]$ with $|\lambda(T_1)|, |\lambda(T_2)| < \delta$, and $\alpha_i \in [r_{i-1}, r_i]$ $(i = 1, \ldots, m)$, $\beta_i \in [s_{i-1}, s_i]$ $(i = 1, \ldots, n)$, such that

$$\left|\sum_{i=1}^{m} f(\alpha_i)\Delta r_i - \sum_{i=1}^{n} f(\beta_i)\Delta s_i\right| \geq \epsilon_0.$$

Now let $T' = \{t_0, \ldots, t_k\}$ be a partition of $[c,b]$ with $|\lambda(T')| < \delta$. Then $T_1 \cup T'$ and $T_2 \cup T'$ are partitions of $[a,b]$ with $\lambda(T_1 \cup T'), \lambda(T_2 \cup T') < \delta$. Then

$$\left|\left(\sum_{i=1}^{m} f(\alpha_i)\Delta r_i + \sum_{i=1}^{k} f(t_i)\Delta t_i\right) - \left(\sum_{i=1}^{n} f(\beta_i)\Delta s_i + \sum_{i=1}^{k} f(t_i)\Delta t_i\right)\right| \geq \epsilon_0.$$

It means $f(x)$ is not integrable on $[a,b]$. A contradiction.

Now let

$$f_1(x) := \begin{cases} f(x) & , \text{ if } x \in [a,c] \\ 0 & , \text{ if } x \in [c,b] \end{cases}, \quad f_2(x) := \begin{cases} 0 & , \text{ if } x \in [a,c] \\ f(x) & , \text{ if } x \in [c,b] \end{cases}.$$

Then $f(x) = f_1(x) + f_2(x)$ for $x \in [a, b]$, and so the equation (2.8.6) follows from the equation (2.8.4).

For (6), let M be the number of elements in $\{x \in [a, b] : f(x) \neq g(x)\}$. For a partition $T = \{x_0, \ldots, x_n\}$ of $[a, b]$, a choice of $\xi_i \in [x_{i-1}, x_i]$, the Riemann sum

$$\sum_{i=1}^{n} (f(\xi_i) - h(\xi_i))\Delta x_i \tag{2.8.7}$$

has at most M terms that are nonzero. Letting $\lambda(T) \to 0$, the Riemann sum (2.8.7) goes to zero as well.

For (7), it would be easier to use *Darboux sum* 达布和 (cf. [1, P.132, Exercise 2.8.7]). Let $f_+(x) = \max\{f(x), 0\}$, $f_-(x) = \max\{-f(x), 0\}$. Then $f(x) = f_+(x) - f_-(x)$, and $|f(x)| = f_+(x) + f_-(x)$. We shall use the integrability of $f(x)$ to show that $f_+(x)$ is integrable.

Assume that $f_+(x)$ is not integrable on $[a, b]$. This means that there exists a $\epsilon_0 > 0$, such that for any $\delta > 0$ there are two partitions $T_1 = \{r_0, \ldots, r_m\}, T_2 = \{s_0, \ldots, s_n\}$ of $[a, b]$ with $|\lambda(T_1)|, |\lambda(T_2)| < \delta$, and $\alpha_i \in [r_{i-1}, r_i]$ $(i = 1, \ldots, m)$, $\beta_i \in [s_{i-1}, s_i]$ $(i = 1, \ldots, n)$, such that

$$f_+(\alpha_i) = \sup_{\alpha \in [r_{i-1}, r_i]} f_+(\alpha), \quad f_+(\beta_i) = \inf_{\beta \in [s_{i-1}, s_i]} f_+(\beta),$$

$$\sum_{i=1}^{m} f_+(\alpha_i)\Delta r_i - \sum_{i=1}^{n} f_+(\beta_i)\Delta s_i \geq \epsilon_0. \tag{2.8.8}$$

On the other hand, since f is integrable, we have

$$\sum_{i=1}^{m} f(\alpha_i')\Delta r_i - \sum_{i=1}^{n} f(\beta_i')\Delta s_i < \frac{1}{2}\epsilon_0, \tag{2.8.9}$$

for sufficiently small $\delta > 0$, where

$$f(\alpha_i') = \sup_{\alpha \in [r_{i-1}, r_i]} f(\alpha), \quad f(\beta_i') = \inf_{\beta \in [s_{i-1}, s_i]} f(\beta).$$

Then we see that on $[r_{i-1}, r_i] \cap [s_{j-1}, s_j]$, we have

$$f(\alpha_i') - f(\beta_j') \geq \max\{f(\alpha_i) - f(\beta_j), 0\}.$$

Then one may check that

$$\max\{f(\alpha_i), 0\} \leq \max\{f(\beta_j), 0\} + (f(\alpha'_i) - f(\beta'_j)).$$

Thus, we get that

$$f_+(\alpha_i) - f_+(\beta_j) = \max\{f(\alpha_i), 0\} - \max\{f(\beta_j), 0\} \leq f(\alpha'_i) - f(\beta'_j).$$

In other words,

$$\begin{aligned}(f_+(\alpha_i) - f_+(\beta_j)) \cdot |[r_{i-1}, r_i] \cap [s_{j-1}, s_j]| \\ \leq (f(\alpha'_i) - f(\beta'_j)) \cdot |[r_{i-1}, r_i] \cap [s_{j-1}, s_j]|.\end{aligned}$$

However, combining (2.8.8) and (2.8.9), we see that

$$\epsilon_0 \leq \sum_{i=1}^{m} f_+(\alpha_i)\Delta r_i - \sum_{i=1}^{n} f_+(\beta_i)\Delta s_i \leq \sum_{i=1}^{m} f(\alpha'_i)\Delta r_i - \sum_{i=1}^{n} f(\beta'_i)\Delta s_i < \frac{1}{2}\epsilon_0,$$

which is a contradiction.

Therefore, we conclude that f_+ is integrable. Similarly, f_-, and so $|f|$ are integrable. Finally, note that $|f| \geq \pm f \geq -|f|$, by the inequality (2.8.5), we have

$$\left|\int f(x)dx\right| \leq \int |f(x)|dx.$$

□

Proposition 2.8.8. *Let f be nonnegative and continuous on $[a, b]$, and satisfy $f(x_0) > 0$ for some $x_0 \in [a, b]$. Then we have $\int_a^b f(x)dx > 0$.*

证明. For $x_0 \in (a, b)$, write $A = f(x_0) > 0$. Then by the continuity of f, there exists $\delta > 0$ such that

$$f(x) \geq \frac{A}{2}$$

for any $x \in [x_0 - \delta, x_0 + \delta]$. Then

$$\int_a^b f(x)dx \geq \int_{x_0-\delta}^{x_0+\delta} f(x)dx \geq \int_{x_0-\delta}^{x_0+\delta} \frac{A}{2}dx = A\delta > 0.$$

□

2.9 First fundamental theorem of calculus

Theorem 2.9.1 (Mean value theorem for the definite integral, [1, §2.9 定理 1])**.** *Let $f(x)$ be continuous on $[a, b]$. Then there is a point $c \in [a, b]$ such that*

$$\int_a^b f(x)dx = f(c)(b-a).$$

证明. Since $f(x)$ is continuous on $[a, b]$, by the *extreme value theorem* (Theorem 1.6.3), it attains a maximum M and a minimum m, i.e.

$$m \leq f(x) \leq M$$

for all $x \in [a, b]$. Then we have

$$\int_a^b m dx \leq \int_a^b f(x)dx \leq \int_a^b M dx.$$

It follows that

$$m \leq \frac{1}{b-a} \int_a^b f(x)dx \leq M.$$

Then by the *intermediate value theorem*, there is a point $c \in [a, b]$ such that

$$f(c) = \frac{1}{b-a} \int_a^b f(x)dx$$

as required. □

One may obtain a generalization of Theorem 2.9.1 (cf. Remark 2.8.3):

Theorem 2.9.1′ (Mean value theorem for the Riemann-Stieltjes integral)**.** Let f be continuous on $[a, b]$, g be increasing (or decreasing) and integrable on $[a, b]$. Then there exists $\xi \in [a, b]$ such that

$$\int_{x=a}^b f(x)dg(x) = f(\xi)(g(b) - g(a)).$$

证明. Assume that $g(x) \geq 0$. Since $f(x)$ is continuous on $[a, b]$, by the *extreme value theorem* (Theorem 1.6.3), it attains a maximum M and a minimum m, i.e.

$$m \leq f(x) \leq M$$

for all $x \in [a, b]$. Then we have

$$\int_{x=a}^{b} m dg(x) \leq \int_{x=a}^{b} f(x) dg(x) \leq \int_{x=a}^{b} M dg(x).$$

It follows that

$$m \leq \frac{1}{g(b) - g(a)} \int_{x=a}^{b} f(x) dg(x) \leq M.$$

Then by the *intermediate value theorem*, there is a point $\xi \in [a, b]$ such that

$$f(\xi) = \frac{1}{g(b) - g(a)} \int_{x=a}^{b} f(x) dg(x)$$

as required. □

Example 2.9.2. Let f be continuous on $[0, 1]$ and satisfy

$$\int_{0}^{1} f(x) dx = \int_{0}^{1} x f(x) dx = 0.$$

Prove that f has at least two zeroes in $[0, 1]$.

证明. By the mean value theorem (Theorem 2.9.1), there is $\xi \in [0, 1]$ so that

$$f(\xi) = \int_{0}^{1} f(x) dx = 0.$$

Suppose that ξ is the unique zero of $f(x)$ on $[0, 1]$. Then by the continuity of f, we may assume

$$\begin{cases} f(x) > 0 & , \text{ if } x \in [0, \xi) \\ f(x) < 0 & , \text{ if } x \in (\xi, 1] \end{cases}.$$

Clearly $\xi \neq 0, 1$; for otherwise, it contradicts Proposition 2.8.8.

Now by Proposition 2.8.8, we have

$$\int_{0}^{\xi} x f(x) dx < \xi \int_{0}^{\xi} f(x) dx = -\xi \int_{\xi}^{1} f(x) dx < -\int_{\xi}^{1} x f(x) dx.$$

This leads to a contradiction. □

Let f be a Riemann integrable function on $[a, b]$. Then for $x \in [a, b]$, the function

$$F(x) = \int_{a}^{x} f(t) dt$$

is called an *integral with variable upper limit*.

Theorem 2.9.3 (First fundamental theorem of calculus, [1, §2.9 定理 2])**.** *Let f be continuous on $[a, b]$. Then*

$$F_a(x) = \int_a^x f(t)dt \tag{2.9.1}$$

is continuous on $[a, b]$, differentiable on (a, b), and

$$\frac{d}{dx} F_a(x) = f(x)$$

for $x \in (a, b)$.

证明. By Theorem 2.9.1, for any $x_0 \in [a, b)$ and $x > x_0$, $x \in (a, b]$, we have

$$F_a(x) - F_a(x_0) = \int_{x_0}^x f(t)dt = f(c)(x - x_0) \to 0 \quad (x \to x_0 + 0)$$

where $c \in (x_0, x)$. Thus, we conclude that F_a is right continuous:

$$\lim_{x \to x_0 + 0} F_a(x) = F_a(x_0)$$

for $x_0 \in [a, b)$. Similarly, one can deduce that F_a is left continuous on $(a, b]$, and so F_a is continuous on $[a, b]$.

Now for $x_0, x \in (a, b)$, by Theorem 2.9.1, there is a point c between x_0, x so that

$$f(c) = \frac{1}{x - x_0} \int_{x_0}^x f(t)dt = \frac{F_a(x) - F_a(x_0)}{x - x_0}.$$

When $x \to x_0$, we have $c \to x_0$, and so $f(c) \to f(x_0)$ by the continuity of f. Thus, we conclude that

$$\lim_{x \to x_0} \frac{F_a(x) - F_a(x_0)}{x - x_0} = f(x_0).$$

□

Remark 2.9.4. One may obtain the same result if $F_a(x)$ in (2.9.1) is replaced by

$$F_b(x) = \int_b^x f(t)dt.$$

In other words, it does not matter if x is greater or smaller than the lower limit of integration.

Remark 2.9.5. Informally, the first fundamental theorem of calculus asserts that

$$f(x) \xmapsto{\int_a^x} \int_a^x f(t)dt \xmapsto{\frac{d}{dx}} \frac{d}{dx}\int_a^x f(t)dt = f(x)$$

given a certain number of assumptions on f. In other words, this means that the derivative of an integral recovers the original function.

Sometimes we may consider the *integral with variable lower limit*:

$$G_b(x) = \int_x^b f(t)dt$$

Note that we have

$$\int_x^b f(t)dt = -\int_b^x f(t)dt.$$

Thus, we obtain

Corollary 2.9.6. *Let f be continuous on $[a, b]$. Then*

$$G_b(x) = \int_x^b f(t)dt$$

is continuous on $[a, b]$, differentiable on (a, b), and

$$\frac{d}{dx}G_b(x) = -f(x)$$

for $x \in (a, b)$.

Now let $\varphi(x)$ be differentiable, f continuous on an appropriate interval, let

$$F_a(x) = \int_a^x f(t)dt.$$

Then

$$\frac{d}{dx}\int_a^{\varphi(x)} f(t)dt = \frac{d}{dx}(F_a \circ \varphi)(x) = f(\varphi(x))\varphi'(x).$$

See also the *Leibniz integral rule* (莱布尼茨积分法则) [2, P.349].

Example 2.9.7 ([1, §2.9 例 3])**.** Find the derivative of the function

$$F(x) = \int_{x^2}^x \sqrt{1+t}dt.$$

证明. Note that $F(x) = \int_0^x \sqrt{1+t}dt + \int_{x^2}^0 \sqrt{1+t}dt$. Then by the chain rule, one calculates

$$\frac{d}{dx}F(x) = \frac{d}{dx}\left(\int_0^x \sqrt{1+t}dt - \int_0^{x^2} \sqrt{1+t}dt\right) = \sqrt{1+x} - 2x\sqrt{1+x^2}.$$

□

Example 2.9.8 ([1, §2.9 例 4])**.** Find the derivative of the function

$$G(x) = \int_0^x xf(t)dt$$

where $f(t)$ is continuous on $\mathbb{R}$.

证明. Write $G(x) = x\int_0^x f(t)dt$. Then by the product rule, one calculates

$$\frac{d}{dx}G(x) = \int_0^x f(t)dt + xf(x).$$

□

Remark 2.9.9. Note that the first fundamental theorem of calculus (Theorem 2.9.3) ensures that every **continuous** function on $[a,b]$ has an antiderivative. For discontinuous functions, the situation is more complicated, and is a graduate-level *real analysis* (实分析) topic.

2.10 Second fundamental theorem of calculus/Newton-Leibniz formula

Now we discuss the reverse of the first fundamental theorem of calculus:

Theorem 2.10.1 (Second fundamental theorem of calculus, Newton-Leibniz formula, [1, §2.10 定理])**.** *Let f, F be continuous on $[a,b]$ so that*

$$F'(x) = f(x)$$

for any $x \in (a,b)$. Then we have

$$\int_a^b f(x)dx = F(b) - F(a).$$

In Section 4.1, we shall see that Theorem 2.10.1 can be strengthened to the following:

Theorem 2.10.1′ (Second fundamental theorem of calculus, Newton-Leibniz formula, [1, P.260 习题 16])**.** Let f be integrable on $[a, b]$. If $F : [a, b] \to \mathbb{R}$ is an antiderivative of f, then we have

$$\int_a^b f(x)dx = F(b) - F(a).$$

Proof of Theorem 2.10.1. Let

$$F_a(x) = \int_a^x f(t)dt.$$

Then by Theorem 2.9.3, we have

$$F_a'(x) = f(x)$$

for $x \in (a, b)$. Thus, there is $C \in \mathbb{R}$ such that

$$F(x) = F_a(x) + C$$

for all $x \in (a, b)$. Then by the continuity of F and F_a, we have

$$F(a) = \lim_{x \to a+0} F(x) = \lim_{x \to a+0} F_a(x) + C = F_a(a) + C = C.$$

Similarly, we have $F(b) = F_a(b) + C$. Therefore, we have

$$F_a(b) = \int_a^x f(t)dt = F(b) - C = F(b) - F(a).$$

□

Remark 2.10.2. As mentioned earlier, the second fundamental theorem of calculus is the reverse of the first:

$$F(x) \xmapsto{\frac{d}{dx}} f(x) \xmapsto{\int_a^b} \int_a^b f(t)dt = F(b) - F(a)$$

given a certain number of assumptions on f. In other words, this means that the integral of a derivative recovers the original function.

Example 2.10.3 ([1, §2.10 例 1]). Find $\int_0^1 (e^x + x)\,dx$.

证明. Since $\int (e^x + x)\,dx = e^x + \frac{1}{2}x^2 + C$, by Theorem 2.10.1, we have

$$\int_0^1 (e^x + x)\,dx = \left(e^x + \frac{1}{2}x^2\right)\Big|_0^1 = e + \frac{1}{2} - 1 = e - \frac{1}{2}.$$

□

Example 2.10.4 ([1, §2.10 例 4]). Find $\lim_{n\to\infty} \sum_{i=1}^{n} \left(\frac{i}{n} + 1\right)^3 \frac{1}{n}$.

证明. Consider the partition $T = \{0, \frac{1}{n}, \ldots, \frac{n-1}{n}, 1\}$ of $[0, 1]$. Then $\Delta x_i = \frac{1}{n}$, $(i = 1, \ldots, n)$. Pick $\xi_i = \frac{i}{n} \in [\frac{i-1}{n}, \frac{i}{n}]$, $(i = 1, \ldots, n)$. Thus, the Riemann sum becomes

$$\sum_{i=1}^{n} (\xi_i + 1)^3 \Delta x_i = \sum_{i=1}^{n} \left(\frac{i}{n} + 1\right)^3 \frac{1}{n}.$$

Since $\int (e^x + x)\,dx = e^x + \frac{1}{2}x^2 + C$, by Theorem 2.10.1, we have

$$\lim_{n\to\infty} \sum_{i=1}^{n} \left(\frac{i}{n} + 1\right)^3 \frac{1}{n} = \int_0^1 (x+1)^3 dx = \frac{1}{4}(x+1)^4\Big|_0^1 = \frac{1}{4}(2^4 - 1) = \frac{15}{4}.$$

□

Example 2.10.5 ([1, §2.10 例 5]). Find the integral $\int_{-1}^{2} |x| \cdot [x] dx$.

证明. Note first that

$$f(x) = |x| \cdot [x] = \begin{cases} x & , \text{ if } x \in [-1, 0) \\ 0 & , \text{ if } x \in [0, 1) \\ x & , \text{ if } x \in [1, 2) \end{cases}.$$

Then by Theorem 2.10.1, we have

$$\begin{aligned} \int_{-1}^{2} f(x)dx &= \int_{-1}^{0} f(x)dx + \int_{0}^{1} f(x)dx + \int_{1}^{2} f(x)dx \\ &= \frac{1}{2}x^2\Big|_{-1}^{0} + 0 + \frac{1}{2}x^2\Big|_{1}^{2} = 1. \end{aligned}$$

□

Example 2.10.6 ([1, §2.10 例 7])**.** Find the area between the curves $y^2 = 4x$ and $4x - 3y = 4$.

证明. The intersections of the curves $y^2 = 4x$ and $4x - 3y = 4$ are $(4,4), (\frac{1}{4}, -1)$. Then by Theorem 2.10.1, the area is

$$\begin{aligned}
\text{Area} &= \int_{-1}^{4} \left(\frac{3y+4}{4} - \frac{y^2}{4} \right) dy = \frac{1}{4} \int_{-1}^{4} \left(3y + 4 - y^2 \right) dy \\
&= \frac{1}{4} \left(\frac{3}{2} y^2 + 4y - \frac{1}{3} y^3 \right) \Big|_{-1}^{4} = \frac{125}{24}.
\end{aligned}$$

$\square$

Chapter 3

Calculations and Applications of Integration

3.1 Change of variables for indefinite integrals

Recall that a continuous function f on $[a, b]$ is Riemann integrable (see Theorem 7.5). In other words, the definite integral $\int_a^b f(t)dt$ of f exists. It follows that $F(x) = \int_a^x f(t)dt$ is well-defined. By the first fundamental theorem of calculus, f has an antiderivative. Further, given an an antiderivative F of f, by the second fundamental theorem of calculus, the definite integral

$$\int_a^b f(t)dt = F(b) - F(a).$$

Therefore, we may compute the definite integrals relatively easily provided that we can find an antiderivative of the integrand f.

In what follows, we introduce some techniques to find the antiderivatives of a given function f. First, recall that if F is differentiable, we have

$$dF(x) = F'(x)dx.$$

Hence, we also write

$$\int dF(x) = \int F'(x)dx = F(x) + C.$$

We see that the integral sign $\int$ and the differential sign d cancel each other out. Therefore, the integration and differentiation are inverse to each other.

Consider a composition of differentiable functions

$$x \mapsto y(x) \mapsto z(y(x)).$$

Then by the chain rule, we have

$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx}.$$

Take the integral on both sides

$$\int \frac{dz}{dx}dx = \int \frac{dz}{dy}\frac{dy}{dx}dx = \int \frac{dz}{dy}dy. \tag{3.1.1}$$

The equation (3.1.1) is called the *change of variables formula* for indefinite integrals.

Moreover, consider the definite integral on $[a, b]$. Then by (3.1.1) and the second fundamental theorem of calculus, we have

Theorem 3.1.1 ([1, §3.4 定理 2])**.** *Suppose that $y : [a, b] \to [A, B]$ and $z : [A, B] \to \mathbb{R}$ have continuous derivatives. Then we have*

$$\int_a^b \frac{dz}{dx}dx = z(y(b)) - z(y(a)) = \int_{y(a)}^{y(b)} \frac{dz}{dy}dy. \tag{3.1.2}$$

The equation (3.1.2) is called the *change of variables formula* for definite integrals. The condition that "z' is continuous " can be replaced by "F is integrable", i.e. for any integrable F, we also have

$$\int_a^b F(y(x))y'(x)dx = \int_{y(a)}^{y(b)} F(y)dy.$$

See Theorem 3.4.3.

Example 3.1.2 ([1, §3.1 例 1])**.** Find $\int \tan x dx$.

证明. Rewrite it as $\int \frac{\sin x}{\cos x}dx$. Note that $\sin x dx = d(-\cos x)$. Letting $y = \cos x$, we have

$$\int \frac{\sin x}{\cos x}dx = -\int \frac{1}{\cos x}d\cos x = -\int \frac{1}{y}dy = -\ln|y| + C = -\ln|\cos x| + C.$$

□

Some frequently use differentials:

$$d(ax+c) = adx, \quad e^x dx = de^x, \quad \cos x dx = d\sin x, \quad \text{etc.}$$

Example 3.1.3 ([1, §3.1 例 3])**.** Find $\int \frac{xdx}{1+x^4}$.

证明.

$$\int \frac{xdx}{1+x^4} = \int \frac{d\frac{1}{2}x^2}{1+x^4} = \frac{1}{2}\arctan x^2 + C.$$

□

Example 3.1.4 ([1, §3.1 例 4])**.** Find $\int \sin nx \sin mx dx$, for $0 < n < m$.

证明.

$$\begin{aligned}
&\int \sin nx \sin mx dx \\
=&\frac{1}{2}\int [\cos(m-n)x - \cos(m+n)x]dx \\
=&\frac{1}{2}\left[\int \cos(m-n)xdx - \int \cos(m+n)xdx\right] \\
=&\frac{1}{2}\left[\frac{1}{m-n}\int \cos(m-n)xd(m-n)x - \frac{1}{m+n}\int \cos(m+n)xd(m+n)x\right] \\
=&\frac{1}{2}\left[\frac{1}{m-n}\sin(m-n)x - \frac{1}{m+n}\sin(m+n)x\right] + C.
\end{aligned}$$

□

Example 3.1.5 ([1, §3.1 例 5])**.** Find $\int \frac{dx}{a^2-x^2}$, for $a > 0$.

证明.

$$\begin{aligned}
\int \frac{dx}{a^2-x^2} =& \int \frac{1}{(a-x)(a+x)}dx \\
=&\frac{1}{2a}\int \frac{1}{a-x} + \frac{1}{a+x}dx \\
=&\frac{-1}{2a}\int \frac{1}{a-x}d(a-x) + \frac{1}{2a}\int \frac{1}{a+x}d(a+x) \\
=&\frac{-1}{2a}\ln|a-x| + \frac{1}{2a}\ln|a+x| + C \\
=&\frac{1}{2a}\ln\frac{|a+x|}{|a-x|} + C.
\end{aligned}$$

□

Example 3.1.6 ([1, §3.1 例 7])**.** Find $\int \frac{dx}{\sqrt{a^2-x^2}}$, for $a>0$.

证明.

$$\int \frac{dx}{\sqrt{a^2-x^2}} = \int \frac{dx}{\sqrt{a^2\left(1-\left(\frac{x}{a}\right)^2\right)}} = \int \frac{d\left(\frac{x}{a}\right)}{\sqrt{1-\left(\frac{x}{a}\right)^2}} = \arcsin\left(\frac{x}{a}\right)+C.$$

□

Example 3.1.7 ([1, §3.1 例 9])**.** Find $\int \frac{dx}{\cos x}$.

证明.

$$\int \frac{dx}{\cos x} = \int \frac{\cos x dx}{\cos^2 x} = \int \frac{d\sin x}{\cos^2 x} = \int \frac{d\sin x}{1-\sin^2 x} = \frac{1}{2}\ln\frac{|1+\sin x|}{|1-\sin x|}+C.$$

The last equality follows from Example 3.1.5. □

Example 3.1.8 ([1, §3.1 例 11])**.** Find $\int \frac{dx}{\sqrt{x+1}+1}$, for $x>-1$.

证明. Let $t=\sqrt{x+1}$. Then $x=t^2-1$ $(t>0)$, and $dx=2tdt$. Then we have

$$\int \frac{dx}{\sqrt{x+1}+1} = \int \frac{2tdt}{t+1} = 2\int\left(1-\frac{1}{t+1}\right)dt = 2(t-\ln(t+1))+C.$$

Change it back, and we get

$$\int \frac{dx}{\sqrt{x+1}+1} = 2\sqrt{x+1}-2\ln(\sqrt{x+1}+1)+C.$$

□

Example 3.1.9 ([1, §3.1 例 12])**.** Find $\int \sqrt{a^2-x^2}dx$, for $a>0$.

证明. Consider the parametric system:

$$\begin{cases} a\sin\theta = x \\ a\cos\theta = \sqrt{a^2-x^2} \end{cases}$$

for $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. Then we have

$$\begin{aligned}\int \sqrt{a^2 - x^2} dx =& \int a\cos\theta d(a\sin\theta) \\ =& a^2 \int \cos^2\theta d\theta \\ =& \frac{a^2}{2} \int (1+\cos 2\theta) d\theta \\ =& \frac{a^2}{2} \left(\theta + \frac{1}{2}\sin 2\theta\right) + C \\ =& \frac{a^2}{2} (\theta + \sin\theta\cos\theta) + C \\ =& \frac{a^2}{2} \left(\arcsin\frac{x}{a} + \frac{1}{a^2} x\sqrt{a^2 - x^2}\right) + C.\end{aligned}$$

□

Example 3.1.10 ([1, §3.1 例 13]). Find $\int \frac{dx}{\sqrt{x^2+a^2}}$, for $a > 0$.

证明. Consider the parametric system:

$$\begin{cases} a\tan\theta = x \\ \cos\theta = \dfrac{a}{\sqrt{a^2+x^2}} \end{cases}$$

for $\theta \in (-\frac{\pi}{2}, \frac{\pi}{2})$. Then we have

$$\begin{aligned}\int \frac{1}{\sqrt{a^2+x^2}} dx =& \int \frac{1}{a}\cos\theta d(a\tan\theta) \\ =& \int \cos\theta \cdot \frac{1}{\cos^2\theta} d\theta \\ \overset{\text{by Example 3.1.7}}{=\!=\!=\!=} & \frac{1}{2}\ln\frac{|1+\sin\theta|}{|1-\sin\theta|} + C \\ =& \frac{1}{2}\ln\frac{|1+\frac{x}{\sqrt{a^2+x^2}}|}{|1-\frac{x}{\sqrt{a^2+x^2}}|} + C \\ =& \ln(x + \sqrt{a^2+x^2}) + C.\end{aligned}$$

Another way: Consider the parametric system

$$\begin{cases} a\sinh t = x \\ a\cosh t = \sqrt{a^2+x^2} \end{cases}$$

for $t \in \mathbb{R}$. Then we have

$$\int \frac{1}{\sqrt{a^2+x^2}}dx = \int \frac{1}{a\cosh t}d(a\sinh t) = \int 1dt = t + C = \ln(x+\sqrt{a^2+x^2}) + C.$$

$\square$

Recall the basic properties of **hyperbolic functions**:

- $(\sinh t)' = \cosh t$, $(\cosh t)' = \sinh t$,
- $\operatorname{ch}^2 x - \operatorname{sh}^2 x = 1$,
- $\operatorname{sh}(x \pm y) = \operatorname{sh}(x)\operatorname{ch}(y) \pm \operatorname{ch}(x)\operatorname{sh}(y)$,
- $\operatorname{ch}(x \pm y) = \operatorname{ch}(x)\operatorname{ch}(y) \pm \operatorname{sh}(x)\operatorname{sh}(y)$,
- $\operatorname{sh}^{-1}(x) = \ln(x+\sqrt{x^2+1})$ for $x \in \mathbb{R}$,
- $\operatorname{ch}^{-1}(x) = \ln(x+\sqrt{x^2-1})$ for $x \in [1,+\infty)$.

Implicit function: Suppose that $y(x)$ is a function determined by the equation $F(x,y) = 0$. To solve the integral of the form

$$\int G(x,y)dx,$$

we can consider a parametric system:

$$\begin{cases} x = \varphi(t) \\ y = \psi(t) \end{cases}$$

so that $F(\varphi(t),\psi(t)) = 0$. Then solve the integral

$$\int G(\varphi(t),\psi(t))d\varphi(t).$$

Example 3.1.11. Suppose that $y(x)$ is a function determined by the equation

$$e^{xy} + x^2y - 1 = 0.$$

Find

$$\int x^2y^2dx.$$

证明. Let $t = xy$, and we get a parametric system

$$\begin{cases} x = \dfrac{1-e^t}{t} \\ \\ y = \dfrac{t^2}{1-e^t} \end{cases}$$

for $t \neq 0$. Then we have

$$\begin{aligned} \int x^2y^2dx =& \int t^2 d\left(\frac{1-e^t}{t}\right) \\ =& \int t^2 \cdot \frac{-te^t - (1-e^t)}{t^2} dt \\ =& -\int te^t dt - \int dt + \int e^t dt \\ \xlongequal{\text{by Example 3.2.2}}& -(te^t - e^t) - t + e^t + C \\ =& -xye^{xy} + 2e^{xy} - xy + C. \end{aligned}$$

□

Example 3.1.12 ([1, §3.1 例 16]). Find $\int \frac{\sqrt{4-x^2}}{x}dx$.

证明. Let $u = \sqrt{4-x^2}$. Then $x^2 = 4-u^2$, and so $2xdx = -2udu$. Then we have

$$\begin{aligned} \int \frac{\sqrt{4-x^2}}{x}dx =& \int \frac{\sqrt{4-x^2}}{x^2}xdx \\ =& \int \frac{u}{4-u^2}(-udu) \\ =& \int \frac{-u^2}{4-u^2}du \\ =& \int 1 - \frac{4}{4-u^2}du \\ =& u - 4\int \frac{1}{4-u^2}du \\ \xlongequal{\text{by Example 3.1.5}}& u - 4 \cdot \frac{1}{4}\ln\left|\frac{2+u}{2-u}\right| + C \\ =& \sqrt{4-x^2} - \ln\left|\frac{2+\sqrt{4-x^2}}{2-\sqrt{4-x^2}}\right| + C. \end{aligned}$$

□

3.2 Integration by parts

Recall the product rule (Leibniz rule):

$$d(uv) = udv + vdu.$$

Take the integrals and we get

$$\int udv = \int d(uv) - \int vdu = uv - \int vdu. \tag{3.2.1}$$

The equation (3.2.1) is called the *integration by parts formula* for indefinite integrals.

For the definite integrals on $[a, b]$, by (3.2.1) and the second fundamental theorem of calculus, we have

Theorem 3.2.1 ([1, §3.4 定理 1]). *Suppose that $u(x)$ and $v(x)$ have continuous derivatives on $[a, b]$. Then we have*

$$\int_{x=a}^{b} u(x)dv(x) = u(x)v(x)\Big|_a^b - \int_{x=a}^{b} v(x)du(x). \tag{3.2.2}$$

The equation (3.2.2) is called the *integration by parts formula* for definite integrals.

Example 3.2.2 ([1, §3.2 例 1]). Find $\int xe^x dx$.

证明.

$$\int xe^x dx = \int xde^x = xe^x - \int e^x dx = xe^x - e^x + C.$$

□

Example 3.2.3 ([1, §3.2 例 3]). Find $\int x^3 \ln x dx$.

证明.

$$\begin{aligned}\int x^3 \ln x dx =& \frac{1}{4}\int \ln x dx^4 \\ =& \frac{1}{4}x^4 \ln x - \frac{1}{4}\int x^4 d\ln x \\ =& \frac{1}{4}x^4 \ln x - \frac{1}{4}\int x^3 dx \\ =& \frac{1}{4}x^4 \ln x - \frac{1}{16}x^4 + C.\end{aligned}$$

□

Example 3.2.4 ([1, §3.2 例 4]). Find $\int \arctan x dx$.

证明.

$$\begin{aligned}\int \arctan x dx =& x \arctan x - \int x d \arctan x \\ =& x \arctan x - \int \frac{x}{1+x^2} dx \\ =& x \arctan x - \frac{1}{2} \int \frac{1}{1+x^2} d(1+x^2) \\ =& x \arctan x - \frac{1}{2} \ln(1+x^2) + C.\end{aligned}$$

□

Example 3.2.5 ([1, §3.2 例 6]). Find $\int \sqrt{a^2+x^2} dx$, for $a > 0$.

证明.

$$\begin{aligned}\int \sqrt{a^2+x^2} dx =& x\sqrt{a^2+x^2} - \int x \cdot \frac{1}{2} \cdot \frac{2x}{\sqrt{a^2+x^2}} dx \\ =& x\sqrt{a^2+x^2} - \int \frac{x^2}{\sqrt{a^2+x^2}} dx \\ =& x\sqrt{a^2+x^2} + \int \frac{a^2}{\sqrt{a^2+x^2}} dx - \int \sqrt{a^2+x^2} dx \\ \xlongequal{\text{by Example3.1.10}}& x\sqrt{a^2+x^2} + a^2 \ln(x+\sqrt{a^2+x^2}) - \int \sqrt{a^2+x^2} dx.\end{aligned}$$

Thus, we conclude that

$$\int \sqrt{a^2+x^2} dx = \frac{1}{2} x\sqrt{a^2+x^2} + \frac{a^2}{2} \ln(x+\sqrt{a^2+x^2}) + C.$$

□

Example 3.2.6 ([1, §3.2 例 8]). Find $\int e^{ax} \cos bx dx$, for $a, b > 0$.

证明.

$$\begin{aligned}\int e^{ax}\cos bx dx =&\frac{1}{a}\int \cos bx de^{ax}\\=&\frac{1}{a}e^{ax}\cos bx-\frac{1}{a}\int e^{ax}d\cos bx\\=&\frac{1}{a}e^{ax}\cos bx+\frac{b}{a}\int e^{ax}\sin bx dx\\=&\frac{1}{a}e^{ax}\cos bx+\frac{b}{a^2}\int \sin bx de^{ax}\\=&\frac{1}{a}e^{ax}\cos bx+\frac{b}{a^2}e^{ax}\sin bx-\frac{b}{a^2}\int e^{ax}d\sin bx\\=&\frac{1}{a}e^{ax}\cos bx+\frac{b}{a^2}e^{ax}\sin bx-\frac{b^2}{a^2}\int e^{ax}\cos bx dx.\end{aligned}$$

Then we obtain

$$\int e^{ax}\cos bx dx=\frac{1}{a^2+b^2}(ae^{ax}\cos bx+be^{ax}\sin bx)+C.$$

□

Example 3.2.7 ([1, §3.2 例 9])**.** Find $I_n=\int \frac{dt}{(t^2+a^2)^n}$, for $a>0$ and integer $n\geq 2$.

证明.

$$\begin{aligned}I_n =&\frac{t}{(t^2+a^2)^n}+2n\int \frac{t^2dt}{(t^2+a^2)^{n+1}}\\=&\frac{t}{(t^2+a^2)^n}+2n\int \frac{t^2+a^2-a^2}{(t^2+a^2)^{n+1}}dt\\=&\frac{t}{(t^2+a^2)^n}+2nI_n-2na^2\int \frac{1}{(t^2+a^2)^{n+1}}dt\\=&\frac{t}{(t^2+a^2)^n}+2nI_n-2na^2I_{n+1}.\end{aligned}$$

Therefore, we get

$$I_{n+1}=\frac{2n-1}{2na^2}I_n+\frac{t}{2na^2(t^2+a^2)^n}. \tag{3.2.3}$$

Note that

$$I_1=\int \frac{dt}{a^2+t^2}=\frac{1}{a}\arctan\frac{t}{a}+C. \tag{3.2.4}$$

The consequence follows from (3.2.3) and (3.2.4). □

3.3 Antiderivatives of rational fractions

3.3.1 Rational fractions

Let us consider the problem of integrating $\int R(x)dx$ where $R(x)$ is a *rational fraction* (有理分式), i.e. $R(x) = \dfrac{P(x)}{Q(x)}$ for polynomials P, Q with real coefficients.

Let

$$\mathbb{R}[x] = \left\{ \sum_{i=0}^{n} a_i x^i : n \in \mathbb{N},\ a_i \in \mathbb{R} \right\}.$$

A rational fraction $R(x) = \dfrac{P(x)}{Q(x)}$ is called *proper* if the degrees (see also Appendix 7.7) $\deg(P) < \deg(Q)$, and *improper* otherwise.

First, by the *fundamental theorem of algebra* (代数基本定理) (see Corollary 7.7.5), we can write

$$Q(x) = c\prod_{i=1}^{l}(x - x_i)^{k_i} \cdot \prod_{i=1}^{n}(x^2 + p_i x + q_i)^{m_i}$$

for $c, x_i, p_i, q_i \in \mathbb{R}$, and $l, n, k_i, m_i \in \mathbb{N}$.

Then by the *partial fraction decomposition* (部分分式分解) (see Theorem 7.7.8), we can rewrite the rational fraction

$$\frac{P(x)}{Q(x)} = p(x) + \sum_{i=1}^{l}\sum_{j=1}^{k_i} \frac{A_{ij}}{(x - x_i)^j} + \sum_{i=1}^{n}\sum_{j=1}^{m_i} \frac{B_{ij}x + D_{ij}}{(x^2 + p_i x + q_i)^j}$$

where $p(x) \in \mathbb{R}[x]$ and $A_{ij}, B_{ij}, D_{ij} \in \mathbb{R}$.

Example 3.3.1. Let $P(x) = 4x^3 + 6x^2 + x + 2$, $Q(x) = x^2(x^2 + 1)$. Then the partial fraction decomposition of $P(x)/Q(x)$ is

$$\frac{P(x)}{Q(x)} = \frac{1}{x} + \frac{2}{x^2} + \frac{3x + 4}{x^2 + 1}.$$

Thus, we can find the integral of the rational fractions step by step:

- $\displaystyle\int \frac{A}{x - a}dx = a\ln|x - a| + C.$
- $\displaystyle\int \frac{A}{(x - a)^n}dx = \frac{A}{1 - n}(x - a)^{1-n} + C \ (n > 1).$

- For the rational fraction $\dfrac{Bx+D}{x^2+px+q}$, we have

$$\begin{aligned}
&\int \frac{Bx+D}{x^2+px+q}dx \\
=&\frac{B}{2}\int \frac{1}{\left(x+\frac{p}{2}\right)^2+\left(q-\frac{p^2}{4}\right)}d\left(x+\frac{p}{2}\right)^2+\left(D-\frac{Bp}{2}\right)\int \frac{1}{\left(x+\frac{p}{2}\right)^2+\left(q-\frac{p^2}{4}\right)}dx \\
=&\frac{B}{2}\ln|x^2+px+q|+\frac{D-\frac{Bp}{2}}{\sqrt{q-\frac{p^2}{4}}}\arctan\frac{x+\frac{p}{2}}{\sqrt{q-\frac{p^2}{4}}}+C.
\end{aligned}$$

- For the rational fraction $\dfrac{Bx+D}{(x^2+px+q)^n}$ with $n \geq 2$, we have

$$\begin{aligned}
&\int \frac{Bx+D}{(x^2+px+q)^n}dx \\
=&\frac{B}{2}\int \frac{1}{\left(\left(x+\frac{p}{2}\right)^2+\left(q-\frac{p^2}{4}\right)\right)^n}d\left(x+\frac{p}{2}\right)^2+\left(D-\frac{Bp}{2}\right)\int \frac{1}{\left(\left(x+\frac{p}{2}\right)^2+\left(q-\frac{p^2}{4}\right)\right)^n}dx \\
=&\frac{B}{2(1-n)}(x^2+px+q)^{1-n}+\left(D-\frac{Bp}{2}\right)\int \frac{1}{\left(\left(x+\frac{p}{2}\right)^2+\left(q-\frac{p^2}{4}\right)\right)^n}dx.
\end{aligned}$$

  The last integral can be obtained from Example 3.2.7.

Therefore, we can find the antiderivative of any rational fraction.

Example 3.3.2 ([1, §3.3 例 2])**.** Find $\displaystyle\int \frac{x^3+1}{x(x-1)^3}dx$.

证明. First, the rational fraction is proper. Then by the partial fraction decomposition, we have

$$\frac{x^3+1}{x(x-1)^3}=\frac{A_{11}}{x}+\frac{A_{21}}{x-1}+\frac{A_{22}}{(x-1)^2}+\frac{A_{23}}{(x-1)^3}.$$

That is,

$$\frac{x^3+1}{x(x-1)^3}=\frac{A_{11}(x-1)^3+A_{21}x(x-1)^2+A_{22}x(x-1)+A_{23}x}{x(x-1)^3}.$$

Consider the numerators only. Take $x=1$ on both sides, and we get

$$1+1=A_{23}.$$

Thus, we can rewrite

$$x^3 - 2x + 1 = A_{11}(x-1)^3 + A_{21}x(x-1)^2 + A_{22}x(x-1).$$

Since $x^3 - 2x + 1 = (x-1)(x^2+x-1)$, one can divide the equation by $(x-1)$, then take $x = 1$, and obtain:

$$1 + 1 - 1 = A_{22}.$$

Then, we can rewrite

$$x^2 - 1 = A_{11}(x-1)^2 + A_{21}x(x-1).$$

Again, one can divide the equation by $(x-1)$, then take $x = 1$, and obtain:

$$1 + 1 = A_{21}.$$

Finally, the equation reduces to

$$-x + 1 = A_{11}(x-1).$$

So $A_{11} = -1$. Therefore, we obtain

$$\frac{x^3+1}{x(x-1)^3} = \frac{-1}{x} + \frac{2}{x-1} + \frac{1}{(x-1)^2} + \frac{2}{(x-1)^3}.$$

Take the integral, and we get

$$\int \frac{x^3+1}{x(x-1)^3}dx = -\ln|x| + 2\ln|x-1| - \frac{1}{x-1} - \frac{1}{(x-1)^2} + C.$$

□

Example 3.3.3 ([1, §3.3 例 3])**.** Find $\int \frac{4}{x^3+4x}dx$.

证明. First, the rational fraction is proper. Also $x^3 + 4x = x(x^2+4)$. Then by the partial fraction decomposition, we have

$$\frac{4}{x^3+4x} = \frac{A_{11}}{x} + \frac{B_{21}x + D_{21}}{x^2+4}.$$

Thus,

$$4 = A_{11}(x^2+4) + B_{21}x^2 + D_{21}x.$$

Letting $x = 0$, we have $A_{11} = 1$. Then the equation reduces to

$$0 = x^2 + B_{21}x^2 + D_{21}x.$$

It follows that $B_{21} = -1$ and $D_{21} = 0$. Thus, we obtain

$$\frac{4}{x^3 + 4x} = \frac{1}{x} - \frac{x}{x^2 + 4}.$$

Take the integral, and we get

$$\int \frac{4}{x^3 + 4x} dx = \ln|x| - \frac{1}{2}\ln|x^2 + 4| + C.$$

□

Example 3.3.4 ([1, §3.3 例 5])**.** Find $\int \frac{x^3}{x^2 + x - 2} dx$.

证明. First, the rational fraction is improper. So we rewrite

$$\frac{x^3}{x^2 + x - 2} = x + \frac{-x^2 + 2x}{x^2 + x - 2} = x - 1 + \frac{3x - 2}{x^2 + x - 2}.$$

Also $x^2 + x - 2 = (x - 1)(x + 2)$. Then by the partial fraction decomposition, we have

$$\frac{3x - 2}{x^2 + x - 2} = \frac{A_{11}}{x - 1} + \frac{A_{21}}{x + 2}.$$

Thus,

$$3x - 2 = A_{11}(x + 2) + A_{21}(x - 1).$$

Letting $x = 1$, we have $A_{11} = \frac{1}{3}$. Letting $x = -2$, we have $A_{21} = \frac{8}{3}$. Thus, we obtain

$$\begin{aligned}\int \frac{x^3}{x^2 + x - 2} dx &= \int x - 1 + \frac{1}{3(x - 1)} + \frac{3}{8(x + 2)} dx \\ &= \frac{1}{2}x^2 - x + \frac{1}{3}\ln|x - 1| + \frac{8}{3}\ln|x + 2| + C.\end{aligned}$$

□

3.3.2 Trigonometric functions

Let $R(x, y)$ be a rational fraction in x and y, i.e. $R(x, y) = \dfrac{P(x, y)}{Q(x, y)}$ for polynomials P, Q, which are linear combinations of monomials $x^m y^n$ for $m, n \in \mathbb{N}$.

Several methods exist for computing the integral

$$\int R(\cos x, \sin x)dx.$$

One of which is completely general, although not always the most efficient: We make the change of variable

$$t = \tan \frac{x}{2}.$$

Then we have the identities

$$\cos x = \frac{1-t^2}{1+t^2}, \quad \sin x = \frac{2t}{1+t^2}, \quad dx = \frac{2dt}{1+t^2}. \tag{3.3.1}$$

Then we have

$$\int R(\cos x, \sin x)dx = \int R\left(\frac{1-t^2}{1+t^2}, \frac{2t}{1+t^2}\right)\frac{2}{1+t^2}dt.$$

Thus, we see that the integrand becomes a rational fraction in t.

Example 3.3.5 ([1, §3.3 例 6])**.** Consider $\displaystyle\int \frac{\cot x}{\sin x + \cos x - 1}dx$.

证明. Make the substitution $t = \tan \frac{x}{2}$. By (3.3.1), we have

$$\int \frac{\cot x}{\sin x + \cos x - 1}dx = \int \frac{1+t}{2t^2}dt.$$

Then we see the problem reduces to Section 3.3.1. (One may solve the integral immediately.) □

There are other methods that may be more efficient in some situations. See also [1, P.206 Exercises 7 and 9].

3.3.3 Radicals

Let $R(x,y)$ be a rational fraction in x and y. Consider the integral of the form

$$\int R\left(x, \sqrt[n]{\frac{ax+b}{cx+d}}\right) dx.$$

We make the change of variable

$$t = \sqrt[n]{\frac{ax+b}{cx+d}}.$$

Then $x(t)$, and so the integrand become rational fractions in t.

Example 3.3.6 ([1, §3.3 例 11])**.** Consider $\int x\sqrt{\frac{x-1}{x+1}}dx$.

证明. Make the substitution $t = \sqrt{\frac{x-1}{x+1}}$. Then

$$x = \frac{1+t^2}{1-t^2}, \quad dx = \frac{4t}{(1-t^2)^2}dt.$$

Thus, we have

$$\int x\sqrt{\frac{x-1}{x+1}}dx = \int \frac{1+t^2}{1-t^2} \cdot t \cdot \frac{4t}{(1-t^2)^2}dt.$$

Then we see the problem reduces to Section 3.3.1. $\square$

3.4 Techniques of definite integrals

3.4.1 Integration by parts in the definite integral

By using Theorem 2.10.1′, we obtain a stronger form of Theorem 3.2.1:

Theorem 3.4.1 (Integration by parts formula)**.** *Let $u, v : [a,b] \to \mathbb{R}$ be differentiable functions on $[a,b]$ such that u', v' are Riemann integrable on $[a,b]$. Then we have*

$$\int_a^b u(x)v'(x)dx = u(x)v(x)\Big|_a^b - \int_a^b v(x)u'(x)dx.$$

Example 3.4.2 (Wallis' integrals 沃利斯公式, [1, §3.4 例 2]). Find

$$I_n = \int_0^{\frac{\pi}{2}} \sin^n x dx$$

for integer $n \geq 2$.

证明. Write

$$I_n = -\int_0^{\frac{\pi}{2}} \sin^{n-1} x d\cos x.$$

Then by the integration by parts formula, we have

$$\begin{aligned}
I_n &= -\sin^{n-1} x \cos x\Big|_0^{\frac{\pi}{2}} + \int_0^{\frac{\pi}{2}} \cos x d\sin^{n-1} x \\
&= (n-1)\int_0^{\frac{\pi}{2}} \cos^2 x \sin^{n-2} x dx \\
&= (n-1)\int_0^{\frac{\pi}{2}} (1-\sin^2 x)\sin^{n-2} x dx \\
&= (n-1)I_{n-2} - (n-1)I_n.
\end{aligned}$$

Thus, we get

$$I_n = \frac{n-1}{n} I_{n-2}.$$

Note that

$$I_0 = \int_0^{\frac{\pi}{2}} 1dx = \frac{\pi}{2}, \quad I_1 = \int_0^{\frac{\pi}{2}} \sin x dx = 1.$$

Thus, we conclude that

$$I_n = \begin{cases} \frac{(n-1)!!}{n!!} \cdot \frac{\pi}{2} & , \text{ if } n = 2k \\ \frac{(n-1)!!}{n!!} & , \text{ if } n = 2k+1 \end{cases}.$$

□

3.4.2 Change of variables in the definite integral

We have seen in Theorem 3.1.1 a weaker form of the following change of variables formula for definite integrals:

Theorem 3.4.3. *Suppose that $y : [a,b] \to [A,B]$ has continuous derivatives, and $F : [A,B] \to \mathbb{R}$ is integrable. Then we have*

$$\int_a^b F(y(x))y'(x)dx = \int_{y(a)}^{y(b)} F(y)dy. \tag{3.4.1}$$

It is instructive to rewrite (3.4.1) as Riemann sums. Consider a partition T of $[a, b]$:

$$a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b.$$

Then the left hand side of (3.4.1) becomes

$$\sum_{i=1}^{n} F(y(\xi_i))y'(\xi_i)\Delta x_i \to \int_a^b F(y(x))y'(x)dx \tag{3.4.2}$$

For simplicity, we assume that y is strictly increasing. Then $y(T)$ is a partition of $[y(a), y(b)]$:

$$y(a) = y(x_0) < y(x_1) < y(x_2) < \cdots < y(x_{n-1}) < y(x_n) = y(b).$$

And we have

$$\sum_{i=1}^{n} F(y(\xi_i))\Delta y_i \to \int_{y(a)}^{y(b)} F(y)dy. \tag{3.4.3}$$

Now note that

$$\Delta y_i = y(x_i) - y(x_{i-1}) = y'(\xi_i)\Delta x_i + o(\Delta x_i). \tag{3.4.4}$$

Then we see that

$$\begin{aligned}\sum_{i=1}^{n} F(y(\xi_i))\Delta y_i &= \sum_{i=1}^{n} F(y(\xi_i)) \cdot (y'(\xi_i)\Delta x_i + o(\Delta x_i)) \\ &= \sum_{i=1}^{n} F(y(\xi_i))y'(\xi_i)\Delta x_i + \sum_{i=1}^{n} F(y(\xi_i))o(\Delta x_i).\end{aligned} \tag{3.4.5}$$

Heuristically, since F is bounded, i.e. $|F(x)| \leq M$ for $x \in [a, b]$, we should have

$$\left|\sum_{i=1}^{n} F(y(\xi_i))o(\Delta x_i)\right| \leq o(M(b-a)) \to 0. \tag{3.4.6}$$

Therefore, combining (3.4.3)(3.4.5)(3.4.6), we have

$$\sum_{i=1}^{n} F(y(\xi_i))y'(\xi_i)\Delta x_i \to \int_{y(a)}^{y(b)} F(y)dy. \tag{3.4.7}$$

Comparing (3.4.2) and (3.4.7), we obtain a proof of the change of variables formula.

Remark 3.4.4. Rigorously, in order to prove (3.4.6), we need $o(\Delta x_i)/\Delta x_i \to 0$ **uniformly/simultaneously** as $\max_i\{\Delta x_i\} \to 0$. This can be done by using the *uniform continuity* of y'.

More precisely, by Theorem 7.1.12, for any $\epsilon > 0$, we have

$$|y'(x) - y'(y)| < \epsilon \tag{3.4.8}$$

for any $x, y \in [a, b]$ with $|x - y| < \delta$. Now let $\lambda(T) = \max_i\{\Delta x_i\} < \delta$. Then (3.4.4) can be rewritten as

$$\Delta y_i \leq \sup_{\xi \in [x_{i-1}, x_i]} y'(\xi)\Delta x_i \leq \inf_{\xi \in [x_{i-1}, x_i]} y'(\xi)\Delta x_i + \epsilon \Delta x_i,$$

and

$$\Delta y_i \geq \inf_{\xi \in [x_{i-1}, x_i]} y'(\xi)\Delta x_i \geq \sup_{\xi \in [x_{i-1}, x_i]} y'(\xi)\Delta x_i - \epsilon \Delta x_i.$$

It follows that

$$|F(y(\xi_i))\Delta y_i - F(y(\xi_i))y'(\xi_i)\Delta x_i| \leq \epsilon |F(y(\xi_i))|\Delta x_i.$$

Then the consequence follows from the fact that F is integrable on $[a, b]$. The proof also shows that Theorem 3.1.1 only requires that F is integrable.

Remark 3.4.5. The proof clearly shows the relationship between **differentiation** and **integration**. In particular, when $F(y) \equiv 1$, we recover the *second fundamental theorem of calculus* (*Newton-Leibniz formula*).

Remark 3.4.6. The best linear approximation (3.4.4) plays a crucial role in the proof. Let $L_i = dy(\xi_i)$ be the linear map

$$L_i : \Delta x \mapsto y'(\xi_i)\Delta x.$$

Locally, the length of $[y(x_{i-1}), y(x_i)]$ is approximated by the length $L_i([x_{i-1}, x_i])$:

$$\text{length}([y(x_{i-1}), y(x_i)]) \sim \text{length}(L_i[x_{i-1}, x_i]) = |y'(\xi_i)| \cdot \text{length}([x_{i-1}, x_i]).$$

The idea will also be used in high-dimensional integration.

More precisely, let $\mathbf{y} : \mathbb{R}^m \to \mathbb{R}^m$ be differentiable. Then the differential $d\mathbf{y}(\xi_i)$ is represented by the *Jacobian matrix* $J_{\mathbf{y}(\xi_i)}$. The high-dimensional Riemann integral considers the Riemann sum of the *volume element* (体积元), i.e. the volume of the cubes:

$$V_i = [x_{i-1}^{(1)}, x_i^{(1)}] \times [x_{i-1}^{(2)}, x_i^{(2)}] \times \cdots \times [x_{i-1}^{(m)}, x_i^{(m)}] \subset \mathbb{R}^m.$$

Then the change of variables formula is deduced by the observation

$$\mathrm{volume}(\mathbf{y}(V_i)) \sim |\det(J_{\mathbf{y}(\xi_i)})| \cdot \mathrm{volume}(V_i)$$

where $\det(\cdot)$ denotes the *determinant* (行列式) of a matrix (See Definition 7.2.32).

Example 3.4.7 ([1, §3.4 例 5]). Prove that $\int_0^{\frac{\pi}{2}} \cos^n x dx = \int_0^{\frac{\pi}{2}} \sin^n x dx$.

证明. Let $x = \frac{\pi}{2} - t$ for $t \in [0, \frac{\pi}{2}]$. Then we have

$$\int_0^{\frac{\pi}{2}} \cos^n x dx = \int_{\frac{\pi}{2}}^0 \sin^n t \cdot (-dt) = \int_0^{\frac{\pi}{2}} \sin^n t dt.$$

□

Example 3.4.8 ([1, §3.4 例 6]). Find $\int_0^a x^4 \sqrt{a^2 - x^2} dx$.

证明. Let $x = a \cos t$ for $t \in [0, \frac{\pi}{2}]$. Then we have

$$\begin{aligned}
\int_0^a x^4 \sqrt{a^2 - x^2} dx &= \int_{\frac{\pi}{2}}^0 a^4 \cos^4 t \cdot a \sin t \cdot (-a \sin t) dt \\
&= a^6 \int_0^{\frac{\pi}{2}} \cos^4 t \cdot (1 - \cos^2 t) dt \\
&\xlongequal{\text{by Example 3.4.7}} a^6 \int_0^{\frac{\pi}{2}} \sin^4 t dt - a^6 \int_0^{\frac{\pi}{2}} \sin^6 t dt \\
&\xlongequal{\text{by Example 3.4.2}} a^6 I_4 - a^6 I_6 = \frac{\pi}{32} a^6.
\end{aligned}$$

□

Example 3.4.9 ([1, §3.4 例 4]). Calculate the area of the ellipse S:

$$S = \left\{ (x, y) \in \mathbb{R}^2 : \frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1 \right\}$$

for $a, b > 0$.

证明. The area can be expressed as

$$\text{Area}(S) = 2\int_{-a}^{a} b\sqrt{1-\frac{x^2}{a^2}}dx.$$

Let $x = a\sin t$ for $t \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. Then

$$\begin{aligned}\text{Area}(S) =& 2b\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos t \cdot a\cos t dt\\ =& 2ab\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^2 t dt\\ =& 2ab\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{1}{2}(1+\cos 2t)dt\\ =& ab\left(t+\frac{1}{2}\sin 2t\right)\Bigg|_{-\frac{\pi}{2}}^{\frac{\pi}{2}} = \pi ab.\end{aligned}$$

□

There is another method, using the linear algebra, to find the area of ellipses.

Proof of Example 3.4.9 using the linear algebra. Consider the 2-dimensional unit ball B:

$$B = \left\{\begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 : x^2+y^2 \le 1\right\}.$$

(Here we consider the *column vectors.*) Define $L : \mathbb{R}^2 \to \mathbb{R}^2$ by

$$L : \begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax \\ by \end{bmatrix}.$$

Then for any $\begin{bmatrix} u \\ v \end{bmatrix} \in L(B)$, we can write $\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} ax \\ by \end{bmatrix}$ with $x^2+y^2 \le 1$. So

$$\frac{u^2}{a^2}+\frac{v^2}{b^2} \le 1.$$

That is, $L(B)$ is an ellipse with width $2a$ and height $2b$. Then by Theorem 7.2.34, we have

$$\text{Area}(L(B)) = |\det L| \cdot \text{Area}(B) = ab\cdot\pi.$$

□

More generally, any linear map $L : \mathbb{R}^2 \to \mathbb{R}^2$ assigns an ellipse to a ball. See also the *singular value decomposition* (奇异值分解, Theorem 7.2.42).

3.4.3 Even, odd, and periodic functions

Recall that a function f on $[-a, a]$ is called *even* if

$$f(x) = f(-x)$$

for any $x \in [-a, a]$; is called *odd* if

$$f(x) = -f(-x)$$

for any $x \in [-a, a]$.

Proposition 3.4.10 ([1, §3.4 命题 1])**.** *Let $f(x)$ be even on $[-a, a]$, $g(x)$ odd on $[-a, a]$. Suppose further that f and g are integrable. Then*

$$\int_{-a}^{a} f(x)dx = 2\int_{0}^{a} f(x)dx, \quad \int_{-a}^{a} g(x)dx = 0.$$

证明. Since f is even, by Theorem 3.4.3, we have

$$\begin{aligned}\int_{-a}^{a} f(x)dx &= \int_{0}^{a} f(x)dx + \int_{-a}^{0} f(x)dx \\ &= \int_{0}^{a} f(x)dx + \int_{0}^{-a} f(-x)d(-x) \\ &= \int_{0}^{a} f(x)dx + \int_{0}^{a} f(t)d(t) = 2\int_{0}^{a} f(x)dx.\end{aligned}$$

For odd g, we have

$$\begin{aligned}\int_{-a}^{a} g(x)dx &= \int_{0}^{a} g(x)dx + \int_{-a}^{0} g(x)dx \\ &= \int_{0}^{a} g(x)dx + \int_{0}^{-a} -g(-x)d(-x) \\ &= \int_{0}^{a} g(x)dx - \int_{0}^{a} g(t)d(t) = 0.\end{aligned}$$

□

Example 3.4.11 ([1, §3.4 例 9])**.** Find $\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^5 x dx$.

证明. Note that $\cos^5 x$ is even. Then

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^5 x dx = 2\int_{0}^{\frac{\pi}{2}} \cos^5 x dx = 2\int_{0}^{\frac{\pi}{2}} \sin^5 x dx \xlongequal{\text{[1, §3.4 Example 2]}} 2I_5 = \frac{16}{15}.$$

□

Example 3.4.12 ([1, §3.4 例 10])**.** Find $\int_{-2}^{2} x\sin^4 x + x^3 - x^4 dx$.

证明. Note that $x\sin^4 x$ and x^3 are odd, and x^4 is even. Then

$$\int_{-2}^{2} x\sin^4 x + x^3 - x^4 dx = -2\int_0^2 x^4 dx = -\frac{2}{5}x^5\Big|_0^2 = -\frac{64}{5}.$$

□

Definition 3.4.13 (Periodic function)**.** Let $f(x)$ be a function on $\mathbb{R}$. If there is a number $T \neq 0$ such that

$$f(x+T) = f(x)$$

for any $x \in \mathbb{R}$, then we say $f(x)$ is a *periodic function*, and T is a *period* of $f(x)$.

If T is a period of f, then nT is also a period of f for $n \in \mathbb{Z}$. Hence, we see that a periodic function always has infinitely many periods. If a periodic function has a smallest positive period, then we say the period is the *smallest period*. Not any periodic function has smallest period, e.g. Dirichlet function.

Proposition 3.4.14 ([1, §3.4 命题 2])**.** *Let $f(x)$ be integrable on any bounded interval, T a period of $f(x)$. Then for any $a \in \mathbb{R}$, we have*

$$\int_0^T f(x)dx = \int_a^{a+T} f(x)dx.$$

证明. We write

$$\int_a^{a+T} f(x)dx = \int_a^T f(x)dx + \int_T^{T+a} f(x)dx.$$

Make the change of variable $t = x - T$. Then by Theorem 3.4.3, we get

$$\int_T^{T+a} f(x)dx = \int_0^a f(t+T)dt = \int_0^a f(x)dx.$$

Thus, we obtain

$$\int_a^{a+T} f(x)dx = \int_a^T f(x)dx + \int_0^a f(x)dx = \int_0^T f(x)dx.$$

□

Corollary 3.4.15. *Let $f(x)$ be integrable on any bounded interval, T a period of $f(x)$. Then for any $a, b \in \mathbb{R}$, we have*

$$\int_{a+T}^{b+T} f(x)dx = \int_a^b f(x)dx.$$

证明. One calculates

$$\begin{aligned}\int_{a+T}^{b+T} f(x)dx &= \int_{a+T}^{a} f(x)dx + \int_a^b f(x)dx + \int_b^{b+T} f(x)dx \\ &\xlongequal{\text{Proposition 3.4.14}} -\int_0^T f(x)dx + \int_a^b f(x)dx + \int_0^T f(x)dx = \int_a^b f(x)dx.\end{aligned}$$

□

Example 3.4.16 ([1, §3.4 例 11])**.** Find $\displaystyle\int_1^{1+\pi} |\cos x|dx$.

证明. Note that π is a period of $|\cos x|$. Then by Proposition 3.4.14, we have

$$\begin{aligned}\int_1^{1+\pi} |\cos x|dx = \int_0^{\pi} |\cos x|dx &= \int_{0-\frac{\pi}{2}}^{\pi-\frac{\pi}{2}} |\cos x|dx \\ &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos x dx = \sin x\Big|_{-\frac{\pi}{2}}^{\frac{\pi}{2}} = 2.\end{aligned}$$

□

Example 3.4.17 ([1, §3.4 例 12])**.** Find $\displaystyle\int_{\frac{\pi}{2}}^{\pi} |\sin 2x|dx$.

证明. Note that $\frac{\pi}{2}$ is a period of $|\sin 2x|$. Then by Corollary 3.4.15, we have

$$\int_{\frac{\pi}{2}}^{\pi} |\sin 2x|dx = \int_{0+\frac{\pi}{2}}^{\frac{\pi}{2}+\frac{\pi}{2}} |\sin 2x|dx = \int_0^{\frac{\pi}{2}} \sin 2x dx = -\frac{1}{2}\cos 2x\Big|_0^{\frac{\pi}{2}} = 1.$$

□

Example 3.4.18 ([1, §3.4 例 13])**.** Find $\displaystyle\int_0^n (x-[x])dx$, where $n \in \mathbb{N}^*$.

证明. Note that 1 is a period of $x-[x]$. Then by Proposition 3.4.14, we have

$$\int_0^n (x-[x])dx = \sum_{i=1}^{n} \int_{i-1}^{i} (x-[x])dx = n\int_0^1 (x-[x])dx = n\int_0^1 xdx = \frac{n}{2}.$$

□

3.5 Applications of integration

3.5.1 Additive interval function and the integral

Definition 3.5.1 (Additive interval function)**.** Let $[a, b]$ be an interval. An *additive interval function* (or *finitely additive measure*) I on $[a, b]$ is a function that assigns a number $I[c, d] \in \mathbb{R}$ to each closed interval $[c, d] \subset [a, b]$:

$$[c, d] \mapsto I[c, d]$$

such that

$$I[c, e] = I[c, d] + I[d, e] \tag{3.5.1}$$

for any $c \le d \le e \in [a, b]$.

Example 3.5.2. Let f be integrable on $[a, b]$. Then

$$I[c, d] = \int_c^d f(t)dt$$

is an additive interval function.

Proposition 3.5.3. *Let I be an additive interval function on $[a, b]$. Suppose that there exists an integrable function f on $[a, b]$ so that*

$$\inf_{x\in[c,d]} f(x)(d - c) \le I[c, d] \le \sup_{x\in[c,d]} f(x)(d - c) \tag{3.5.2}$$

holds for any $[c, d] \subset [a, b]$. Then

$$I[a, b] = \int_a^b f(t)dt.$$

Remark 3.5.4. In terms of differentials, (3.5.2) can be represented as

$$dI = f(x)dx$$

where $dx = \Delta x = d - c$ and $dI = I[c, d]$.

Proof of Proposition 3.5.3. Let T be an partition of $[a, b]$:

$$a = x_0 < x_1 < \cdots < x_n = b.$$

Write $\Delta x_i = x_i - x_{i-1}$, $\lambda(T) = \max\{\Delta x_i : i = 1, 2, \ldots, n\}$. Let

$$m_i = \inf_{x \in [x_{i-1}, x_i]} f(x), \quad M_i = \sup_{x \in [x_{i-1}, x_i]} f(x).$$

By (3.5.2), we have

$$m_i \Delta x_i \leq I[x_{i-1}, x_i] \leq M_i \Delta x_i.$$

Then by (3.5.1), we have

$$\sum_{i=1}^{n} m_i \Delta x_i \leq I[a, b] \leq \sum_{i=1}^{n} M_i \Delta x_i.$$

Then we see that $I[a, b]$ is dominated by the upper and lower Darboux sums of the function f (cf. [1, P.132, Exercise 7]). Therefore, letting $\lambda(T) \to 0$, we establish the proposition. □

3.5.2 Arc length

Definition 3.5.5 (Arc)**.** An *arc* $\overset{\frown}{AB}$ in a plane is a continuous map $\Gamma : [\alpha, \beta] \to \mathbb{R}^2$ given by a parametric system:

$$\Gamma : t \mapsto \begin{cases} x = x(t) \\ y = y(t) \end{cases}$$

with $\Gamma(\alpha) = A$, $\Gamma(\beta) = B$. An arc $\overset{\frown}{AB}$ is called *smooth* if $x(t)$ and $y(t)$ have continuous derivatives.

Let $s[\alpha, \beta]$ be the length of the arc Γ. Then one may show that s is an additive interval function. Let $\Gamma(c) = M$, $\Gamma(d) = N$. Then

$$s[c, d] = |MN| + o(d - c) = \sqrt{(x(d) - x(c))^2 + (y(d) - y(c))^2} + o(d - c).$$

Since the arc Γ is smooth, we have

$$x(d) - x(c) \sim x'(d)(d - c), \quad y(d) - y(c) \sim y'(d)(d - c), \quad (d \to c)$$

and so

$$s[c, d] = \sqrt{(x'(d))^2 + (y'(d))^2} \cdot (d - c) + o(d - c).$$

In terms of differentials, we obtain:

$$ds = \sqrt{(x'(t))^2 + (y'(t))^2}dt.$$

Also, one may check that the additive interval function s is dominated by the absolute value of the velocity function

$$|\Gamma'(t)| = \sqrt{(x'(t))^2 + (y'(t))^2}.$$

(See also Section 5.4.) Therefore, by Proposition 3.5.3, we conclude

$$s = \int_{\alpha}^{\beta} \sqrt{(x'(t))^2 + (y'(t))^2}dt. \tag{3.5.3}$$

Remark 3.5.6. In particular, note that

- if an arc $\Gamma : [\alpha, \beta] \mapsto \mathbb{R}^2$ is given by

  $$x \mapsto (x, f(x))$$

  then we have

  $$s = \int_{\alpha}^{\beta} \sqrt{1 + (f'(x))^2}dx;$$

- if an arc $\Gamma : [\alpha, \beta] \mapsto \mathbb{R}^2$ is given by the polar coordinate

  $$\theta \mapsto (r(\theta)\cos\theta, r(\theta)\sin\theta)$$

  where $r = r(\theta)$ has a continuous derivative on (a, b), we have

  $$s = \int_{\alpha}^{\beta} \sqrt{r^2(\theta) + (r'(\theta))^2}d\theta.$$

Example 3.5.7 ([1, §3.5 例 2])**.** Find the circumference of the ellipse:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \tag{3.5.4}$$

where $a \geq b > 0$.

证明. Consider the parametric system of (3.5.4)

$$\begin{cases} x = a\sin\theta \\ y = b\cos\theta \end{cases}$$

for $\theta \in [0, 2\pi]$. Then the differential form

$$ds = \sqrt{(a\cos\theta)^2 + (-b\sin\theta)^2}d\theta = \sqrt{a^2 - (a^2 - b^2)\sin^2\theta}d\theta.$$

Thus, the circumference of (3.5.4) is

$$s = \int_0^{2\pi} \sqrt{a^2 - k^2\sin^2\theta}d\theta, \quad k^2 = a^2 - b^2.$$

If $a = b$, then the ellipse (3.5.4) becomes a circle, and

$$s = \int_0^{2\pi} \sqrt{a^2}d\theta = 2\pi a.$$

If $a > b$, then $k^2 \neq 0$, and the integral

$$E(k, \varphi) = \int_0^{\varphi} \sqrt{a^2 - k^2\sin^2\theta}d\theta$$

is not an elementary function. $E(k, \varphi)$ is the *incomplete elliptic integral of second kind* in the Legendre's trigonometric form. □

3.5.3 Volume of a solid of revolution

Let $y = f(x)$ be nonnegative and continuous on $[a, b]$. Then $x \mapsto (x, f(x))$ is a curve in $\mathbb{R}^2$. Consider the solid generated by revolving the curve about the x-axis. We shall find the volume of the solid, and the surface area of the solid.

Let us first discuss the volume. Let $V[a, b]$ be the volume of the solid. Then V is an additive interval function. Moreover, for any $[c, d] \subset [a, b]$, we have

$$\inf_{x\in[c,d]} \pi f^2(x)(d - c) \leq V[c, d] \leq \sup_{x\in[c,d]} \pi f^2(x)(d - c).$$

In terms of differentials, we obtain:

$$dV = \pi f^2(x)dx.$$

Also, by Proposition 3.5.3, we conclude

$$V[a, b] = \pi \int_a^b f^2(x)dx.$$

Example 3.5.8. Find the volume of a 3-dimensional ball of radius R.

证明. By revolving about the x-axis the semicircle $y = \sqrt{R^2 - x^2}$ for $x \in [-R, R]$, we obtain the volume

$$V = \pi \int_{-R}^{R} (R^2 - x^2) dx = \frac{4}{3}\pi R^3.$$

□

3.5.4 Area of a surface of revolution

To see the area of a surface of revolution, let $F[a, b]$ be the area of the surface. Then F is an additive interval function. Moreover, for any $[c, d] \subset [a, b]$, we have

$$\inf_{x\in[c,d]} 2\pi f(x)s[c, d] \le F[c, d] \le \sup_{x\in[c,d]} 2\pi f(x)s[c, d]$$

where $s[c, d]$ is the length of the arc $(x, f(x))$ for $x \in [c, d]$ (cf. Section 3.5.2).

In terms of differentials, we obtain:

$$dF = 2\pi f(x)\sqrt{1 + f'(x)}dx.$$

Then, by Proposition 3.5.3, we conclude

$$F[a, b] = 2\pi \int_a^b f(x)\sqrt{1 + f'(x)}dx.$$

3.5.5 Areas in polar coordinates

Suppose a region is bounded by the rays $\theta = \alpha$ and $\theta = \beta$ and the curve $r = r(\theta)$. Let $S[\alpha, \beta]$ be the area of the region. Then S is an additive interval function. Moreover, for any $[c, d] \subset [\alpha, \beta]$, we have

$$\inf_{\theta\in[c,d]} \frac{1}{2}r^2(\theta)(d - c) \le S[c, d] \le \sup_{\theta\in[c,d]} \frac{1}{2}r^2(\theta)(d - c).$$

In terms of differentials, we obtain:

$$dS = \frac{1}{2}r^2(\theta)d\theta.$$

Then, by Proposition 3.5.3, we conclude

$$S[\alpha, \beta] = \frac{1}{2}\int_\alpha^\beta r^2(\theta)d\theta. \tag{3.5.5}$$

Example 3.5.9 ([1, §3.5 例 7])**.** Find the area of the rose with 3-petals:

$$r(\theta) = a\sin(3\theta)$$

for $\theta \in [0, 2\pi]$.

证明. Note that the area S is $3\times$ the area of $0 \le \theta \le \frac{\pi}{3}$. By (3.5.5), we have

$$S = 3 \times \frac{1}{2}\int_0^{\frac{\pi}{3}} a^2 \sin^2(3\theta) d\theta = \frac{1}{4}\pi a^2.$$

□

Chapter 4

Differential Calculus

4.1 Mean value theorems

The *Fermat's lemma* (费马引理) states that a differentiable function at the maximum or minimum has the derivative zero.

Lemma 4.1.1 (Fermat's lemma, [1, §4.5 定理 1])**.** *If $f : (a, b) \to \mathbb{R}$ attains a maximum or minimum at $c \in (a, b)$, and differentiable at c, then $f'(c) = 0$.*

证明*.* Assume that f attains a maximum at c. Then

$$f(x) - f(c) \leq 0$$

for $x \in [a, b]$. It follows that

$$f'(c) = \lim_{x \to c+0} \frac{f(x) - f(c)}{x - c} \leq 0,$$

$$f'(c) = \lim_{x \to c-0} \frac{f(x) - f(c)}{x - c} \leq 0.$$

Therefore, we conclude that $f'(c) = 0$. □

Together with the extreme value theorem (Theorem 1.6.3), we obtain the *Rolle's theorem* (罗尔中值定理).

Theorem 4.1.2 (Rolle's theorem, [1, §4.1 定理 1])**.** *Let $f(x)$ be continuous on $[a, b]$, differentiable on (a, b). If $f(a) = f(b)$, then there is a point $c \in (a, b)$ such that*

$$f'(c) = 0.$$

证明. Since f is continuous, by the extreme value theorem (Theorem 1.6.3), it attains a maximum M and a minimum m. These can only occur

(1) at endpoints a and b,

(2) at interior points $c \in (a, b)$.

If both the M and m occur at the endpoints, then because $f(a) = f(b)$, it must be the case that f is a constant function. Then $f'(x) \equiv 0$ for $x \in (a, b)$, and the point c can be taken anywhere in (a, b).

If either m or M occurs at a point $c \in (a, b)$. By Fermat's lemma (Lemma 4.1.1), we conclude that $f'(c) = 0$. □

Remark 4.1.3. By the assumption of Theorem 4.1.2, we have

$$\int f'(t)dt = f(x) + C.$$

Note that the proof relies on the continuity of f. This is in contrast to Theorem 2.9.1, where the proof relies on the continuity of f'. So we shall see that the *mean value theorem* deduced by Rolle's theorem is stronger than Theorem 2.9.1.

Example 4.1.4. Let $f : [0, 1] \to \mathbb{R}$ be continuous on $[0, 1]$, differentiable on $(0, 1)$, and $f(1) = 0$. Show that there is a $\xi \in (0, 1)$, such that $f'(\xi) = (1 - \xi^{-1})f(\xi)$.

证明. Solve the differential equation $f'(x) = (1 - x^{-1})f(x)$. Then

$$f(x) = \frac{e^x}{x}$$

is a solution of the equation. Consider $F(x) = xe^{-x}f(x)$. Then $F(0) = F(1) = 0$. By Rolle's theorem, there is a $\xi \in (0, 1)$ such that

$$0 = F'(\xi) = e^{-\xi}(f'(\xi) - (1 - \xi^{-1})f(\xi)).$$

The consequence follows. □

Example 4.1.5. Let $f : [0,1] \to \mathbb{R}$ be continuous on $[0,1]$, differentiable on $(0,1)$, $f(0) = f(1) = 0$, $f(\frac{1}{2}) = 1$. Show that for any $\lambda \in \mathbb{R}$, there is a $\xi \in (0,1)$, such that $f'(\xi) - \lambda(f(\xi) - \xi) = 1$.

证明. Solve the differential equation $(f(x) - x)' = \lambda(f(x) - x)$. Then

$$f(x) - x = ce^{\lambda x}.$$

is a solution of the equation for $c \in \mathbb{R}$. Consider $F(x) = e^{-\lambda x}(f(x) - x)$. By the extreme value theorem (Theorem 1.6.3), F attains a maximum at $\xi \in [0,1]$. Since $F(0) = 0$, $F(\frac{1}{2}) > 0$, $F(1) < 0$, we conclude $\xi \in (0,1)$. By Fermat's lemma, we have

$$0 = F'(\xi) = e^{-\xi}(f'(\xi) - 1 - \lambda(f(\xi) - \xi)).$$

The consequence follows. □

By removing the requirement that $f(a) = f(b)$, we obtain the *Lagrange's mean value theorem* (拉格朗日中值定理).

Theorem 4.1.6 (Lagrange's mean value theorem, [1, §4.1 定理 2])**.** *Let $f(x)$ be continuous on $[a,b]$, differentiable on (a,b). Then there is a point $c \in (a,b)$ such that*

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

证明. Let

$$h(x) = x(f(b) - f(a)) - f(x)(b - a).$$

Then $h(x)$ is continuous on $[a,b]$, differentiable on (a,b), and satisfies

$$h(a) = h(b).$$

Thus, by Rolle's theorem (Theorem 4.1.2), there is a point $c \in (a,b)$ such that

$$0 = h'(c) = f(b) - f(a) - f'(c)(b - a).$$

This establishes the theorem. □

As discussed in Remark 4.1.3, Lagrange's mean value theorem (Theorem 4.1.6) is stronger than Theorem 2.9.1. One consequence of this is that we can obtain a stronger form of the Newton-Leibniz formula (Theorem 2.10.1′):

Proof of Theorem 2.10.1′. For any closed interval $[c, d] \subset [a, b]$, let

$$I[c, d] = F(d) - F(c).$$

Then I is additive. Also, by the Lagrange's mean value theorem (Theorem 4.1.6), we have

$$\inf_{x\in[c,d]} f(x)(d-c) \le I[c, d] \le \sup_{x\in[c,d]} f(x)(d-c).$$

Therefore, by Proposition 3.5.3, we conclude Theorem 2.10.1′. □

Another consequence of Lagrange's mean value theorem is to show that the only functions f that have the derivative $f' \equiv 0$ is the constant functions.

Corollary 4.1.7 ($f' \equiv 0$ implies $f \equiv C$, [1, §4.1 推论])**.** *Suppose that f is continuous on $[a, b]$, differentiable on (a, b) with $f'(x) = 0$ for any $x \in (a, b)$. Then $f(x) \equiv C$ for some $C \in \mathbb{R}$.*

证明. Let $x_1, x_2 \in (a, b)$ be two different points. Then by Lagrange's mean value theorem, there is a point c between x_1, x_2 such that

$$f(x_1) - f(x_2) = f'(c)(x_1 - x_2) = 0.$$

Therefore, we see that $f(x_1) = f(x_2)$ and so f is a constant function. □

Similarly, Lagrange's mean value theorem indicates that the only functions f that have the nonnegative derivative $f' \ge 0$ is the increasing functions.

Corollary 4.1.8 ($f' \ge 0$ implies f is increasing, [1, §4.1 定理 3])**.** *Suppose that f is continuous on $[a, b]$, differentiable on (a, b) with*

$$f'(x) \ge 0 \quad (\text{or } f'(x) > 0)$$

for any $x \in (a, b)$. Then $f(x)$ is increasing (or strictly increasing) on $[a, b]$. Similarly, if

$$f'(x) \le 0 \quad (\text{or } f'(x) < 0)$$

for any $x \in (a, b)$, then $f(x)$ is decreasing (or strictly decreasing) on $[a, b]$.

证明. Let $[x_1, x_2] \subset (a, b)$. Then by Lagrange's mean value theorem, there is a point $c \in (x_1, x_2)$ such that

$$f(x_1) - f(x_2) = f'(c)(x_1 - x_2).$$

If $f'(c) \geq 0$, then $f(x_1) \leq f(x_2)$; if $f'(c) > 0$, then $f(x_1) < f(x_2)$. The consequence follows immediately. □

Example 4.1.9 ([1, §4.1 例 4])**.** Prove that $e^x > 1 + x$ for $x \neq 0$.

证明. Let $f(x) = e^x - 1 - x$. Then f is differentiable on $\mathbb{R}$, and

$$f'(x) = e^x - 1.$$

For $x \in (0, +\infty)$, $f'(x) > 0$ and so $f(x) > f(0) = 0$. For $x \in (-\infty, 0)$, $f'(x) < 0$ and so $f(x) > f(0) = 0$. □

The following *Cauchy's mean value theorem* is a useful generalization of Lagrange's theorem, and is also deduced by Rolle's theorem.

Theorem 4.1.10 (Cauchy's mean value theorem, [1, §4.2 定理 1])**.** *Let $f(x)$, $g(x)$ be continuous on $[a, b]$, differentiable on (a, b). Then there is a point $c \in (a, b)$ such that*

$$g'(c)(f(b) - f(a)) = f'(c)(g(b) - g(a)).$$

证明. Let

$$h(x) = g(x)(f(b) - f(a)) - f(x)(g(b) - g(a)).$$

Then $h(x)$ is continuous on $[a, b]$, differentiable on (a, b), and satisfies

$$h(a) = h(b).$$

Thus, by Rolle's theorem (Theorem 4.1.2), there is a point $c \in (a, b)$ such that

$$0 = h'(c) = g'(c)(f(b) - f(a)) - f'(c)(g(b) - g(a)).$$

This establishes the theorem. □

Remark 4.1.11. If we regard $t \mapsto (f(t), g(t))$ for $t \in [a, b]$ as a motion of a particle, then $(f'(t), g'(t))$ is its velocity at time t, and $(f(b) - f(a), g(b) - f(a))$ is its displacement. Then Cauchy's mean value theorem (Theorem 4.1.10) asserts that at some time $c \in (a, b)$, we have

$$\begin{cases} (-g'(c), f'(c)) \cdot (f'(c), f'(c)) = 0 \\ (-g'(c), f'(c)) \cdot (f(b) - f(a), g(b) - g(a)) = 0 \end{cases}.$$

Then we see that the vectors $(f'(c), f'(c))$ and $(f(b) - f(a), g(b) - g(a))$ are *parallel* (平行).

Example 4.1.12 ([1, P.259, Exercise 12])**.** Let f be differentiable on $(x_0, +\infty)$ so that $\lim_{x \to +\infty} f'(x) = 0$. Prove that

$$\lim_{x \to +\infty} \frac{f(x)}{x} = 0.$$

证明. Since $\lim_{x \to +\infty} f'(x) = 0$, for any $\epsilon > 0$, there is an $x_1 > x_0$ such that $|f'(x)| < \epsilon$ for $x > x_1$. Then

$$\left| \frac{f(x)}{x} \right| = \left| \frac{f(x) - f(x_1) + f(x_1)}{x} \right| = \left| \frac{f'(\xi)(x - x_1) + f(x_1)}{x} \right| \leq |f'(\xi)| + \left| \frac{f(x_1)}{x} \right|$$

for $\xi \in (x_1, x)$. Now let $x_2 > x_0$ be so large that $\left| \frac{f(x_1)}{x} \right| < \epsilon$ for $x > x_2$. Thus, we have

$$\left| \frac{f(x)}{x} \right| \leq 2\epsilon$$

for $x > \max\{x_1, x_2\}$. □

Theorem 4.1.13 (Darboux's theorem, 达布中值定理)**.** *Let $f : (A, B) \to \mathbb{R}$ be a differentiable function. Then the derivative f' has the intermediate value property: Let $[a, b] \subset (A, B)$. Then for every η between $f'(a)$ and $f'(b)$, there is $\xi \in [a, b]$ such that*

$$f'(\xi) = \eta.$$

证明. Assume that $f'(a) < f'(b)$, and $\eta \in (f'(a), f'(b))$. Consider the function $F : [a, b] \to \mathbb{R}$ defined by

$$F(t) = f(t) - \eta t.$$

Now by the extreme value theorem (Theorem 1.6.3), F attain a minimum M at some point $\xi \in [a, b]$.

- If $\xi = a$, then $\lim_{t \to a+0} \dfrac{F(t) - F(a)}{t - a} \geq 0$. It contradicts $F'(a) = f'(a) - \eta < 0$.
- If $\xi = b$, then $\lim_{t \to b-0} \dfrac{F(t) - F(b)}{t - b} \leq 0$. It contradicts $F'(b) = f'(b) - \eta > 0$.
- If $\xi \in (a, b)$, then by Fermat's lemma (Lemma 4.1.1), $0 = F'(\xi) = f'(\xi) - \eta$.

This establishes the theorem. $\square$

4.2 L'Hôpital's rule

A major consequence of Cauchy's mean value theorem is the following useful device for finding the limit of a ratio of functions, known as *L'Hôpital's rule* (洛必达法则).

Theorem 4.2.1 (L'Hôpital's rule, [1, §4.2 定理 2, 3, 4])**.** *Suppose that f and g are differentiable and $g'(x) \neq 0$ on an interval $(a, a + R)$, and that*

$$\lim_{x \to a+0} \frac{f'(x)}{g'(x)}$$

exists. Suppose further that one of the following holds:

(1) $f(x) \to 0$, $g(x) \to 0$ as $x \to a + 0$,

(2) $g(x) \to \infty$ as $x \to a + 0$,

Then we have

$$\lim_{x \to a+0} \frac{f(x)}{g(x)} = \lim_{x \to a+0} \frac{f'(x)}{g'(x)}.$$

A similar assertion holds for $x \to a - 0$, $x \to a$, $x \to +\infty$, $x \to -\infty$, $x \to \infty$.

证明. First, for the case that the variable $x \to +\infty$ goes to infinity, we make the change of variables $z = x^{-1}$. Then it turns to the case that the variable $z \to 0 + 0$ goes to a finite number.

Next, since $g'(x) \neq 0$ on the neighborhood, by Corollary 4.1.8, g is strictly monotone. Hence, shrinking the neighborhood if necessary, we can assume that $g(x) \neq 0$ for all $x \in (a, a+R)$.

Then we discuss the two different situations:

(i) In the case (1), for any $y \in (a, a+R)$, there is an $x = x(y) < y$ such that

$$|f(x)| \leq |g(y)|^3, \quad |g(x)| \leq |g(y)|^3.$$

It follows that

$$\frac{f(x)}{g(y)} = o(g(y)), \quad \frac{g(x)}{g(y)} = o(g(y)) \quad (y \to a+0). \tag{4.2.1}$$

(ii) In the case (2), for any $y \in (a, a+R)$, there is an $x = x(y) < y$ such that

$$|g(x)|^{\frac{1}{3}} \geq |f(y)|, \quad |g(x)|^{\frac{1}{3}} \geq |g(y)|.$$

It follows that

$$\frac{f(y)}{g(x)} = o(g(x)^{-\frac{1}{2}}), \quad \frac{g(y)}{g(x)} = o(g(x)^{-\frac{1}{2}}) \quad (y \to a+0). \tag{4.2.2}$$

In both cases, Cauchy's mean value theorem (Theorem 4.1.10) asserts that there is $c = c(y) \in (x(y), y)$ such that

$$f(y) - f(x) = \frac{f'(c)}{g'(c)}(g(y) - g(x)). \tag{4.2.3}$$

Note in particular that $x(y) \to a+0$, $c(y) \to a+0$, as $y \to a+0$.

(i) For the case (1), we divide both sides of (4.2.3) by $g(y)$ and use (4.2.1):

$$\frac{f(y)}{g(y)} = \frac{f'(c)}{g'(c)}\left(1 - \frac{g(x)}{g(y)}\right) + \frac{f(x)}{g(y)} = \frac{f'(c)}{g'(c)}\left(1 - o(g(y))\right) + o(g(y)).$$

The consequence follows from letting $y \to a+0$.

(ii) For the case (2), we divide both sides of (4.2.3) by $g(x)$ and use (4.2.2):

$$\frac{f(x)}{g(x)} = \frac{f'(c)}{g'(c)}\left(1 - \frac{g(y)}{g(x)}\right) + \frac{f(y)}{g(x)} = \frac{f'(c)}{g'(c)}\left(1 - o(g(x)^{-\frac{1}{2}})\right) + o(g(x)^{-\frac{1}{2}}).$$

The consequence follows from letting $y \to a+0$.

Hence, we establish the L'Hôpital's rule. $\square$

Remark 4.2.2. The limit that has the expression (1) or (2) (and $f(x) \to \infty$) of Theorem 4.2.1 is called an *indeterminate form* (未定式), sometimes written $\frac{0}{0}$, $\frac{\infty}{\infty}$, respectively.

Remark 4.2.3. One should double-check the conditions of L'Hôpital's rule before applying it. In doing so, one must **not forget** to verify the existence of the limit of the ratio of the derivatives, as well as the condition (1) or (2).

For instance, for $f(x) = x^2 \sin(x^{-1})$ and $g(x) = x$, we have

$$\lim_{x\to 0} \frac{f(x)}{g(x)} = 0, \quad \lim_{x\to 0} \frac{f'(x)}{g'(x)} = \lim_{x\to 0} \left(2x \sin \frac{1}{x} - \cos \frac{1}{x}\right) \text{ does not exist.}$$

We cannot apply L'Hôpital's rule here, because the limit of the ratio of the derivatives does not exist.

Example 4.2.4 (Important limits)**.** One may use L'Hôpital's rule and show that

$$\lim_{x\to\infty} \frac{\sin x}{x} = \lim_{x\to\infty} \frac{\cos x}{1} = 1.$$

However, this is a *circular argument* (循环论证). In fact, here we use $(\sin x)' = \cos x$, which is attributed to the important limit itself (see Example 2.1.8). Similarly, one should not apply L'Hôpital's rule to Example 1.3.18.

Example 4.2.5 ([1, §4.2 例 1])**.** Find $\lim_{x\to 0} \frac{\ln(1+x) - x}{x^2}$.

证明. By L'Hôpital's rule, we have

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

$\square$

Example 4.2.6 ([1, §4.2 例 7])**.** Find $\lim_{x\to+\infty} \frac{P(x)}{e^x}$.

证明. Let $P(x) = a_0 + a_1 x + \cdots + a_n x^n$. Then $P^{(n)}(x) = a_n n!$, $(e^x)^{(n)} = e^x$ and so

$$\lim_{x\to+\infty} \frac{P^{(n)}(x)}{(e^x)^{(n)}} = 0.$$

By L'Hôpital's rule, we have

$$\lim_{x\to+\infty}\frac{P^{(n-1)}(x)}{(e^x)^{(n-1)}}=\lim_{x\to+\infty}\frac{P^{(n)}(x)}{(e^x)^{(n)}}=0.$$

By induction, we conclude

$$\lim_{x\to+\infty}\frac{P(x)}{e^x}=\lim_{x\to+\infty}\frac{P^{(n)}(x)}{(e^x)^{(n)}}=0.$$

□

Example 4.2.7 ([1, §4.2 例 8])**.** Find $\lim\limits_{x\to0+0}x^x$.

证明. Consider

$$\lim_{x\to0+0}\ln x^x=\lim_{x\to0+0}x\ln x=\lim_{x\to0+0}\frac{\ln x}{x^{-1}}.$$

Then it is an indeterminate form (2) of Theorem 4.2.1. Thus, by L'Hôpital's rule, we have

$$\lim_{x\to0+0}\frac{\ln x}{x^{-1}}=\lim_{x\to+\infty}\frac{x^{-1}}{-x^{-2}}=0.$$

By the continuity of the exponential function, we have

$$\lim_{x\to0+0}e^{\ln x^x}=e^{\lim_{x\to0+0}\ln x^x}=1.$$

□

Example 4.2.8. Suppose that $f(x)$ is twice differentiable. Prove that

$$\lim_{h\to0}\frac{f(x+2h)+f(x)-2f(x+h)}{h^2}=f''(x).$$

证明. By L'Hôpital's rule, we have

$$\lim_{h\to0}\frac{f(x+2h)+f(x)-2f(x+h)}{h^2}=\lim_{h\to0}\frac{f'(x+2h)-f'(x+h)}{h}.$$

Then by the definition of derivatives, we have

$$\lim_{h\to0}\frac{f'(x+2h)-f'(x+h)}{h}=\lim_{h\to0}\left(\frac{f'(x+2h)-f'(x)}{h}-\frac{f'(x+h)-f'(x)}{h}\right)=f''(x).$$

□

4.3 Series

After developing a reasonable theory of limits, we will use that theory to discuss infinite series.

Definition 4.3.1 (Formal series)**.** A *formal series* (or *series* for short) (形式级数) is any expression of the form

$$\sum_{k=m}^{\infty} a_k$$

where m is an integer or $-\infty$, and a_k is a number for any integer $k \geq m$.

At present, this series is only defined **formally**; we have not set this sum equal to any real number. To rigorously define what the series actually sums to, we need another definition.

Definition 4.3.2 (Convergence of series)**.** Let $m \in \mathbb{Z}$ and $\sum_{k=m}^{\infty} a_k$ be a formal series. For any $n \geq m$, we define the n-th *partial sum* (部分和) S_n of this series to be

$$S_n \coloneqq \sum_{k=m}^{n} a_k.$$

Suppose that the limit exists:

$$\lim_{n \to \infty} S_n = S.$$

Then we say that the series $\sum_{k=m}^{\infty} a_k$ is *convergent*, and *converges to* S. We also write

$$S = \sum_{k=m}^{\infty} a_k$$

and say that S is the *sum* of the series $\sum_{k=m}^{\infty} a_k$.

If the partial sums $\{S_n\}$ diverge, then we say that the series $\sum_{k=m}^{\infty} a_k$ is *divergent*, and we do not assign any number value to that series.

For $m = -\infty$, we can consider two series $\sum_{k=0}^{\infty} a_k$ and $\sum_{k=1}^{\infty} a_{-k}$ simultaneously, and say that $\sum_{k=-\infty}^{\infty} a_k$ is *convergent* if both of them are convergent, and so on.

Remark 4.3.3. A convergent series $\sum_{k=m}^{\infty} a_k$ can be considered as a Lebesgue integral on $S = \{-m, -m+1, \ldots\}$. One also writes

$$\sum_{k=m}^{\infty} a_k = \int_S a_k dk.$$

4.3.1 Decimal numeral system

Definition 4.3.4 (Decimals). A *decimal* (十进制小数) is any sequence of digits, and a decimal point, arranged as

$$\pm a_m \ldots a_0.a_{-1}a_{-2}\ldots$$

for $a_i \in \{0, 1, \ldots, 9\}$, $i \in \{m, \ldots, 0, -1, \ldots\}$. The decimal is equal to the series

$$\pm\, a_m \ldots a_0.a_{-1}a_{-2}\ldots = \pm 1 \times \sum_{i=-m}^{\infty} a_{-i} \times 10^{-i}. \tag{4.3.1}$$

A *decimal fraction* is a decimal $\pm a_m \ldots a_0.a_{-1}a_{-2}\ldots$ so that there is $n \geq -m$ such that $a_{-i} = 0$ for any $i \geq n$.

Remark 4.3.5. For an integer $p \geq 2$, one may similarly define the *p-adic number* (p-进制小数) by replacing 10 by p.

Proposition 4.3.6. *The series of any decimal (4.3.1) is convergent.*

证明. For integers $n \geq -m$, let

$$S_n = \sum_{i=-m}^{n} a_{-i} \times 10^{-i}.$$

Then $\{S_n\}$ is monotonically increasing as $n \to \infty$.

Moreover, let $M = 10^{m+1}$. For any $n \geq -m$, suppose that $M - S_{n-1} \geq 10^{-n+1}$. Then

$$\begin{aligned} M - S_{n+1} &= M - S_n + S_n - S_{n+1} \\ &\geq 10^{-n+1} - a_{-n} \times 10^{-n} \\ &= (10 - a_{-n}) \times 10^{-n} \geq 10^{-n}. \end{aligned}$$

By induction, we see that M is larger than S_n for any $n \geq -m$; in other words, M is an upper bound of S_n. Then by Theorem 1.1.4, we conclude that the series (4.3.1) is convergent. □

Example 4.3.7.

- $\frac{1}{3} = 0.3333\ldots$
- $\pi = 3.1415926\ldots$
- By the definition, the complex number i does not have a decimal representation.

Theorem 4.3.8 (Existence of decimal representations). *Every real number $x \in \mathbb{R}$ has at least one decimal representation:*

$$x = \pm a_m \ldots a_0.a_{-1}a_{-2}\ldots$$

for some $a_i \in \{0, 1, \ldots, 9\}$, $i \in \{m, \ldots, 0, -1, \ldots\}$.

证明. We only show the decimal representation for positive number $x \in \mathbb{R}^+$. Then there exists an integer $m \in \mathbb{N}$ such that

$$x < 10^{m+1} = 10 \cdot 10^m.$$

Then by the *Euclidean division*, there exist unique integers $a_m \in \{0, 1, \ldots, 9\}$ and x_{m-1} such that

$$x = a_m 10^m + x_{m-1} \quad \text{and} \quad 0 \leq x_{m-1} < 10^m.$$

Replacing the x by x_{m-1}, we can find unique integers $a_m \in \{0, 1, \ldots, 9\}$ and x_{m-2} such that

$$x_{m-1} = a_{m-1} 10^{m-1} + x_{m-2} \quad \text{and} \quad 0 \leq x_{m-2} < 10^{m-1}.$$

Repeating the process, we obtain

$$x = a_m 10^m + \cdots + a_0 10^0 + a_{-1} 10^{-1} + \cdots .$$

□

In other words, any real number can be approximated by a sequence of decimal fractions:

Corollary 4.3.9. *Let x be a real number. Then there is a sequence $\{y_n\}$ of decimal fractions so that*

$$0 \leq x - y_n \leq 10^{-n}$$

for $n \in \mathbb{N}$.

证明. Pick $y_n = x_{m-(n+m)}$ as in the proof of Theorem 4.3.8. □

Example 4.3.10 (Failure of uniqueness of decimal representations, [1, §1.3 例 8])**.** One may check that the number 1 has two different decimal representations:

$$1.000\ldots \quad \text{and} \quad 0.999\ldots.$$

4.3.2 Power series

We now discuss an important class of series of functions, that of *power series.*

Definition 4.3.11 (Formal power series)**.** Let x_0 be a real number. The *formal power series* (形式幂级数) is a series of the form

$$\sum_{k=0}^{\infty} c_k (x - x_0)^k$$

for $c_k \in \mathbb{R}$, $x \in \mathbb{R}$.

After introducing the notion of a formal power series, we then focus on when the series converges to a meaningful function, and what one can say about the function obtained in this manner.

As shown in Corollary 4.3.9, if a power series converges to a function, its partial sums P_n should **approach** this function. Recall that when a function $f : \mathbb{R} \to \mathbb{R}$ is differentiable, it can be approximated by a linear function

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + o(x - x_0) \quad (x \to x_0).$$

These enlighten us to consider the approximations of a function f in the neighborhood of x_0 by a polynomial

$$P_n(x) = \sum_{k=0}^{n} c_k(x - x_0)^k.$$

Suppose now that $f : \mathbb{R} \to \mathbb{R}$ is n times differentiable. We are going to find the coefficients c_k so that $P_n(x)$ is the **best polynomial approximation** of f in a neighborhood of x_0 with degree $\leq n$.

Firstly, we write

$$f(x) = \sum_{k=0}^{n} c_k(x - x_0)^k + r(x) \tag{4.3.2}$$

where $r(x) = f(x) - P_n(x)$ is the remainder term. Since P_n approximates f in a neighborhood of x_0, we further assume $r(x) \to 0$ as $x \to x_0$. By (4.3.2), it is equivalent to say that

$$f(x_0) = c_0.$$

Secondly, we differentiate both sides of (4.3.2) and obtain

$$f'(x) = \sum_{k=0}^{n-1} c_{k+1}(k+1)(x - x_0)^k + r'(x). \tag{4.3.3}$$

Now note that

$$r'(x_0) = \lim_{x \to x_0} \frac{r(x)}{x - x_0}.$$

Because we want $r(x)$ to be as small as possible, we naturally assume that $r(x)$ is an infinitesimal of higher order than $(x - x_0)$ as $x \to x_0$, i.e. $r'(x_0) = 0$. By (4.3.3), it is equivalent to say that

$$f'(x_0) = c_1 \cdot 1.$$

Next, we differentiate both sides of (4.3.3) and obtain

$$f''(x) = \sum_{k=0}^{n-2} c_{k+2}(k+1)(k+2)(x - x_0)^k + r''(x). \tag{4.3.4}$$

Now note by the L'Hôpital's rule (Theorem 4.2.1) that

$$r''(x_0) = \lim_{x \to x_0} \frac{r'(x)}{x - x_0} = \lim_{x \to x_0} \frac{r(x)}{\frac{1}{2}(x - x_0)^2}.$$

Again, because we want $r(x)$ to be as small as possible, we naturally assume that $r(x)$ is an infinitesimal of higher order than $(x-x_0)^2$ as $x \to x_0$, i.e. $r''(x_0) = 0$. By (4.3.4), it is equivalent to say that

$$f''(x_0) = c_2 \cdot 2!.$$

Repeating the above argument, we obtain the following:

Theorem 4.3.12 (Taylor's theorem, [1, §4.3 定理 1]). *Let $n \in \mathbb{N}$, $x_0 \in (A, B)$, and $f : (A, B) \to \mathbb{R}$ be n times differentiable at x_0. Then in a neighborhood of x_0, $f(x)$ can be approximated by a polynomial $P_n(x)$ of order n so that*

$$\lim_{x \to x_0} \frac{f(x) - P_n(x)}{(x-x_0)^n} = 0. \tag{4.3.5}$$

The polynomial is given by

$$P_n(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k. \tag{4.3.6}$$

Remark 4.3.13.

- A polynomial $P_n(x)$ of degree $\leq n$ satisfies (4.3.5) is called a *Taylor's polynomial* (泰勒多项式) of order n of $f(x)$ at x_0. The Taylor's theorem asserts that Taylor's polynomials of order n exist if $f(x)$ is n times differentiable.

- The limit (4.3.5) indicates that

$$h_n(x) = \frac{f(x) - P_n(x)}{(x-x_0)^n}$$

  is well-defined on $\mathbb{R}$ and assigns 0 to x_0. The remainder term

$$h_n(x)(x-x_0)^n \tag{4.3.7}$$

  is called the *Peano form of the remainder* (佩亚诺余项).

- Also, the equation

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k + h_n(x)(x-x_0)^n \tag{4.3.8}$$

  is called the *Taylor's formula* (泰勒公式) of order n of $f(x)$ at x_0. In particular, when $x_0 = 0$, the Taylor's formula (4.3.8) is also called the *Maclaurin's formula* (麦克劳林公式).

In addition, note that the Taylor's polynomial of a function, if exists, is uniquely determined:

Theorem 4.3.14 (Uniqueness of Taylor's polynomials, [1, §4.3 定理 2])**.** *Let $n \in \mathbb{N}$, $x_0 \in (A, B)$, and $f : (A, B) \to \mathbb{R}$ be n times differentiable at x_0. Suppose that $P_n(x)$ and $Q_n(x)$ are two Taylor's polynomials of order n of $f(x)$ at x_0. Then*

$$P_n = Q_n.$$

证明. Suppose that

$$P_n(x) - Q_n(x) = \sum_{k=0}^{n} a_k (x - x_0)^k.$$

By (4.3.5), we have

$$\lim_{x \to x_0} \frac{P_n(x) - Q_n(x)}{(x - x_0)^n} = 0.$$

It forces $a_0 = \cdots = a_n = 0$. □

Example 4.3.15 ([1, §4.3 例 1])**.** Find the Taylor's formula of $f(x) = e^x$ at $x = 0$.

证明. Since $f^{(n)}(x) = e^x$ for all $n \in \mathbb{N}$, by Theorem 4.3.12, we have

$$\begin{aligned} e^x =& e^0 + \frac{e^0}{1!}(x-0) + \frac{e^0}{2!}(x-0)^2 + \cdots + \frac{e^0}{n!}(x-0)^n + o((x-0)^n) \\ =& 1 + \frac{1}{1!}x + \frac{1}{2!}x^2 + \cdots + \frac{1}{n!}x^n + o(x^n), \quad (x \to 0). \end{aligned}$$

□

Example 4.3.16 ([1, §4.3 例 2])**.** Find the Taylor's formula of $f(x) = \sin x$ at $x = 0$.

证明. Recall from Exercise 2.6.2 that

$$f^{(n)}(x) = (\sin x)^{(n)} = \sin\left(x + \frac{n\pi}{2}\right).$$

for all $n \in \mathbb{N}$. Then by Theorem 4.3.12, we have

$$\begin{aligned} \sin x =& \sin 0 + \frac{\sin\frac{\pi}{2}}{1!}(x-0) + \cdots + \frac{\sin\frac{n\pi}{2}}{n!}(x-0)^n + o((x-0)^n) \\ =& \frac{1}{1!}x - \frac{1}{3!}x^3 + \cdots + \frac{(-1)^k}{(2k+1)!}x^{2k+1} + o(x^{2k+1}), \quad (x \to 0) \end{aligned}$$

for $2k + 1 \le n \le 2k + 2$. □

Example 4.3.17 ([1, §4.3 例 3])**.** Find the Taylor's formula of $f(x) = \cos x$ at $x = 0$.

证明. Note that

$$\cos x = 1 - \int_0^x \sin t dt.$$

Let $k \in \mathbb{N}$, n be an integer so that $2k+1 \le n-1 \le 2k+2$, and

$$P_k(x) = \frac{1}{1!}x - \frac{1}{3!}x^3 + \cdots + \frac{(-1)^k}{(2k+1)!}x^{2k+1}$$

be the Taylor's polynomial of order $(n-1)$ of $\sin x$ at 0. Then heuristically, the Taylor's polynomial of order n of $\cos x$ at 0 is given by

$$Q_k(x) = 1 - \int_0^x P_k(t)dt = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 + \cdots + \frac{(-1)^k}{(2k+2)!}x^{2k+2}.$$

In fact, by L'Hôpital's rule (Theorem 4.2.1), one calculates

$$\lim_{x\to x_0} \frac{\cos x - Q_k(x)}{(x-x_0)^n} = \lim_{x\to x_0} \frac{-\sin x + P_k(x)}{(x-x_0)^{n-1}} = 0.$$

Thus, we prove that $Q_k(x)$ is the Taylor's polynomial of order n of $\cos x$ at 0 by Theorem 4.3.14. $\square$

Example 4.3.18 ([1, §4.3 例 6])**.** Find the Taylor's formula of $f(x) = e^{-x^2}$ at $x = 0$.

证明. By Example 4.3.15, we have

$$e^x = 1 + \frac{1}{1!}x + \frac{1}{2!}x^2 + \cdots + \frac{1}{n!}x^n + o(x^n), \quad (x \to 0)$$

for all $n \in \mathbb{N}$. Then by Theorem 4.3.14, we easily obtain the Taylor's formula of e^{-x^2}:

$$\begin{aligned} e^{-x^2} =& 1 + \frac{1}{1!}(-x^2) + \frac{1}{2!}(-x^2)^2 + \cdots + \frac{1}{n!}(-x^2)^n + o(x^{2n}) \\ =& 1 - \frac{1}{1!}x^2 + \frac{1}{2!}x^4 + \cdots + \frac{(-1)^n}{n!}x^{2n} + o(x^{2n}), \quad (x \to 0). \end{aligned}$$

$\square$

Example 4.3.19 ([1, §4.3 例 5])**.** Find the Taylor's formula of $f(x) = \ln(1+x)$ at $x = 0$.

证明. Note that

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + \cdots + (-1)^n x^n + o(x^n). \tag{4.3.9}$$

Then by Theorem 4.3.14, we easily obtain the Taylor's formula of $\ln(1+x)$ by integrating both sides of (4.3.9):

$$\ln(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + \cdots + \frac{(-1)^{n-1}}{n}x^n + o(x^n).$$

□

The Peano form of the remainder (4.3.7) is not easy to observe the behavior x in a deleted neighborhood of x_0. We therefore will conduct further study on the estimation of the remainder $r_n(x)$ of the Taylor's formula of order n of $f(x)$ at x_0:

Theorem 4.3.20. *Let $n \in \mathbb{N}$, $x_0 \in (A, B)$, and $f : (A, B) \to \mathbb{R}$ be $n+1$ times differentiable. Then we have the Taylor's expansion:*

$$f(x) = P_n(x) + r_n(x), \quad x \in (A, B)$$

where $P_n(x)$ is the Taylor's polynomial of order n at x_0. Moreover, the remainder term r_n has the expressions:

(1) Integral form of the remainder (积分型余项):

$$r_n(x) = \int_{x_0}^{x} \frac{f^{(n+1)}(t)}{n!}(x-t)^n dt. \tag{4.3.10}$$

(2) Lagrange form of the remainder (拉格朗日余项, cf. [1, §4.4 定理]):

$$r_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x - x_0)^{n+1}. \tag{4.3.11}$$

for some $\xi = \xi(x_0, x)$ between x_0 and x.

(3) Cauchy form of the remainder (柯西余项, cf. [2, §10.6 命题]):

$$r_n(x) = \frac{f^{(n+1)}(\xi)}{n!}(x - \xi)^n (x - x_0) \tag{4.3.12}$$

for some $\xi = \xi(x_0, x)$ between x_0 and x.

证明. Let $n \in \mathbb{N}$, $x_0 \in (A, B)$, and $f : (A, B) \to \mathbb{R}$ be $n+1$ times differentiable. For $x, t \in (A, B)$, let $P_{n,t}(x)$ be the Taylor's polynomial of order n of $f(x)$ at t, and

$$R(t) = f(x) - P_{n,t}(x).$$

Then we have $R(x) = 0$, $R(x_0) = r(x)$ and

$$\frac{d}{dt}R(t) = -\frac{f^{(n+1)}(t)}{n!}(x-t)^n.$$

Then by the second fundamental theorem of calculus (Theorem 2.10.1), we obtain

$$r_n(x) = R(x_0) - R(x) = \int_{x_0}^{x} \frac{f^{(n+1)}(t)}{n!}(x-t)^n dt.$$

This is the integral form of the remainder (4.3.10).

Moreover, if $g : (A, B) \to \mathbb{R}$ is differentiable so that $g'(t) \neq 0$ for $t \in (A, B)$, then by Cauchy's mean value theorem (Theorem 4.1.10), there is a point $\xi = \xi(x_0, x)$ between x_0 and x such that

$$\frac{0 - r_n(x)}{g(x) - g(x_0)} = \frac{R(x) - R(x_0)}{g(x) - g(x_0)} = \frac{R'(\xi)}{g'(\xi)} = -\frac{f^{(n+1)}(\xi)}{n! g'(\xi)}(x-\xi)^n. \tag{4.3.13}$$

In particular, for appropriate choices of $g(t)$, we obtain:

- Letting $g(t) = (x-t)^{n+1}$, (4.3.13) becomes

  $$r_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x - x_0)^{n+1}.$$

  This is the Lagrange form of the remainder (4.3.11).

- Letting $g(t) = x - t$, (4.3.13) becomes

  $$r_n(x) = \frac{f^{(n+1)}(\xi)}{n!}(x-\xi)^n(x - x_0).$$

  This is the Cauchy form of the remainder (4.3.12).

This establishes the theorem. □

With the help of the estimation of the remainder term, we are able to discuss which functions are lucky enough to be representable as power series.

Definition 4.3.21 (Taylor series)**.** Suppose a power series $\sum_{k=0}^{\infty} c_k(x-x_0)^k$ converges to an infinitely differentiable function $f(x)$ on a neighborhood of x_0. Suppose further that the partial sums P_n of the series $\sum_{k=0}^{\infty} c_k(x-x_0)^k$ are Taylor polynomials. Then we say that the power series $\sum_{k=0}^{\infty} c_k(x-x_0)^k$ is the *Taylor series* of $f(x)$ at x_0.

Remark 4.3.22. The function f that has a power series expansion is called (real) *analytic* (解析函数, Definition 7.8.5). In Appendix 7.8, we shall discuss the properties of real analytic functions, and show that the second assumption in Definition 4.3.21, namely the partial sums are required to be Taylor polynomials, is redundant.

Example 4.3.23 (Taylor series of basic elementary functions, [2, §10.6.3])**.**

- Exponential functions: Let $M > 0$, $f(x) = e^x$. By Example 4.3.15 and the Lagrange form of the remainder (4.3.11), we have

$$|r_n(x)| = \left|\frac{e^{\xi}x^{n+1}}{(n+1)!}\right| \le \left|\frac{e^M M^{n+1}}{(n+1)!}\right| \to 0, \quad n \to \infty$$

  for $\xi, x \in [-M, M]$. Thus, we have shown that the Taylor polynomials $\{P_n(x)\}$ are a Cauchy sequence, and converge to e^x. Then we have

$$e^x = 1 + \frac{1}{1!}x + \frac{1}{2!}x^2 + \cdots + \frac{1}{n!}x^n + \cdots$$

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

- Trigonometric functions: Let $M > 0$, $f(x) = \sin x$. By Examples 2.6.2, 4.3.16 and the Lagrange form of the remainder (4.3.11), we have

$$|r_n(x)| = \left|\frac{\sin(\xi + \frac{(n+1)\pi}{2})x^{n+1}}{(n+1)!}\right| \le \left|\frac{M^{n+1}}{(n+1)!}\right| \to 0, \quad n \to \infty$$

  for $\xi, x \in [-M, M]$. It follows that

$$\begin{aligned}\sin x =& x - \frac{1}{3!}x^3 + \cdots + \frac{(-1)^k}{(2k+1)!}x^{2k+1} + \cdots \\ \cos x =& 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 + \cdots + \frac{(-1)^k}{(2k+2)!}x^{2k+2} + \cdots,\end{aligned}$$

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

- Real power functions: Let $M > 0$, $\alpha \in \mathbb{R} \setminus \{0\}$, $f(x) = (1+x)^\alpha$. Then

$$f^{(n)}(x) = \alpha(\alpha-1)\cdots(\alpha-n+1)(1+x)^{\alpha-n}$$

for $x \in [-1+M^{-1}, 1-M^{-1}]$, $n \in \mathbb{N}$. By the Cauchy form of the remainder (4.3.12), we have

$$\begin{aligned} |r_n(x)| &= \left| \frac{\alpha(\alpha-1)\cdots(\alpha-n)(1+\xi)^{\alpha-n-1}}{n!}(x-\xi)^n x \right| \\ &= |\alpha(1+\xi)^{\alpha-1}x| \left| \prod_{k=1}^{n} \left(\frac{\alpha}{k} - 1\right) \left(\frac{x-\xi}{1+\xi}\right)^n \right| \\ &\le C_{\alpha,M} \left| \prod_{k=1}^{n} \left(\frac{\alpha}{k} - 1\right) \left(\frac{x-\xi}{(-1)-\xi}\right)^n \right| \to 0, \quad n \to \infty \end{aligned}$$

for $C_{\alpha,M} > 0$, $x \in [-1+M^{-1}, 1-M^{-1}]$ and ξ between 0 and x. It follows that

$$(1+x)^\alpha = 1 + \alpha x + \frac{\alpha(\alpha-1)}{2!}x^2 + \cdots + \frac{\alpha(\alpha-1)\cdots(\alpha-n+1)}{n!}x^n + \cdots$$

for $x \in (-1, 1)$. In fact, by the theory of uniform convergence (Definition 4.3.27), we may further deduce that the above equation also holds for $x = \pm 1$ for certain α (cf. [2, §10.6.3(5)]).

- Logarithm functions: Let $M > 0$, $f(x) = \ln(1+x)$. Then

$$f^{(n)}(x) = [\ln(1+x)]^{(n)} = (-1)^n(n-1)!(1+x)^{1-n}$$

for $x \in [-1+M^{-1}, 1-M^{-1}]$, $n \in \mathbb{N}$. By the Cauchy form of the remainder (4.3.12), we have

$$|r_n(x)| = \left| \frac{(-1)^{n+1}n!(1+\xi)^{-n}(x-\xi)^n x}{n!} \right| \le \left| \frac{x-\xi}{(-1)-\xi} \right|^n \to 0, \quad n \to \infty$$

for $x \in [-1+M^{-1}, 1-M^{-1}]$ and ξ between 0 and x. It follows that

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots + (-1)^{n+1}\frac{x^n}{n} + \cdots$$

for $x \in (-1, 1)$. In fact, by the theory of uniform convergence (Definition 4.3.27), we can further deduce that the above equation also holds for $x = 1$ (cf. [2, §10.5 例 7]).

Example 4.3.24 (An infinite differentiable function that fails to have a Taylor series expansion)**.** Consider the function

$$f(x) = \begin{cases} e^{-\frac{1}{x^2}} & , \text{ if } x \neq 0 \\ 0 & , \text{ if } x = 0 \end{cases} .$$

Note first that $f'(0) = 0$. In fact,

$$\lim_{\Delta x \to 0} \frac{f(\Delta x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{e^{-\frac{1}{(\Delta x)^2}}}{\Delta x} = \lim_{\Delta x \to 0} \frac{\frac{1}{\Delta x}}{e^{\frac{1}{(\Delta x)^2}}} = 0.$$

It follows that

$$f'(x) = \begin{cases} 2x^{-3}e^{-\frac{1}{x^2}} & , \text{ if } x \neq 0 \\ 0 & , \text{ if } x = 0 \end{cases} .$$

Similarly, one may show that f is infinitely differentiable and $f^{(n)}(0) = 0$. Then the Taylor polynomials $P_n(x)$ of order n of $f(x)$ at 0 are all equal to 0. Thus, $f(x)$ fails to have a Taylor series expansion at 0.

In particular, Example 4.3.24 shows that infinite differentiable functions are not always real analytic (Definition 7.8.5).

Example 4.3.25. Suppose that $f(x)$ is continuous on $[a, b]$ and three times differentiable on (a, b). Prove that there exists a $\xi \in (a, b)$ such that

$$f(b) = f(a) + f'\left(\frac{a+b}{2}\right)(b-a) + \frac{1}{24}f'''(\xi)(b-a)^3. \tag{4.3.14}$$

证明. Consider the Taylor's formula of $f(x)$ at $p = \frac{a+b}{2}$:

$$f(x) = f(p) + f'(p)(x-p) + \frac{f''(p)}{2}(x-p)^2 + \frac{f'''(\xi)}{6}(x-p)^3$$

where ξ is between x and p. Thus, letting $x = a$ and $x = b$, there are $\xi_1 \in (a, p)$, $\xi_2 \in (p, b)$ such that

$$f(a) = f(p) + f'(p)(a-p) + \frac{f''(p)}{2}(a-p)^2 + \frac{f'''(\xi_1)}{6}(a-p)^3, \tag{4.3.15}$$

$$f(b) = f(p) + f'(p)(b-p) + \frac{f''(p)}{2}(b-p)^2 + \frac{f'''(\xi_2)}{6}(b-p)^3. \tag{4.3.16}$$

Let (4.3.16) minus (4.3.15). Then since $(b-p)^3 = -(a-p)^3 = \frac{(b-a)^3}{8}$, one deduces

$$\begin{aligned} f(b) =& f(a) + f'(p)(b-a) + \frac{f'''(\xi_2)}{6}(b-p)^3 - \frac{f'''(\xi_1)}{6}(a-p)^3 \\ =& f(a) + f'(p)(b-a) + \frac{f'''(\xi_1) + f'''(\xi_2)}{48}(b-a)^3. \end{aligned}$$

On the other hand, by Darboux's theorem (Theorem 4.1.13), we conclude that there is a $\xi \in (\xi_1, \xi_2)$ such that

$$f'''(\xi) = \frac{f'''(\xi_1) + f'''(\xi_2)}{2}.$$

Thus, the remainder term becomes

$$\frac{f'''(\xi)}{24}(b-a)^3.$$

This establishes (4.3.14). □

4.3.3 Uniform convergence

Definition 4.3.26 (Uniform norm)**.** Let $S \subset \mathbb{R}$. Let $f : S \to \mathbb{R}$ be a function on S. The *uniform norm* (or *infinity norm* (∞-范数), *maximum norm*, *supremum norm*) of f is defined by

$$\|f\|_\infty \coloneqq \sup_{x \in S} |f(x)|.$$

The uniform norm $\|\cdot\|_\infty$ is a *norm* (范数). In other words, it behaves similarly to the absolute value $|\cdot|$ on $\mathbb{R}$ or $\mathbb{R}^n$. In particular, for two functions $f, g : S \to \mathbb{R}$, the number $\|f - g\|_\infty$ measures the distance between f and g.

Definition 4.3.27 (Uniform convergence)**.** Let $S \subset \mathbb{R}$ be a subset. For $n \in \mathbb{N}$, let $f_n : S \to \mathbb{R}$ be a function on S. Let $f : S \to \mathbb{R}$ be another function on S. We say that the sequence $\{f_n\}_{n \in \mathbb{N}}$ of functions *converges uniformly* (一致收敛) to f on S if for any $\epsilon > 0$, there exists $N > 0$ such that

$$\|f_n - f\|_\infty < \epsilon$$

for any $n \geq N$. We also say the function f is the *uniform limit* of the functions f_n.

Remark 4.3.28. For a series $\sum_{k=0}^{\infty} a_k(x)$, we say that the series *uniformly converges* to a function $f(x)$ if its partial sums $S_n(x) = \sum_{k=0}^{n} a_k(x)$ uniformly converges to the function $f(x)$ on the domain.

In terms of the uniform convergence, the polynomials can approach any continuous function. See Appendix 7.9 for more details.

Theorem 4.3.29 (Weierstrass approximation theorem 魏尔斯特拉斯逼近定理)**.** *Let $f : [A, B] \to \mathbb{R}$ be a continuous function. Then for any $\epsilon > 0$, there is a polynomial $P : [A, B] \to \mathbb{R}$ such that*

$$\|f - P\|_\infty < \epsilon.$$

4.3.4 Fourier series

In this section, we briefly introduce the power series on a circle, namely the *Fourier series*. One may see Appendix 7.10 for more details.

We are mainly interested in the *complex-valued functions*, i.e. functions with codomain $\mathbb{C}$. Note that a complex-valued function $F : \mathbb{R} \to \mathbb{C}$ can always be decomposed as two real-valued functions $a, b : \mathbb{R} \to \mathbb{R}$:

$$F(x) = a(x) + ib(x).$$

Next, recall from Definition 3.4.13 that a periodic function $F : \mathbb{R} \to \mathbb{C}$ with period T means

$$F(x + T) = F(x)$$

for any $x \in \mathbb{R}$. It can be identified as a function on a circle

$$S^1 = \{e^{2\pi ix} \in \mathbb{C} : x \in [0, 1]\} = \{\cos(2\pi x) + i \sin(2\pi x) \in \mathbb{C} : x \in [0, 1]\}$$

where the second equality follows from the *Euler's formula* (欧拉公式):

$$e^{i\theta} = \cos\theta + i \sin\theta, \quad \theta \in \mathbb{R}.$$

More precisely, given a periodic function $F : \mathbb{R} \to \mathbb{C}$ with period T, we define $f_F : S^1 \to \mathbb{C}$ by

$$f_F : e^{2\pi i x} \mapsto F(T \cdot x).$$

Then we see that f_F inherits all the information of F. On the other hand, any $f : S^1 \to \mathbb{C}$ induces a periodic function $F_f : \mathbb{R} \to \mathbb{C}$ by

$$F_f : x \mapsto f(e^{2\pi i x}).$$

Therefore, we may consider the continuity, differentiability, and integrability of $f : S^1 \to \mathbb{C}$ by its induced periodic function F_f. For instance, the integral

$$\int_{S^1} f(z)dz := \int_0^1 F_f(x)dx = \int_0^1 f(e^{2\pi i x})dx.$$

Here dz denotes the infinitesimal arc length $\dfrac{ds}{2\pi}$ of S^1 so that the circumference is equal to 1; Or more deeply, dz is the normalized *Haar measure* (哈尔测度) on S^1.

The space of complex valued continuous functions on S^1 is denoted by $C(S^1, \mathbb{C})$. Clearly, we can add, subtract, multiply, and take limits of continuous functions on S^1. The simplest functions in $C(S^1, \mathbb{C})$ are z^n ($z \in S^1$, $n \in \mathbb{Z}$). We are going to use these functions to approach general periodic functions:

Definition 4.3.30 (Fourier series)**.** A function $f \in C(S^1, \mathbb{C})$ is said to be a *trigonometric polynomial* (三角多项式) if we can write

$$f(z) = \sum_{n=-N}^{N} c_n z^n$$

for some integer $N \geq 0$ and some complex numbers $c_n \in \mathbb{C}$. A *Fourier series* (傅里叶级数) is a series so that the partial sums are trigonometric polynomials, i.e. a series of the form

$$\sum_{n=-\infty}^{\infty} c_n z^n$$

for $c_n \in \mathbb{C}$ and $z \in S^1$.

Remark 4.3.31. Recall the *Euler's formula* (欧拉公式):

$$e^{i\theta} = \cos\theta + i\sin\theta, \quad \theta \in \mathbb{R}.$$

Thus, for $z = e^{2\pi inx} \in S^1$, we usually rewrite the Fourier series as:

$$\begin{aligned}
\sum_{n\in\mathbb{Z}} c_n z^n &= \sum_{n\in\mathbb{Z}} c_n e^{2\pi inx} \\
&= \sum_{n\in\mathbb{Z}} c_n \cos 2\pi nx + i \sum_{n\in\mathbb{Z}} c_n \sin 2\pi nx \\
&= c_0 + \sum_{n=1}^{\infty} (c_n + c_{-n}) \cos 2\pi nx + i \sum_{n=1}^{\infty} (c_n - c_{-n}) \sin 2\pi nx.
\end{aligned}$$

Next, we discuss how to find the coefficients of the Fourier series of a function.

Definition 4.3.32 (Fourier transform)**.** For any $f \in C(S^1, \mathbb{C})$, we define the *Fourier transform* (傅里叶变换) $\hat{f} : \mathbb{Z} \to \mathbb{C}$ by the formula

$$\hat{f}(n) := \int_{S^1} f(z) z^{-n} dz = \int_0^1 f(e^{2\pi ix}) e^{-2\pi inx} dx. \tag{4.3.17}$$

Example 4.3.33. For $m, n \in \mathbb{Z}$, we have

$$\int_{S^1} z^m z^{-n} dz = \begin{cases} 1 & , \text{ if } m = n \\ 0 & , \text{ if } m \neq n \end{cases}.$$

Thus, if $f(z) = \sum_{m=-N}^{N} c_m z^m$ is a trigonometric polynomial on the circle S^1, then

$$\hat{f}(n) = \int_{S^1} f(z) z^{-n} dz = \int_{S^1} \left(\sum_{m=-N}^{N} c_m z^m \right) z^{-n} dz = c_n.$$

In other words, the Fourier transform finds the coefficients of this polynomial.

Heuristically, Fourier transform should be able to capture all the information about the Fourier series of a function (if it has a Fourier series expansion).

Theorem 4.3.34 (Fourier inversion theorem, 傅里叶反演定理)**.** *Let $f \in C(S^1, \mathbb{C})$, and suppose that the series $\sum_{n=-\infty}^{\infty} |\hat{f}(n)|$ is convergent. Then f has the Fourier expansion*

$$f(z) = \sum_{n=-\infty}^{\infty} \hat{f}(n) z^n \tag{4.3.18}$$

for $z \in S^1$*. Moreover, the Fourier series* $\sum_{n=-\infty}^{\infty} \hat{f}(n)z^n$ *converges uniformly to* f *on* S^1*, i.e.*

$$\lim_{N\to\infty} \left\| f - \sum_{n=-N}^{N} \hat{f}(n)z^n \right\|_\infty = 0.$$

In addition, the *Plancherel theorem* (普朗歇尔定理), also known as *Parseval's identity* (帕塞瓦尔恒等式), states that the Fourier coefficients satisfy the *Pythagoras theorem* (勾股定理):

Theorem 4.3.35 (Plancherel theorem, [2, §12.3 推论])**.** *Let* $f \in C(S^1, \mathbb{C})$*. Then the series*

$$\sum_{n=-\infty}^{\infty} |\hat{f}(n)|^2$$

is convergent. Moreover, we have

$$\int_{S^1} |f(z)|^2 dz = \sum_{n=-\infty}^{\infty} |\hat{f}(n)|^2. \tag{4.3.19}$$

Remark 4.3.36. As mentioned in Remark 4.3.3, the Fourier series can be considered as a Lebesgue integral. In particular, (4.3.18) can be rewritten as

$$f(e^{2\pi ix}) = f(z) = \sum_{n=-\infty}^{\infty} \hat{f}(n)z^n = \int_{\mathbb{Z}} \hat{f}(n)z^n dn = \int_{\mathbb{Z}} \hat{f}(n)e^{2\pi ixn} dn. \tag{4.3.20}$$

for $z = e^{2\pi ix} \in S^1$. Comparing the integral (4.3.20) with (4.3.17), one may say:

$f : S^1 \to \mathbb{C}$ is the (inverse) Fourier transform of the function $\hat{f} : \mathbb{Z} \to \mathbb{C}$.

This explain the terminology "inversion theorem". In fact, we observe the *Pontryagin duality* (庞特里亚金对偶性) between the *time domain* S^1 (时域) and *frequency domain* $\mathbb{Z}$ (频域):

$$f : S^1 \to \mathbb{C} \quad \underset{\text{(Inverse) Fourier Transform}}{\overset{\text{Fourier Transform}}{\Longleftrightarrow}} \quad \hat{f} : \mathbb{Z} \to \mathbb{C}.$$

Similarly, (4.3.19) can be rewritten as

$$\int_{S^1} |f(z)|^2 dz = \sum_{n=-\infty}^{\infty} |\hat{f}(n)|^2 = \int_{\mathbb{Z}} |\hat{f}(n)|^2 dn.$$

Thus, Plancherel theorem (Theorem 4.3.35) also reflects the Pontryagin duality:

$$\int_{S^1} |f(z)|^2 dz \quad \underset{\text{(Inverse) Fourier Transform}}{\overset{\text{Fourier Transform}}{\Longleftrightarrow}} \quad \int_{\mathbb{Z}} |\hat{f}(n)|^2 dn.$$

4.4 Local maxima and local minima

A very common application of using derivatives is to locate maxima and minima. We now present it in a rigorous manner.

Definition 4.4.1 (Local maxima and local minima)**.** Let $f : (A, B) \to \mathbb{R}$ be a function, and let $x_0 \in (A, B)$. We say that f attains a *local maximum* (极大值) of f at x_0 if there exists a $\delta > 0$ such that f attains a maximum at x_0 on $U_\delta(x_0) \subset (A, B)$, i.e.

$$f(x) \leq f(x_0)$$

for any $x \in U_\delta(x_0)$.

Similarly, we say that f attains a *local minimum* (极小值) of f at x_0 if there exists a $\delta > 0$ such that f attains a minimum at x_0 on $U_\delta(x_0) \subset (A, B)$, i.e.

$$f(x) \geq f(x_0)$$

for any $x \in U_\delta(x_0)$. A local maximum or minimum is also called a *local extremum* (极值).

Remark 4.4.2. If $f : (A, B) \to \mathbb{R}$ attains a extremum at x_0, we sometimes say that f attains a *global extremum* (最值) at x_0, in order to distinguish it from the local extrema. Note however that base on our textbook's definitions ([1, §4.5]), if a function $g : [A, B] \to \mathbb{R}$ attains a extremum at the endpoints, then it is **not** even a local extremum, since it is not a extremum **on** a neighborhood.

Recall the Fermat's lemma (Lemma 4.1.1) states that if $f : (A, B) \to \mathbb{R}$ attains a local extremum at x_0, then

$$f'(x_0) = 0. \tag{4.4.1}$$

We say that the points x_0 that satisfies (4.4.1) are *stationary* (稳定点).

Example 4.4.3 ([1, §4.5 例 1])**.** Find the local extrema of the function $f(x) = x^3 - 6x^2 - 15x + 4$.

证明. Since $f'(x) = 3x^2 - 12x - 15 = 3(x + 1)(x - 5)$, the stationary points of f are $x = -1$ and $x = 5$. Also, $f'(x) > 0$ for $x \in (-\infty, -1)$; $f'(x) < 0$ for $x \in (-1, 5)$;

$f'(x) > 0$ for $x \in (5, +\infty)$. Then $x = -1$ is a local maximum and $x = 5$ is a local minimum. $\square$

Theorem 4.4.4 ([1, §4.5 定理 2]). *Suppose $f : (A, B) \to \mathbb{R}$ satisfies*

$$f'(x_0) = \cdots = f^{(n-1)}(x_0) = 0, \quad f^{(n)}(x_0) \neq 0$$

for some $x_0 \in (A, B)$, $n \in \mathbb{N}^*$.

- *If n is odd, then there is no extremum at x_0.*
- *If n is even, the point x_0 is a strict local minimum if $f^{(n)}(x_0) > 0$, and a strict local maximum if $f^{(n)}(x_0) < 0$.*

证明. By the Taylor's formula (with respect to the Peano form of the remainder), we have

$$\begin{aligned} f(x) - f(x_0) &= \frac{1}{n!} f^{(n)}(x_0)(x - x_0)^n + h_n(x)(x - x_0)^n \\ &= \left(\frac{1}{n!} f^{(n)}(x_0) + h_n(x) \right)(x - x_0)^n, \end{aligned} \tag{4.4.2}$$

where $\lim_{x \to x_0} h_n(x) = 0$. We shall argue similar to the proof of Fermat's lemma (Lemma 4.1.1).

- If n is odd, the factor $(x - x_0)^n$ changes sign when x passes through x_0, and then the sign of (4.4.2) also changes. Consequently, $f(x) - f(x_0)$ also changes sign, and so there is no extremum.
- If n is even, the factor $(x - x_0)^n > 0$ for $x \neq x_0$, and then the sign of (4.4.2) does not change on a deleted neighborhood of x_0. Consequently, $f(x) - f(x_0)$ also does not change sign, and so there is an extremum.

This establishes the theorem. $\square$

Example 4.4.5 ([1, §4.5 例 2]). Study the local extrema of the function

$$f(x) = x^3 e^{-x}.$$

证明. Since $f'(x) = (3x^2 - x^3)e^{-x}$, the stationary points of f are $x = 0$ and $x = 3$. Further, one calculates

$$f''(x) = (x^3 - 6x^2 + 6x)e^{-x}.$$

Then $f''(x) = 0$, $f''(3) = -9e^{-3} < 0$. Then by Theorem 4.4.4, $x = 3$ is a (strict) local maximum.

Also, one calculates

$$f'''(x) = (-x^3 + 9x^2 - 18x + 6)e^{-x}.$$

Then $f'''(0) > 0$, and so by Theorem 4.4.4, $x = 0$ is not an extremum. □

Example 4.4.6 (Snell's law, 光的折射定律, [1, §4.5 例 5])**.** Consider two media, and suppose that light propagates from point A to B. Then we have

$$\frac{\sin\alpha}{\sin\beta} = \frac{v_1}{v_2}$$

where α is the *angle of incidence* (入射角), β is the *angle of refraction* (折射角), v_1 is the velocity of light in the first medium, v_2 is the velocity of light in the second medium (see Figure 4.1).

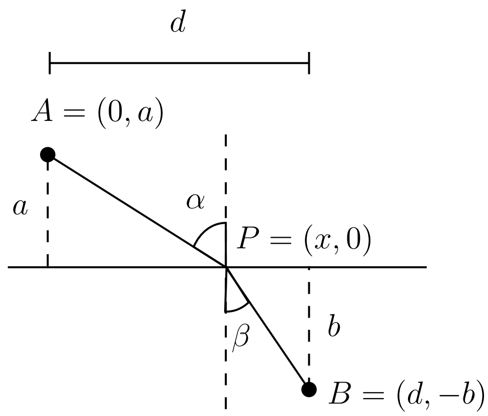


图 4.1: Snell's law.

证明. If v_1 and v_2 are the velocities of light in the two media, the time required to travel the path is

$$T(x) = \frac{1}{v_1}\sqrt{a^2 + x^2} + \frac{1}{v_2}\sqrt{b^2 + (d - x)^2}.$$

According to *Fermat's principle* (费马原理), the actual orbit of a light ray between A and B is such that the ray requires minimum time to pass from A to B compared with all paths joining A and B.

We now find the extremum of the function $T(x)$:

$$T'(x) = \frac{1}{v_1}\frac{x}{\sqrt{a^2+x^2}} - \frac{1}{v_2}\frac{d-x}{\sqrt{b^2+(d-x)^2}} = 0.$$

It yields a stationary point $x_0 \in \mathbb{R}$:

$$\frac{1}{v_1}\sin\alpha = \frac{1}{v_1}\frac{x_0}{\sqrt{a^2+x_0^2}} = \frac{1}{v_2}\frac{d-x_0}{\sqrt{b^2+(d-x_0)^2}} = \frac{1}{v_2}\sin\beta.$$

Then one calculates

$$T''(x_0) = \frac{1}{v_1}\frac{a^2}{(a^2+x_0^2)^{\frac{3}{2}}} + \frac{1}{v_2}\frac{b^2}{(b^2+(d-x_0)^2)^{\frac{3}{2}}} > 0.$$

Hence, we conclude that f attains a local minimum at x_0.

On the other hand, since $\lim\limits_{x\to\infty} T(x) = +\infty$, $T(x)$ has at least one global minimum on $\mathbb{R}$. Thus, we conclude that T attains the global minimum at x_0, i.e. $P = (x_0, 0)$. $\square$

Example 4.4.7 (Least squares, 最小二乘法, [1, §4.5 例 6])**.** Suppose that the n observations have been made, resulting in n values $a_1, a_2, \dots, a_n$. The problem is to determine a number in a reasonable way from these observations.

Gauss and Legendre discovered a simple and effective method by taking the squares of residuals:

$$f(x) = \sum_{i=1}^{n}(x-a_i)^2$$

and find the minimum of f. Then the minimum is the **best estimation value** for the observations.

One may calculates

$$f'(x) = 2\sum_{i=1}^{n}(x-a_i) = 0.$$

It follows that there is only one stationary point

$$x_0 = \frac{1}{n}(a_1 + \cdots + a_n).$$

On the other hand, since $\lim\limits_{x\to\infty} f(x) = +\infty$, f has at least one global minimum on $\mathbb{R}$. Thus, we conclude that f attains the global minimum at x_0.

4.5 Convexity of a function

Definition 4.5.1. A function $f:(a,b)\to\mathbb{R}$ is said to be *convex* (下凸) if we have the inequalities

$$f(\lambda x_1+(1-\lambda)x_2)\le \lambda f(x_1)+(1-\lambda)f(x_2) \tag{4.5.1}$$

for any $x_1,x_2\in(a,b)$ and any $\lambda\in[0,1]$. Similarly, a is said to be *concave* (上凸) if the reversed inequalities (4.5.1) hold.

Via the definition, one easily deduces the *Jensen's inequality* (琴生不等式):

Theorem 4.5.2 (Jensen's inequality)**.** *Let $f:(a,b)\to\mathbb{R}$ be a convex function. Then for any $x_1,\dots,x_n\in(a,b)$, $\lambda_1,\dots,\lambda_n\in[0,1]$ with $\lambda_1+\cdots+\lambda_n=1$, we have*

$$f\left(\sum_{i=1}^n\lambda_i x_i\right)\le\sum_{i=1}^n\lambda_i f(x_i).$$

The equality holds if and only if either

(1) $x_1=x_2=\cdots=x_n$, or

(2) f is linear on a domain containing $x_1,x_2,\dots,x_n$.

Example 4.5.3 (AM-GM inequality)**.** For $x_1,\dots,x_n\in(0,+\infty)$, we have

$$\sqrt[n]{x_1x_2\cdots x_n}\le\frac{x_1+x_2+\cdots+x_n}{n}.$$

The equality holds if and only if $x_1=x_2=\cdots=x_n$.

证明. Note that $-\ln x$ is a strictly convex function on $(0,+\infty)$. Then by Jensen's inequality, for any $x_1,\dots,x_n\in(0,+\infty)$, we have

$$\ln\sqrt[n]{x_1x_2\cdots x_n}=\frac{1}{n}\sum_{i=1}^n\ln(x_i)\le\ln\left(\frac{1}{n}\sum_{i=1}^n x_i\right)=\ln\frac{x_1+x_2+\cdots+x_n}{n}.$$

The equality holds if and only if $x_1=x_2=\cdots=x_n$. □

The convexity of a differentiable function depends on the derivative of this function.

Proposition 4.5.4 (Convexity of a differentiable function)**.** *A differentiable function $f : (a, b) \to \mathbb{R}$ is (strictly) convex if and only if its derivative f' is (strictly) increasing.*

证明. Assume that $x_1, x, x_2 \in (a, b)$ with $x_1 < x < x_2$. Let $A = x - x_1$, $B = x_2 - x$, $\lambda = \frac{B}{A+B}$. Then by (4.5.1), we have

$$f(x) \le \frac{B}{A+B} f(x_1) + \frac{A}{A+B} f(x_2).$$

It follows that

$$B(f(x) - f(x_1)) \le A(f(x_2) - f(x))$$

and so

$$\frac{f(x) - f(x_1)}{x - x_1} \le \frac{f(x_2) - f(x)}{x_2 - x} \tag{4.5.2}$$

for any $x_1, x, x_2 \in (a, b)$ with $x_1 < x < x_2$.

Now letting x in (4.5.2) tend first to x_1, then to x_2, we obtain

$$f'(x_1) \le \frac{f(x_2) - f(x_1)}{x_2 - x_1} \le f'(x_2).$$

In particular, we see that f' is increasing. To see the strict convexity, applying Lagrange's mean value theorem (Theorem 4.1.6) to (4.5.2), there are $\xi_1 \in (x_1, x)$, $\xi_2 \in (x, x_2)$ such that

$$f'(x_1) \le f'(\xi_1) = \frac{f(x) - f(x_1)}{x - x_1} < \frac{f(x_2) - f(x)}{x_2 - x} = f'(\xi_2) \le f'(x_2).$$

On the other hand, assume that f' is (strictly) increasing. Then for $a < x_1 < x < x_2 < b$, by Lagrange's mean value theorem (Theorem 4.1.6), there are $\xi_1 \in (x_1, x)$, $\xi_2 \in (x, x_2)$ such that

$$f'(\xi_1) = \frac{f(x) - f(x_1)}{x - x_1}, \quad f'(\xi_2) = \frac{f(x_2) - f(x)}{x_2 - x}.$$

Then the (strict) monotonicity of f' implies (4.5.2), and so the (strict) convexity of f. □

Corollary 4.5.5. *A differentiable function $f : (a, b) \to \mathbb{R}$ is convex if and only if for any $x_0 \in (a, b)$, we have*

$$f(x) \geq f(x_0) + f'(x_0)(x - x_0) \tag{4.5.3}$$

for any $x \in (a, b)$ with $x \neq x_0$. The statement for strict convex functions is similar.

证明. By Proposition 4.5.4, f' is increasing. If $x > x_0$, by Lagrange's mean value theorem (Theorem 4.1.6), there exists $\xi \in (x_0, x)$ such that

$$f(x) - f(x_0) = f'(\xi)(x - x_0) \geq f'(x_0)(x - x_0).$$

The case when $x < x_0$ is similar.

On the contrary, if (4.5.3) holds, then

$$f'(x_0) \leq \frac{f(x_2) - f(x_0)}{x_2 - x_0}$$

for any $x_2 > x_0$, and

$$\frac{f(x_0) - f(x_1)}{x_0 - x_1} \leq f'(x_0)$$

for any $x_1 < x_0$. Then we obtain (4.5.2) for $x = x_0$. The consequence follows. □

Corollary 4.5.6 ([1, §4.6 定理 1])**.** *A twice differentiable function $f : (a, b) \to \mathbb{R}$ is convex if and only if for any $x \in (a, b)$, we have*

$$f''(x) \geq 0.$$

In addition, if $f''(x) > 0$ for all $x \in (a, b)$, then f is strictly convex.

证明. It is a direct consequence of Corollary 4.1.8 and Proposition 4.5.4. □

It is useful to study the points so that the convexity of a function changes.

Definition 4.5.7 (Inflection point)**.** Let $x_0 \in (a, b)$, $f : (a, b) \to \mathbb{R}$ be differentiable. If there is an $\epsilon > 0$ so that the function is convex (resp. concave) on $(x_0 - \epsilon, x_0)$, and concave (resp. convex) on $(x_0, x_0 + \epsilon)$, then x_0 is called a *point of inflection* (拐点).

Theorem 4.5.8 ([1, §4.6 定理 2])**.** *Let $f:(a,b)\to\mathbb{R}$ be twice differentiable. If $x_0\in(a,b)$ is an inflection point of f, then $f''(x_0)=0$.*

证明. Since x_0 is an inflection point of f, by Proposition 4.5.4 and Definition 4.5.7, we see that f' attains either a local maximum or minimum. Then by Fermat's lemma (Lemma 4.1.1), we have $f''(x_0)=0$. □

Finally, we introduce another useful concept, namely the asymptotes of a function:

Definition 4.5.9 (Asymptotes)**.** A line $L(x)=kx+b$ is called an *asymptote* (渐近线) of the graph of a function $f:(a,+\infty)\to\mathbb{R}$ as $x\to+\infty$ if

$$\lim_{x\to+\infty}(f(x)-L(x))=0.$$

The case when $x\to-\infty$ can be similarly defined. If an asymptote has the form $L(x)=b$, then we say that L is a *horizontal asymptote* (水平渐近线).

If $|f(x)|\to\infty$ as $x\to a-0$ or $x\to a+0$, then the line $x=a$ is called the *vertical asymptote* (垂直渐近线) of the graph of f.

Theorem 4.5.10 ([1, §4.6 定理 3])**.** *Let $f:(a,+\infty)\to\mathbb{R}$, $L(x)=kx+b$. Then L is an asymptote of the graph of f as $x\to+\infty$ if and only if*

$$\lim_{x\to+\infty}\frac{f(x)}{x}=k,\qquad \lim_{x\to+\infty}(f(x)-kx)=b.$$

证明. If L is an asymptote of f, then

$$\lim_{x\to+\infty}(f(x)-kx-b)=0.$$

It follows that $\lim\limits_{x\to+\infty}(f(x)-kx)=b$. Moreover,

$$\lim_{x\to+\infty}\left(\frac{f(x)}{x}-k\right)=\lim_{x\to+\infty}\frac{f(x)-kx}{x}=0.$$

Then we have $\lim\limits_{x\to+\infty}\dfrac{f(x)}{x}=k$.

On the contrary, $\lim\limits_{x\to+\infty}(f(x)-kx)=b$ clearly means $\lim\limits_{x\to+\infty}(f(x)-L(x))=0$. □

Example 4.5.11 ([1, §4.6 例 3])**.** Sketch the graph of $f(x) = \dfrac{x^2}{x-1}$.

证明. f is not defined at $x = 1$. Besides,

$$\lim_{x\to 1+0} f(x) = +\infty, \quad \lim_{x\to 1-0} f(x) = -\infty.$$

Then $x = 1$ is a vertical asymptote of f.

Next, one calculates

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

Then f has stationary points $x = 0$ and $x = 2$. Moreover,

$$f''(x) = \frac{2}{(x-1)^3}.$$

Then f is convex on $(1, +\infty)$, and concave on $(-\infty, 1)$.

Finally, consider

$$\lim_{x\to\infty} \frac{f(x)}{x} = \lim_{x\to\infty} \frac{x^2}{x(x-1)} = 1.$$

and

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

Then $L(x) = x + 1$ is an asymptote of f as $x \to \infty$.

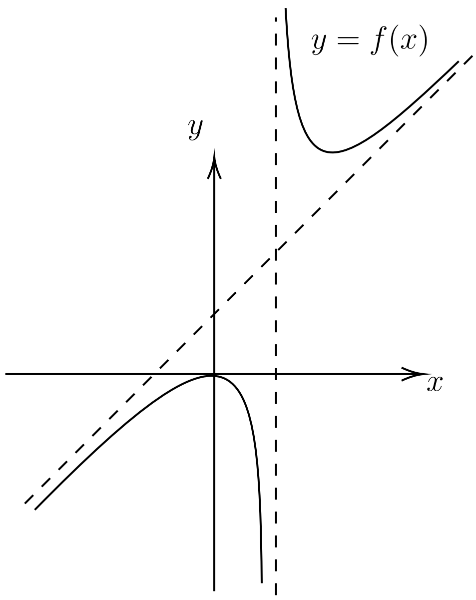


图 4.2: Example 4.5.11.

Hence, the graph of f is shown in Figure 4.2. $\square$

4.6 Curvature

In mathematics, *curvature* (曲率) is any of several strongly related notions in geometry that intuitively measure the amount by which a geometric object (e.g. curves, surfaces, n-dimensional manifolds) deviates from being the Euclidean space (e.g. $\mathbb{R}$, $\mathbb{R}^2$, $\mathbb{R}^n$). In this section, we briefly introduce the curvature of curves.

First, let $\Gamma : [0, T] \to \mathbb{R}^n$ be a twice differentiable curve parametrized by arc length. Then the velocity

$$|\mathbf{v}| = |\Gamma'(t)| = 1$$

is a unit vector. Moreover, by [1, §5.5 Exercises 3, 4], the acceleration $\mathbf{a}$ satisfies

$$\langle \mathbf{a}, \mathbf{v} \rangle = \Gamma''(t) \cdot \Gamma'(t) = 0.$$

Thus, $\mathbf{a}$ and $\mathbf{v}$ are orthogonal. Intuitively, the larger the quantity $|\mathbf{a}|$ is, the easier the curve of Γ is to bend in the next moment. In other words, the quantity $|\mathbf{a}|$ reflects the deviation of the curve from being the straight line in the direction $\mathbf{v}$.

Definition 4.6.1 (Curvature of curves)**.** For $n \in \mathbb{N}^*$, let $\Gamma : [0, T] \to \mathbb{R}^n$ be a twice differentiable curve parametrized by arc length, we define the *curvature* (曲率) of Γ at $\Gamma(t)$ by

$$\kappa(t) = |\Gamma''(t)|.$$

Since curves are not always parametrized by arc length, in the following, we deduce a more applicable formula for the curvature of a *plane curve* (or an *arc* in a plane). Let $\Gamma : [A, B] \to \mathbb{R}^2$ be a plane curve given by a parametric system:

$$\Gamma : t \mapsto \begin{cases} x = x(t) \\ y = y(t) \end{cases}.$$

Suppose that $x(t)$ and $y(t)$ are twice differentiable.

Now find the tangential and normal components $\mathbf{a}_t$ and $\mathbf{a}_n$ respectively of the acceleration $\mathbf{a} = (x''(t), y''(t))$ of the curve Γ at $\Gamma(t)$. Thus, $\mathbf{a} = \mathbf{a}_t + \mathbf{a}_n$, where $\mathbf{a}_t$ is parallel with the velocity $\mathbf{v}(t) = (x'(t), y'(t))$ (i.e. tangent to the curve), and $\mathbf{a}_n$ is directed along the normal to the curve.

Then for any $t_0 \in [A, B]$, we consider the quantity

$$\kappa_\Gamma(t_0) = \frac{|\mathbf{a}_n(t_0)|}{|\mathbf{v}(t_0)|^2}. \tag{4.6.1}$$

and call it the *curvature* of Γ at $\Gamma(t_0)$.

Now note that for any parametrization of a curve $\Gamma : [A, B] \to \mathbb{R}^2$, we may calculate the arc length $s : [A, B] \to [0, T]$ via (3.5.3):

$$s(t) = \int_A^t \sqrt{(x'(r))^2 + (y'(r))^2} dr.$$

Then we obtain a parametrization of the curve by arc length

$$\Gamma_s = \Gamma \circ s^{-1},$$

and $\Gamma = \Gamma_s \circ s$. The following proposition justifies the terminology of (4.6.1):

Proposition 4.6.2. *Let $\Gamma_s : [0, T] \to \mathbb{R}^2$ be a twice differentiable plane curve parametrized by arc length, i.e. $|\Gamma_s'| \equiv 1$. Let $\gamma : [A, B] \to [0, T]$ be a twice differentiable function. Let $\Gamma = \Gamma_s \circ \gamma$. Then for any $t_0 \in [A, B]$, we have*

$$\kappa_\Gamma(t_0) = \kappa(\gamma(t_0))$$

where $\kappa(\gamma(t_0))$ is the curvature of Γ_s at $\Gamma_s(\gamma(t_0))$.

证明. Since Γ_s is parametrized by arc length, we have

$$0 = \frac{d}{dt}\langle \Gamma_s'(t), \Gamma_s'(t)\rangle = 2\langle \Gamma_s''(t), \Gamma_s'(t)\rangle$$

for $t \in [0, T]$ (cf. [1, §5.5 Exercises 3, 4]). Also, one calculates

$$\begin{aligned}\frac{d\Gamma}{dt}(t_0) =& \Gamma_s'(\gamma(t_0)) \cdot \gamma'(t_0), \\ \frac{d^2\Gamma}{dt^2}(t_0) =& \Gamma_s''(\gamma(t_0)) \cdot (\gamma'(t_0))^2 + \Gamma_s'(\gamma(t_0)) \cdot \gamma''(t_0).\end{aligned}$$

Then we have

$$\kappa_\Gamma(t_0) = \frac{|\Gamma_s''(\gamma(t_0)) \cdot (\gamma'(t_0))^2|}{|\Gamma_s'(\gamma(t_0)) \cdot \gamma'(t_0)|^2} = |\Gamma_s''(\gamma(t_0))| = \kappa(\gamma(t_0)).$$

This establishes the proposition. $\square$

Remark 4.6.3. In particular, Proposition 4.6.2 shows that the curvature (4.6.1) does not depend on the parametrization of the curve.

We define the *radius of curvature* (曲率半径) at $\Gamma(t)$ by the quantity

$$R(t) = \frac{1}{\kappa_\Gamma(t)} = \frac{|\mathbf{v}(t)|^2}{|\mathbf{a}_n(t)|}.$$

The motivation of the quantity $R(t)$ is to find the **best circular approximation** of the curve at a given point, namely the *osculating circle* (曲率圆):

Example 4.6.4 ([1, §4.7 例 2])**.** Let $r > 0$ and $\Gamma(\theta) = (r\cos\theta, r\sin\theta),\ \theta \in [0, 2\pi]$. Then one calculates

$$\begin{aligned} \mathbf{v}(\theta) =& (-r\sin\theta, r\cos\theta), \\ \mathbf{a}_n(\theta) =& (-r\cos\theta, -r\sin\theta). \end{aligned}$$

Then the radius of curvature is

$$R(\theta) = \frac{|\mathbf{v}(\theta)|^2}{|\mathbf{a}_n(\theta)|} = r.$$

Example 4.6.5 ([1, §4.7 例 1])**.** Let $\Gamma(t)$, $t \in [A, B]$ be a straight line. Then

$$\mathbf{a}_n(t) = \mathbf{0}.$$

In particular, the curvature of a straight line vanishes: $\kappa \equiv 0$. Of course, how could a straight line deviate from being straight?

Proposition 4.6.6. *In terms of the parametric system* $\Gamma(t) = (x(t), y(t))$*, the curvature at* $\Gamma(t)$ *is given by*

$$\kappa(t) = \frac{|x'(t)y''(t) - x''(t)y'(t)|}{[(x'(t))^2 + (y'(t))^2]^{\frac{3}{2}}}$$

In particular, if $\Gamma(x) = (x, f(x))$ *is a graph of a function, then*

$$\kappa(t) = \frac{|f''(t)|}{[1 + (f'(t))^2]^{\frac{3}{2}}}.$$

证明. One calculates

$$\mathbf{v} = \Gamma' = (x', y'), \quad |\mathbf{v}| = \sqrt{(x')^2 + (y')^2}.$$

Then by the *Gram-Schmidt process* (施密特正交化), we have

$$\mathbf{a} = \Gamma'' = (x'', y''), \quad \mathbf{a}_t = \frac{\mathbf{a} \cdot \mathbf{v}}{|\mathbf{v}|^2}\mathbf{v} = \frac{x'x'' + y'y''}{(x')^2 + (y')^2}(x', y'),$$

and so

$$\begin{aligned}\mathbf{a}_n = \mathbf{a} - \mathbf{a}_t =& (x'', y'') - \frac{x'x'' + y'y''}{(x')^2 + (y')^2}(x', y') \\ =& \left(\frac{y'(y'x'' - x'y'')}{(x')^2 + (y')^2}, \frac{x'(x'y'' - y'x'')}{(x')^2 + (y')^2}\right).\end{aligned}$$

It follows that

$$\kappa = \frac{|\mathbf{a}_n|}{|\mathbf{v}|^2} = \frac{|x'y'' - x''y'|}{[(x')^2 + (y')^2]^{\frac{3}{2}}}.$$

Thus, we establish the formula. □

In high dimensional, curvature is a main topic of *differential geometry*. For integer $n \geq 2$, the curvature of n-dimensional Riemannian manifolds can be defined *intrinsically* (内蕴) without reference to a larger space.

- For dimension $n = 2$, one may study the *principal curvature* (主曲率), *mean curvature* (平均曲率), *Gaussian curvature* (高斯曲率), etc.
- In mathematics the infinitesimal geometry of Riemannian manifolds with dimension $n > 2$ is too complicated to be described by a single number at a given point. One may study the *sectional curvature* (截面曲率), *Riemann curvature tensor* (黎曼曲率张量), *Ricci curvature tensor* (里奇曲率张量), etc.

Chapter 5

Linear Algebra

5.1 Euclidean space

For $n \in \mathbb{N}^*$, the n-dimensional *Euclidean space* (n 维欧式空间) is the set

$$\mathbb{R}^n = \{(x_1, \ldots, x_n) : x_1, \ldots, x_n \in \mathbb{R}\}$$

of ordered n-real numbers, equipped with the addition, scalar multiplication, and dot product (See also [1, §6.1 Exercise 4]). The elements in $\mathbb{R}^n$ are called the *vectors* (向量). $\mathbb{R}^n$ is a fundamental *vector space*. In the following, we shall quickly review the idea of the structure of Euclidean spaces. One may see Appendix 7.2 for more details.

5.1.1 Norm

A (real) *vector space* (向量空间) V is some space that behaves like a Euclidean space $\mathbb{R}^n$. It has two operations:

- addition $+ : V \times V \to V$,
- scalar multiplication $\cdot : \mathbb{R} \times V \to V$,

defined in a compatible manner. That is to say, the calculation of vectors satisfies the associativity, commutativity, distributivity, etc. (See Definition 7.2.5 for more details.)

A *norm* (范数) $\|\cdot\|: V \to [0, +\infty)$ on a (real) vector space V is a function similar to the absolute value $|\cdot|$ in $\mathbb{R}$ (see also Definition 7.2.11). It measures the length of a vector. In $\mathbb{R}^n$, the common norms include

- *Euclidean norm* (also called *l^2 norm*): For $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$,

$$\|\mathbf{x}\|_2 \coloneqq \sqrt{x_1^2 + \cdots + x_n^2}.$$

- *l^p norm* for $p \geq 1$: For $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$,

$$\|\mathbf{x}\|_p \coloneqq (|x_1|^p + \cdots + |x_n|^p)^{\frac{1}{p}}.$$

- *Maximum norm* (or *uniform norm*, *infinite norm*): For $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$,

$$\|\mathbf{x}\|_\infty \coloneqq \max_i |x_i|.$$

  The maximum norm $\|\cdot\|_\infty$ can be thought of approached by the l^p norm $\|\cdot\|_p$ as $p \to +\infty$.

Note that all these norms are equivalent (Proposition 7.2.15). For instance, we have

Lemma 5.1.1. *For $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$, we have*

$$\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_1 \leq \sqrt{n}\|\mathbf{x}\|_2 \leq n\|\mathbf{x}\|_\infty.$$

证明. $\|\mathbf{x}\|_1 \leq \sqrt{n}\|\mathbf{x}\|_2$ follows from Cauchy-Schwarz inequality. □

A norm must satisfy the triangle inequality:

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|.$$

A vector $\mathbf{x} \in V$ induces a *unit vector* $\mathbf{x}^\circ = \frac{\mathbf{x}}{\|\mathbf{x}\|}$. In what follows, when we say a norm $\|\cdot\|$ on $\mathbb{R}^n$, we refer to the **Euclidean norm** $\|\cdot\| = \|\cdot\|_2$, unless otherwise specified.

The prototypical representation of elements in $\mathbb{R}^n$ is the *Cartesian coordinate system* (笛卡尔坐标系). For instance, in 3-dimensional, for vectors $\mathbf{v}_1 = (x_1, y_1, z_1), \mathbf{v}_2 = (x_2, y_2, z_2) \in \mathbb{R}^3$, we have

$$\mathbf{v}_1 + \mathbf{v}_2 = (x_1 + x_2, y_1 + y_2, z_1 + z_2).$$

Also for $\mathbf{v}_1 = (x_1, y_1, z_1) \in \mathbb{R}^3$, $\lambda \in \mathbb{R}$, we have

$$\lambda\mathbf{v} = (\lambda x, \lambda y, \lambda z).$$

Then let $\mathbf{i}, \mathbf{j}, \mathbf{k} \in \mathbb{R}^3$ be unit vectors on x-axis, y-axis, z-axis, respectively. Let $O = (0, 0, 0) \in \mathbb{R}^3$ be the origin. Then an element $P = (x, y, z) \in \mathbb{R}^3$ corresponds bijectively to

$$\begin{aligned}\overrightarrow{OP} &= (x, 0, 0) + (0, y, 0) + (0, 0, z)\\ &= x(1, 0, 0) + y(0, 1, 0) + z(0, 0, 1)\\ &= x\mathbf{i} + y\mathbf{j} + z\mathbf{k}.\end{aligned}$$

5.1.2 Dot product

An *inner product* (内积) $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ on a (real) vector space V is an operation that allow formal definitions of geometric notions, such as lengths, angles, and *orthogonality* (垂直/正交). In particular, an inner product $\langle \cdot, \cdot \rangle$ naturally induces an associated norm $\| \cdot \|$:

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$$

for any $\mathbf{v} \in V$. (see also 7.2.19.)

The most commonly used inner product on $\mathbb{R}^n$ is the *dot product* (点乘): For $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, the dot product is given by

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\| \|\mathbf{w}\| \cos(\mathbf{v}, \mathbf{w})$$

where $(\mathbf{v}, \mathbf{w}) \in [0, \pi]$ is the angle between v and w. The dot product satisfies:

(1) (Commutativity) For any $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, we have $\mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v}$.

(2) (Associativity) For any $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, $\lambda \in \mathbb{R}$, we have $(\lambda\mathbf{v}) \cdot \mathbf{w} = \lambda(\mathbf{v} \cdot \mathbf{w})$.

(3) (Distributivity) For any $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, we have $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$.

In addition, the dot product induces the Euclidean norm: $\sqrt{\mathbf{v}\cdot\mathbf{v}} = \|\mathbf{v}\|$ for $\mathbf{v} \in \mathbb{R}^n$. Also, for $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, we have

$$\mathbf{v} \text{ and } \mathbf{w} \text{ are orthogonal/perpendicular (正交/垂直)} \quad \Leftrightarrow \quad \mathbf{v}\cdot\mathbf{w} = 0.$$

In terms of Cartesian coordinates, for vectors $\mathbf{v}_1 = (x_1, y_1, z_1), \mathbf{v}_2 = (x_2, y_2, z_2) \in \mathbb{R}^3$, the dot product is given by the formula

$$\begin{aligned}
\mathbf{v}_1 \cdot \mathbf{v}_2 &= (x_1, y_1, z_1) \cdot (x_2, y_2, z_2) \\
&= (x_1\mathbf{i} + y_1\mathbf{j} + z_1\mathbf{k}) \cdot (x_2\mathbf{i} + y_2\mathbf{j} + z_2\mathbf{k}) \\
&= x_1x_2 + y_1y_2 + z_1z_2.
\end{aligned}$$

Example 5.1.2 ([1, §5.2 例 2])**.** Let $A(1,0,1)$, $B(0,1,1)$, $C(1,-1,1) \in \mathbb{R}^3$. Find the angle between $\overrightarrow{AB}$ and $\overrightarrow{AC}$.

证明. One calculates $\overrightarrow{AB} = (-1,1,0)$, $\overrightarrow{AC} = (0,-1,0)$. Then the angle

$$\cos(\overrightarrow{AB}, \overrightarrow{AC}) = \frac{\overrightarrow{AB}\cdot\overrightarrow{AC}}{|\overrightarrow{AB}||\overrightarrow{AC}|} = \frac{-1}{\sqrt{2}\cdot 1} = -\frac{\sqrt{2}}{2}.$$

Thus, $(\overrightarrow{AB}, \overrightarrow{AC}) = \frac{3\pi}{4}$. □

Here we introduce a useful concepts - *direction cosine* (方向余弦). Let $\mathbf{a} = (x, y, z) \in \mathbb{R}^3$. Then we have the direction cosine:

$$\begin{aligned}
\cos(\mathbf{a}, \mathbf{i}) &= \frac{\mathbf{a}\cdot\mathbf{i}}{\|\mathbf{a}\|} = \frac{x}{\sqrt{x^2+y^2+z^2}}, \\
\cos(\mathbf{a}, \mathbf{j}) &= \frac{\mathbf{a}\cdot\mathbf{j}}{\|\mathbf{a}\|} = \frac{y}{\sqrt{x^2+y^2+z^2}}, \\
\cos(\mathbf{a}, \mathbf{k}) &= \frac{\mathbf{a}\cdot\mathbf{k}}{\|\mathbf{a}\|} = \frac{z}{\sqrt{x^2+y^2+z^2}}.
\end{aligned}$$

Then we have

$$\cos^2(\mathbf{a}, \mathbf{i}) + \cos^2(\mathbf{a}, \mathbf{j}) + \cos^2(\mathbf{a}, \mathbf{k}) = 1$$

and the unit vector

$$\frac{\mathbf{a}}{\|\mathbf{a}\|} = \cos(\mathbf{a}, \mathbf{i})\mathbf{i} + \cos(\mathbf{a}, \mathbf{j})\mathbf{j} + \cos(\mathbf{a}, \mathbf{k})\mathbf{k}.$$

5.1.3 Cross product

A *cross product* (叉乘) $\times : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$ is an operation on two vectors to find a vector that is perpendicular to both two vectors. For $\mathbf{v}, \mathbf{w} \in \mathbb{R}^3$, the cross product is defined by the formula

$$\mathbf{v} \times \mathbf{w} = (\|\mathbf{v}\| \|\mathbf{w}\| \sin(\mathbf{v}, \mathbf{w}))\mathbf{n},$$

where

- $(\mathbf{v}, \mathbf{w}) \in [0, \pi]$ is the angle between $\mathbf{v}$ and $\mathbf{w}$,
- $\mathbf{n}$ is the unit vector perpendicular to the plane containing $\mathbf{v}$ and $\mathbf{w}$, with direction such that the ordered set $(\mathbf{v}, \mathbf{w}, \mathbf{n})$ is *positively oriented* (正定向).

The cross product satisfies:

(1) (Anticommutativity) For any $\mathbf{v}, \mathbf{w} \in \mathbb{R}^3$, we have $\mathbf{v} \times \mathbf{w} = -\mathbf{w} \times \mathbf{v}$.

(2) (Associativity) For any $\mathbf{v}, \mathbf{w} \in \mathbb{R}^3$, $\lambda \in \mathbb{R}$, we have

$$\lambda(\mathbf{v} \times \mathbf{w}) = (\lambda \mathbf{v}) \times \mathbf{w} = \mathbf{v} \times (\lambda \mathbf{w}).$$

(3) (Distributivity) For any $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$, we have $(\mathbf{u} + \mathbf{v}) \times \mathbf{w} = \mathbf{u} \times \mathbf{w} + \mathbf{v} \times \mathbf{w}$.

(4) (*Jacobi identity* 雅可比恒等式) For any $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$, we have

$$(\mathbf{u} \times \mathbf{v}) \times \mathbf{w} + (\mathbf{v} \times \mathbf{w}) \times \mathbf{u} + (\mathbf{w} \times \mathbf{u}) \times \mathbf{v} = \mathbf{0}.$$

Remark 5.1.3. Therefore, $(\mathbb{R}^3, \times)$ forms a *Lie algebra* (李代数) - the Lie algebra $\mathfrak{so}(3)$ of the *special orthogonal group/rotation group* (特殊正交群/旋转群) in 3 dimensions, SO(3). Each vector $\mathbf{v} \in \mathbb{R}^3$ may be pictured as an infinitesimal rotation around the axis $\mathbf{v}$, with angular speed equal to the magnitude of $\mathbf{v}$. The cross product is a measure of the non-commutativity between two rotations.

For $\mathbf{v}, \mathbf{w} \in \mathbb{R}^3$, we have

$$\mathbf{v} \text{ and } \mathbf{w} \text{ are } \textit{parallel/collinear} \text{ (平行/共线)} \quad \Leftrightarrow \quad \mathbf{v} \times \mathbf{w} = \mathbf{0}.$$

Next, we introduce the coordinate expression of cross products. Note that

$$\mathbf{i}\times\mathbf{i}=\mathbf{j}\times\mathbf{j}=\mathbf{k}\times\mathbf{k}=\mathbf{0},\quad \mathbf{i}\times\mathbf{j}=\mathbf{k},\quad \mathbf{j}\times\mathbf{k}=\mathbf{i},\quad \mathbf{k}\times\mathbf{i}=\mathbf{j}.$$

Then for vectors $\mathbf{v}_1=(x_1,y_1,z_1),\mathbf{v}_2=(x_2,y_2,z_2)\in\mathbb{R}^3$, one calculates

$$\begin{aligned}\mathbf{v}_1\times\mathbf{v}_2&=(x_1\mathbf{i}+y_1\mathbf{j}+z_1\mathbf{k})\times(x_2\mathbf{i}+y_2\mathbf{j}+z_2\mathbf{k})\\&=(y_1z_2-y_2z_1)\mathbf{i}+(z_1x_2-z_2x_1)\mathbf{j}+(x_1y_2-x_2y_1)\mathbf{k}\\&=\det\begin{bmatrix}\mathbf{i}&\mathbf{j}&\mathbf{k}\\x_1&y_1&z_1\\x_2&y_2&z_2\end{bmatrix}.\end{aligned}\tag{5.1.1}$$

Example 5.1.4 ([1, §5.2 例 4])**.** Are $\mathbf{a}=(5,6,0)$, $\mathbf{b}=(1,2,3)$ collinear?

证明. Since

$$\mathbf{a}\times\mathbf{b}=\det\begin{bmatrix}\mathbf{i}&\mathbf{j}&\mathbf{k}\\5&6&0\\1&2&3\end{bmatrix}=18\mathbf{i}-15\mathbf{j}+4\mathbf{k}\neq\mathbf{0},$$

we conclude that $\mathbf{a}$ and $\mathbf{b}$ are not collinear. □

5.1.4 Scalar triple product

For $\mathbf{u},\mathbf{v},\mathbf{w}\in\mathbb{R}^3$, the *scalar triple product* (or *mixed product*, 标量三重积/混合积) is defined as

$$\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w}).$$

Geometrically, the scalar triple product $\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})$ is the (signed) volume of the parallelepiped defined by the three vectors $\mathbf{u},\mathbf{v},\mathbf{w}$.

Note that for any $\mathbf{u},\mathbf{v},\mathbf{w}\in\mathbb{R}^3$, we have

$$\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})=\mathbf{v}\cdot(\mathbf{w}\times\mathbf{u})=\mathbf{w}\cdot(\mathbf{u}\times\mathbf{v}).$$

In particular, we have

$$\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})=(\mathbf{u}\times\mathbf{v})\cdot\mathbf{w}.$$

Also, for $u,v,w\in\mathbb{R}^3$, we have

$$\mathbf{u},\mathbf{v},\mathbf{w}\text{ are } \textit{coplanar}\ (\text{共面})\quad\Leftrightarrow\quad \mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})=0.$$

In terms of coordinates, the scalar triple product can be expressed as follows. Let $\mathbf{u} = (x_1, y_1, z_1), \mathbf{v} = (x_2, y_2, z_2), \mathbf{w} = (x_3, y_3, z_3) \in \mathbb{R}^3$. One calculates

$$\begin{aligned}(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w} &= \det \begin{bmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{bmatrix} \cdot (x_3\mathbf{i} + y_3\mathbf{j} + z_3\mathbf{k}) \\ &= \det \begin{bmatrix} x_3 & y_3 & z_3 \\ x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{bmatrix} = \det \begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{bmatrix}. \end{aligned} \tag{5.1.2}$$

Remark 5.1.5. It is instructive to recall Theorem 7.2.34. Let $L : \mathbb{R}^3 \to \mathbb{R}^3$ be defined by

$$L : \begin{bmatrix} a \\ b \\ c \end{bmatrix} \mapsto \begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix}.$$

Then the unit cube $[0,1]^3 = \left\{ \begin{bmatrix} a \\ b \\ c \end{bmatrix} \in \mathbb{R}^3 : a, b, c \in [0,1] \right\}$ is sent by the map L to the parallelepiped V defined by the three vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$. Then by Theorem 7.2.34, we conclude the volume of the parallelepiped V is

$$\operatorname{vol}(V) = \operatorname{vol}(L([0,1]^3)) = |\det(L)| \operatorname{vol}([0,1]^3) = \left| \det \begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{bmatrix} \right|.$$

This is exactly what we claimed at the beginning.

Example 5.1.6 ([1, §5.2 例 5])**.** Are $\mathbf{a} = (3, 0, 5)$, $\mathbf{b} = (1, 2, 3)$, $\mathbf{c} = (5, 4, 11)$ coplanar?

证明. Since

$$(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} = \det \begin{bmatrix} 3 & 0 & 5 \\ 1 & 2 & 3 \\ 5 & 4 & 11 \end{bmatrix} = 0,$$

we conclude that $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ are coplanar. □

One may also deduce the following identity:

Lemma 5.1.7. *For any* $\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^3$*, we have*

$$[\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})][\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})] = \det \begin{bmatrix} \mathbf{u} \cdot \mathbf{a} & \mathbf{u} \cdot \mathbf{b} & \mathbf{u} \cdot \mathbf{c} \\ \mathbf{v} \cdot \mathbf{a} & \mathbf{v} \cdot \mathbf{b} & \mathbf{v} \cdot \mathbf{c} \\ \mathbf{w} \cdot \mathbf{a} & \mathbf{w} \cdot \mathbf{b} & \mathbf{w} \cdot \mathbf{c} \end{bmatrix}.$$

证明. In terms of Cartesian coordinates, we write

$$\mathbf{u} = (x_1, y_1, z_1), \quad \mathbf{v} = (x_2, y_2, z_2), \quad \mathbf{w} = (x_3, y_3, z_3),$$
$$\mathbf{a} = (x_4, y_4, z_4), \quad \mathbf{b} = (x_5, y_5, z_5), \quad \mathbf{c} = (x_6, y_6, z_6).$$

Then by (5.1.2), we obtain

$$\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = \det \begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{bmatrix}, \quad \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \det \begin{bmatrix} x_4 & y_4 & z_4 \\ x_5 & y_5 & z_5 \\ x_6 & y_6 & z_6 \end{bmatrix}.$$

Then by Propositions 7.2.33 and 7.2.36, we have

$$\begin{aligned} [\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})][\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})] &= \det \begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{bmatrix} \cdot \det \begin{bmatrix} x_4 & y_4 & z_4 \\ x_5 & y_5 & z_5 \\ x_6 & y_6 & z_6 \end{bmatrix} \\ &= \det \begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{bmatrix} \begin{bmatrix} x_4 & x_5 & x_6 \\ y_4 & y_5 & y_6 \\ z_4 & z_5 & z_6 \end{bmatrix} \\ &= \det \begin{bmatrix} \mathbf{u} \cdot \mathbf{a} & \mathbf{u} \cdot \mathbf{b} & \mathbf{u} \cdot \mathbf{c} \\ \mathbf{v} \cdot \mathbf{a} & \mathbf{v} \cdot \mathbf{b} & \mathbf{v} \cdot \mathbf{c} \\ \mathbf{w} \cdot \mathbf{a} & \mathbf{w} \cdot \mathbf{b} & \mathbf{w} \cdot \mathbf{c} \end{bmatrix}. \end{aligned}$$

This establishes the identity. □

For vectors $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^3$, one may also study the *vector triple product* (向量三重积):

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}).$$

The vector triple product has the following expansion:

Lemma 5.1.8 (Lagrange)**.** *For* $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^3$*, we have*

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}. \tag{5.1.3}$$

证明. In terms of Cartesian coordinates, we write

$$\mathbf{a} = (x_1, y_1, z_1), \quad \mathbf{b} = (x_2, y_2, z_2), \quad \mathbf{c} = (x_3, y_3, z_3).$$

By (5.1.1), we have

$$\mathbf{b} \times \mathbf{c} = \left(\det \begin{bmatrix} y_2 & z_2 \\ y_3 & z_3 \end{bmatrix}, \det \begin{bmatrix} z_2 & x_2 \\ z_3 & x_3 \end{bmatrix}, \det \begin{bmatrix} x_2 & y_2 \\ x_3 & y_3 \end{bmatrix} \right).$$

Then the x component of $\mathbf{a} \times (\mathbf{b} \times \mathbf{c})$ is given by:

$$\begin{aligned}
[\mathbf{a} \times (\mathbf{b} \times \mathbf{c})]_x =& y_1 \cdot \det \begin{bmatrix} x_2 & y_2 \\ x_3 & y_3 \end{bmatrix} - z_1 \cdot \det \begin{bmatrix} z_2 & x_2 \\ z_3 & x_3 \end{bmatrix} \\
=& y_1(x_2 y_3 - x_3 y_2) - z_1(z_2 x_3 - x_2 z_3) \\
=& x_2(y_1 y_3 + z_1 z_3) - x_3(y_1 y_2 + z_1 z_2) \\
=& x_2(x_1 x_3 + y_1 y_3 + z_1 z_3) - x_3(x_1 x_2 + y_1 y_2 + z_1 z_2) \\
=& (\mathbf{a} \cdot \mathbf{c}) x_2 - (\mathbf{a} \cdot \mathbf{b}) x_3 = [(\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}]_x.
\end{aligned}$$

This establishes the identity. □

A direct consequence of (5.1.3) is the *Jacobi identity*:

$$(\mathbf{a} \times \mathbf{b}) \times \mathbf{c} + (\mathbf{b} \times \mathbf{c}) \times \mathbf{a} + (\mathbf{c} \times \mathbf{a}) \times \mathbf{b} = \mathbf{0}$$

for any $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^3$.

5.2 Lines and planes in $\mathbb{R}^3$

5.2.1 Planes

A *plane* is a 2-dimensional (Euclidean) space that extends indefinitely. In $\mathbb{R}^n$, a plane is determined by a rank-2 linear map $A : \mathbb{R}^2 \to \mathbb{R}^n$ (corresponding to an

$(n \times 2)$-matrix) and a translation $\mathbf{b} \in \mathbb{R}^n$:

$$\begin{bmatrix} u \\ v \end{bmatrix} \mapsto A \begin{bmatrix} u \\ v \end{bmatrix} + \mathbf{b}, \quad \begin{bmatrix} u \\ v \end{bmatrix} \in \mathbb{R}^2.$$

Thus, $A(\mathbb{R}^2) + \mathbf{b}$ is a plane passing through $\mathbf{b}$.

In $\mathbb{R}^3$, if $\mathbf{n} = (A, B, C)$ is the vector normal to a plane, and $P_0(x_0, y_0, z_0)$, $P(x, y, z)$ are two points in this plane, then we have

$$\mathbf{n} \cdot \overrightarrow{P_0P} = 0.$$

That is

$$A(x - x_0) + B(y - y_0) + C(z - z_0) = 0. \tag{5.2.1}$$

This is the *point-normal form* of the equation of a plane (点法式方程).

Clearly, (5.2.1) shows that the equation of a plane can also be written as

$$Ax + By + Cz + D = 0 \tag{5.2.2}$$

where at least one of A, B, C is nonzero, and (A, B, C) is the normal vector of this plane. This is the *general form* of the equation of a plane (一般方程).

Example 5.2.1 ([1, §5.3 例 3])**.** Suppose that the equation of a plane S is

$$3x + 4y + 6z = 1.$$

Find the point-normal form of S.

证明*.* Clearly, $(-3, 1, 1) \in S$ and $(3, 4, 6)$ is the normal vector of S. Then the point normal form of S is

$$3(x + 3) + 4(y - 1) + 6(z - 1) = 0.$$

□

Now suppose that there are three points $P_1(x_1, y_1, z_1), P_2(x_2, y_2, z_2), P_3(x_3, y_3, z_3)$ that are not collinear, i.e.

$$\overrightarrow{P_1P_2} \times \overrightarrow{P_1P_3} \neq \mathbf{0}.$$

Then they determine a plane. To find the equation, let $P(x,y,z) \in \mathbb{R}^3$ be a point in this plane. Since $\overrightarrow{P_1P_2} \times \overrightarrow{P_1P_3}$ is a normal vector, we have

$$\overrightarrow{P_1P} \cdot (\overrightarrow{P_1P_2} \times \overrightarrow{P_1P_3}) = 0.$$

This leads to the equation of the plane:

$$\det \begin{bmatrix} x - x_1 & y - y_1 & z - z_1 \\ x_2 - x_1 & y_2 - y_1 & z_2 - z_1 \\ x_3 - x_1 & y_3 - y_1 & z_3 - z_1 \end{bmatrix} = 0.$$

Example 5.2.2 ([1, §5.3 例 4])**.** Suppose that a plane S pass through $P_1(0,0,1)$, $P_2(1,1,0)$, $P_3(1,0,1)$. The equation of S is given by

$$\det \begin{bmatrix} x - 0 & y - 0 & z - 1 \\ 1 & 1 & -1 \\ 1 & 0 & 0 \end{bmatrix} = 0.$$

That is

$$y + z - 1 = 0.$$

When A, B, C of (5.2.2) are all nonzero, then the x, y, z-intercepts of the plane (5.2.2) are $-\frac{D}{A}$, $-\frac{D}{B}$, $-\frac{D}{C}$, respectively. If $D \neq 0$, then the equation of the plane can be rewritten as

$$\frac{x}{-\frac{D}{A}} + \frac{y}{-\frac{D}{B}} + \frac{z}{-\frac{D}{C}} = 1.$$

Now note that two planes are parallel if their normal vectors are parallel. In other words, if two planes are given by

$$A_1x + B_1y + C_1z + D_1 = 0,$$
$$A_2x + B_2y + C_2z + D_2 = 0,$$

then they are **parallel** if and only if (A_1, B_1, C_1) and (A_2, B_2, C_2) are parallel, i.e. there is a $\lambda \in \mathbb{R}$ such that

$$(A_1, B_1, C_1) = \lambda(A_2, B_2, C_2).$$

Moreover, the two planes **coincide** if and only if there is a $\lambda \in \mathbb{R}$ such that

$$(A_1, B_1, C_1, D_1) = \lambda(A_2, B_2, C_2, D_2).$$

Further, the angle θ between two planes are just the angle between the normal vectors (A_1, B_1, C_1) and (A_2, B_2, C_2), i.e.

$$\cos\theta = \frac{A_1A_2 + B_1B_2 + C_1C_2}{\sqrt{A_1^2 + B_1^2 + C_1^2}\sqrt{A_2^2 + B_2^2 + C_2^2}}.$$

Thus, two planes are **perpendicular** if and only if

$$A_1A_2 + B_1B_2 + C_1C_2 = 0.$$

Example 5.2.3 ([1, §5.3 例 8])**.** Determine l and k so that the plane S:

$$x + ly + kz = 1$$

and the plane $x + y - z = 8$ are perpendicular, and $(1, 1, -\frac{2}{3}) \in S$.

证明. Since two planes are perpendicular, we have

$$1 + l - k = 0.$$

Since $(1, 1, -\frac{2}{3}) \in S$, we have

$$1 + l - \frac{2}{3}k = 1.$$

Thus, we solve that $l = 2$ and $k = 3$. □

5.2.2 Lines

Similar to planes, a *line* is a 1-dimensional (Euclidean) space that extends indefinitely. In $\mathbb{R}^n$, a line is determined by a rank-1 linear map $A : \mathbb{R} \to \mathbb{R}^n$ (corresponding to an $(n \times 1)$-matrix) and a translation $\mathbf{b} \in \mathbb{R}^n$:

$$t \mapsto At + \mathbf{b}, \quad t \in \mathbb{R}. \tag{5.2.3}$$

Thus, $A(\mathbb{R}) + \mathbf{b}$ is a line passing through $\mathbf{b}$.

In $\mathbb{R}^3$, write

$$A = \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix}.$$

Then the point $P(x, y, z)$ in the line (5.2.3) can be rewritten as

$$\begin{cases} x - x_0 = ta \\ y - y_0 = tb \\ z - z_0 = tc \end{cases}, \quad t \in \mathbb{R} \tag{5.2.4}$$

where at least one of a, b, c is nonzero. This is the *parametric system* of a line. And (a, b, c) is called the *direction vector* (方向向量). If a, b, c are all nonzero, then (5.2.4) can be rewritten as

$$\frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c}.$$

This is the *standard form* of a line (标准方程).

A line can also be considered as the intersection of two planes:

$$\begin{cases} A_1x + B_1y + C_1z + D_1 = 0 \\ A_2x + B_2y + C_2z + D_2 = 0 \end{cases}$$

where the normal vectors (A_1, B_1, C_1) and (A_2, B_2, C_2) are not parallel (i.e. linearly independent). This is the *general form* of the equations of a line (一般方程). The direction vector of this line is normal to (A_1, B_1, C_1) and (A_2, B_2, C_2), and so can be obtained by

$$(A_1, B_1, C_1) \times (A_2, B_2, C_2) = \det \begin{bmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \end{bmatrix}.$$

Example 5.2.4 ([1, §5.3 例 11])**.** Let L be a line given by

$$\begin{cases} x - y + z - 1 = 0 \\ 2x + y - 3z = 0 \end{cases}.$$

Find the standard form of L.

证明. One easily checks that $(\frac{1}{3},-\frac{2}{3},0)\in L$. And

$$(1,-1,1)\times(2,1,-3)=\det\begin{bmatrix}\mathbf{i}&\mathbf{j}&\mathbf{k}\\1&-1&1\\2&1&-3\end{bmatrix}=(2,5,3).$$

Thus, the standard form of L is

$$\frac{x-\frac{1}{3}}{2}=\frac{y+\frac{2}{3}}{5}=\frac{z}{3}.$$

□

5.3 Quadric

In Section 5.2.1, we have seen that a linear equation with three variables determines a plane in $\mathbb{R}^3$. Now we discuss the surface determine by a quadratic equation with three variables:

$$Ax^2+By^2+Cz^2+2Dxy+2Eyz+2Fxz+Gx+Hy+Iz+J=0, \qquad (5.3.1)$$

where $A,B,\ldots,J\in\mathbb{R}$ and at least one of A,B,C,D,E,F is nonzero. A surface determined by the equation (5.3.1) is called a *quadric* or *quadric surface* (二次曲面). It is a generalization of *conic sections* (ellipses, parabolas, and hyperbolas).

More generally, in coordinates $x_1,x_2,\ldots,x_{d+1}$, a *general quadric* is a d-dimensional hypersurface in $\mathbb{R}^{d+1}$ defined by the algebraic equation

$$\sum_{i,j=1}^{d+1}x_iQ_{ij}x_j+\sum_{i=1}^{d+1}P_ix_i+R=0$$

which may be rewritten in vector and matrix notation as

$$\mathbf{x}Q\mathbf{x}^T+P\mathbf{x}^T+R=0 \qquad (5.3.2)$$

where $\mathbf{x}=(x_1,\ldots,x_{d+1}),P=(P_1,\ldots,P_{d+1})\in\mathbb{R}^{d+1}$ are row vectors, $Q=(Q_{ij})$ is a $(d+1)\times(d+1)$ real symmetric matrix, $R\in\mathbb{R}$ is a scalar constant. The term $\mathbf{x}Q\mathbf{x}^T$

is called a *quadratic form* (二次型), and $P\mathbf{x}^T$ is called a *linear form*. For instance, the equation (5.3.2) can be rewritten as

$$\begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} A & D & F \\ D & B & E \\ F & E & C \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} G & H & I \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + J = 0.$$

Then by the spectral theorem (Theorem 7.2.41), Q is orthogonally diagonalizable. In other words, there exists a suitable change of Cartesian coordinates such that Q becomes a diagonal matrix:

$$\mathbf{x} \begin{bmatrix} Q_{11} & 0 & \cdots & 0 \\ 0 & Q_{22} & \cdots & 0 \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & Q_{d+1,d+1} \end{bmatrix} \mathbf{x}^T + \begin{bmatrix} P_1 & P_2 & \cdots & P_{d+1} \end{bmatrix} \mathbf{x}^T + R = 0.$$

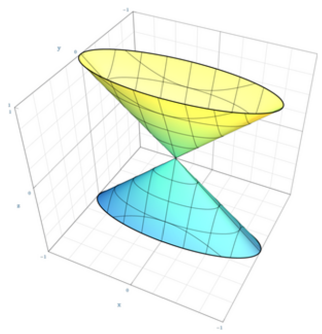
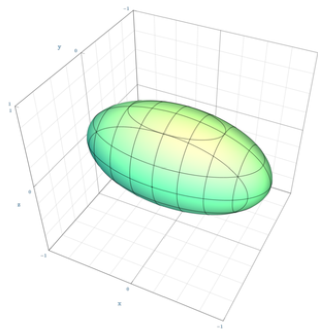
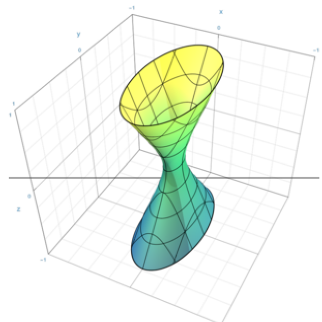
Together with a translation, we may further assume that $P_i = 0$ if $Q_{ii} \neq 0$. Also, the *inertia* of Q can easily be obtained by looking at the sign of its diagonal elements. The number n_+ of positive Q_{ii} is called the *positive index of inertia* of Q (正惯性指数), and the number n_- of negative Q_{ii} is called the *negative index of inertia* of Q (负惯性指数).

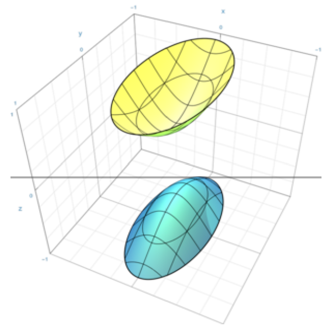
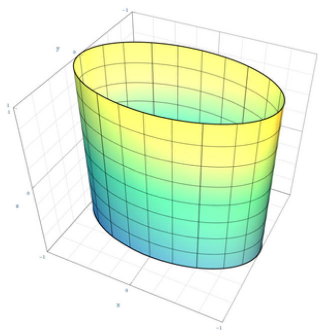
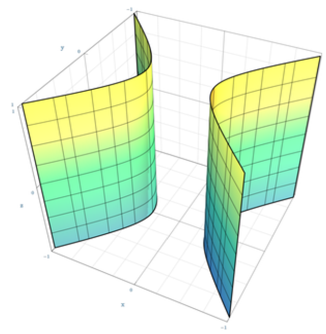
Under the above change of Cartesian coordinates, we are allowed to put the equation of the quadric into a unique simple form. This form is called the *normal form* (规范型) of the equation. The existence of the normal form of a quadric is also called the *principal axis theorem* (主轴定理).

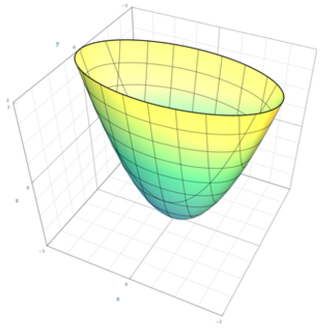
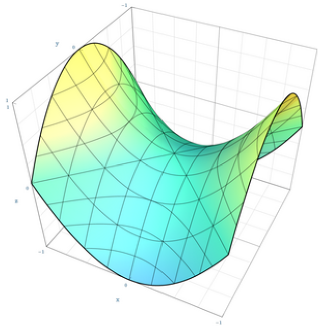
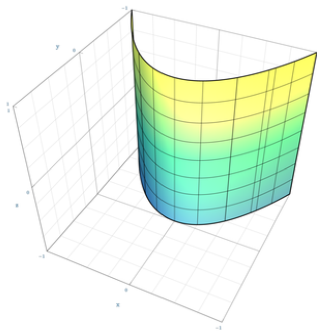
In $\mathbb{R}^3$, the normal forms of (5.3.1) are as follows:

$$\begin{aligned} &\frac{x^2}{a^2} + \frac{y^2}{b^2} + \varepsilon_1 \frac{z^2}{c^2} + \varepsilon_2 = 0, \\ &\frac{x^2}{a^2} - \frac{y^2}{b^2} + \varepsilon_3 = 0, \\ &\frac{x^2}{a^2} + \varepsilon_4 \frac{y^2}{b^2} - z = 0, \\ &\frac{x^2}{a^2} + \varepsilon_5 = 0, \end{aligned}$$

where $\varepsilon_1, \varepsilon_2, \varepsilon_4, \varepsilon_5 \in \{-1, 0, 1\}$ and $\varepsilon_3 \in \{0, 1\}$. Thus, there are 17 normal forms. 9 of them are real quadrics, and the 8 remaining are the imaginary quadrics and reducible quadrics (which are decomposed in two planes). We introduce the 9 real quadrics as follows.

Quadric	Equation	Graph
椭圆锥面 *Elliptic cone* or *Conical quadric*	$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=0,$	
椭球面 *Ellipsoid*	$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1.$ Parametrization $\begin{cases} x=a\sin\varphi\cos\theta \\ y=b\sin\varphi\sin\theta \\ z=c\cos\theta \end{cases},$ $\varphi\in[0,\pi]$, $\theta\in[0,2\pi)$.	
单叶双曲面 *Hyperboloid of one sheet* or *Hyperbolic hyperboloid*	$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=1.$	

Quadric	Equation	Graph
双叶双曲面 *Hyperboloid of two sheet* or *Elliptic hyperboloid*	$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=-1.$	
椭圆柱面 *Elliptic cylinder*	$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1.$	
双曲柱面 *Hyperbolic cylinder*	$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1.$	

Quadric	Equation	Graph
椭圆抛物面 *Elliptic paraboloid*	$\frac{x^2}{a^2} + \frac{y^2}{b^2} - z = 0.$	
双曲抛物面/马鞍面 *Hyperbolic paraboloid* or *Saddle*	$\frac{x^2}{a^2} - \frac{y^2}{b^2} - z = 0.$ Remark: The Gaussian curvature at saddle points is negative.	
抛物柱面 *Parabolic cylinder*	$x^2 + 2ay = 0.$	

Example 5.3.1. Find the normal form of

$$2x^2 + 3y^2 - z^2 + 4x - 12y + 5 = 0. \tag{5.3.3}$$

证明. The equation (5.3.3) is equivalent to

$$2(x+1)^2 + 3(y-2)^2 - z^2 = 9.$$

Letting $\tilde{x} = x + 1$, $\tilde{y} = y - 2$ and $\tilde{z} = z$, we find the normal form

$$\frac{\tilde{x}^2}{a^2} + \frac{\tilde{y}^2}{b^2} - \frac{\tilde{z}^2}{c^2} = 1$$

where $a = \frac{3}{\sqrt{2}}$, $b = \sqrt{3}$, $c = 3$. This is a hyperboloid of one sheet. □

5.4 Vector calculus

Let $\Gamma : [A, B] \to \mathbb{R}^3$ be a *smooth curve*. In other words, we have the parametric system:

$$\Gamma(t) = \begin{cases} x = x(t) \\ y = y(t) \\ z = z(t) \end{cases}, \quad t \in [A, B]$$

so that $x(t), y(t), z(t)$ have continuous derivatives, and that at least one of $x'(t)$, $y'(t)$, $z'(t)$ is nonzero for all $t \in [A, B]$.

Then for any $t \in [A, B]$, we have the tangent vector

$$\begin{aligned} \lim_{\Delta t \to 0} \frac{\Gamma(t + \Delta t) - \Gamma(t)}{\Delta t} &= \lim_{\Delta t \to 0} \frac{x(t + \Delta t) - x(t)}{\Delta t}\mathbf{i} + \lim_{\Delta t \to 0} \frac{y(t + \Delta t) - y(t)}{\Delta t}\mathbf{j} \\ &\quad + \lim_{\Delta t \to 0} \frac{z(t + \Delta t) - z(t)}{\Delta t}\mathbf{k} \\ &= x'(t)\mathbf{i} + y'(t)\mathbf{j} + z'(t)\mathbf{k}. \end{aligned}$$

Since Γ is smooth, the vector $\Gamma'(t) = (x'(t), y'(t), z'(t))$ is nonzero for all $t \in [A, B]$. The vector $\Gamma'(t)$ is called the *tangent vector* of Γ at $\Gamma(t)$. Then the equation of tangent line is given by

$$T(u) = \Gamma(t_0) + \Gamma'(t_0)u, \quad u \in \mathbb{R}.$$

In terms of Cartesian coordinates, the tangent line at $\Gamma(t_0)$ is given by

$$T(u) = \begin{cases} x = x(t_0) + x'(t_0)u \\ y = y(t_0) + y'(t_0)u \\ z = z(t_0) + z'(t_0)u \end{cases}, \quad u \in \mathbb{R}.$$

Or in the standard form, we have

$$\frac{x - x(t_0)}{x'(t_0)} = \frac{y - y(t_0)}{y'(t_0)} = \frac{z - z(t_0)}{z'(t_0)}.$$

The plane that is orthogonal to the tangent line T is called the *normal plane* of Γ at $\Gamma(t_0)$. The equation of the normal plane is given by

$$x'(t_0)(x - x(t_0)) + y'(t_0)(y - y(t_0)) + z'(t_0)(z - z(t_0)) = 0.$$

Example 5.4.1 ([1, §5.5 例 1])**.** Find the tangent line and the normal plane at $(0, a, \frac{b\pi}{2})$ of the curve

$$\begin{cases} x = a\cos t \\ y = a\sin t \\ z = bt \end{cases}, \quad a > 0,\ b > 0,\ t \in [0, 2\pi].$$

证明. The point $(0, a, \frac{b\pi}{2})$ corresponds to $t = \frac{\pi}{2}$. Then the tangent vector is $(-a, 0, b)$, and the tangent line is

$$\begin{cases} x - 0 = -au \\ y - a = 0 \\ z - \frac{b\pi}{2} = bu \end{cases}, \quad u \in \mathbb{R}.$$

The normal plane is

$$-a(x - 0) + b\left(z - \frac{b\pi}{2}\right) = 0.$$

□

We can discuss the arc length of a smooth curve $\Gamma : [A, B] \to \mathbb{R}^3$. In fact, the arc length s is approximated by the length of piecewise line segments:

$$s = \lim_{\lambda \to 0} \sum_{i=1}^{m} |\Gamma(t_i) - \Gamma(t_{i-1})|$$

where $A = t_0 < t_1 < \cdots < t_{m-1} < t_m = B$ is any partition of $[A, B]$, $\lambda = \{t_i - t_{i-1} : i = 1, 2, \ldots, m\}$. Note that

$$\begin{aligned}|\Gamma(t_i) - \Gamma(t_{i-1})| &= \sqrt{(x(t_i) - x(t_{i-1}))^2 + (y(t_i) - y(t_{i-1}))^2 + (z(t_i) - z(t_{i-1}))^2} \\ &= \sqrt{(x'(t_{i-1}))^2 + (y'(t_{i-1}))^2 + (z'(t_{i-1}))^2}(t_i - t_{i-1}) + o(t_i - t_{i-1}) \\ &= |\Gamma'(t_{i-1})|\Delta t_i + o(\Delta t_i).\end{aligned}$$

Thus, we have

$$\sum_{i=1}^{m} |\Gamma(t_i) - \Gamma(t_{i-1})| = \sum_{i=1}^{m} |\Gamma'(t_{i-1})|\Delta t_i + o(B - A).$$

Moreover, one may check that the arc length s is an additive function on $[A, B]$. Also, for any $[c, d] \subset [A, B]$, the arc length $s[c, d]$ is dominated by the maximum and minimum speeds within the same time period:

$$\inf_{t\in[c,d]} |\Gamma'(t)|(d - c) \leq s[c, d] \leq \sup_{t\in[c,d]} |\Gamma'(t)|(d - c).$$

Therefore, by Proposition 3.5.3, we conclude

$$s = \int_A^B |\Gamma'(t)|dt = \int_A^B \sqrt{(x'(t))^2 + (y'(t))^2 + (z'(t))^2}dt.$$

In terms of differentials, we have

$$ds = \sqrt{(x'(t))^2 + (y'(t))^2 + (z'(t))^2}dt.$$

A similar result holds for $\Gamma : [A, B] \to \mathbb{R}^n$ for any integer $n \geq 2$.

Example 5.4.2 ([1, §5.5 例 2])**.** Find the arc length s of the curve

$$\begin{cases} x = a\cos t \\ y = a \sin t \\ z = bt \end{cases}, \quad a > 0,\ b > 0,\ t \in [0, \pi].$$

证明. By the formula, we have

$$s = \int_0^\pi \sqrt{a^2 \sin^2 t + a^2 \cos^2 t + b^2}dt = \pi\sqrt{a^2 + b^2}.$$

□

Chapter 6

Differential Calculus in Several Variables

6.1 Function of several real variables

Definition 6.1.1 ([1, §6.1 定义 1]). For any nonempty subset $D \subset \mathbb{R}^n$, a map $f : D \to \mathbb{R}$ is called a (real-valued) *function of* n (real) *variables* (多元函数).

Usually, if a function is defined by a formula, such as $z = f(x, y)$, then we say the *domain* of f to be the set of points so that the formula makes sense.

For a function of two variables $f(x, y)$, the set

$$\{(x, y, z) \in \mathbb{R}^3 : z = f(x, y),\ (x, y) \in D\}$$

is called the *graph* of f.

Example 6.1.2 ([1, §6.1 例 1]). The domain of the function $z = \sqrt{r^2 - (x^2 + y^2)}$ is

$$D = \{(x, y) : x^2 + y^2 \leq r^2\}.$$

Its graph is a hemisphere of radius r centered at the origin O.

In this section, we shall also study the differentiation of maps of the form $f : \mathbb{R}^n \to \mathbb{R}^m$, i.e. maps from one Euclidean space to another. For a subset $D \subset \mathbb{R}^n$, let $f : D \to \mathbb{R}^m$ be a map. Then f maps any point $(x_1, \ldots, x_n) \in D$ to a unique point $(y_1, \ldots, y_m) \in \mathbb{R}^m$, where $y_i = f_i(x_1, \ldots, x_n)$ $(i = 1, \ldots, m)$ is determined by $(x_1, \ldots, x_n)$. It follows that

$$f : \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \mapsto \begin{bmatrix} f_1(x_1, \ldots, x_n) \\ \vdots \\ f_m(x_1, \ldots, x_n) \end{bmatrix}.$$

Here we write the elements in Euclidean spaces as column vectors.

To study the differentiation of functions on $\mathbb{R}^n$, we analyze the metric structure of Euclidean spaces:

Definition 6.1.3 (Metric space)**.** Let X be a set. A function $d : X \times X \to [0, +\infty)$ is called a *metric* (度量) or *distance* (距离) if it satisfies the following conditions:

(1) $d(P, Q) = 0$ if and only if $P = Q$.

(2) $d(P, Q) = d(Q, P)$ for any $P, Q \in X$.

(3) $d(P, Q) \leq d(P, R) + d(R, Q)$ for any $P, R, Q \in X$.

A set X equipped with a metric d, written (X, d), is called a *metric space* (度量空间).

Euclidean space $\mathbb{R}^n$ is a metric space, with the Euclidean metric

$$d(P, Q) = |P - Q| = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2}$$

where $P(p_1, \ldots, p_n), Q(q_1, \ldots, q_n) \in \mathbb{R}^n$. In general, any normed vector space $(V, \|\cdot\|)$ induces a metric, and so is a metric space. Thus, other norms on $\mathbb{R}^n$ also induce metrics (Section 5.1.1). For instance, l^1 norm induces the metric

$$d_1(P, Q) = \|P - Q\|_1 = \sum_{i=1}^{n} |p_i - q_i|$$

for $P(p_1, \ldots, p_n), Q(q_1, \ldots, q_n) \in \mathbb{R}^n$. Moreover, since norms on $\mathbb{R}^n$ are equivalent (Lemma 5.1.1), we have

$$n^{-1} \sum_{i=1}^{n} |p_i - q_i| \leq \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2} \leq n \sum_{i=1}^{n} |p_i - q_i|$$

for $P(p_1, \dots, p_n), Q(q_1, \dots, q_n) \in \mathbb{R}^n$.

In the following, we mainly study the Euclidean metric unless otherwise specified.

Definition 6.1.4 ([1, §6.1 定义 2])**.** Let (X, d) be a metric space, $P \in X$ be a point, $r > 0$ a number. The set

$$U_r(P) \coloneqq \{Q \in X : d(P, Q) < r\}$$

is called the r-*neighborhood* of P, or the (open) *ball* of radius r centered at P. The *deleted neighborhood* is then defined by $\mathring{U}_r(P) \coloneqq U_r(P) \setminus \{P\}$.

The shape of balls $U_r(P)$ depends on the choice of the metric d.

Example 6.1.5. Let d be a metric on $\mathbb{R}^n$, $U_r(P)$ be a ball in $\mathbb{R}^n$.

- If $d = d_2$ is induced by the Euclidean norm $\|\cdot\|_2$, then the ball $U_r(P) = U_r^{(2)}(P)$ is the n-dimensional **round ball** of radius r centered at P.
- If $d = d_1$ is induced by the l_1 norm $\|\cdot\|_1$, then the ball $U_r(P) = U_r^{(1)}(P)$ is an inscribed n-dimensional **regular polyhedron** of centered at P, inscribed in an n-dimensional round ball of radius r.
- If $d = d_\infty$ is induced by the maximum norm $\|\cdot\|_\infty$, then the ball $U_r(P) = U_r^{(\infty)}(P)$ is an n-dimensional **cube** of centered at P, circumscribed around an n-dimensional round ball of radius r.

Since any pair of norms in $\mathbb{R}^n$ are equivalent (Lemma 5.1.1), we obtain a nested sequence of balls induced by different metrics/norms:

$$U_r^{(\infty)}(P) \subset U_{\sqrt{n}r}^{(2)}(P) \subset U_{nr}^{(1)}(P) \subset U_{nr}^{(\infty)}(P)$$

for any $r > 0$. So, when we discuss limit-related problems, these containment relations of the balls mean that choosing different balls does not affect the limitation.

Definition 6.1.6 (Interior, exterior and boundary)**.** Let (X, d) be a metric space, $S \subset X$ a subset, $P \in X$ a point.

- P is said to be an *interior point* (内点) of S if there is some $r > 0$ such that

$$U_r(P) \subset S.$$

  The set $\operatorname{int} S$ (also written S°) of interior points of S is called the *interior* (内部) of S.

- P is said to be an *exterior point* (外点) of S if there is some $r > 0$ such that

$$U_r(P) \subset X \setminus S.$$

  The set $\operatorname{ext} S$ of exterior points of S is called the *exterior* (外部) of S.

- P is said to be a *boundary point* (边界点) of S if it is neither an interior nor exterior point of S. The set ∂S of boundary points of S is called the *boundary* (边界) of S.

Definition 6.1.7 (Open and closed sets)**.** Let (X, d) be a metric space, $S \subset X$ a subset.

- S is called *open* (开集) if each point $x \in S$ is an interior point, i.e. $\operatorname{int} S = S$.
- S is called *closed* (闭集) if the complement $S^c = X \setminus S$ is open.

Example 6.1.8. In $\mathbb{R}^n$, the only sets that are both open and closed are $\mathbb{R}^n$ and $\emptyset$.

Example 6.1.9. For $a, b \in \mathbb{R}$ with $a < b$,

- (a, b) is open;
- $[a, b]$ is closed;
- $[a, b)$ is neither open nor closed.

Proposition 6.1.10 (Open sets)**.** *Let (X, d) be a metric space.*

(1) For any open set $S \subset X$, we have $\partial S \cap S = \emptyset$.

(2) For a family of open sets $S_\alpha \subset X$ $(\alpha \in A)$, the union $\bigcup_{\alpha \in A} S_\alpha$ is open.

(3) For finitely many open sets $S_1, \ldots, S_n \subset X$, the intersection $S_1 \cap \cdots \cap S_n$ is open.

Proposition 6.1.11 (Closed sets)**.** *Let (X, d) be a metric space.*

(1) For any closed set $S \subset X$, we have $\partial S \subset S$.

(2) For a family of closed sets $S_\alpha \subset X$ $(\alpha \in A)$, the intersection $\bigcap_{\alpha \in A} S_\alpha$ is closed.

(3) For finitely many open sets $S_1, \ldots, S_n \subset X$, the union $S_1 \cup \cdots \cup S_n$ is closed.

Definition 6.1.12 (Limit points and closure)**.** Let (X, d) be a metric space, $S \subset X$ a subset. A point $x \in X$ is called a *limit point* (聚点) of S if $U_r(x) \cap S$ is an infinite set for any $r > 0$. The union of S and the set of all its limit points is called the *closure* (闭包) of S, written $\overline{S}$.

Proposition 6.1.13. *Let (X, d) be a metric space, $S \subset X$ a subset. Then S is closed if and only if S contains all its limit points, i.e. $S = \overline{S}$.*

Definition 6.1.14 (Bounded sets)**.** Let (X, d) be a metric space, $S \subset X$ a subset. Then S is called *bounded* (有界) if there exists a point $x \in X$ and $r > 0$ such that $S \subset U_r(x)$. Otherwise, S is called *unbounded* (无界).

We next discuss the *connectedness* (连通性) of the Euclidean spaces.

Definition 6.1.15 (Connectedness)**.** Let $S \subset \mathbb{R}^n$ be a subset. S is said to be *connected*[1] if for any two points $x, y \in S$, there is a (continuous) curve $\Gamma : [A, B] \to S$ such that $\Gamma(A) = x$ and $\Gamma(B) = y$.

Definition 6.1.16. An open connected subset of $\mathbb{R}^n$ is called a *domain* or *region* (区域). The closure of a domain is called a *closed domain* or *closed region* (闭区域).

Example 6.1.17.

- $\mathbb{R}^n$ itself is a region.

[1] In the theory of topology, it is the notion of *path connectedness* (道路连通性).

- The ball $B(x,r) = \{y \in \mathbb{R}^n : |y - x| < r\}$ is a region for any $r > 0$, its closure is a closed region.
- $\{(x,y) \in \mathbb{R}^2 : x^2 + y^2 < 1 \text{ or } x^2 + y^2 > 2\}$ is open but not connected, so is not a region.

Remark 6.1.18. Open regions behave similarly to open intervals in $\mathbb{R}$; Bounded closed regions behave similarly to bounded closed intervals in $\mathbb{R}$.

6.2 Limits of functions of several variables

6.2.1 Definitions

Definition 6.2.1 (Limit)**.** Let (X, d_X), (Y, d_Y) be two metric spaces. Let $P \in X$, $\mathring{U}_r(P) \subset X$ be a deleted neighborhood of X, $y \in Y$ and $f : \mathring{U}_r(P) \to Y$ a map. Suppose that for any $\epsilon > 0$, there is $\delta = \delta(\epsilon) > 0$ such that

$$d_Y(f(x), y) < \epsilon$$

for any x with $d_X(P, x) < \delta$ (i.e. $x \in \mathring{U}_\delta(P)$). Then we say that $f(x)$ has the *limit* y as $x \to P$, and write

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

We also say that $f(x)$ tends to y, as $x \to P$, and write

$$f(x) \to y \quad (x \to P).$$

For a given f, if there is y so that $\lim_{x \to P} f(x) = y$, then we say f has a limit (is convergent) as $x \to P$ (at P).

Definition 6.2.2 (Limits on Euclidean spaces, [1, §6.2 定义 1&2])**.** Let $(X, d_X) = (\mathbb{R}^n, \|\cdot\|_X)$, $(Y, d_Y) = (\mathbb{R}^m, \|\cdot\|_Y)$. Let $P \in \mathbb{R}^n$, $\mathring{U}_r(P) \subset \mathbb{R}^n$ be a deleted neighborhood of $\mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^m$ and $f : \mathring{U}_r(P) \to \mathbb{R}^m$ a map. Then the limit

$$\lim_{\mathbf{x} \to P} f(\mathbf{x}) = \mathbf{y}$$

can be interpreted as follows:

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

$$\|f(\mathbf{x}) - \mathbf{y}\|_Y < \epsilon$$

for any

$$0 < \|\mathbf{x} - P\|_X < \delta. \tag{6.2.1}$$

Note that the choice of the norms on $\mathbb{R}^n$ does **not** affect the limit, since all norms in $\mathbb{R}^n$ are equivalent (Lemma 5.1.1).

- If $\|\cdot\| = \|\cdot\|_2$ is the Euclidean norm, $\mathbf{x} = (x_1, \ldots, x_n), P = (P_1, \ldots, P_n) \in \mathbb{R}^n$, then (6.2.1) means

$$0 < \sqrt{(x_1 - P_1)^2 + \cdots + (x_n - P_n)^2} < \delta.$$

- If $\|\cdot\| = \|\cdot\|_1$ is the l^1 norm, $\mathbf{x} = (x_1, \ldots, x_n), P = (P_1, \ldots, P_n) \in \mathbb{R}^n$, then (6.2.1) means

$$0 < |x_1 - P_1| + \cdots + |x_n - P_n| < \delta.$$

- If $\|\cdot\| = \|\cdot\|_\infty$ is the maximum norm, $\mathbf{x} = (x_1, \ldots, x_n), P = (P_1, \ldots, P_n) \in \mathbb{R}^n$, then (6.2.1) means

$$0 < \max_i |x_i - P_i| < \delta.$$

Remark 6.2.3. Informally, $x \to P$ means x can approach P along any different path. In other words, if there are two paths so that x approaching P along these paths have different limits, then f is not convergent at P. More precisely, let $\Gamma_1 : [0, T] \to D$, $\Gamma_2 : [0, T] \to D$ be different curves on D with $\Gamma_1(T) = \Gamma_2(T) = P$. Suppose that

$$\lim_{t \to T} f(\Gamma_1(t)) \neq \lim_{t \to T} f(\Gamma_2(t)).$$

Then $\lim\limits_{x \to P} f(x)$ does not exist.

Example 6.2.4 ([1, §6.2 例 1])**.** Let $f(x, y) = x^2 + \sin y$. Prove that $f(x, y) \to 9$ as $(x, y) \to (3, 0)$.

证明. Note that

$$|f(x,y)-9| \le |x^2-9| + |\sin y| \le |x+3||x-3| + |y|.$$

Also, for $|x-3|<1$, $|x+3||x-3| < 7|x-3|$.

Thus, for any $\epsilon > 0$, let $\delta = \min\{1, \frac{\epsilon}{8}\}$. Then we have

$$|f(x,y)-9| \le 7|x-3| + |y| < \epsilon$$

for $\max\{|x-3|, |y|\} < \delta$. □

Example 6.2.5 ([1, §6.2 例 2])**.** Consider the function

$$f(x,y) = \frac{x \sin y}{\sqrt{x^2+y^2}}.$$

Prove that $f(x,y) \to 0$ as $(x,y) \to (0,0)$.

证明. Note that

$$|x \sin y| \le |xy| \le \frac{1}{2}(x^2+y^2).$$

It follows that

$$|f(x,y)| \le \frac{1}{2}\sqrt{x^2+y^2}.$$

Thus, for any $\epsilon > 0$, let $\delta = 2\epsilon$. Then we have

$$|f(x,y) - 0| < \epsilon$$

for $0 < \sqrt{x^2+y^2} < \delta$. □

Example 6.2.6 ([1, §6.2 例 3])**.** Determine whether the following limit exists or not:

$$\lim_{(x,y)\to(0,0)} \frac{|x|}{\sqrt{x^2+y^2}}. \tag{6.2.2}$$

证明. Let $y = kx$ for some $k \in \mathbb{R}$. Then $(x,y) = (x,kx) \to (0,0)$ as $x \to 0$. Then

$$\frac{|x|}{\sqrt{x^2+y^2}} = \frac{|x|}{\sqrt{x^2+k^2x^2}} = \frac{1}{\sqrt{1+k^2}}.$$

It follows that for different k, the limits are different. We conclude that (6.2.2) does not exist. □

6.2.2 Theorems of limits of functions

The 1-dimensional versions of the following theorems have already been studied in Section 1.4. One may easily obtain the proofs after replacing the absolute value $|\cdot|$ on $\mathbb{R}$ by the norms $\|\cdot\|$ on $\mathbb{R}^n$.

Theorem 6.2.7 (Algebraic limit theorem, [1, §6.2 定理 1])**.** *Let* (X,d) *be a metric space,* $\mathring{U}_r(P)\subset X$ *be a deleted neighborhood. Let* $f,g:\mathring{U}_r(P)\to\mathbb{R}$ *be functions. If*

$$\lim_{\mathbf{x}\to P} f(\mathbf{x}) = y_1, \quad \lim_{\mathbf{x}\to P} g(\mathbf{x}) = y_2,$$

then

$$\lim_{x\to P}(f(\mathbf{x})\pm g(\mathbf{x})) = y_1\pm y_2, \quad \lim_{\mathbf{x}\to P} f(\mathbf{x})g(\mathbf{x}) = y_1y_2,$$

and for $y_2\neq 0$*, we have*

$$\lim_{\mathbf{x}\to P}\frac{f(\mathbf{x})}{g(\mathbf{x})} = \frac{y_1}{y_2}.$$

Theorem 6.2.8 ([1, §6.2 定理 2])**.** *Let* (X,d) *be a metric space,* $\mathring{U}_r(P)\subset X$ *be a deleted neighborhood. Let* $f,g:\mathring{U}_r(P)\to\mathbb{R}$ *be functions so that*

$$f(\mathbf{x})\geq g(\mathbf{x})$$

for $\mathbf{x}\in\mathring{U}_r(P)$*. Suppose that as* $\mathbf{x}\to P$*,* $f(\mathbf{x})$ *and* $g(\mathbf{x})$ *are both convergent. Then*

$$\lim_{\mathbf{x}\to P} f(\mathbf{x})\geq\lim_{\mathbf{x}\to P} g(\mathbf{x}).$$

Theorem 6.2.9 (Squeeze theorem, [1, §6.2 定理 3])**.** *Let* (X,d) *be a metric space,* $\mathring{U}_r(P)\subset X$ *be a deleted neighborhood. Let* $f,g,h:\mathring{U}_r(P)\to\mathbb{R}$ *be functions such that*

$$h(\mathbf{x})\leq f(\mathbf{x})\leq g(\mathbf{x})$$

for $\mathbf{x}\in\mathring{U}_r(P)$*. If* $\lim_{\mathbf{x}\to P} g(\mathbf{x}) = \lim_{\mathbf{x}\to P} h(\mathbf{x}) = y$*, then* $\lim_{\mathbf{x}\to P} f(\mathbf{x}) = y$*.*

Theorem 6.2.10. *Let* (X,d) *be a metric space,* $\mathring{U}_r(P)\subset X$ *be a deleted neighborhood. Let* $f,g:\mathring{U}_r(P)\to\mathbb{R}$ *be functions. Suppose that*

$$\lim_{\mathbf{x}\to P} f(\mathbf{x}) = l_1, \quad \lim_{\mathbf{x}\to P} g(\mathbf{x}) = l_2.$$

Also, $l_1 > l_2$. Then there is $\delta > 0$ such that

$$f(\mathbf{x}) > g(\mathbf{x})$$

for $\mathbf{x} \in \mathring{U}_\delta(P)$.

Theorem 6.2.11 (Composition, [1, §6.2 定理 4&5])**.** *Let (X, d_X), (Y, d_Y), (Z, d_Z) be metric spaces. Let $\mathring{U}_r(P) \subset X$, $\mathring{U}_s(Q) \subset Y$ be deleted neighborhoods. Let $g : \mathring{U}_s(Q) \to Z$ be a map such that the following limit exists:*

$$\lim_{\mathbf{y}\to Q} g(\mathbf{y}).$$

Let $f : \mathring{U}_r(P) \to \mathring{U}_s(Q)$ be a map such that

$$\lim_{\mathbf{x}\to P} f(\mathbf{x}) = Q.$$

Then

$$\lim_{\mathbf{x}\to P} g(f(\mathbf{x})) = \lim_{\mathbf{y}\to Q} g(\mathbf{y}).$$

Example 6.2.12 ([1, §6.2 例 6])**.** Prove that

$$\lim_{(x,y)\to(0,0)} (1 + x^2 + y^2)^{\frac{1}{x^2+y^2}} = e.$$

证明. Let $u(x, y) = x^2 + y^2$. Then

$$(1 + x^2 + y^2)^{\frac{1}{x^2+y^2}} = (1 + u)^{\frac{1}{u}} =: g(u).$$

Then by Theorem 6.2.11, we have

$$\lim_{(x,y)\to(0,0)} (1 + x^2 + y^2)^{\frac{1}{x^2+y^2}} = \lim_{(x,y)\to(0,0)} g(u(x, y)) = \lim_{u\to 0} g(u) = \lim_{u\to 0} (1 + u)^{\frac{1}{u}} = e.$$

□

Example 6.2.13 ([1, §6.2 例 7])**.** Find the limit

$$I = \lim_{(u,v)\to(0,0)} \left(\frac{\sin 2(u^2 + v^2)}{u^2 + v^2}\right)^{\frac{u \sin v}{\sqrt{u^2+v^2}}}.$$

证明. Let

$$x(u,v) = \frac{\sin 2(u^2+v^2)}{u^2+v^2}, \quad y(u,v) = \frac{u\sin v}{\sqrt{u^2+v^2}}.$$

Then by Theorem 6.2.11 and Example 6.2.5, we have

$$\lim_{(u,v)\to(0,0)} x(u,v) \overset{t=u^2+v^2}{=\!=\!=\!=} \lim_{t\to 0} \frac{\sin 2t}{t} = 2, \qquad \lim_{(u,v)\to(0,0)} y(u,v) = 0.$$

Then by Theorems 6.2.11 and 6.2.7, we have

$$I = \lim_{(u,v)\to(0,0)} x(u,v)^{y(u,v)} = \lim_{(u,v)\to(0,0)} e^{y(u,v)\ln x(u,v)} = e^{0\ln 2} = 1.$$

□

6.2.3 Iterated limits

In multivariable calculus, an *iterated limit* (累次极限) is a limit of a sequence or a limit of a function in the form

$$\lim_{y\to b}\lim_{x\to a} f(x,y),$$

or other similar forms. To differentiate, we refer to the limit discussed in Example 6.2.2 as the *ordinary limit* (全面极限).

One should not think that the ordinary limit of a function of several variables can be found by computing the iterated limit.

Example 6.2.14 ([1, §6.2 例 8])**.** For $\mathbf{0} \neq (x,y) \in \mathbb{R}^2$, define the function

$$f(x,y) = \frac{xy}{x^2+y^2}.$$

Then the iterated limits

$$\lim_{y\to 0}\lim_{x\to 0} f(x,y) = \lim_{x\to 0}\lim_{y\to 0} f(x,y) = 0.$$

However, the function has no ordinary limit as $(x,y) \to (0,0)$.

Example 6.2.15. For $(x,y) \in \mathbb{R}^2$, define the function

$$f(x,y) = \begin{cases} y\sin\frac{1}{x} & , \text{ if } x \neq 0 \\ 0 & , \text{ if } x = 0 \end{cases}$$

Then we have

$$\lim_{(x,y)\to(0,0)} f(x,y) = 0, \quad \lim_{x\to 0}\lim_{y\to 0} f(x,y) = 0.$$

However, the limit $\lim_{y\to 0}\lim_{x\to 0} f(x,y)$ does not exist.

Theorem 6.2.16. *Let $D \subset \mathbb{R}^2$ be a region, $(a,b) \in D$, and $f : D \to \mathbb{R}$ a function. Suppose that*

- *the ordinary limit $\lim_{(x,y)\to(a,b)} f(x,y)$ exists,*
- *the limit $\lim_{x\to a} f(x,y)$ exists for each y near b,*
- *the limit $\lim_{y\to b} f(x,y)$ exists for each x near a.*

Then we have

$$\lim_{x\to a}\lim_{y\to b} f(x,y) = \lim_{y\to b}\lim_{x\to a} f(x,y) = \lim_{(x,y)\to(a,b)} f(x,y).$$

6.3 Continuity of a function of several variables

Definition 6.3.1 (Continuous functions)**.** Let (X, d_X), (Y, d_Y) be metric spaces, $S \subset X$ a subset, $f : S \to Y$ a function. Then we say the function f is *continuous* at $x_0 \in S$ if for any $\epsilon > 0$, there exists $\delta = \delta(\epsilon) > 0$ such that

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

for any $x \in S$ with $d_X(x, x_0) < \delta$. If f is continuous at any point of S, then we say f is *continuous* on S, or f is a *continuous function* on S.

Definition 6.3.2 (Continuous functions on Euclidean spaces, [1, §6.3 定义 1&2])**.** Let $(X, d_X) = (\mathbb{R}^n, \|\cdot\|_X)$, $(Y, d_Y) = (\mathbb{R}^m, \|\cdot\|_Y)$. Let $S \subset \mathbb{R}^n$ be a subset, $f : S \to \mathbb{R}^m$ a function. Then the function f is *continuous* at $\mathbf{x}_0 \in S$ if for any $\epsilon > 0$, there exists $\delta = \delta(\epsilon) > 0$ such that

$$\|f(\mathbf{x}) - f(\mathbf{x}_0)\|_Y < \epsilon$$

for any $\mathbf{x} \in S$ with $\|\mathbf{x} - \mathbf{x}_0\|_X < \delta$.

We mainly focus on the cases when $S = D$ is a region or $S = \overline{D}$ a closed region.

- When $S = D$ is a region, f is *continuous* at $x_0 \in D$ means
$$\lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x}) = f(\mathbf{x}_0).$$
- When $S = \overline{D}$ is a closed region, f is *continuous* at $x_0 \in \overline{D}$ means
$$\lim_{\substack{\mathbf{x}\to\mathbf{x}_0\\ \mathbf{x}\in\overline{D}}} f(\mathbf{x}) = f(\mathbf{x}_0).$$

Example 6.3.3 (Projection)**.** For $1 \le i \le n$, the projection $\pi_i : \mathbb{R}^n \to \mathbb{R}$ defined by
$$\pi_i : (x_1, \ldots, x_i, \ldots, x_n) \mapsto x_i$$
is continuous on $\mathbb{R}^n$. In fact, for any $\mathbf{x}_0 \in \mathbb{R}^n$, any $\epsilon > 0$, let $\delta = \epsilon > 0$. Then we have
$$|\pi_i(\mathbf{x}) - \pi_i(\mathbf{x}_0)| < \epsilon$$
for any $\|\mathbf{x} - \mathbf{x}_0\|_\infty = \max_i |\pi_i(\mathbf{x}) - \pi_i(\mathbf{x}_0)| < \delta$.

Similar to 1-dimensional continuous functions, we have:

Theorem 6.3.4 ([1, §6.3 定理 1])**.** *Let f and g be functions continuous at $\mathbf{x}_0 \in \mathbb{R}^n$. Then $f(\mathbf{x}) \pm g(\mathbf{x})$, $f(\mathbf{x})g(\mathbf{x})$, and $\frac{f(\mathbf{x})}{g(\mathbf{x})}$ $(g(\mathbf{x}_0) \neq 0)$ are continuous at $\mathbf{x}_0$.*

Theorem 6.3.5 (Composition, [1, 6.3 定理 2])**.** *Let $D_X \subset \mathbb{R}^n$, $D_Y \subset \mathbb{R}^m$ be subsets, and let $\mathbf{x}_0 \in D_X$. Let $f : D_X \to D_Y$ be continuous at $\mathbf{x}_0$, and $g : D_Y \to \mathbb{R}^k$ be continuous at $f(\mathbf{x}_0)$. Then the composition $g \circ f$ is continuous at $\mathbf{x}_0$.*

Clearly, Theorem 6.3.4 follows from Theorem 6.2.7, and Theorem 6.3.5 follows from Theorem 6.2.11.

Remark 6.3.6. Note that Theorem 6.3.5 and Theorem 6.2.11 imply that the limit sign and continuous functions commute. More precisely, let $\mathring{U}_r(P) \subset \mathbb{R}^n$, $Q \in \mathbb{R}^m$, and let $f : \mathring{U}_r(P) \to R^m$ be so that
$$\lim_{\mathbf{x}\to P} f(\mathbf{x}) = Q.$$
Let $U_s(Q) \subset \mathbb{R}^n$ and $g : U_s(Q) \to \mathbb{R}^k$ be continuous at Q. Then we have
$$\lim_{\mathbf{x}\to P} g(f(\mathbf{x})) = g(Q) = g\left(\lim_{\mathbf{x}\to P} f(\mathbf{x})\right).$$

Example 6.3.7 ([1, §6.3 例 1])**.** The function $f(x,y) = \sin(x+y) + |x+y+1|$ is continuous on $\mathbb{R}^2$. This is because $\sin z$ and $|z|$ are continuous on $\mathbb{R}$, $x+y$ and $x+y+1$ are continuous on $\mathbb{R}^2$ (by Example 6.3.3 and Theorem 6.3.4), and $f(x,y)$ is the composition of these continuous functions (Theorem 6.3.5).

Definition 6.3.8 (Elementary functions)**.** Let $D \subset \mathbb{R}^n$ be a region. A function $f : D \to \mathbb{R}$ is said to be *elementary*, if it is defined as taking sums, products, roots and compositions of finitely many elementary functions of a single variable.

Similar to Example 6.3.7, Theorem 6.3.4 and Theorem 6.3.5 imply elementary functions are continuous.

Theorem 6.3.9 ([1, §6.3 定理 3])**.** *Let $D \subset \mathbb{R}^n$ be a region. Then any elementary function $f : D \to \mathbb{R}$ is continuous on D.*

Example 6.3.10 ([1, §6.3 例 2])**.** Discuss the continuity of the function

$$f(x,y) = \begin{cases} \dfrac{x^2y^2}{x^4+y^4} & , \text{ if } (x,y) \neq (0,0) \\ \\ 0 & , \text{ if } (x,y) = (0,0) \end{cases}.$$

证明. By Theorem 6.3.9, $f(x,y)$ is continuous at any $(x,y) \neq (0,0)$.

For $(x,y) = (0,0)$, consider $(x,y) = (t,kt)$ for $k,t \in \mathbb{R}$ with $t \to 0$. Then

$$\lim_{t\to 0} f(t,kt) = \lim_{t\to 0} \frac{t^2(kt)^2}{t^4+(kt)^4} = \frac{k^2}{1+k^4}$$

depending on k. Thus, we conclude that f is not continuous at $(0,0)$. □

Proposition 6.3.11. *Let $D \subset \mathbb{R}^n$ be a region, $f : D \to \mathbb{R}^m$ a function, $\pi_i : \mathbb{R}^m \to \mathbb{R}$ projections. Then f is continuous if and only if $f_i = \pi_i \circ f$ is continuous for any $1 \leq i \leq m$.*

证明. If f is continuous, then $f_i = \pi_i \circ f$ is continuous by Example 6.3.3 and Theorem 6.3.5.

On the contrary, if $f_i = \pi_i \circ f$ is continuous for any $1 \leq i \leq m$, then $f = f_1\mathbf{e}_1 + \cdots + f_m\mathbf{e}_m$ is continuous by Theorem 6.3.4. □

Similar to closed intervals, continuous functions on closed regions (more generally, *compact sets* 紧集) have good properties.

Theorem 6.3.12 (Boundedness theorem, [1, §6.3 定理 4])**.** *Let $\overline{D} \subset \mathbb{R}^n$ be a closed region, $f : \overline{D} \to \mathbb{R}$ a continuous function. Then f is bounded.*

证明. The idea is similar to the 1-dimensional case (Theorem 1.6.4). Suppose that f does not have an upper bound. Then we can find a sequence $\mathbf{x}_n \in \overline{D}$, such that

$$f(\mathbf{x}_n) > n.$$

Since $\{\mathbf{x}_n\} \subset \overline{D}$ is bounded, by finding the subsequences of each coordinate step by step, we can find a subsequence $\{\mathbf{x}_{n_k}\}$ so that $\mathbf{x}_{n_k} \to \ell \in \overline{D}$ $(k \to \infty)$. Then by the continuity of f,

$$\lim_{k\to\infty} f(\mathbf{x}_{n_k}) = f(\ell) \neq +\infty.$$

A contradiction. Thus, we see that f is bounded on $\overline{D}$. □

Theorem 6.3.13 (Extreme value theorem, [1, §6.3 定理 5])**.** *Let $\overline{D} \subset \mathbb{R}^n$ be a closed region, $f : \overline{D} \to \mathbb{R}$ a continuous function. Then f must attain a maximum, and a minimum, each at least once. That is, there are $\mathbf{x}_1, \mathbf{x}_2 \in \overline{D}$ such that*

$$f(\mathbf{x}_1) \leq f(\mathbf{x}) \leq f(\mathbf{x}_2)$$

for any $\mathbf{x} \in \overline{D}$.

证明. This again is analogous to the 1-dimensional case (Theorem 1.6.3). Consider the image

$$E = \{f(\mathbf{x}) : \mathbf{x} \in \overline{D}\}.$$

By Theorem 1.6.4, we see that E is bounded. Thus, by Theorem 7.1.7, $\sup(E)$ exists. Then for any $n \in \mathbb{N}^*$, there exists $y_n \in E$ such that

$$\sup(E) - \frac{1}{n} \leq y_n \leq \sup(E). \tag{6.3.1}$$

Then by the definition of E, there exists $\mathbf{x}_n \in [a, b]$ such that $f(\mathbf{x}_n) = y_n$. By finding the subsequences of each coordinate step by step, we can find a subsequence $\{\mathbf{x}_{n_k}\}$ so that $\mathbf{x}_{n_k} \to \ell \in \overline{D}$ $(k \to \infty)$. Then by the continuity of f, we conclude

$$\lim_{k\to\infty} f(\mathbf{x}_{n_k}) = f(\ell).$$

By Theorem 1.3.8 and (6.3.1), we have $f(\ell) = \sup(E)$. One can similarly discuss the infima. $\square$

Theorem 6.3.14 (Intermediate value theorem, [1, §6.3 定理 6]). *Let $\overline{D} \subset \mathbb{R}^n$ be a closed region, $f : \overline{D} \to \mathbb{R}$ be a continuous function attaining a maximum M and a minimum m. Then for any value $\eta \in [m, M]$, there is $\xi \in \overline{D}$, such that*

$$f(\xi) = \eta.$$

证明. By the extreme value theorem (Theorem 6.3.13), f attains a minimum, say at $\mathbf{a} \in \overline{D}$, and a maximum, say at $\mathbf{b} \in \overline{D}$. Since $\overline{D}$ is connected, there is a (continuous) curve $\Gamma : [0, T] \to \overline{D}$ such that $\Gamma(0) = \mathbf{a}$, $\Gamma(T) = \mathbf{b}$. Then via Theorem 6.3.5, the function $f \circ \Gamma : [0, T] \to [m, M]$ is continuous and attains the minimum and the maximum

$$m = f(\mathbf{a}) = f(\Gamma(0)), \quad M = f(\mathbf{b}) = f(\Gamma(T)).$$

The consequence now follows from the 1-dimensional intermediate value theorem (Theorem 1.6.1). $\square$

6.4 Differentials in several variable calculus

6.4.1 Partial derivatives

Definition 6.4.1 (Partial derivatives). Let $D \subset \mathbb{R}^n$ be a region, $\mathbf{x}_0 \in D$ a point, $f : D \to \mathbb{R}$ a function. For $1 \leq j \leq n$, let

$$\mathbf{e}_j = (0, \ldots, 0, \underbrace{1}_{j}, 0, \ldots, 0)$$

be the unit vector of the j-th coordinate. Suppose that the following limit exists:

$$\lim_{t \to 0} \frac{f(\mathbf{x}_0 + t\mathbf{e}_j) - f(\mathbf{x}_0)}{t} = L.$$

Then we say the limit is the *partial derivative of f with respect to the x_j variable at* $\mathbf{x}_0$. We write the limit as

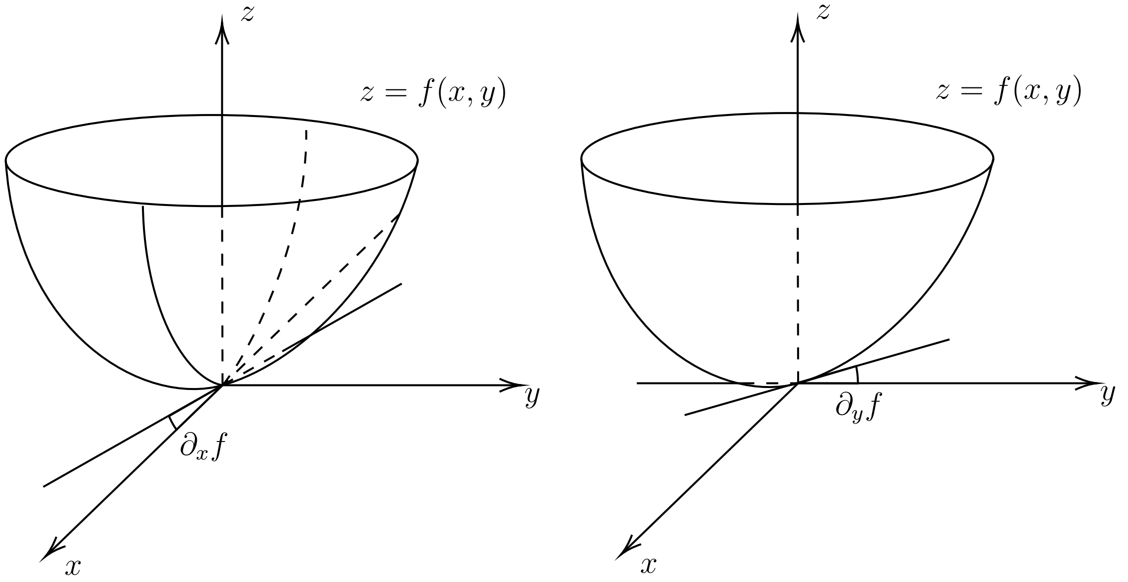
$$L = f_{x_j}(\mathbf{x}_0) = \frac{\partial f}{\partial x_j}(\mathbf{x}_0) = \left.\frac{\partial f}{\partial x_j}\right|_{\mathbf{x}_0} = \partial_{x_j} f(\mathbf{x}_0).$$

Example 6.4.2 ([1, §6.4 例 2]). Let $f(x, y) = x^y$ $(x, y > 0)$. Find $\dfrac{\partial f}{\partial x}$ and $\dfrac{\partial f}{\partial y}$.

证明.

$$\begin{aligned}\frac{\partial f}{\partial x} =& yx^{y-1},\\ \frac{\partial f}{\partial y} =& x^y \ln x.\end{aligned}$$

□



Geometrically, the partial derivative $\partial_{x_j} f(P)$ of a function $f : D \subset \mathbb{R}^n \to \mathbb{R}$ with respect to the x_j variable at $P = (p_1, \ldots, p_n)$ represents the slope of the tangent line at P of the curve Γ_j determined by the system of equations:

$$\begin{cases} x_{n+1} = f(x_1, \ldots, x_n), \\ x_1 = p_1, \\ \vdots \\ x_{j-1} = p_{j-1}, \\ x_{j+1} = p_{j+1}, \\ \vdots \\ x_n = p_n, \end{cases}.$$

Example 6.4.3. Let $f(x,y) = \mathrm{sgn}(xy)$ where $\mathrm{sign}(\cdot)$ is the sign function (Example 1.2.12). Then one checks that $\partial_x f(0,0) = \partial_y f(0,0) = 0$. However, f is not continuous at $(0,0)$.

6.4.2 Higher-order partial derivatives

Let $D \subset \mathbb{R}^n$ be a region, $\mathbf{x}_0 \in D$ a point, $f : D \to \mathbb{R}$ a function. For $1 \leq j \leq n$, if $\partial_{x_j} f$ exists on D, then it can be seen as another function defined on D, and can again be partially differentiated. Similar to the 1-dimensional derivative (Definition 2.6.1), the partial derivative of $\partial_{x_j} f$ is called the *second partial derivative* (二阶偏导数). If the direction of derivative is **not** repeated, it is called a *mixed partial derivative*:

$$\frac{\partial^2 f}{\partial x_i \partial x_j}$$

Theorem 6.4.4 (Clairaut's theorem 克莱罗定理, [1, §6.4 定理 1])**.** *Let $D \subset \mathbb{R}^n$ be a region, $P \in D$ a point, $f : D \to \mathbb{R}$ a function. If f has continuous second partial derivatives on that neighborhood of P, then for all $i, j \in \{1, \dots, n\}$, we have*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(P) = \frac{\partial^2 f}{\partial x_j \partial x_i}(P).$$

证明. We assume without loss of generality that $i \neq j$ and $P = \mathbf{0}$. Let

$$\partial_{x_i} \partial_{x_j} f(\mathbf{0}) = A_{ij}, \quad \partial_{x_j} \partial_{x_i} f(\mathbf{0}) = A_{ji}.$$

Let $\epsilon > 0$. Because the double derivatives of f are continuous, we can find a $\delta > 0$ such that

$$|\partial_{x_i} \partial_{x_j} f(\mathbf{x}) - A_{ij}| < \epsilon, \quad |\partial_{x_j} \partial_{x_i} f(\mathbf{x}) - A_{ji}| < \epsilon$$

for any $\|\mathbf{x}\| \leq 2\delta$.

Now we consider the quantity

$$F = f(\delta \mathbf{e}_i + \delta \mathbf{e}_j) + f(\mathbf{0}) - f(\delta \mathbf{e}_i) - f(\delta \mathbf{e}_j)$$

where $\mathbf{e}_i$ is the unit vector of the i-th coordinate.

Now note that

$$f(\delta\mathbf{e}_i + \delta\mathbf{e}_j) - f(\delta\mathbf{e}_j) = \int_0^\delta f_{x_i}(t\mathbf{e}_i + \delta\mathbf{e}_j)dt,$$
$$f(\delta\mathbf{e}_i) - f(\mathbf{0}) = \int_0^\delta f_{x_i}(t\mathbf{e}_i)dt.$$

It follows that

$$F = \int_0^\delta \left(f_{x_i}(t\mathbf{e}_i + \delta\mathbf{e}_j) - f_{x_i}(t\mathbf{e}_i)\right) dt.$$

On the other hand, by the mean value theorem, for each x_i, we have

$$f_{x_i}(t\mathbf{e}_i + \delta\mathbf{e}_j) - f_{x_i}(t\mathbf{e}_i) = \delta \cdot \partial_{x_j}\partial_{x_i} f(t\mathbf{e}_i + s\mathbf{e}_j)$$

for some $s \in (0, \delta)$. By our construction of δ, we thus have

$$|f_{x_i}(t\mathbf{e}_i + \delta\mathbf{e}_j) - f_{x_i}(t\mathbf{e}_i) - \delta A_{ji}| \leq \epsilon\delta.$$

Integrating this from 0 to δ, we thus obtain

$$|F - \delta^2 A_{ji}| \leq \epsilon\delta^2.$$

Via the same argument with the role of i and j reversed, we obtain

$$|F - \delta^2 A_{ji}| \leq \epsilon\delta^2.$$

It follows that

$$|\delta^2 A_{ij} - \delta^2 A_{ji}| \leq 2\epsilon\delta^2$$

and thus

$$|A_{ij} - A_{ji}| \leq 2\epsilon.$$

Since ϵ was chosen arbitrarily, we conclude that $A_{ij} = A_{ji}$. □

Remark 6.4.5. The proof of Clairaut's theorem (Theorem 6.4.4) shows that

$$\frac{\partial^2 f}{\partial x \partial y}(0,0) \approx \frac{f(\delta,\delta) + f(0,0) - f(\delta,0) - f(0,\delta)}{\delta^2}, \quad \delta \to 0$$

for $f : \mathbb{R}^2 \to \mathbb{R}$ having continuous partial derivatives.

Definition 6.4.6 (Differentiability class)**.** For a region $D \subset \mathbb{R}^n$, $k \in \mathbb{N}^*$, let $C^k(D)$ be the class of functions $f : D \to \mathbb{R}$ that have continuous k-th partial derivatives in any direction. In other words,

$$C^k(D) \coloneqq \left\{ f : D \to \mathbb{R} \,\middle|\, \begin{array}{l} \dfrac{\partial^k f}{\partial x_1^{k_1} \partial x_2^{k_2} \cdots \partial x_n^{k_n}} \text{ is continuous on } D \text{ for} \\ \text{every } k_1, \ldots, k_n \in \mathbb{N} \text{ with } k_1 + \ldots + k_n = k \end{array} \right\}.$$

A function f is said to be of *class* C^k, if $f \in C^k(D)$ for some region $D \subset \mathbb{R}^n$. In general, a map $f = (f_1, \ldots, f_m) : D \to \mathbb{R}^m$ is said to be of *class* C^k, if $f_i \in C^k(D)$ for any $1 \leq i \leq m$.

Definition 6.4.7 (Gradient and Laplacian)**.** The *gradient* (梯度) is a vector operator defined by

$$\nabla = \mathrm{grad} \coloneqq \left(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n} \right).$$

That is, for a region $D \subset \mathbb{R}^n$, $f \in C^1(D)$, we have

$$\nabla f = \mathrm{grad}(f) = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right).$$

The *Laplace operator* or *Laplacian* (拉普拉斯算子) is defined by

$$\Delta \coloneqq \nabla \cdot \nabla = \frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_n^2}.$$

In other words, let $D \subset \mathbb{R}^n$ be a region, $f \in C^2(D)$. We have

$$\Delta f = \frac{\partial^2 f}{\partial x_1^2} + \cdots + \frac{\partial^2 f}{\partial x_n^2}.$$

Definition 6.4.8 (Laplace's equation)**.** The *Laplace's equation* (拉普拉斯方程) is

$$\Delta f = 0.$$

The *Poisson's equation* (泊松方程) is

$$\Delta f = h$$

for some function $h : D \to \mathbb{R}$.

Definition 6.4.9 (Harmonic functions)**.** Let $D \subset \mathbb{R}^n$ be a region. A function $f \in C^2(D)$ is called *harmonic* (调和函数) if it satisfies the Laplace's equation, i.e.

$$\Delta f = 0.$$

Example 6.4.10 ([1, §6.4 例 8])**.** The function $z(x, y) = \ln \sqrt{x^2 + y^2}$ is harmonic. In fact, one calculates

$$\frac{\partial^2 z}{\partial x^2} = \frac{y^2 - x^2}{(x^2 + y^2)^2}, \quad \frac{\partial^2 z}{\partial y^2} = \frac{x^2 - y^2}{(x^2 + y^2)^2}.$$

It follows that $\Delta z = 0$.

Theorem 6.4.11 (The mean value property)**.** *Let $D \subset \mathbb{R}^n$ be a region, $f \in C^2(D)$ a harmonic function, and $U_r(P) \subset D$. Then we have*

$$f(P) = \frac{1}{\mathrm{vol}(U_r(P))} \int_{U_r(P)} f(\mathbf{x}) d\mathbf{x} = \frac{1}{\mathrm{vol}(\partial U_r(P))} \int_{\partial U_r(P)} f(\mathbf{x}) d\mathbf{x}$$

where $U_r(P)$ is the round ball centered at P of radius r, and $\partial U_r(P)$ is the sphere centered at P of radius r.

In mathematics, a *partial differential equation* (*PDE* for short, 偏微分方程) is an equation which computes a function between various partial derivatives of a multivariable function. Laplace's equation and Poisson's equation are classic PDEs.

6.4.3 Differentials

Definition 6.4.12 (Fréchet derivative)**.** Let $(X, \|\cdot\|_X)$, $(Y, \|\cdot\|_Y)$ be normed vector spaces, and $U \subset X$ be an open subset of X. A map $f : U \to Y$ is called *Fréchet differentiable* (or *differentiable* for short) at $\mathbf{x}_0 \in U$ if there exists a continuous linear map $L : X \to Y$ such that

$$\lim_{\|\mathbf{h}\|_X \to 0} \frac{\|f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) - L(\mathbf{h})\|_Y}{\|\mathbf{h}\|_X} = 0.$$

The *differential* (or *Fréchet derivative* 弗雷歇微分) Df of f at $\mathbf{x}_0$ is then defined by

$$Df = Df(\mathbf{x}_0) = L.$$

One can easily show that linear maps between finite dimensional vector spaces are continuous. Thus, we have:

Definition 6.4.13 (Differentials on Euclidean spaces, [1, §6.4 定义 2])**.** Let $(X, \|\cdot\|_X) = (\mathbb{R}^n, \|\cdot\|_X)$, $(Y, \|\cdot\|_Y) = (\mathbb{R}^m, \|\cdot\|_Y)$. Let $D \subset \mathbb{R}^n$ be a region, $f : D \to \mathbb{R}^m$ a map. Then the function f is *differentiable* at $\mathbf{x}_0 \in D$ if there exists a linear map $L : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{\|\mathbf{h}\|_X \to 0} \frac{\|f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) - L(\mathbf{h})\|_Y}{\|\mathbf{h}\|_X} = 0.$$

In other words, f is differentiable at $\mathbf{x}_0$ if

$$f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + L(\mathbf{h}) + o(\mathbf{h}), \quad \|\mathbf{h}\|_X \to 0.$$

The *differential* df of f at $\mathbf{x}_0$ is given by

$$df = df(\mathbf{x}_0, \mathbf{h}) = L(\mathbf{h}).$$

Remark 6.4.14. As discussed in Remark 2.3.9, $df = df(\mathbf{x}_0, \mathbf{h})$ depends on both $\mathbf{x}_0$ and $\mathbf{h}$. Note that we commonly refer to df as a vector $df(\mathbf{x}_0, \mathbf{h}) \in \mathbb{R}^m$, or a linear map $df(\mathbf{x}_0, \cdot) : \mathbb{R}^n \to \mathbb{R}^m$, which is the **best linear approximation** to f at $\mathbf{x}_0$. Usually, we write $\mathbf{h} = d\mathbf{x} = (dx_1, \ldots, dx_n)$.

Similar to 1-dimensional functions (Proposition 2.1.5), a differentiable map must be continuous.

Theorem 6.4.15 ([1, §6.4 定理 2])**.** *Let $D \subset \mathbb{R}^n$ be a region, $f : D \to \mathbb{R}^m$ a map. If f is differentiable at $\mathbf{x}_0$, then f is continuous at $\mathbf{x}_0$.*

Theorem 6.4.16 ([1, §6.4 定理 3])**.** *Let $D \subset \mathbb{R}^n$ be a region, $f : D \to \mathbb{R}^m$ a map. If f is differentiable at $\mathbf{x}_0$, then all partial derivatives $\partial_{x_j}(\pi_i f)$ exist at $\mathbf{x}_0$, and that*

$$\pi_i(df(\mathbf{x}_0, \mathbf{e}_j)) = \partial_{x_j}(\pi_i f)(\mathbf{x}_0)$$

where $\pi_i : \mathbb{R}^m \to \mathbb{R}$ is the projection to the i-th coordinate, $\mathbf{e}_j$ is the unit vector of the j-th coordinate in $\mathbb{R}^n$. In other words, write $f = (f_1, \ldots, f_m)$, and we have

$$df(\mathbf{x}_0, \cdot) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}_0) & \frac{\partial f_1}{\partial x_2}(\mathbf{x}_0) & \cdots & \frac{\partial f_1}{\partial x_n}(\mathbf{x}_0) \\ \frac{\partial f_2}{\partial x_1}(\mathbf{x}_0) & \frac{\partial f_2}{\partial x_2}(\mathbf{x}_0) & \cdots & \frac{\partial f_2}{\partial x_n}(\mathbf{x}_0) \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{x}_0) & \frac{\partial f_m}{\partial x_2}(\mathbf{x}_0) & \cdots & \frac{\partial f_m}{\partial x_n}(\mathbf{x}_0) \end{bmatrix} = \begin{bmatrix} \nabla f_1(\mathbf{x}_0) \\ \nabla f_2(\mathbf{x}_0) \\ \vdots \\ \nabla f_n(\mathbf{x}_0) \end{bmatrix}. \tag{6.4.1}$$

证明. For simplicity, we assume $m = 1$. Since f is differentiable,

$$\lim_{\|\mathbf{h}\|_X \to 0} \frac{|f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) - df(\mathbf{x}_0, \mathbf{h})|}{\|\mathbf{h}\|_X} = 0.$$

In particular, for $\mathbf{h} = t\mathbf{e}_j$ with $t \to 0$, we have

$$\lim_{t \to 0} \frac{|f(\mathbf{x}_0 + t\mathbf{e}_j) - f(\mathbf{x}_0) - df(\mathbf{x}_0, t\mathbf{e}_j)|}{\|t\mathbf{e}_j\|_X} = 0.$$

This is equivalent to say

$$\lim_{t \to 0} \frac{f(\mathbf{x}_0 + t\mathbf{e}_j) - f(\mathbf{x}_0) - t df(\mathbf{x}_0, \mathbf{e}_j)}{t} = 0$$

and so $\partial_{x_j} f(\mathbf{x}_0) = df(\mathbf{x}_0, \mathbf{e}_j)$ for any $1 \leq j \leq n$. □

Definition 6.4.17 (Jacobian matrix and determinant)**.** Let $D \subset \mathbb{R}^n$ be a region, $f : D \to \mathbb{R}^m$ a map, $\mathbf{x}_0 \in D$ a point. Then the matrix in (6.4.1) is called the *Jacobian matrix* or *Jacobian* (雅可比矩阵) of f at $\mathbf{x}_0$, denoted $\mathbf{J}_f = \mathbf{J}_f(\mathbf{x}_0)$. Thus, $\mathbf{J}_f(\mathbf{x}_0)$ is defined such that its (i, j)-th entry is

$$\frac{\partial f_i}{\partial x_j}(\mathbf{x}_0).$$

When $m = n$, the Jacobian is a square matrix. Then we refer to the determinant of the Jacobian matrix as the *Jacobian determinant* (雅可比行列式). We also write

$$\det \mathbf{J}_f(\mathbf{x}_0) = \frac{D(f_1, \ldots, f_n)}{D(x_1, \ldots, x_n)}(\mathbf{x}_0).$$

Example 6.4.18 ([1, §6.4 例 5])**.** Consider the function

$$f(x, y) = \begin{cases} \dfrac{xy}{x^2 + y^2} & , \text{ if } (x, y) \neq (0, 0) \\ \\ 0 & , \text{ if } (x, y) = (0, 0) \end{cases}.$$

By definition, we know that at $(x, y) \neq (0, 0)$, we have

$$\begin{aligned} f_x(x, y) =& \frac{y(x^2 + y^2) - xy \cdot 2x}{(x^2 + y^2)^2} = \frac{y^3 - x^2 y}{(x^2 + y^2)^2}, \\ f_x(x, y) =& \frac{x(x^2 + y^2) - xy \cdot 2y}{(x^2 + y^2)^2} = \frac{x^3 - y^2 x}{(x^2 + y^2)^2}. \end{aligned}$$

For $(x,y)=(0,0)$, we have

$$f_x(0,0)=\lim_{\Delta x\to 0}\frac{f(0+\Delta x,0)-f(0,0)}{\Delta x}=0,$$
$$f_x(0,0)=\lim_{\Delta x\to 0}\frac{f(0+\Delta y,0)-f(0,0)}{\Delta y}=0.$$

However, f is not continuous at $(x,y)=(0,0)$, as shown in Example 6.3.10,

$$\lim_{t\to 0} f(t,kt)=\frac{1}{k}.$$

for $k\in\mathbb{R}\setminus\{0\}$. Thus, we conclude that f is not differentiable at $(x,y)=(0,0)$ by Theorem 6.4.15, even if the Jacobian $\mathbf{J}_f$ of f exists.

Theorem 6.4.19 ([1, §6.4 定理 4])**.** *Let $D\subset\mathbb{R}^n$ be a region, $f=(f_1,\dots,f_m):D\to\mathbb{R}^m$ a map, $\mathbf{x}_0\in D$ a point. If all partial derivatives $\partial_{x_j}f_i$ exist on D and continuous at $\mathbf{x}_0$, then f is differentiable at $\mathbf{x}_0$ and*

$$df(\mathbf{x}_0)=\mathbf{J}_f(\mathbf{x}_0).$$

证明. Let $\|\cdot\|_X=\|\cdot\|_1$, $\|\cdot\|_Y=\|\cdot\|_\infty$. Since all partial derivatives $\partial_{x_j}f_i$ exist, the Jacobian is well-defined:

$$\mathbf{J}_f=(\partial_{x_j}f_i).$$

Let $\epsilon>0$, we shall find a $\delta>0$ such that

$$\|f(\mathbf{x}_0+\mathbf{h})-f(\mathbf{x}_0)-J_f(\mathbf{h})\|_\infty\le\epsilon\|\mathbf{h}\|_1 \tag{6.4.2}$$

for any $0<\|\mathbf{h}\|_1<\delta$.

First, by the continuity of $\partial_{x_j}f_i$ at $\mathbf{x}_0$, we choose $\delta>0$ such that

$$|\partial_{x_j}f_i\,(\mathbf{x}_0+\mathbf{h})-\partial_{x_j}f_i(\mathbf{x}_0)|<\epsilon \tag{6.4.3}$$

for any $0<\|\mathbf{h}\|_1<\delta$, and any i, j.

Write $\mathbf{h}=t_1\mathbf{e}_1+\cdots+t_n\mathbf{e}_n$, $\mathbf{h}_j=t_1\mathbf{e}_1+\cdots+t_j\mathbf{e}_j$. Then

$$\pi_i(\mathbf{J}_f(\mathbf{h}))=t_1\partial_{x_1}f_i(\mathbf{x}_0)+\cdots+t_n\partial_{x_n}f_i(\mathbf{x}_0).$$

Now for any $1 \le i \le m$, we have

$$
\begin{aligned}
& |f_i(\mathbf{x}_0 + \mathbf{h}) - f_i(\mathbf{x}_0) - \pi_i(\mathbf{J}_f(\mathbf{h}))| \\
\le & \sum_{j=1}^{n} |f_i(\mathbf{x}_0 + \mathbf{h}_j) - f_i(\mathbf{x}_0 + \mathbf{h}_{j-1}) - t_j \partial_{x_j} f_i(\mathbf{x}_0)| \\
= & \sum_{j=1}^{n} |f_i(\mathbf{x}_0 + \mathbf{h}_{j-1} + t_j \mathbf{e}_j) - f_i(\mathbf{x}_0 + \mathbf{h}_{j-1}) - t_j \partial_{x_j} f_i(\mathbf{x}_0)| \\
\xlongequal[s_j \in (0, t_j)]{\text{mean value theorem}} & \sum_{j=1}^{n} |t_j| \cdot |\partial_{x_j} f_i(\mathbf{x}_0 + \mathbf{h}_{j-1} + s_j \mathbf{e}_j) - \partial_{x_j} f_i(\mathbf{x}_0)| \\
\le & \sum_{j=1}^{n} |t_j| \cdot \epsilon = \epsilon \|\mathbf{h}\|_1
\end{aligned}
$$

where the last inequality follows from (6.4.3). Thus, we conclude (6.4.2). □

Corollary 6.4.20 ([1, §6.4 推论]). *For any region $D \subset \mathbb{R}^n$,*

(1) any $f \in C^1(D)$ is differentiable,

(2) any elementary function that has the first partial derivative on D in any direction is differentiable.

Example 6.4.21 ([1, §6.4 例 11]). Let $z(x, y) = x^y$ $(x, y > 0)$. Find dz.

证明. By Example 6.4.2, the Jacobian of z is

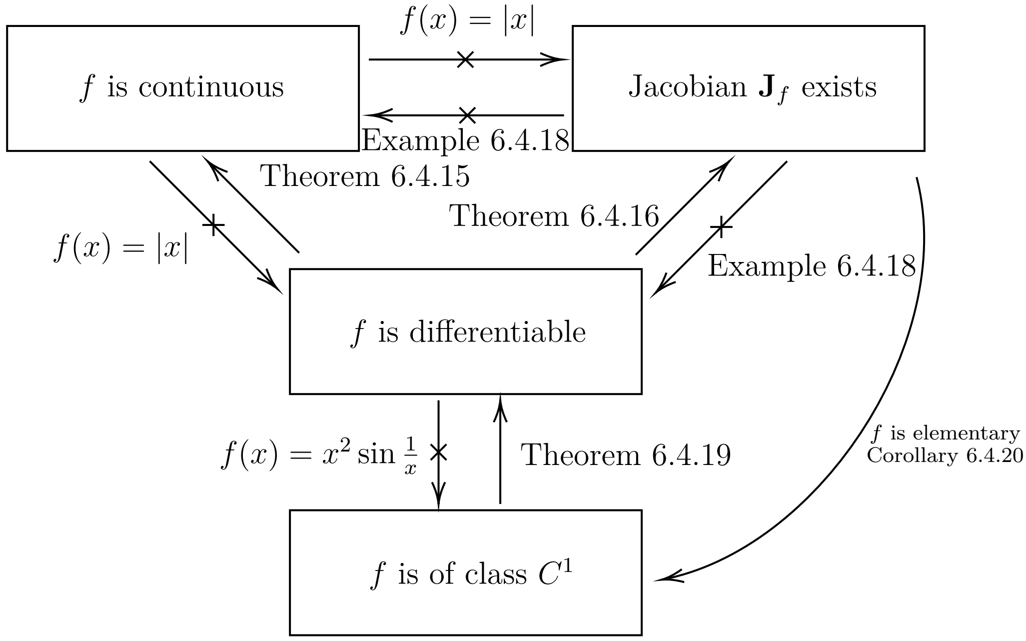
$$\mathbf{J}_z = \begin{bmatrix} yx^{y-1} & x^y \ln x \end{bmatrix}.$$

Then one calculates

$$dz = \begin{bmatrix} yx^{y-1} & x^y \ln x \end{bmatrix} \begin{bmatrix} dx \\ dy \end{bmatrix} = yx^{y-1}dx + x^y \ln x dy.$$

□

In summary, we have:



6.4.4 Higher-order differentials

We now discuss the *higher-order differentials*. Suppose that $f : \mathbb{R}^n \to \mathbb{R}^m$ is k-times differentiable. Recall from Remark 6.4.23 that a differential $df = F(\mathbf{x}, d\mathbf{x}) = f'(\mathbf{x})d\mathbf{x}$ is considered as function depending on both $\mathbf{x}$ and $d\mathbf{x}$. We fix $d\mathbf{x}$. Then the differential is again a map $df(\cdot, d\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}^m$. Then we can differentiate df again:

$$d^2 f = d(df(\cdot, d\mathbf{x})) = d(\mathbf{J}_f(d\mathbf{x})) = \mathbf{J}_{\mathbf{J}_f(d\mathbf{x})}(d\mathbf{x}) = \mathbf{J}_{df}(d\mathbf{x}).$$

Again, $d^2 f(\cdot, d\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}^m$ is a map from $\mathbb{R}^n$ to $\mathbb{R}^m$:

$$f : \mathbb{R}^n \to \mathbb{R}^m \quad \xmapsto{d} \quad df(\cdot, d\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}^m \quad \xmapsto{d} \quad d^2 f(\cdot, d\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}^m.$$

Similarly, we can define the higher-order differential:

$$d^k f = d(d^{k-1} f) = \mathbf{J}_{d^{k-1} f}(d\mathbf{x}).$$

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

For $u : \mathbb{R}^n \to \mathbb{R}$, $d\mathbf{x} = (dx_1, \ldots, dx_n)$, one calculates

$$\begin{aligned}
du(\cdot, d\mathbf{x}) &= \begin{bmatrix} \frac{\partial u}{\partial x_1} & \frac{\partial u}{\partial x_2} & \cdots & \frac{\partial u}{\partial x_n} \end{bmatrix} \begin{bmatrix} dx_1 \\ dx_2 \\ \vdots \\ dx_n \end{bmatrix} \\
&= dx_1 \frac{\partial u}{\partial x_1} + \cdots + dx_n \frac{\partial u}{\partial x_n} = \left(dx_1 \frac{\partial}{\partial x_1} + \cdots + dx_n \frac{\partial}{\partial x_n} \right) u.
\end{aligned}$$

Thus, inductively, we obtain

$$\begin{aligned}
d^k u(\cdot, d\mathbf{x}) &= d(d^{k-1} u(\cdot, d\mathbf{x})) \\
&= \left(dx_1 \frac{\partial}{\partial x_1} + \cdots + dx_n \frac{\partial}{\partial x_n} \right) d^{k-1} u(\cdot, d\mathbf{x}) \\
&= \left(dx_1 \frac{\partial}{\partial x_1} + \cdots + dx_n \frac{\partial}{\partial x_n} \right)^k u \qquad (6.4.4)
\end{aligned}$$

for $k \in \mathbb{N}$. In particular, for $k = 2$, we have

$$d^2 u(\cdot, d\mathbf{x}) = \begin{bmatrix} dx_1 & dx_2 & \cdots & dx_n \end{bmatrix} \mathbf{H}_u \begin{bmatrix} dx_1 \\ dx_2 \\ \vdots \\ dx_n \end{bmatrix}$$

where $\mathbf{H}_u$ is the *Hessian matrix* (黑塞矩阵):

$$\mathbf{H}_u = \mathbf{J}_{\mathbf{J}_u} = \mathbf{J}_{\nabla u} = \begin{bmatrix} \frac{\partial^2 u}{\partial x_1^2} & \frac{\partial^2 u}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 u}{\partial x_1 \partial x_n} \\ \frac{\partial^2 u}{\partial x_2 \partial x_1} & \frac{\partial^2 u}{\partial x_2^2} & \cdots & \frac{\partial^2 u}{\partial x_2 \partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 u}{\partial x_n \partial x_1} & \frac{\partial^2 u}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 u}{\partial x_n^2} \end{bmatrix}. \qquad (6.4.5)$$

Example 6.4.22 ([1, §6.5 例 9])**.** Let $f(x, y) = e^x y^2$ $(x, y > 0)$. Find $d^3 f$.

证明. Note that

$$\begin{aligned}
d^3 f &= \left(dx \frac{\partial}{\partial x} + dy \frac{\partial}{\partial y} \right)^3 f \\
&= dx^3 \frac{\partial^3 f}{\partial x^3} + 3 dx^2 dy \frac{\partial^3 f}{\partial x^2 \partial y} + 3 dx dy^2 \frac{\partial^3 f}{\partial x \partial y^2} + dy^3 \frac{\partial^3 f}{\partial y^3}.
\end{aligned}$$

Then one calculates

$$\frac{\partial^3 f}{\partial x^3} = e^x y^2, \quad \frac{\partial^3 f}{\partial x^2 \partial y} = 2e^x y, \quad \frac{\partial^3 f}{\partial x \partial y^2} = 2e^x, \quad \frac{\partial^3 f}{\partial y^3} = 0.$$

Thus, $d^3 f = e^x y^2 dx^3 + 6e^x y dx^2 dy + 6e^x dx dy^2$. □

6.4.5 Chain rule

We are now ready to state the several variable calculus chain rule. It is instructive to review the 1-dimensional chain rule (Theorem 2.2.1).

Theorem 6.4.23 (Chain rule, [1, §6.5 定理 1])**.** *Let $D_X \subset \mathbb{R}^n$, $D_Y \subset \mathbb{R}^m$ be regions. Let $\mathbf{x}_0 \in D_X$, $f : D_X \to D_Y$ and $g : D_Y \to \mathbb{R}^k$. Suppose that $\mathbf{J}_f$ exists at $\mathbf{x}_0$, g is differentiable at $f(\mathbf{x}_0)$. Then $\mathbf{J}_{g\circ f}$ exists at $\mathbf{x}_0$, and*

$$\mathbf{J}_{g\circ f}(\mathbf{x}_0) = \mathbf{J}_g(f(\mathbf{x}_0)) \cdot \mathbf{J}_f(\mathbf{x}_0). \tag{6.4.6}$$

证明. For $1 \le i \le n$, we write

$$\frac{\partial f}{\partial x_j}(\mathbf{x}_0) = \begin{bmatrix} \frac{\partial f_1}{\partial x_j}(\mathbf{x}_0) \\ \frac{\partial f_2}{\partial x_j}(\mathbf{x}_0) \\ \vdots \\ \frac{\partial f_m}{\partial x_j}(\mathbf{x}_0) \end{bmatrix}, \quad \frac{\partial (g\circ f)}{\partial x_j}(\mathbf{x}_0) = \begin{bmatrix} \frac{\partial (g\circ f)_1}{\partial x_j}(\mathbf{x}_0) \\ \frac{\partial (g\circ f)_2}{\partial x_j}(\mathbf{x}_0) \\ \vdots \\ \frac{\partial (g\circ f)_m}{\partial x_j}(\mathbf{x}_0) \end{bmatrix}.$$

Since g is differentiable at $f(\mathbf{x}_0)$, we have

$$g(f(\mathbf{x}_0) + \mathbf{h}) = g(f(\mathbf{x}_0)) + \mathbf{J}_g(\mathbf{h}) + o(\mathbf{h}), \quad \|\mathbf{h}\|_Y \to 0.$$

Let $\mathbf{h} = f(\mathbf{x}_0 + \mathbf{t}) - f(\mathbf{x}_0)$. Then we have

$$\frac{g(f(\mathbf{x}_0 + \mathbf{t})) - g(f(\mathbf{x}_0))}{\|\mathbf{t}\|_Y} = \mathbf{J}_g\left(\frac{\mathbf{h}}{\|\mathbf{t}\|_Y}\right) + o\left(\frac{\mathbf{h}}{\|\mathbf{t}\|_Y}\right), \quad \|\mathbf{h}\|_Y \to 0. \tag{6.4.7}$$

For $1 \le j \le n$, let $\mathbf{t} = t\mathbf{e}_j$. We have

$$\frac{\mathbf{h}}{\|\mathbf{t}\|_Y} = \frac{f(\mathbf{x}_0 + \mathbf{t}) - f(\mathbf{x}_0)}{\|\mathbf{t}\|_Y} \to \frac{\partial f}{\partial x_j}(\mathbf{x}_0), \quad \|\mathbf{t}\|_X \to 0,$$

$$\|\mathbf{h}\|_Y = \|f(\mathbf{x}_0 + \mathbf{t}) - f(\mathbf{x}_0)\|_Y \to 0, \quad \|\mathbf{t}\|_X \to 0,$$

$$\frac{g(f(\mathbf{x}_0+\mathbf{t}))-g(f(\mathbf{x}_0))}{\|\mathbf{t}\|_Y} \to \frac{\partial(g\circ f)}{\partial x_j}(\mathbf{x}_0), \quad \|\mathbf{t}\|_X \to 0.$$

Thus, by (6.4.7), we have

$$\frac{\partial(g\circ f)}{\partial x_j}(\mathbf{x}_0) - \mathbf{J}_g\left(\frac{\partial f}{\partial x_j}(\mathbf{x}_0)\right) \to \mathbf{0}, \quad \|\mathbf{t}\|_X \to 0.$$

Therefore, we obtain (6.4.6). $\square$

Example 6.4.24 ([1, §6.5 例 7]). Let $f(x,y,w) \in C^1(\mathbb{R}^3)$, and $z(x,y) = f(x,y,x^2y)$. Find $\dfrac{\partial z}{\partial x}$ and $\dfrac{\partial z}{\partial y}$.

证明. Let $g : (x,y) \mapsto (x,y,x^2y)$. Then $z = f\circ g$. One calculates

$$\mathbf{J}_f = \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial w} \end{bmatrix}, \quad \mathbf{J}_g = \begin{bmatrix} \frac{\partial g_1}{\partial x} & \frac{\partial g_1}{\partial y} \\ \frac{\partial g_2}{\partial x} & \frac{\partial g_2}{\partial y} \\ \frac{\partial g_3}{\partial x} & \frac{\partial g_3}{\partial y} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2xy & x^2 \end{bmatrix}.$$

By the chain rule (Theorem 6.4.23), we obtain

$$\begin{bmatrix} \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \end{bmatrix} = \mathbf{J}_z = \mathbf{J}_f\mathbf{J}_g = \begin{bmatrix} \frac{\partial f}{\partial x} + 2xy\frac{\partial f}{\partial w} & \frac{\partial f}{\partial y} + x^2\frac{\partial f}{\partial w} \end{bmatrix}.$$

$\square$

Example 6.4.25 ([1, P.343, Exercise 10]). Let $D \subset \mathbb{R}^2$ be a convex region, $f \in C^1(D)$. Suppose that

$$xf_x(x,y) + yf_y(x,y) = 0 \tag{6.4.8}$$

for $(x,y) \in D$. Prove that there is a function $F(\theta)$ such that $f(r\cos\theta, r\sin\theta) = F(\theta)$. In particular, if $(0,0) \in D$, then $f \equiv C$ for some $C \in \mathbb{R}$.

Remark 6.4.26. (6.4.8) implies that f is *homogeneous of degree* 0. See Exercise 8.12.8 for more details.

证明. Let $g : (r,\theta) \mapsto (r\cos\theta, r\sin\theta)$. Then by the chain rule and (6.4.8), we have

$$\frac{\partial(f\circ g)}{\partial r}(r,\theta) = f_x(g(r,\theta))\cos\theta + f_y(g(r,\theta))\sin\theta = 0$$

for any $g(r,\theta) \in D$. Now fix θ. For any $0 \le r_1 < r_2$, by the Lagrange's mean value theorem, there is a $\xi \in (r_1, r_2)$ such that

$$f(g(r_2,\theta)) - f(g(r_1,\theta)) = (r_2 - r_1)\frac{\partial(f\circ g)}{\partial r}(\xi,\theta) = 0.$$

Thus, one defines $F(\theta) = f(g(r_1,\theta))$. □

Example 6.4.27 ([1, P.343, Exercise 11]). Let $D \subset \mathbb{R}^2$ be a convex region, $f \in C^1(D)$. Suppose that

$$yf_x(x,y) - xf_y(x,y) = 0 \tag{6.4.9}$$

for $(x,y) \in D$. Prove that there is a function $G(r)$ such that $f(r\cos\theta, r\sin\theta) = G(r)$.

证明. Let $g : (r,\theta) \mapsto (r\cos\theta, r\sin\theta)$. Then by the chain rule and (6.4.9), we have

$$\frac{\partial(f\circ g)}{\partial\theta}(r,\theta) = -f_x(g(r,\theta))r\sin\theta + f_y(g(r,\theta))r\cos\theta = 0$$

for any $g(r,\theta) \in D$. Now fix r. For any $\theta_1 < \theta_2$, by the Lagrange's mean value theorem, there is a $\xi \in (\theta_1, \theta_2)$ such that

$$f(g(r,\theta_2)) - f(g(r,\theta_1)) = (\theta_2 - \theta_1)\frac{\partial(f\circ g)}{\partial\theta}(r,\xi) = 0.$$

Thus, one defines $G(r) = f(g(r,\theta_1))$. □

Example 6.4.28 (2023 Calculus B, Final). Let $z(x,y) : \mathbb{R}^2 \to \mathbb{R}$ be a function satisfied

$$x^2 z_x + y^2 z_y = z^2.$$

Let

$$x = t, \quad y = \frac{t}{1+tu}, \quad z = \frac{t}{1+tW}.$$

Show that $W_t = 0$.

证明. One writes

$$\frac{1}{z} = \frac{1}{t} + W.$$

Then

$$\frac{1}{t^2} - \frac{z_t}{z^2} = W_t.$$

Now

$$z_t = z_x x_t + z_y y_t = z_x + z_y \frac{1}{(1+tu)^2} = \frac{z^2}{t^2}.$$

We get $W_t = 0$. $\square$

Example 6.4.29. Consider

$$f(x,y) = \begin{cases} \frac{x^2 y}{x^2+y^2} & , \text{ if } x^2+y^2 \neq 0 \\ 0 & , \text{ if } x^2+y^2 = 0 \end{cases}$$

(1) Show that $f_x(0,0)$, $f_y(0,0)$ exist, but f is not differentiable at $(0,0)$.

(2) Let $x = y = t$. Then $g(t) = f(t,t) = \frac{t}{2}$. Explain why the chain rule fails: $g'(t) \neq f_x(t) + f_y(t)$.

证明. One calculates

$$f_x(0,0) = 0, , \quad f_y(0,0) = 0.$$

However,

$$A(x,y) = \frac{f(x,y) - f(0,0) - f_x(0,0)x - f_y(0,0)y}{\sqrt{x^2+y^2}} = \frac{x^2 y}{(x^2+y^2)^{\frac{3}{2}}}.$$

As $(t,0) \to 0$, $A(t,0) \to 0$. As $(t,t) \to 0$, $A(t,t) \to \frac{\sqrt{2}}{4}$. Thus, the limit does not exist.

(2) $g'(0) = \frac{1}{2}$. $f_x(0,0) + f_y(0,0) = 0$. The chain rule fails because f is not differentiable at $(0,0)$. $\square$

Via the chain rule (Theorem 6.4.23), we immediately obtain the change of variables formula:

Theorem 6.4.30 (Change of variables, [1, §6.5 定理 2])**.** *Suppose that* $\mathbf{y}$ *is a differentiable map of* $\mathbf{x}$*, and* $\mathbf{z}$ *is a differentiable map of* $\mathbf{y}$*. Consider the composition*

$$\mathbf{x} \mapsto \mathbf{y}(\mathbf{x}) \mapsto \mathbf{z}(\mathbf{y}(\mathbf{x})).$$

Then the differentials satisfy

$$d\mathbf{z}(\mathbf{y}(\mathbf{x}), d\mathbf{y}) = \mathbf{J}_{\mathbf{z}}(d\mathbf{y}) = \mathbf{J}_{\mathbf{z}}\mathbf{J}_{\mathbf{y}}(d\mathbf{x}) = \mathbf{J}_{\mathbf{z}\circ\mathbf{y}}(d\mathbf{x}) = d(\mathbf{z}\circ\mathbf{y})(\mathbf{x}, d\mathbf{x}).$$

We also write

$$d\mathbf{z} = \mathbf{J}_{\mathbf{z}}(d\mathbf{y}) = \mathbf{J}_{\mathbf{z}\circ\mathbf{y}}(d\mathbf{x}). \tag{6.4.10}$$

Remark 6.4.31. As in Remark 2.6.6, since the higher-order differential of the composition of two functions does not have the chain rule, the (6.4.10) does not hold for the higher-order differentials. In other words, for a function

$$\mathbf{x} \mapsto \mathbf{y}(\mathbf{x}) \mapsto \mathbf{z}(\mathbf{y}(\mathbf{x})),$$

we usually do **NOT** have

$$\mathbf{J}_{\mathbf{J}_{\mathbf{z}}(d\mathbf{y})}(d\mathbf{y}) = \mathbf{J}_{\mathbf{J}_{\mathbf{z}\circ\mathbf{y}}(d\mathbf{x})}(d\mathbf{x}).$$

In application, when we study the (multiple) integrals, we commonly change the variables to simplify the calculation. More precisely, let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a differentiable map, and let Z be an n-dimensional geometric object (e.g. a square in $\mathbb{R}^2$, a cube in $\mathbb{R}^3$, etc.) We want to know the volume change between Z and $f(Z)$. The simplest situation is that f is linear, for which Theorem 7.2.34 implies

$$\mathrm{vol}(f(Z)) = |\det(f)|\,\mathrm{vol}(Z).$$

Now suppose that f is a differentiable function. Then locally, f can be approximated by the linear map $\mathbf{J}_f$. Thus, in a small neighborhood $U_r(\mathbf{z})$ of Z, we abuse notation and denote by $d\mathbf{z}$ the volume of $U_r(\mathbf{z})$, say the *volume element* (体积元). Also, denote by $df(\mathbf{z})$ the volume of $f(U_r(\mathbf{z}))$. Then we have

$$df(\mathbf{z}) = |\det \mathbf{J}_f(\mathbf{z})| d\mathbf{z}.$$

This is the *change of variables formula* for (multiple) integrals.

Similar to 1-dimensional functions, one may easily show:

Proposition 6.4.32. *Let $D \subset \mathbb{R}^n$ be a region, and let $u, v : D \to \mathbb{R}$ be differentiable functions on D. Then we have*

(1) $d(u \pm v) = du \pm dv$,

(2) $d(cu) = cdu$ for $c \in \mathbb{R}$,

(3) (product rule) $d(u \cdot v) = udv + vdu$,

(4) $d\left(\frac{u}{v}\right) = \frac{vdu - udv}{v^2}$,

(5) (chain rule) $df(u) = f'(u)du$ *for differentiable* $f : \mathbb{R} \to \mathbb{R}$.

证明. We only show (1), and the other properties are similar. Let $h(u, v) = u \pm v$. Then by Theorem 6.4.30, we have

$$dh = \mathbf{J}_h \begin{bmatrix} du \\ dv \end{bmatrix} = \begin{bmatrix} 1 & \pm 1 \end{bmatrix} \begin{bmatrix} du \\ dv \end{bmatrix} = du \pm dv.$$

□

Example 6.4.33 ([1, §6.5 例 8])**.** Let $u = \sin(x^2 + y^2) + e^{xz}$. Find the differential du at $(1, 0, 1)$.

证明. By the chain rule, we have

$$\begin{aligned} du &= \cos(x^2 + y^2)d(x^2 + y^2) + e^{xz}d(xz) \\ &= \cos(x^2 + y^2)(2xdx + 2ydy) + e^{xz}(dz + dx). \end{aligned}$$

Letting $(x, y, z) = (1, 0, 1)$, we have

$$du = \cos 1 \cdot (2dx + 0) + e(dz + dx) = (2\cos 1 + e)dx + edz.$$

□

6.5 Directional derivatives

We now generalize the notion of partial derivatives (Definition 6.4.1):

Definition 6.5.1 (Directional derivatives, [1, §6.6 定义])**.** Let $D \subset \mathbb{R}^n$ be a region, $\mathbf{x}_0 \in D$ a point, $f : D \to \mathbb{R}$ a function, $\mathbf{v} \in \mathbb{R}^n$ a vector. Suppose that the following limit exists:

$$\lim_{t \to 0} \frac{f(\mathbf{x}_0 + t\mathbf{v}) - f(\mathbf{x}_0)}{t} = L.$$

Then we say the limit is the *directional derivative of* f *in the direction* $\mathbf{v}$ *at* $\mathbf{x}_0$. We denote the limit by

$$L = \nabla_{\mathbf{v}} f(\mathbf{x}_0).$$

In particular, let $\mathbf{v}^\circ = \frac{\mathbf{v}}{\|\mathbf{v}\|}$ be the unit vector of $\mathbf{v}$. The directional derivative of f in the direction $\mathbf{v}^\circ$ at $\mathbf{x}_0$ is also written by

$$\nabla_{\mathbf{v}^\circ} f(\mathbf{x}_0) = \frac{\partial f}{\partial \mathbf{v}}(\mathbf{x}_0) = \partial_{\mathbf{v}} f(\mathbf{x}_0).$$

The following theorem explains the notation:

Theorem 6.5.2 ([1, §6.6 定理]). *Let $D \subset \mathbb{R}^n$ be a region, $\mathbf{x}_0 \in D$ a point, $f : D \to \mathbb{R}$ a function differentiable at $\mathbf{x}_0$. Let $\nabla f = df(\mathbf{x}_0, \cdot)$ be the gradient of f at $\mathbf{x}_0$. Then*

$$\nabla_{\mathbf{v}} f = df(\mathbf{x}_0, \mathbf{v}) = (\nabla f)(\mathbf{v}) = (\nabla f) \cdot \mathbf{v}$$

for any $\mathbf{v} \in \mathbb{R}^n$.

证明. By the definition of the differential df, we have

$$f(\mathbf{x}_0 + t\mathbf{v}) - f(\mathbf{x}_0) = df(\mathbf{x}_0, t\mathbf{v}) + o(t\mathbf{v}), \quad t \to 0.$$

It follows that $\nabla_{\mathbf{v}} f = df(\mathbf{x}_0, \mathbf{v})$. □

Note in particular that, for any $\mathbf{v} \in \mathbb{R}^n$, we have

$$\nabla_{-\mathbf{v}} f = (\nabla f) \cdot (-\mathbf{v}) = -(\nabla f) \cdot \mathbf{v} = -\nabla_{\mathbf{v}} f.$$

Example 6.5.3 ([1, §6.6 例 3]). Let $u = xy + yz + zx$, and $\mathbf{v} = (1, 3, 1)$. Find $\partial_{\mathbf{v}} u(1, 1, 1)$.

证明. One calculates

$$\begin{aligned} \mathbf{v}^\circ &= \frac{1}{\|\mathbf{v}\|}(1,3,1) = \frac{1}{\sqrt{11}}(1,3,1), \\ \nabla f(1,1,1) &= \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} \end{bmatrix}(1,1,1) \\ &= \begin{bmatrix} y+z & x+z & x+y \end{bmatrix}(1,1,1) = \begin{bmatrix} 2 & 2 & 2 \end{bmatrix}. \end{aligned}$$

Then we have

$$\partial_{\mathbf{v}} u(1,1,1) = (\nabla f) \cdot \mathbf{v}^\circ = \begin{bmatrix} 2 & 2 & 2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{11}} \\ \frac{3}{\sqrt{11}} \\ \frac{1}{\sqrt{11}} \end{bmatrix} = \frac{10}{\sqrt{11}}.$$

□

Note also that by Theorem 6.5.2 and the Cauchy-Schwarz inequality, we have

$$|\nabla_{\mathbf{v}} f| \leq \|\nabla f\| \cdot \|\mathbf{v}\|.$$

The equality holds if and only if ∇f and $\mathbf{v}$ are linearly dependent. In particular, for a unit vector $\mathbf{v} = \mathbf{v}^{\circ}$, $\nabla_{\mathbf{v}} f$ attains the maximum if ∇f and $\mathbf{v}$ are in the same direction, and the minimum if ∇f and $\mathbf{v}$ are in the opposite direction.

Example 6.5.4 (Electric potential)**.** Let q be a point charge at the origin. Then the potential of the electrostatic field at point (x, y, z) is

$$V = \frac{q}{4\pi\varepsilon} \cdot \frac{1}{\sqrt{x^2 + y^2 + z^2}}$$

where ε is the permittivity. Then one calculates

$$\nabla V = (V_x, V_y, V_z) = \frac{q}{4\pi\varepsilon} \cdot \frac{-1}{(x^2 + y^2 + z^2)^{\frac{3}{2}}}(x, y, z)$$

where the right hand side is equal to the opposite value of the electric field $\mathbf{E}$ at (x, y, z). This is analogous to the gravitational potential. The fact that $\nabla V = -\mathbf{E}$ indicates that V increases most quickly in the direction $-\mathbf{E}$.

By Proposition 6.4.32, we immediately obtain:

Proposition 6.5.5. *Let $D \subset \mathbb{R}^n$ be a region, and let $u, v : D \to \mathbb{R}$ be differentiable functions on D. Then we have*

(1) $\nabla(u \pm v) = \nabla u \pm \nabla v$,

(2) (product rule) $\nabla(u \cdot v) = u\nabla v + v\nabla u$,

(3) $\nabla\left(\frac{u}{v}\right) = \frac{v\nabla u - u\nabla v}{v^2}$,

(4) (chain rule) $\nabla f(u, v) = f_u \nabla u + f_v \nabla v$ *for differentiable* $f : \mathbb{R}^2 \to \mathbb{R}$.

6.6 Mean value theorem for multivariate functions

Theorem 6.6.1 (Lagrange's mean value theorem, [1, §6.7 定理 1])**.** *Let $D \subset \mathbb{R}^n$ be a region, $f : D \to \mathbb{R}$ a function and let*

$$[\mathbf{x}, \mathbf{x} + \mathbf{h}] = \{\mathbf{v} \in \mathbb{R}^n : \mathbf{v} = \mathbf{x} + t\mathbf{h},\ t \in [0, 1]\} \subset D$$

be a closed line segment with endpoints $\mathbf{x}, \mathbf{x} + \mathbf{h} \in D$. Suppose f is continuous on $[\mathbf{x}, \mathbf{x}+\mathbf{h}]$, and differentiable on $(\mathbf{x}, \mathbf{x}+\mathbf{h}) = [\mathbf{x}, \mathbf{x}+\mathbf{h}] \setminus \{\mathbf{x}, \mathbf{x}+\mathbf{h}\}$. Then there exists a time $\theta \in (0, 1)$ such that

$$f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) = df(\mathbf{x} + \theta\mathbf{h}, \mathbf{h}). \tag{6.6.1}$$

证明. Let $\Gamma : [0, 1] \to [\mathbf{x}, \mathbf{x} + \mathbf{h}]$ be the line given by $t \mapsto \mathbf{x} + t\mathbf{h}$. Then Γ is differentiable on $(0, 1)$. It follows that $f \circ \Gamma : [0, 1] \to \mathbb{R}$ is continuous on $[0, 1]$ and differentiable on $(0, 1)$. Then by Lagrange's mean value theorem (Theorem 4.1.6), there is a time $\theta \in (0, 1)$ such that

$$f(\Gamma(1)) - f(\Gamma(0)) = (f \circ \Gamma)'(\theta). \tag{6.6.2}$$

Since $f' = \mathbf{J}_{f\circ\Gamma}$, by the chain rule (Theorem 6.4.23), we have

$$(f \circ \Gamma)'(\theta) = \mathbf{J}_{f\circ\Gamma}(\theta) = \mathbf{J}_f(\Gamma(\theta)) \cdot \mathbf{J}_\Gamma(\theta) = \mathbf{J}_f(\Gamma(\theta))\mathbf{h} = df(\mathbf{x} + \theta\mathbf{h}, \mathbf{h}). \tag{6.6.3}$$

Combining (6.6.2) with (6.6.3), we conclude (6.6.1). □

In particular, write $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{h} = (h_1, \ldots, h_n)$. Then (6.6.1) becomes

$$\begin{aligned} f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) = df(\mathbf{x} + \theta\mathbf{h}, \mathbf{h}) &= \begin{bmatrix} f_{x_1}(\mathbf{x} + \theta\mathbf{h}) & \cdots & f_{x_n}(\mathbf{x} + \theta\mathbf{h}) \end{bmatrix} \begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix} \\ &= h_1 f_{x_1}(\mathbf{x} + \theta\mathbf{h}) + \cdots + h_n f_{x_n}(\mathbf{x} + \theta\mathbf{h}). \end{aligned}$$

One may deduce the high-dimensional analogy of Corollary 4.1.7:

Corollary 6.6.2 ([1, §6.7 推论])**.** *Let $D \subset \mathbb{R}^n$ be a region, and let $f \in C^1(D)$ be a function so that $\mathbf{J}_f \equiv 0$ on D. Then $f \equiv C$ for some $C \in \mathbb{R}$.*

证明. Let $P_0, P \in D$ be two points. Since D is connected, there is a curve $\Gamma : [0,1] \to D$ such that $\Gamma(0) = P_0$, $\Gamma(1) = P$. Define

$$T = \{t \in [0,1] : f(\Gamma(t)) = f(\Gamma(0))\}.$$

Note that $0 \in T$, and $T \subset [0,1]$. Then by the least-upper-bound property (Theorem 7.1.7), let $t_m = \sup(T) \in [0,1]$. Suppose that $t_m \neq 1$. Then since $\Gamma(t_m) \in D$ is an interior point of D, there is a small neighborhood $U_r(\Gamma(t_m)) \subset D$. Further, there is a time $t' \in (t_m, 1]$ such that $\Gamma(t') \in U_r(\Gamma(t_m))$. It follows that the line $[\Gamma(t_m), \Gamma(t')] \subset U_r(\Gamma(t_m)) \subset D$. By the Lagrange's mean value theorem (Theorem 6.6.1), we conclude that

$$f(\Gamma(t')) - f(\Gamma(t_m)) = 0.$$

This contradicts that $t_m = \sup(T)$, and so $t_m = 1$. □

6.7 Taylor's theorem for multivariate functions

Next, we discuss the **best polynomial approximation** of a multivariate function. It is instructive to review Theorems 4.3.12, 4.3.14 and 4.3.20.

Theorem 6.7.1 (Taylor's theorem, [1, §6.7 定理 2, 3])**.** *Let $D \subset \mathbb{R}^n$ a region, $\mathbf{x}_0 \in D$, $m \in \mathbb{N}$, and $f : D \to \mathbb{R}$ be m times differentiable at $\mathbf{x}_0$. Then for $\mathbf{x} = \mathbf{x}_0 + \mathbf{h}$ near $\mathbf{x}_0$, $f(\mathbf{x})$ can be approximated by a polynomial $P_m(\mathbf{x})$ of order m in the sense that*

$$\lim_{\mathbf{x} \to \mathbf{x}_0} \frac{|f(\mathbf{x}) - P_m(\mathbf{x})|}{\|\mathbf{x} - \mathbf{x}_0\|^m} = 0. \tag{6.7.1}$$

The polynomial is given by

$$P_m(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + \sum_{k=1}^{m} \frac{1}{k!} d^k f(\mathbf{x}_0, \mathbf{h}). \tag{6.7.2}$$

Moreover, $P_m(\mathbf{x})$ is the unique polynomial of order m that satisfies (6.7.1). This is the Taylor's polynomial.

Further, if $f \in C^{m+1}(D)$, then the remainder term $r_m = f - P_m$ has the expression:

$$r_m(\mathbf{x}_0 + \mathbf{h}) = \frac{1}{(m+1)!} d^{m+1} f(\mathbf{x}_0 + \theta \mathbf{h}, \mathbf{h}) \tag{6.7.3}$$

for some $\theta = \theta(\mathbf{x}_0, \mathbf{h}) \in (0,1)$. This is the Lagrange form of the remainder.

证明. Since $\mathbf{x}_0 \in D$ is an interior point in D, $U_r(\mathbf{x}_0) \subset D$ for some $r > 0$. Then for any $\mathbf{x} = \mathbf{x}_0 + \mathbf{h} \in D$, the line segment $\Gamma : [-1,1] \to [\mathbf{x}_0 - \mathbf{h}, \mathbf{x}_0 + \mathbf{h}] \subset U_r(\mathbf{x}_0) \subset D$. Then $f \circ \Gamma : [-1,1] \to \mathbb{R}$ is m times differentiable at 0. Then we apply Theorem 4.3.12 and obtain the taylor's formula

$$f \circ \Gamma(t) = \sum_{k=0}^{m} \frac{(f \circ \Gamma)^{(k)}(0)}{k!} t^k + \tilde{r}_n(t). \tag{6.7.4}$$

Note that $f(\mathbf{x}_0 + s\mathbf{h}) = f \circ \Gamma(s)$. Now suppose that $f \circ \Gamma$ is $m-1$ times differentiable on a neighborhood of 0. Moreover, for $k \leq m-1$, we assume

$$d^{k-1} f(\Gamma(s), t\mathbf{h}) = (f \circ \Gamma)^{(k-1)}(s) t^{k-1}. \tag{6.7.5}$$

Then one calculates

$$\begin{aligned} d^k f(\Gamma(s), t\mathbf{h}) &= \mathbf{J}_{d^{k-1}f(\cdot, t\mathbf{h})}(\Gamma(s), t\mathbf{h}) = \mathbf{J}_{d^{k-1}f(\cdot, t\mathbf{h})}(\Gamma(s)) \cdot \mathbf{J}_\Gamma(s) t \\ &= \mathbf{J}_{d^{k-1}f(\Gamma(\cdot), t\mathbf{h})}(s) t = \mathbf{J}_{(f\circ\Gamma)^{(k-1)}(\cdot) t^{k-1}}(s) t = (f \circ \Gamma)^{(k)}(s) t^k. \end{aligned}$$

Thus, by induction, (6.7.5) holds for all $0 \leq k \leq m-1$. Finally, since f is m times differentiable at $\mathbf{x}_0$, a similar argument implies

$$\begin{aligned} d^m f(\mathbf{x}_0, t\mathbf{h}) &= \mathbf{J}_{d^{m-1}f(\cdot, t\mathbf{h})}(\Gamma(0), t\mathbf{h}) = \mathbf{J}_{d^{m-1}f(\cdot, t\mathbf{h})}(\Gamma(0)) \cdot \mathbf{J}_\Gamma(0) t \\ &= \mathbf{J}_{d^{m-1}f(\Gamma(\cdot), t\mathbf{h})}(0) t = \mathbf{J}_{(f\circ\Gamma)^{(m-1)}(\cdot) t^{m-1}}(0) t = (f \circ \Gamma)^{(m)}(0) t^m. \end{aligned} \tag{6.7.6}$$

Then (6.7.4) and (6.7.6) implies (6.7.2).

The proof of the uniqueness of the Taylor's polynomial is the same as Theorem 4.3.14.

Now we show the Lagrange form of the remainder. If $f \in C^{m+1}(D)$, then (6.7.5) holds for all $0 \leq k \leq m+1$. Also, by Theorem 4.3.20, the remainder term has the expression:

$$\tilde{r}_m(t) = \frac{(f \circ \Gamma)^{(m+1)}(\theta)}{(m+1)!} t^{m+1}.$$

for some $\theta \in (0,t)$. Then by (6.7.5), we have

$$\tilde{r}_m(t) = \frac{1}{(m+1)!} d^{m+1} f(\mathbf{x}_0 + \theta\mathbf{h}, t\mathbf{h})$$

for some $\theta \in (0,t)$. Thus we establish (6.7.3). $\square$

In particular, if we write $\mathbf{h} = (dx_1, \ldots, dx_n)$, then via (6.4.4), we can rewrite (6.7.2) and (6.7.3) as

$$P_m(\mathbf{x}_0 + \mathbf{h}) = \sum_{k=1}^{m} \frac{1}{k!} \left(dx_1 \frac{\partial}{\partial x_1} + \cdots + dx_n \frac{\partial}{\partial x_n} \right)^k f(\mathbf{x}_0),$$
$$r_m(\mathbf{x}_0 + \mathbf{h}) = \frac{1}{(m+1)!} \left(dx_1 \frac{\partial}{\partial x_1} + \cdots + dx_n \frac{\partial}{\partial x_n} \right)^{m+1} f(\mathbf{x}_0 + \theta \mathbf{h}).$$

Remark 6.7.2. Suppose that there is a constant $M > 0$ such that

$$\left| \frac{\partial^{m+1} f}{\partial x_1^{m_1} \partial x_2^{m_2} \cdots \partial x_n^{m_n}} \right| < M$$

for any $m_1 + \cdots + m_n = m + 1$, $m \in \mathbb{N}$. Then we have

$$|r_m(\mathbf{x}_0 + \mathbf{h})| \leq \frac{M}{(m+1)!} (|dx_1| + \cdots + |dx_n|)^{m+1} = \frac{M \cdot \|\mathbf{h}\|_1^{m+1}}{(m+1)!} \to 0, \quad m \to \infty.$$

Thus, f has a Taylor series expansion.

Example 6.7.3 ([1, §6.7 例 1])**.** Find the Taylor's polynomial of $f(x, y) = \sin(\frac{\pi}{2} x^2 y)$ of order 2 at $(1, 1)$ with the Peano form of the remainder.

证明. One calculates

$$f(1,1) = 1, \quad f_x(1,1) = 0, \quad f_y(1,1) = 0,$$

$$f_{xx}(1,1) = -\pi^2, \quad f_{xy}(1,1) = -\frac{\pi^2}{2}, \quad f_{yy}(1,1) = -\frac{\pi^2}{4}.$$

Let $dx = x - 1$, $dy = y - 1$. Then we have

$$\sin\left(\frac{\pi}{2} x^2 y\right) = 1 - \frac{\pi^2}{2}(x-1)^2 - \frac{\pi^2}{2}(x-1)(y-1) - \frac{\pi^2}{8}(y-1)^2$$
$$+ o((x-1)^2 + (y-1)^2), \quad x \to 1, \ y \to 1.$$

□

Example 6.7.4 ([1, §6.7 例 3])**.** Find the Taylor's polynomial of $f(x, y) = e^x \cos y$ of order 2 at $(0, 0)$ with the Peano form of the remainder.

证明. One calculates

$$
\begin{aligned}
e^x &= 1 + x + \frac{1}{2}x^2 + o(x^2), \quad x \to 0, \\
\cos y &= 1 - \frac{1}{2}y^2 + o(y^2), \quad y \to 0.
\end{aligned}
$$

Via a multiplication, we have

$$
e^x \cos y = 1 + x + \frac{1}{2}x^2 - \frac{1}{2}y^2 + o(x^2 + y^2), \quad x \to 0,\ y \to 0.
$$

Then by the uniqueness, we obtain the Taylor's polynomial. □

6.8 Implicit function theorem

6.8.1 Introduction

In this section, we shall prove the implicit function theorem. This is a fundamental theorem for solving equations of functions. We start with a simple one:

$$
x^2 + y^2 - 1 = 0, \quad (x, y) \in \mathbb{R}^2.
$$

The points that satisfies the equation form a circle. While the entire circle is not a graph of a function of a single variable, certain **local portions** of it are. Thus, the equation **locally** determines an implicit function. The implicit function theorem is the generalization of this observation.

Let $F : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a function defined by

$$
F : (x, y) \mapsto F(x, y).
$$

Then we consider the equation

$$
F(x, y) = 0 \tag{6.8.1}
$$

and ask whether the equation actually indicates a graph of some function, and whether that function is continuous or differentiable. Geometrically, equation (6.8.1) is the intersection of the plane $z = 0$ and the graph of the function F:

$$
z = F(x, y).
$$

To answer these questions, we first consider the simplest example

$$F(x, y) = ax + by + c.$$

Then $F_x = a$ and $F_y = b$. By (6.8.1), one calculates

$$-F_y y = F_x x + c.$$

Now suppose that $F_y \neq 0$. Then the graph $z = F(x, y)$ is not parallel to the plane $z = 0$, and so intersects $z = 0$ at at least one point (x_0, y_0). Moreover, in a neighborhood of (x_0, y_0), the equation (6.8.1) determines an implicit function

$$f(x) = -\frac{F_x}{F_y}x - \frac{c}{F_y}.$$

Further, near (x_0, y_0), one calculates the derivative of the implicit function

$$f'(x) = -\frac{F_x}{F_y}. \tag{6.8.2}$$

Now consider $F(x, y) \in C^1(D)$ such that $F(x_0, y_0) = 0$ and $F_y(x_0, y_0) \neq 0$. Then near (x_0, y_0), F can be approximated by the tangent plane. Heuristically, the above results then imply that there is an implicit function $f(x)$ satisfies the equation (6.8.1) in a neighborhood of (x_0, y_0), provided $F_y(x_0, y_0) \neq 0$. Further, the derivative f' can be found via (6.8.2).

Now we are in the position to state the implicit function theorem. Let $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ be a function defined by

$$F : (\mathbf{x}, \mathbf{y}) \mapsto F(\mathbf{x}, \mathbf{y})$$

or write

$$F : (x_1, \ldots, x_n, y_1, \ldots, y_m) \mapsto (f_1, \ldots, f_m).$$

Then the Jacobian of F is given by

$$\mathbf{J}_F = \left[\begin{array}{ccc|ccc} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} & \frac{\partial f_1}{\partial y_1} & \cdots & \frac{\partial f_1}{\partial y_m} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_m}{\partial x_m} & \cdots & \frac{\partial f_m}{\partial x_n} & \frac{\partial f_m}{\partial y_1} & \cdots & \frac{\partial f_m}{\partial y_m} \end{array}\right] = [F_{\mathbf{x}}, F_{\mathbf{y}}].$$

Note that $F_{\mathbf{x}}$ is an $(m \times n)$-matrix, and $F_{\mathbf{y}}$ is an $(m \times m)$-matrix (i.e. a square matrix).

Theorem 6.8.1 (Implicit function theorem, [1, §6.8 定理 1, 2, 3])**.** *Let $D \subset \mathbb{R}^n \times \mathbb{R}^m$ be a region, $(\mathbf{x}_0, \mathbf{y}_0) \in D$, and $F : D \to \mathbb{R}^m$ a function. Suppose that F satisfies:*

(1) $F(\mathbf{x}_0, \mathbf{y}_0) = \mathbf{0}$.

(2) $\mathbf{J}_F = [F_{\mathbf{x}}, F_{\mathbf{y}}]$ *is continuous, and* $\det F_{\mathbf{y}}(\mathbf{x}_0, \mathbf{y}_0) \neq \mathbf{0}$.

Then there exists a neighborhood $U_r(\mathbf{x}_0) \subset \mathbb{R}^n$, and a unique implicit function $\mathbf{y} = f(\mathbf{x}) : U_r(\mathbf{x}_0) \to \mathbb{R}^m$, such that:

(1) $\mathbf{y}_0 = f(\mathbf{x}_0)$.

(2) For $\mathbf{x} \in U_r(\mathbf{x}_0)$, we have

$$F(\mathbf{x}, f(\mathbf{x})) = \mathbf{0}. \tag{6.8.3}$$

(3) The derivatives of f are continuous on $U_r(\mathbf{x}_0)$, and can be solved by differentiating the equation (6.8.3):

$$\mathbf{0} = \mathbf{J}_{F,\mathbf{x}} = F_{\mathbf{x}}(\mathbf{x}, f(\mathbf{x})) + F_{\mathbf{y}}(\mathbf{x}, f(\mathbf{x})) \cdot \mathbf{J}_f(\mathbf{x}), \quad \mathbf{x} \in U_r(\mathbf{x}_0)$$

where $\mathbf{J}_{F,\mathbf{x}}$ denotes the Jacobian of the map $\mathbf{x} \mapsto (\mathbf{x}, f(\mathbf{x})) \mapsto F(\mathbf{x}, f(\mathbf{x}))$. In other words, we have

$$\mathbf{J}_f(\mathbf{x}) = -(F_{\mathbf{y}}(\mathbf{x}, f(\mathbf{x})))^{-1} \cdot F_{\mathbf{x}}(\mathbf{x}, f(\mathbf{x})), \quad \mathbf{x} \in U_r(\mathbf{x}_0). \tag{6.8.4}$$

Example 6.8.2 ([1, §6.8 例 3])**.** Let $F(u, v) \in C^1(D)$. Find the partial derivatives of the implicit function $z = z(x, y)$ determined by the equation

$$F(x - y, y - z) = 0.$$

证明. Let $G(x, y, z) = F(x - y, y - z)$. There are two approaches. One may directly solve the equation $\mathbf{0} = \mathbf{J}_{G,z}$, where

$$\mathbf{J}_{G,z} = \begin{bmatrix} G_x + G_z \cdot z_x & G_y + G_z \cdot z_y \end{bmatrix} = \begin{bmatrix} F_u - F_v \cdot z_x & -F_u + F_v - F_v \cdot z_y \end{bmatrix}.$$

For the other method, one calculates

$$G_z = -F_v, \quad G_{(x,y)} = [F_u, -F_u + F_v].$$

Then (6.8.4) implies

$$[z_x, z_y] = \mathbf{J}_z = -G_z^{-1} G_{(x,y)} = \frac{1}{F_v}[F_u, -F_u + F_v].$$

$\square$

Example 6.8.3 ([1, §6.8 例 5]). Consider the system of equations:

$$\begin{cases} x^2 + y^2 - uv = 0 \\ xy + u^2 - v^2 = 0 \end{cases}.$$

Discuss whether there is an implicit function $(u, v) = f(x, y)$ determined by the system. Find u_x, u_y, v_x, v_y if possible.

证明. Let $F = (F_1, F_2)$, where $F_1(x, y, u, v) = x^2 + y^2 - uv$, $F_2(x, y, u, v) = xy + u^2 - v^2$. Then

$$\det F_{(u,v)} = \frac{D(F_1, F_2)}{D(u, v)} = \det \begin{bmatrix} -v & -u \\ 2u & -2v \end{bmatrix} = 2(u^2 + v^2).$$

Thus, if $(x, y) \neq (0, 0)$, then $(u, v) \neq (0, 0)$, and so $\det \mathbf{J}_F \neq 0$. Then by the implicit function theorem (Theorem 6.8.1), the system determines an implicit function

$$f(x, y) = (u(x, y), v(x, y)).$$

To see the partial derivatives, there are again two approaches. One may solve the equation $\mathbf{0} = \mathbf{J}_{F,(x,y)}$, where

$$\mathbf{J}_{F,(x,y)} = \begin{bmatrix} 2x - u_x v - u v_x & 2y - u_y v - u v_y \\ x + 2u u_x - 2v v_x & y + 2u u_y - 2v v_y \end{bmatrix}.$$

The second method is to apply the formula (6.8.4). Thus, one calculates

$$F_{(u,v)}^{-1} = \frac{1}{2(u^2 + v^2)} \begin{bmatrix} -2v & u \\ -2u & -v \end{bmatrix}, \quad F_{(x,y)} = \begin{bmatrix} 2x & 2y \\ y & x \end{bmatrix}.$$

Then (6.8.4) implies

$$\begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix} = -F_{(u,v)}^{-1} F_{(x,y)} = \frac{-1}{2(u^2 + v^2)} \begin{bmatrix} -4vx + uy & -4vy + ux \\ -4ux - vy & -4uy - vx \end{bmatrix}.$$

$\square$

Example 6.8.4 ([1, §6.8 例 6])**.** Consider the system of equations:

$$\begin{cases} x = e^u + v \\ xy = e^u + u \end{cases}. \tag{6.8.5}$$

It determines an implicit function $(u, v) = (u(x, y), v(x, y))$. Find u_x, u_{xx}.

证明*.* Let $F = (F_1, F_2)$, where $F_1(x, y, u, v) = x - e^u - v$, $F_2(x, y, u, v) = xy - e^u - u$. Then

$$\det F_{(u,v)} = \frac{D(F_1, F_2)}{D(u, v)} = \det \begin{bmatrix} -e^u & -1 \\ -e^u - 1 & 0 \end{bmatrix} = e^u + 1.$$

Then by the implicit function theorem (Theorem 6.8.1), the system determines an implicit function

$$f(x, y) = (u(x, y), v(x, y)).$$

To see the partial derivatives, one differentiates both sides of (6.8.5) by x:

$$\begin{cases} 1 = e^u u_x + v_x \\ y = e^u u_x + u_x \end{cases}. \tag{6.8.6}$$

It follows that

$$u_x = \frac{y}{e^u + 1}.$$

Differentiate both sides of (6.8.6) by x again:

$$\begin{cases} 0 = e^u u_x u_x + e^u u_{xx} + v_{xx} \\ 0 = e^u u_x u_x + e^u u_{xx} + u_{xx} \end{cases}.$$

We get

$$u_{xx} = \frac{-e^u u_x u_x}{e^u + 1} = \frac{-e^u y^2}{(e^u + 1)^3}.$$

□

6.8.2 Inverse function theorem

A direct consequence of the implicit function theorem (Theorem 6.8.1) is the following important and useful theorem:

Theorem 6.8.5 (Inverse function theorem, [1, §6.8 定理 4]). *Let $D \subset \mathbb{R}^n$ be a region, $\mathbf{x}_0 \in D$, and let $f : D \to \mathbb{R}^n$ be continuous differentiable on D. Suppose that $\mathbf{J}_f(\mathbf{x}_0)$ is invertible. Then there is a neighborhood $U = U_s(f(\mathbf{x}_0)) \subset \mathbb{R}^n$ such that*

$$f : f^{-1}(U) \to U$$

is bijective, where $f^{-1}(U) = \{\mathbf{x} \in D : f(\mathbf{x}) \in U\}$. In particular, there is an inverse map $f^{-1} : U \to f^{-1}(U)$. Moreover, this inverse map is differentiable on U, and

$$\mathbf{J}_{f^{-1}}(f(\mathbf{x})) = (\mathbf{J}_f(\mathbf{x}))^{-1}, \quad \mathbf{x} \in f^{-1}(U).$$

Remark 6.8.6. As we shall see in the proofs, the inverse function theorem does not strongly rely on the underlying space $\mathbb{R}^n$. There are also versions of this theorem for *analytic functions* (解析函数), for differentiable maps between *manifolds* (流形), for differentiable functions between *Banach spaces* (巴拿赫空间, Definition 7.2.16), and so forth.

The inverse function theorem is an implication of the implicit function theorem.

Proof of Theorem 6.8.5 via Theorem 6.8.1. Let $F(\mathbf{y}, \mathbf{x}) = \mathbf{y} - f(\mathbf{x})$. Then by the implicit function theorem (Theorem 6.8.1), the equation

$$\mathbf{y} - f(\mathbf{x}) = \mathbf{0}$$

determines an implicit function $g(\mathbf{y})$ on a neighborhood $U = U_s(f(\mathbf{x}_0))$. Then $\mathbf{y} = f(g(\mathbf{y}))$ means that f is bijective on $f^{-1}(U)$, and $g = f^{-1}$. Moreover, f^{-1} is differentiable on U, and so

$$I = \mathbf{J}_{f \circ f^{-1}}(f(\mathbf{x})) = \mathbf{J}_f(\mathbf{x})\mathbf{J}_{f^{-1}}(f(\mathbf{x}))$$

for $\mathbf{x} \in f^{-1}(U)$. $\square$

In fact, the inverse function theorem (Theorem 6.8.5) and the implicit function theorem (Theorem 6.8.1) are equivalent:

Proof of Theorem 6.8.1 via Theorem 6.8.5. Let $(\mathbf{x}_0, \mathbf{y}_0) \in D \subset \mathbb{R}^n \times \mathbb{R}^m$ and let $F(\mathbf{x}, \mathbf{y}) : D \to \mathbb{R}^m$ be as in Theorem 6.8.1. Let $G : D \to \mathbb{R}^n \times \mathbb{R}^m$ be defined by

$$G : (\mathbf{x}, \mathbf{y}) \mapsto (\mathbf{x}, F(\mathbf{x}, \mathbf{y})). \tag{6.8.7}$$

Then $G \in C^1(D)$ is continuously differentiable, and

$$G(\mathbf{x}_0, \mathbf{y}_0) = (\mathbf{x}_0, F(\mathbf{x}_0, \mathbf{y}_0)) = (\mathbf{x}_0, \mathbf{0}).$$

It follows that

$$\mathbf{J}_G(\mathbf{x}, \mathbf{y}) = \begin{bmatrix} I & 0 \\ F_{\mathbf{x}} & F_{\mathbf{y}} \end{bmatrix}$$

where I is the $(n \times n)$ identity matrix. Since $\det F_{\mathbf{y}}(\mathbf{x}_0, \mathbf{y}_0) \neq 0$, we have

$$\det \mathbf{J}_G(\mathbf{x}_0, \mathbf{y}_0) \neq 0$$

and so the matrix $\mathbf{J}_G(\mathbf{x}_0, \mathbf{y}_0)$ is invertible. Thus, by the inverse function theorem (Theorem 6.8.5), there is a neighborhood $U = U_s(G(\mathbf{x}_0, \mathbf{y}_0)) \subset \mathbb{R}^n$ such that

$$G : G^{-1}(U) \to U$$

is bijective, where $G^{-1}(U) = \{(\mathbf{x}, \mathbf{y}) \in D : G(\mathbf{y}) \in U\}$. In particular, there is an inverse map $G^{-1} : U \to G^{-1}(U)$. Moreover, G^{-1} is differentiable on U, and

$$\mathbf{J}_{G^{-1}}(G(\mathbf{x}, \mathbf{y})) = (\mathbf{J}_G(\mathbf{x}, \mathbf{y}))^{-1}, \quad (\mathbf{x}, \mathbf{y}) \in G^{-1}(U).$$

Now let $\pi_{\mathbf{x}} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$, $\pi_{\mathbf{y}} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ be the projections to the $\mathbf{x}$ and $\mathbf{y}$ coordinates, respectively. Let $g^{\mathbf{x}} = \pi_{\mathbf{x}} \circ G^{-1}$, $g^{\mathbf{y}} = \pi_{\mathbf{y}} \circ G^{-1}$. Then $g^{\mathbf{x}}$ and $g^{\mathbf{y}}$ are continuously differentiable, and $G^{-1} = (g^{\mathbf{x}}, g^{\mathbf{y}})$. Then by (6.8.7), we have

$$(\mathbf{x}, \mathbf{y}) = G^{-1}(G(\mathbf{x}, \mathbf{y})) = (g^{\mathbf{x}}(G(\mathbf{x}, \mathbf{y})), g^{\mathbf{y}}(G(\mathbf{x}, \mathbf{y}))), \quad (\mathbf{x}, \mathbf{y}) \in G^{-1}(U). \tag{6.8.8}$$

Observing the $\mathbf{x}$ coordinate of (6.8.8), we have

$$\mathbf{x} = g^{\mathbf{x}}(G(\mathbf{x}, \mathbf{y})), \quad (\mathbf{x}, \mathbf{y}) \in G^{-1}(U).$$

Differentiating both sides, we obtain

$$[I, \mathbf{0}] = \left[\partial_{\mathbf{x}} g^{\mathbf{x}} + (\partial_{\mathbf{y}} g^{\mathbf{x}}) F_{\mathbf{x}}, (\partial_{\mathbf{y}} g^{\mathbf{x}}) F_{\mathbf{y}}\right], \quad (\mathbf{x}, \mathbf{y}) \in G^{-1}(U).$$

Since $F_{\mathbf{y}}$ is invertible, we conclude that $\partial_{\mathbf{y}} g^{\mathbf{x}} = \mathbf{0}$, and so $g^{\mathbf{x}}(\mathbf{x}, \mathbf{y}) = \mathbf{x}$ for $(\mathbf{x}, \mathbf{y}) \in U$.

Then by (6.8.7) again, we have

$$(\mathbf{x}, \mathbf{y}) = G(G^{-1}(\mathbf{x}, \mathbf{y})) = (\mathbf{x}, F(G^{-1}(\mathbf{x}, \mathbf{y}))) = (\mathbf{x}, F(\mathbf{x}, g^{\mathbf{y}}(\mathbf{x}, \mathbf{y}))) \tag{6.8.9}$$

for $(\mathbf{x}, \mathbf{y}) \in U$. Observing the $\mathbf{y}$ coordinate of (6.8.9), we obtain

$$\mathbf{y} = F(\mathbf{x}, g^{\mathbf{y}}(\mathbf{x}, \mathbf{y})), \quad (\mathbf{x}, \mathbf{y}) \in U. \tag{6.8.10}$$

Now let $U_0 = \{\mathbf{x} \in \mathbb{R}^n : (\mathbf{x}, \mathbf{0}) \in U\}$. and define $f : U_0 \to \mathbb{R}^m$ by

$$f : \mathbf{x} \mapsto g^{\mathbf{y}}(\mathbf{x}, \mathbf{0}).$$

Then (6.8.10) implies

$$F(\mathbf{x}, f(\mathbf{x})) = \mathbf{0}, \quad \mathbf{x} \in U_0.$$

Note in particular that $\mathbf{x}_0 \in U_0$.

Finally, f is continuously differentiable on U_0. By the chain rule, we have

$$\mathbf{0} = F_{\mathbf{x}}(\mathbf{x}, f(\mathbf{x})) + F_{\mathbf{y}}(\mathbf{x}, f(\mathbf{x})) \cdot \mathbf{J}_f(\mathbf{x})$$

for $\mathbf{x} \in U_0$. Thus, we establish the implicit function theorem. □

Example 6.8.7. Consider the system of equations:

$$\begin{cases} u = e^x \cos y \\ v = e^x \sin y \end{cases}. \tag{6.8.11}$$

It determines an implicit function $(x, y) = (x(u, v), y(u, v))$ on a small neighborhood $U \subset \mathbb{R}^2 \setminus \{(0, 0)\}$. Find x_u, x_v, y_u, y_v.

证明. One calculates

$$\det \mathbf{J}_{(u,v)} = \frac{D(u, v)}{D(x, y)} = \det \begin{bmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{bmatrix} = e^{2x}.$$

Then by the inverse function theorem (Theorem 6.8.5), for any $(u_0, v_0) \in \mathbb{R}^2 \setminus \{(0, 0)\}$, the system determines an inverse function

$$(x, y) = (x(u, v), y(u, v))$$

on a neighborhood U of (u_0, v_0). To see the partial derivatives, we find the inverse matrix

$$\mathbf{J}_{(x,y)} = \mathbf{J}_{(u,v)}^{-1} = e^{-2x} \begin{bmatrix} e^x \cos y & e^x \sin y \\ -e^x \sin y & e^x \cos y \end{bmatrix} = \begin{bmatrix} e^{-x} \cos y & e^{-x} \sin y \\ -e^{-x} \sin y & e^{-x} \cos y \end{bmatrix}.$$

□

Example 6.8.8. Let $F : (x, y) \mapsto (u(x, y), v(x, y))$ be defined by

$$\begin{cases} u = x + y - 1 \\ v = x^2 + 2xy + y^2 + 3y + 2 \end{cases}. \tag{6.8.12}$$

Discuss whether there is an inverse map of F.

证明. One calculates

$$\det \mathbf{J}_F = \frac{D(u, v)}{D(x, y)} = \det \begin{bmatrix} 1 & 1 \\ 2x + 2y & 2x + 2y + 3 \end{bmatrix} = 3. \tag{6.8.13}$$

Then by the inverse function theorem (Theorem 6.8.5), the system (6.8.12) locally determines an inverse function.

In fact, one may check that $G : (u, v) \mapsto (x, y)$ defined by

$$\begin{cases} x = \frac{1}{3}u^2 + \frac{5}{3}u - \frac{1}{3}v + 2 \\ y = -\frac{1}{3}u^2 - \frac{2}{3}u + \frac{1}{3}v - 1 \end{cases} \tag{6.8.14}$$

is the inverse of F on $\mathbb{C}^2$. □

As shown in Example 6.8.8, if $F = (F_1, \ldots, F_n) : \mathbb{C}^n \to \mathbb{C}^n$ is a *polynomial map* (i.e. each F_i is a polynomial in n variables), then its jacobian determinant is a polynomial. Suppose further that F has a global inverse $F^{-1} : \mathbb{C}^n \to \mathbb{C}^n$. Then by the chain rule, $\det \mathbf{J}_F \neq 0$ on $\mathbb{C}^n$. This is possible only when $\det \mathbf{J}_F$ is a nonzero constant. Thus, (6.8.13) is not a coincidence.

In 1939, Keller asked: is the converse true? This is a very difficult problem in the field of *algebraic geometry* (代数几何), now known as the *Jacobian conjecture* (雅可比猜想). As of today, there are no plausible claims to have proved it. Even the two-variable case has resisted all efforts.

Conjecture 6.8.9 (Jacobian conjecture)**.** *Let* $F = (F_1, \dots, F_n) : \mathbb{C}^n \to \mathbb{C}^n$ *be a polynomial map, i.e. each* F_i *is a polynomial in* n *variables. Suppose that* $\det \mathbf{J}_F$ *is a nonzero constant. Then* F *has a global polynomial inverse* $G = (G_1, \dots, G_n)$*, i.e.* $F \circ G = I$ *on* $\mathbb{C}^n$*, and each* G_i *is a polynomial in* n *variables.*

6.8.3 Proofs

In this section, we discuss the proofs of implicit and inverse function theorems.

We first show the existence of the implicit function by a very intuitive method. However, the method is not very constructive, since it is adapted only to the case of real-valued functions of real variables.

Sketch proof of Theorem 6.8.1 for $n = m = 1$. We only show the existence of the implicit function. Consider $F(x, y) \in C^1(D)$ such that $F(x_0, y_0) = 0$ and $F_y(x_0, y_0) > 0$. Then for $(x, y) \in U_r(x_0) \times U_r(y_0)$, $F_y(x, y) > 0$ as well. In particular, we have

$$F(x_0, y_0 - h) < 0, \quad F(x_0, y_0 + h) > 0,$$

for $h \in (0, \delta)$. Using the continuity of F, by choosing a smaller $r > 0$ if necessary, we have

$$F(x, y - h) < 0, \quad F(x, y + h) > 0,$$

for $(x, y) \in U_r(x_0) \times U_r(y_0)$. Thus, by the monotonicity and the continuity, for any x near x_0, there is a unique $y = f(x)$ near y_0 such that

$$F(x, f(x)) = 0.$$

In other words, we find an implicit function of the equation (6.8.1) in a small neighborhood of x_0. □

Next, we give a complete proof of the inverse function theorem (Theorem 6.8.5). We shall use the idea of the *Banach fixed-point theorem* (Theorem 7.2.18).

Lemma 6.8.10. *Let* $U_r(\mathbf{0}) \subset \mathbb{R}^n$ *be a ball centered at* $\mathbf{0}$ *with radius* $r \in (0, 1)$*, and let* $g : U_r(\mathbf{0}) \to \mathbb{R}^n$ *be a map such that* $g(\mathbf{0}) = \mathbf{0}$ *and*

$$\|g(\mathbf{x}) - g(\mathbf{y})\| \le \frac{1}{3}\|\mathbf{x} - \mathbf{y}\|$$

for any $\mathbf{x}, \mathbf{y} \in U_r(\mathbf{0})$. *Then the function* $f : U_r(\mathbf{0}) \to \mathbb{R}^n$ *defined by*

$$f : \mathbf{x} \mapsto \mathbf{x} + g(\mathbf{x})$$

is injective. Moreover, $f(U_r(\mathbf{0})) \supset U_{r/2}(\mathbf{0})$.

证明. Now assume that there are $\mathbf{x}, \mathbf{y} \in U_r(\mathbf{0})$ so that $f(\mathbf{x}) = f(\mathbf{y})$. Then $\mathbf{x} - \mathbf{y} = g(\mathbf{y}) - g(\mathbf{x})$ and so

$$\|\mathbf{y} - \mathbf{x}\| = \|g(\mathbf{y}) - g(\mathbf{x})\| \leq \frac{1}{3}\|\mathbf{y} - \mathbf{x}\|.$$

It forces $\mathbf{x} = \mathbf{y}$, and thus f is injective.

Next we shall show that $f(U_r(\mathbf{0})) \supset U_{r/2}(\mathbf{0})$. Let $\mathbf{y} \in U_{r/2}(\mathbf{0})$, $G(\mathbf{x}) = \mathbf{y} - g(\mathbf{x})$. Then for any $\mathbf{x} \in \overline{U_{\frac{5}{6}r}(\mathbf{0})}$,

$$\|G(\mathbf{x})\| \leq \|\mathbf{y}\| + \|g(\mathbf{x})\| \leq \frac{r}{2} + \|g(\mathbf{x}) - g(\mathbf{0})\| \leq \frac{r}{2} + \frac{1}{3}\|\mathbf{x} - \mathbf{0}\| \leq \frac{5}{6}r.$$

Thus, $G : \overline{U_{\frac{5}{6}r}(\mathbf{0})} \to \overline{U_{\frac{5}{6}r}(\mathbf{0})}$. Also, for $\mathbf{x}, \mathbf{y} \in \overline{U_{\frac{5}{6}r}(\mathbf{0})}$, we have

$$\|G(\mathbf{x}) - G(\mathbf{y})\| = \|g(\mathbf{x}) - g(\mathbf{y})\| \leq \frac{1}{3}\|\mathbf{x} - \mathbf{y}\|.$$

Then G is a strict contraction on $\overline{U_{\frac{5}{6}r}(\mathbf{0})}$. Then *Banach fixed-point theorem* (Theorem 7.2.18) asserts that G has a unique fixed point $\mathbf{x}' \in \overline{U_{\frac{5}{6}r}(\mathbf{0})}$, i.e. $\mathbf{x}' = \mathbf{y} - g(\mathbf{x}')$. Thus, we see that $f(\mathbf{x}') = \mathbf{y}$, and so $f(U_r(\mathbf{0})) \supset U_{r/2}(\mathbf{0})$. □

Now we are in the position to prove the inverse function theorem.

Proof of Theorem 6.8.5. By replacing f by $\mathbf{x} \mapsto f(\mathbf{x} + \mathbf{x}_0) - f(\mathbf{x}_0)$ if necessary, we may assume that $\mathbf{x}_0 = \mathbf{0}$ and $f(\mathbf{x}_0) = \mathbf{0}$. Moreover, by replacing f by $(\mathbf{J}_f(\mathbf{0}))^{-1} f$, we assume that $\mathbf{J}_f(\mathbf{0}) = I$.

Let $g : D \to \mathbb{R}^n$ be defined by

$$g : \mathbf{x} \mapsto f(\mathbf{x}) - \mathbf{x}.$$

Then $g(\mathbf{x}) = \mathbf{0}$, $\mathbf{J}_g(\mathbf{0}) = \mathbf{0}$. Then since f and g are continuously differentiable, there exists a radius $r \in (0, 1)$ such that

$$\|\mathbf{J}_g(\mathbf{x})\|_\infty = \max_{i,j} |\partial_{x_j}(g_i)(\mathbf{x})| \leq \frac{1}{3n}, \quad \det \mathbf{J}_f(\mathbf{x}) > 0$$

for all $\mathbf{x} \in U_r(\mathbf{0})$. In particular, for any $\mathbf{x} \in U_r(\mathbf{0})$, $\mathbf{h} = (h_1, \dots, h_n) \in \mathbb{R}^n$, we have

$$\|dg(\mathbf{x}, \mathbf{h})\|_\infty \leq \|\mathbf{J}_g(\mathbf{x})\|_\infty |h_1| + \cdots + \|\mathbf{J}_g(\mathbf{x})\|_\infty |h_n| \leq \frac{1}{3}\|\mathbf{h}\|_\infty.$$

Then by the Lagrange's mean value theorem for multivariate functions (Theorem 6.6.1), we have

$$\|g(\mathbf{x} + \mathbf{h}) - g(\mathbf{x})\|_\infty = \|dg(\mathbf{x} + \theta\mathbf{h}, \mathbf{h})\|_\infty \leq \frac{1}{3}\|\mathbf{h}\|_\infty. \tag{6.8.15}$$

for $\mathbf{x}, \mathbf{h} \in U_r(\mathbf{0})$. In other words, g is a contraction on $U_r(\mathbf{0})$.

Then by Lemma 6.8.10, the map $f(\mathbf{x}) = \mathbf{x} + g(\mathbf{x})$ is injective and $f(U_r(\mathbf{0})) \supset U_{r/2}(\mathbf{0})$. In particular, we have an inverse map $f^{-1} : U_{r/2}(\mathbf{0}) \to U_r(\mathbf{0})$. It remains to show that f is differentiable at $\mathbf{y} \in U_{r/2}(\mathbf{0})$.

First, by (6.8.15) and the triangle inequality, we have

$$\|f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x})\|_\infty = \|\mathbf{h} + g(\mathbf{x} + \mathbf{h}) - g(\mathbf{x})\|_\infty \geq \frac{2}{3}\|\mathbf{h}\|_\infty = \frac{2}{3}\|(\mathbf{x} + \mathbf{h}) - \mathbf{x}\|_\infty.$$

Letting $\mathbf{x} = f^{-1}(\mathbf{y})$ and $\mathbf{x} + \mathbf{h} = f^{-1}(\mathbf{y} + \mathbf{t})$, we have

$$\|(\mathbf{y} + \mathbf{t}) - \mathbf{y}\|_\infty \geq \frac{2}{3}\|f^{-1}(\mathbf{y} + \mathbf{t}) - f^{-1}(\mathbf{y})\|_\infty.$$

It follows that f^{-1} is continuous at $\mathbf{y}$. Moreover, we have

$$\frac{\|f^{-1}(\mathbf{y} + \mathbf{t}) - f^{-1}(\mathbf{y})\|_\infty}{\|\mathbf{t}\|_\infty} \leq \frac{3}{2}. \tag{6.8.16}$$

Next, we shall show that f^{-1} is differentiable at $f(\mathbf{x})$. The idea is similar to the proof of the chain rule (Theorem 6.4.23). Since f is differentiable at $\mathbf{x} \in U_r(\mathbf{0})$,

$$f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + \mathbf{J}_f(\mathbf{h}) + o(\mathbf{h}), \quad \|\mathbf{h}\|_\infty \to 0.$$

Let $\mathbf{x} = f^{-1}(\mathbf{y})$, $\mathbf{h} = f^{-1}(\mathbf{y} + \mathbf{t}) - f^{-1}(\mathbf{y})$. Then we have

$$\frac{\mathbf{t}}{\|\mathbf{t}\|_\infty} = \frac{f(f^{-1}(\mathbf{y} + \mathbf{t})) - f(f^{-1}(\mathbf{y}))}{\|\mathbf{t}\|_\infty} = \mathbf{J}_f\left(\frac{\mathbf{h}}{\|\mathbf{t}\|_\infty}\right) + o\left(\frac{\mathbf{h}}{\|\mathbf{t}\|_\infty}\right), \quad \|\mathbf{t}\|_\infty \to 0.$$

By (6.8.16), $\frac{\|\mathbf{h}\|_\infty}{\|\mathbf{t}\|_\infty}$ is bounded as $\|\mathbf{t}\|_\infty \to 0$. Then we conclude

$$\lim_{\|\mathbf{t}\|_\infty \to 0} \mathbf{J}_f\left(\frac{\mathbf{h}}{\|\mathbf{t}\|_\infty}\right) = \frac{\mathbf{t}}{\|\mathbf{t}\|_\infty}.$$

Since $\mathbf{J}_f$ is invertible at $\mathbf{x}$, we have

$$\lim_{\|\mathbf{t}\|_\infty \to 0} \frac{f^{-1}(\mathbf{y}+\mathbf{t}) - f^{-1}(\mathbf{y})}{\|\mathbf{t}\|_\infty} = \mathbf{J}_f^{-1}\left(\frac{\mathbf{t}}{\|\mathbf{t}\|_\infty}\right).$$

Therefore, f^{-1} is differentiable at $\mathbf{y}$ with

$$\mathbf{J}_{f^{-1}}(\mathbf{y}) = (\mathbf{J}_f(f^{-1}(\mathbf{y})))^{-1}$$

for $\mathbf{y} \in U_{r/2}(\mathbf{0})$. Thus, since f is continuously differentiable, we conclude that f^{-1} is also continuously differentiable. $\square$

6.9 Extrema of functions of several variables

6.9.1 Local extrema

It is instructive to review the local extrema of functions of a single variable (Definition 4.4.1).

Definition 6.9.1 (Local extrema, [1, §6.9 定义])**.** Let $D \subset \mathbb{R}^n$ be a region, $\mathbf{x}_0 \in D$, and $f : D \to \mathbb{R}$ a function. We say that f attains a *local maximum* (极大值) of f at $\mathbf{x}_0$ if there exists a neighborhood $U_\delta(\mathbf{x}_0) \subset D$ such that

$$f(\mathbf{x}) \le f(\mathbf{x}_0)$$

for any $\mathbf{x} \in U_\delta(\mathbf{x}_0)$.

Similarly, we say that f attains a *local minimum* (极小值) of f at $\mathbf{x}_0$ if there exists a neighborhood $U_\delta(\mathbf{x}_0) \subset D$ such that

$$f(\mathbf{x}) \ge f(\mathbf{x}_0)$$

for any $\mathbf{x} \in U_\delta(\mathbf{x}_0)$. A local maximum or minimum is also called a *local extremum* (极值).

Theorem 6.9.2 ([1, §6.9 定理 1])**.** *Let $D \subset \mathbb{R}^n$ be a region, $\mathbf{x}_0 \in D$, $f : D \to \mathbb{R}$ be differentiable at $\mathbf{x}_0$. If f attains a local extremum at $\mathbf{x}_0$, then*

$$df(\mathbf{x}_0) = \mathbf{J}_f(\mathbf{x}_0) = \nabla f(\mathbf{x}_0) = \mathbf{0}. \tag{6.9.1}$$

证明. Let $\Gamma_i(t) = \mathbf{x}_0 + t\mathbf{e}_i$. Since f attains a local extremum at $\mathbf{x}_0$, $f \circ \Gamma_i(t)$ does also at $t = 0$. Then by Fermat's lemma (Lemma 4.1.1), we have

$$0 = (f \circ \Gamma_i)'(0) = \frac{\partial f}{\partial x_i}(\mathbf{x}_0).$$

The consequence follows. $\square$

Again, we say that the points $\mathbf{x}_0$ that satisfies (6.9.1) are *stationary* (稳定点).

Next, we deduce a sufficient condition for functions to attain local extrema. Recall from (6.4.5) that for a twice differentiable function $f : D \to \mathbb{R}$, the *Hessian matrix* (黑塞矩阵) of f is given by:

$$\mathbf{H}_f = \mathbf{J}_{\nabla f} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

One may consider the Hessian as a generalization of the second derivative of single variable functions. If $f \in C^2(D)$, then Clairaut's theorem (Theorem 6.4.4) implies that $\mathbf{H}_f$ is a real symmetric matrix. Then by the spectral theorem (Theorem 7.2.41), $\mathbf{H}_f$ is diagonalizable. Let $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ be the eigenvalues of $\mathbf{H}_f$. Recall that the number n_+ of positive eigenvalues is called the *positive index of inertia* of $\mathbf{H}_f$ (正惯性指数), and the number n_- of negative eigenvalues is called the *negative index of inertia* of $\mathbf{H}_f$ (负惯性指数).

Definition 6.9.3 (Definite matrix)**.** An $n \times n$ real symmetric matrix M is called *positive-definite* if all its eigenvalues are positive, i.e. the positive index of inertia $n_+ = n$. M is called *negative-definite* if all its eigenvalues are negative, i.e. the negative index of inertia $n_- = n$.

The following result is analogous to Theorem 4.4.4.

Theorem 6.9.4 ([1, §6.9 定理 2, 3])**.** *Let $D \subset \mathbb{R}^n$ be a region, $\mathbf{x}_0 \in D$, $f \in C^2(D)$. Suppose that $\nabla f(\mathbf{x}_0) = \mathbf{0}$.*

(1) If the Hessian $\mathbf{H}_f(\mathbf{x}_0)$ is positive-definite, then f attains a (strict) local minimum at $\mathbf{x}_0$. If the Hessian $\mathbf{H}_f(\mathbf{x}_0)$ is negative-definite, then f attains a (strict) local maximum at $\mathbf{x}_0$.

(2) If the Hessian $\mathbf{H}_f(\mathbf{x}_0)$ has a positive and a negative eigenvalue, (equivalently, $\mathbf{h}(\mathbf{H}_f(\mathbf{x}_0))\mathbf{h}^T$ can be positive for some $\mathbf{h} \in \mathbb{R}^n$, negative for another $\mathbf{x}$), then f does not have an extremum at $\mathbf{x}_0$.

(3) Otherwise, f is undetermined. In other words, f may or may not attain an extremum at $\mathbf{x}_0$.

证明. Since $df(\mathbf{x}_0) = \mathbf{0}$, by the Taylor's expansion (Theorem 6.7.1), we have

$$f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) = \mathbf{h}(\mathbf{H}_f(\mathbf{x}_0))\mathbf{h}^T + o(\|\mathbf{h}\|^2), \quad \mathbf{h} \to \mathbf{0}. \tag{6.9.2}$$

On the other hand, using the spectral theorem (Theorem 7.2.41), there is an orthogonal matrix P and a real diagonal matrix D such that

$$\mathbf{H}_f(\mathbf{x}_0) = PDP^T.$$

Since $\|\mathbf{h}\| = \|\mathbf{h}P\|$ (Remark 7.2.39), by replacing $\mathbf{h}$ by $\mathbf{h}P$ if necessary, we assume that $\mathbf{H}_f(\mathbf{x}_0) = D$ is diagonal. Write $\mathbf{h} = (h_1, \ldots, h_n)$, and

$$\mathbf{H}_f(\mathbf{x}_0) = D = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix}.$$

Then $\mathbf{h}(\mathbf{H}_f(\mathbf{x}_0))\mathbf{h}^T = d_1 h_1^2 + \cdots + d_n h_n^2$. Thus, the situations (1), (2) of the theorem follow from (6.9.2) immediately.

For (3), consider $f_1(x, y) = x^2 + y^3$ and $f_2(x, y) = x^2 + y^4$ at $(0, 0)$. □

Example 6.9.5 ([1, §6.9 例 3])**.** Find the stationary points of $f(x, y) = \frac{1}{3}(x^3 + y^3) + xy$, and determine whether f attains extrema at these points.

证明. First, we calculate

$$\mathbf{0} = \nabla f = [x^2 + y, y^2 + x].$$

Then there are two stationary points $(0,0)$, $(-1,-1)$. Now calculate the Hessian

$$\mathbf{H}_f = \begin{bmatrix} 2x & 1 \\ 1 & 2y \end{bmatrix}.$$

Then at $(0,0)$, $\det \mathbf{H}_f < 0$, and we conclude that f does not have an extremum at $(0,0)$. At $(-1,-1)$, $\det \mathbf{H}_f > 0$, and $\mathbf{H}_f$ is negative definite. We conclude that f attains a maximum at $(-1,-1)$. □

Example 6.9.6 (Least squares, 最小二乘法, [1, §6.9 例 5])**.** Suppose that the n observations have been made, resulting in n values $(x_1,y_1),(x_2,y_2),\dots,(x_n,y_n)$ for $x_i \neq x_j$ when $i \neq j$. The problem is to determine a line $L(x) = ax+b$ in a reasonable way from these observations.

Gauss and Legendre discovered a simple and effective method by taking the squares of residuals:

$$u(a,b) = \sum_{i=1}^{n}(L(x_i) - y_i)^2 = \sum_{i=1}^{n}(ax_i + b - y_i)^2$$

and find the minimum of f. Then the minimum is the **best estimation value** for the observations. See also Example 4.4.7.

First, we find the stationary points of $u(a,b)$:

$$\mathbf{0} = \nabla u(a,b) = \begin{bmatrix} u_a & u_b \end{bmatrix} = \begin{bmatrix} 2\sum_{i=1}^{n}(ax_i + b - y_i)x_i & 2\sum_{i=1}^{n}(ax_i + b - y_i) \end{bmatrix}.$$

It follows that

$$\begin{bmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & n \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} y_i \end{bmatrix}. \tag{6.9.3}$$

Thus, if (a,b) is a stationary point of $u(a,b)$, then (6.9.3) has a solution, and so we must have

$$\det \begin{bmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & n \end{bmatrix} \neq 0.$$

But this is the case because by Cauchy-Schwarz inequality,

$$\left(\sum_{i=1}^{n} x_i\right)^2 \leq n \sum_{i=1}^{n} x_i^2$$

and the equality holds only when $(1, \ldots, 1)$ and $(x_1, \ldots, x_n)$ are linearly dependent, which is clearly impossible by the assumption. Moreover, one concludes that there is a unique solution $P_0 = (a_0, b_0)$ of the system (6.9.3). It follows that $u(a, b)$ has a unique stationary point P_0.

Further, one may calculate

$$\mathbf{H}_u = \begin{bmatrix} u_{aa} & u_{ba} \\ u_{ab} & u_{bb} \end{bmatrix} = \begin{bmatrix} 2\sum_{i=1}^n x_i^2 & 2\sum_{i=1}^n x_i \\ 2\sum_{i=1}^n x_i & 2n \end{bmatrix}.$$

Then $\mathbf{H}_u$ is positive definite. Thus, we conclude that $u(a, b)$ attains the minimum at P_0.

6.9.2 Extrema with constraint

Let $D \subset \mathbb{R}^n$ be a region, $f : D \to \mathbb{R}$ a function. The problem of an extremum with constraint usually is to find an extremum for $f(\mathbf{x})$ under the condition that the variable $\mathbf{x}$ must satisfy a system

$$F(\mathbf{x}) = \mathbf{0}$$

for some $F : D \to \mathbb{R}^m$. Put another way, let

$$S = \{\mathbf{x} \in D : F(\mathbf{x}) = \mathbf{0}\}.$$

Then our goal is to find the extremum of the function $f|_S : S \to \mathbb{R}$ defined by

$$f : \mathbf{x} \mapsto f(\mathbf{x}), \quad \mathbf{x} \in S.$$

We give a necessary condition for the existence of extrema with constraint.

Theorem 6.9.7 (Lagrange multiplier 拉格朗日乘数法)**.** *Let $D \subset \mathbb{R}^n$ be a region, $f \in C^1(D)$. Let $F = (F_1, \ldots, F_m) : D \to \mathbb{R}^m$ be continuous differentiable, and*

$$S = \{\mathbf{x} \in D : F(\mathbf{x}) = \mathbf{0}\}.$$

Suppose that $f|_S$ attains an extremum at $\mathbf{x}_0$. Then $\nabla f(\mathbf{x}_0)$ is a linear combination of $\nabla F_1(\mathbf{x}_0), \ldots, \nabla F_m(\mathbf{x}_0)$, i.e. there exist $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ such that

$$\nabla f(\mathbf{x}_0) = \lambda_1 \nabla F_1(\mathbf{x}_0) + \cdots + \lambda_m \nabla F_m(\mathbf{x}_0).$$

证明. Assume that $f|_S$ attains an extremum at $\mathbf{x}_0$. Assume for simplicity that the rank of $\mathbf{J}_F(\mathbf{x}_0)$ is m. Moreover, letting

$$\pi_1 : (x_1, \dots, x_n) \mapsto (x_1, \dots, x_{n-m}), \quad \pi_2 : (x_1, \dots, x_n) \mapsto (x_{n-m+1}, \dots, x_n),$$

we assume for simplicity that the Jacobian $F_{\pi_2}(\mathbf{x}_0)$ is invertible. Then by the implicit function theorem (Theorem 6.8.1), there exists a neighborhood $U \subset \mathbb{R}^{n-m}$ and a map $\varphi : \mathbb{R}^{n-m} \to \mathbb{R}^m$ such that

$$F(\pi_1(\mathbf{x}), \varphi(\pi_1(\mathbf{x}))) = \mathbf{0}, \quad \pi_1(\mathbf{x}) \in U.$$

Then the problem reduces to find the extrema of

$$f(\mathbf{v}, \varphi(\mathbf{v})), \quad \mathbf{v} \in U.$$

In fact, one may apply the above process and conclude that S is an $(n-m)$-dimensional C^1 *manifold* 流形 (see Remark 6.10.2(4)).

Then suppose that there is a smooth curve $\Gamma : [-1, 1] \to S$ such that

$$\Gamma(0) = \mathbf{x}_0, \quad F(\Gamma(t)) = \mathbf{0}, \quad t \in [-1, 1].$$

It follows that $\Gamma'(0)$ is a solution of the equation

$$\mathbf{J}_f(\mathbf{x}_0) \cdot \mathbf{v} = \mathbf{0}, \quad \mathbf{v} \in \mathbb{R}^n. \tag{6.9.4}$$

Moreover, any solution $\mathbf{v}$ of (6.9.4) can be obtained this way, namely $\mathbf{v} = \Gamma'(0)$ for some smooth curve Γ. More precisely, any solution $\mathbf{v}$ is the derivative of a linear combination of the curves

$$\Gamma_j : t \mapsto (t\mathbf{e}_j, \varphi(t\mathbf{e}_j))$$

for $j = 1, \dots, n-m$.

Now by Fermat's lemma (Lemma 4.1.1), we have $(f \circ \Gamma)'(0) = 0$, i.e.

$$\nabla f(\mathbf{x}_0) \cdot \mathbf{v} = 0. \tag{6.9.5}$$

Thus, we see that every solution of (6.9.4) satisfies (6.9.5). If we write

$$\mathbf{J}_F(\mathbf{x}_0) = \begin{bmatrix} \nabla F_1(\mathbf{x}_0) \\ \nabla F_2(\mathbf{x}_0) \\ \vdots \\ \nabla F_m(\mathbf{x}_0) \end{bmatrix}$$

then ∇f and $\nabla F_1, \ldots, \nabla F_m$ are linearly dependent. This implies the theorem. $\square$

Enlighten by Theorem 6.9.7, we define the *Lagrange function*:

$$L(\mathbf{x}, \lambda_1, \ldots, \lambda_m) = f(\mathbf{x}) - \sum_{i=1}^{m} \lambda_i F_i(\mathbf{x}).$$

One then calculates

$$\begin{cases} L_{\mathbf{x}} = \nabla f - \lambda_1 \nabla F_1 - \cdots - \lambda_m \nabla F_m \\ L_{\lambda_1} = F_1 \\ \quad \vdots \\ L_{\lambda_m} = F_m \end{cases}. \tag{6.9.6}$$

Thus, in order to find the extrema of $f|_S$, one may look at the stationary points of L. More precisely, by Theorem 6.9.7, we have

$$\left\{ \mathbf{x} \in S : \begin{array}{c} f|_S \text{ attains an} \\ \text{extremum at } \mathbf{x} \end{array} \right\} \subset \left\{ \mathbf{x} \in S : \begin{array}{c} (\mathbf{x}, \lambda_1, \ldots, \lambda_m) \text{ is} \\ \text{a stationary point of } L \end{array} \right\}.$$

This is the method of *Lagrange multipliers*.

Example 6.9.8 ([1, §6.9 例 6])**.** Find the maximum of the function $f(x, y, z) = xyz$ on the sphere $x^2 + y^2 + z^2 = R^2$ for $x, y, z, R > 0$.

证明*.* When tending to the boundary, $f \to 0$. Thus, the maximum must be attained in the interior. Let $L(x, y, z, \lambda) = xyz + \lambda(x^2 + y^2 + z^2 - R^2)$. Then one calculates

$$\begin{cases} L_x = yz + 2x\lambda = 0 \\ L_y = xz + 2y\lambda = 0 \\ L_z = xy + 2z\lambda = 0 \\ L_\lambda = x^2 + y^2 + z^2 - R^2 = 0 \end{cases}. \tag{6.9.7}$$

Then (6.9.7) implies

$$\lambda = -\frac{yz}{2x} = -\frac{xz}{2y} = -\frac{xy}{2z}.$$

It follows that $x^2 = y^2 = z^2$, and so we find the unique stationary point

$$P = \left(\frac{R}{\sqrt{3}}, \frac{R}{\sqrt{3}}, \frac{R}{\sqrt{3}} \right).$$

Then, f attains the maximum at P, i.e.

$$xyz \leq f\left(\frac{R}{\sqrt{3}}, \frac{R}{\sqrt{3}}, \frac{R}{\sqrt{3}}\right) = \left(\frac{R}{\sqrt{3}}\right)^3 = \left(\frac{x^2+y^2+z^2}{3}\right)^{\frac{3}{2}}.$$

This leads to the *AM-GM inequality* (Example 4.5.3). □

Example 6.9.9 ([1, §6.9 例 7])**.** Find the maximum and minimum distance between the origin $O(0,0,0) \in \mathbb{R}^3$, and the point P on the curve

$$C: \begin{cases} x+y+z=1 \\ x^2+y^2=1 \end{cases}.$$

证明. Let $f(x,y,z) = x^2+y^2+z^2$. The problem reduces to find the extrema of $f|_C$. Define the Lagrange function

$$L(x,y,z,\lambda_1,\lambda_2) = (x^2+y^2+z^2) + \lambda_1(x+y+z-1) + \lambda_2(x^2+y^2-1).$$

Solve the system

$$\begin{cases} L_x = 2x + \lambda_1 + 2x\lambda_2 = 0 \\ L_y = 2y + \lambda_1 + 2y\lambda_2 = 0 \\ L_z = 2z + \lambda_1 = 0 \\ L_{\lambda_1} = x+y+z-1 = 0 \\ L_{\lambda_2} = x^2+y^2-1 = 0 \end{cases}. \tag{6.9.8}$$

We get

$$\begin{cases} x(1+\lambda_2) = z \\ y(1+\lambda_2) = z \end{cases}. \tag{6.9.9}$$

Now we discuss the stationary points:

- If $\lambda_2 = -1$, then (6.9.9) implies $z = 0$. Then the system (6.9.8) implies that $x = 1$, $y = 0$ or $x = 0$, $y = 1$. Then we find the stationary points of L:

$$(1,0,0), \quad (0,1,0).$$

- If $\lambda_2 \neq -1$, then (6.9.9) implies

$$x = y = \frac{z}{1+\lambda_2}.$$

Then the system (6.9.8) implies the stationary points of L:

$$\left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, 1-\sqrt{2}\right), \quad \left(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}, 1+\sqrt{2}\right).$$

One then calculates the value $f(x, y, z)$ of the stationary points, and concludes that f attains the minimum at $(1, 0, 0)$, $(0, 1, 0)$, and the maximum at $\left(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}, 1+\sqrt{2}\right)$.

□

6.10 Regular surfaces

Recall in Definition 3.5.5 that a curve is a continuous map $\Gamma : [a, b] \to \mathbb{R}^n$. Analogously, we can study the geometry in higher dimensions.

Definition 6.10.1 (Surface)**.** Let $I_1, I_2 \subset \mathbb{R}$ be (open or closed) intervals. A (2-dimensional) *surface* $M = M_\Gamma$ is a continuous map $\Gamma : I_1 \times I_2 \to \mathbb{R}^n$ defined by a parametric system:

$$\Gamma : (t, s) \mapsto \begin{cases} x_1 = x_1(t, s) \\ \vdots \\ x_n = x_n(t, s) \end{cases}.$$

We also refer to the image $\Gamma([a_1, b_1] \times [a_2, b_2])$ of Γ as the surface M. A surface is said to be *regular* if Γ is of class C^1, and the rank of its Jacobian $\mathbf{J}_\Gamma$ is 2 for any interior point of $I_1 \times I_2$.

Remark 6.10.2.

(1) One may similarly define k-dimensional *surfaces* and *regular surfaces* for any $k \in \mathbb{N}^*$.

(2) Let $\Gamma = \Gamma(x_1, \ldots, x_k)$ be a k-dimensional regular surface, M be its image, and $P \in M$. By the assumption of the Jacobian, One may obtain k linearly independent tangent vectors of M at P, namely

$$\Gamma_{x_1}(P), \ldots, \Gamma_{x_k}(P). \tag{6.10.1}$$

The vector space $T_P M = \{r_1 \Gamma_{x_1}(P) + \cdots + r_k \Gamma_{x_k}(P) : r_1, \ldots, r_k \in \mathbb{R}\}$ generated by (6.10.1) is called the *tangent space* of M at P.

(3) For any point $P = \Gamma(p_1, \dots, p_k) \in M$, we define

$$\Gamma_P : (r_1, \dots, r_k) \mapsto \Gamma(p_1 + r_1, \dots, p_k + r_k).$$

Since $\mathbf{J}_\Gamma(P)$ is of rank k, one may argue via the implicit function theorem (recall the proof of Theorem 6.9.7) that Γ_P is injective on a neighborhood $U_P \subset \mathbb{R}^k$ containing $\mathbf{0}$. Moreover, $\Gamma_P : U_P \to \Gamma_P(U_P)$ is bijective. The pair (Γ_P, U_P) is called a *chart*, or a *local coordinate* of M at P.

(4) For any two points $P, Q \in M$, let $W = \Gamma_P(U_P) \cap \Gamma_Q(U_Q)$. Then the coordinate change

$$\Gamma_Q^{-1} \circ \Gamma_P : \Gamma_P^{-1}(W) \to \Gamma_Q^{-1}(W)$$

is of class C^1, and has a C^1 inverse. A k-dimensional surface that have local coordinates at any point and the coordinate changes satisfy these, is called a k-dimensional *smooth surface* or a k-dimensional *smooth manifold*. Thus, regular surfaces are smooth manifolds.

Example 6.10.3 ([1, §6.10 例 1])**.** Let $D \subset \mathbb{R}^2$ be a region, $F : D \to \mathbb{R}$ be continuously differentiable. Then the surface

$$z = F(x, y)$$

is regular. In fact, it is given by the map

$$\Gamma : (t, s) \mapsto \begin{cases} x = t \\ y = s \\ z = F(t, s) \end{cases}.$$

Then one calculates

$$\mathbf{J}_\Gamma = [\Gamma_t, \Gamma_s] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ F_x & F_y \end{bmatrix}.$$

Clearly, the rank of $\mathbf{J}_\Gamma$ is 2.

Let $M \subset \mathbb{R}^3$ be a (2-dimensional) regular surface, $P \in M$. The line that passes through P and is perpendicular to the tangent plane T_pM, is called the *normal line*. A vector $\mathbf{v} \in \mathbb{R}^3$ that generates the normal line is called a *normal vector*.

Proposition 6.10.4 ([1, §6.10 命题]). *Let S be a surface determined by the equation*

$$F(x, y, z) = 0$$

for some $F : \mathbb{R}^3 \to \mathbb{R}$. Let $P \in S$. Suppose that $\nabla F(P) \neq \mathbf{0}$. Then the vector $\nabla F(P)$ is a normal vector of S at P.

证明. Let $\Gamma : [-1, 1] \mapsto S$ be a smooth curve with $\Gamma(0) = P$. Then $F \circ \Gamma \equiv 0$. Thus, by the chain rule, we have

$$\nabla F(P) \cdot \Gamma'(0) = 0.$$

Since Γ is chosen arbitrarily, $\Gamma'(0) \in T_P S$ spans the tangent plane. Thus, we conclude that $\nabla F(P)$ is perpendicular to $T_P S$. □

Therefore, the equation of the tangent plane $T_P S$ at $P(p_1, p_2, p_3)$ is

$$F_x(P)(x - p_1) + F_y(P)(y - p_2) + F_z(P)(z - p_3) = 0.$$

The equation of the normal line of $T_P S$ at $P(p_1, p_2, p_3)$ is

$$\frac{x - p_1}{F_x(P)} = \frac{y - p_2}{F_y(P)} = \frac{z - p_3}{F_z(P)}.$$

Now let $f : \mathbb{R}^2 \to \mathbb{R}$ is differentiable at $P(x_0, y_0)$, and $\mathbf{h} = (x - x_0, y - y_0)$. Then the differential

$$df(P, \mathbf{h}) = \nabla f(P) \cdot \mathbf{h} = f_x(P)(x - x_0) + f_y(P)(y - y_0).$$

Let $df(P, \mathbf{h}) = z - z_0$. Then we obtain a plane

$$z - z_0 = f_x(P)(x - x_0) + f_y(P)(y - y_0).$$

This is exactly the tangent plane $T_P S$ at P of the surface S defined by $z = f(x, y)$.

Let $I_1, I_2 \subset \mathbb{R}$ be intervals, and $\Gamma : I_1 \times I_2 \to \mathbb{R}^n$ be a regular surface given by

$$\Gamma : (t, s) \mapsto \begin{cases} x = x(t, s) \\ y = y(t, s) \\ z = z(t, s) \end{cases}.$$

Then one calculates

$$\mathbf{J}_\Gamma = [\Gamma_t, \Gamma_s] = \begin{bmatrix} x_t & x_s \\ y_t & y_s \\ z_t & z_s \end{bmatrix}.$$

To find a normal vector of Γ, we calculate

$$\mathbf{n} = \Gamma_t \times \Gamma_s = \det \begin{bmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_t & y_t & z_t \\ x_s & y_s & z_s \end{bmatrix}.$$

In terms of Jacobian determinants, we have

$$\mathbf{n} = \left(\frac{D(y,z)}{D(t,s)}, \frac{D(z,x)}{D(t,s)}, \frac{D(x,y)}{D(t,s)} \right).$$

The tangent plane at $P(x_0, y_0, z_0)$ then can be expressed as

$$(x - x_0, y - y_0, z - z_0) \cdot \mathbf{n} = 0.$$

In other words,

$$\det \begin{bmatrix} x - x_0 & y - y_0 & z - z_0 \\ x_t(P) & y_t(P) & z_t(P) \\ x_s(P) & y_s(P) & z_s(P) \end{bmatrix} = 0.$$

Example 6.10.5 ([1, §6.10 例 3])**.** Find the tangent plane and a normal vector of the sphere $x^2 + y^2 + z^2 = R^2$ at $P(0, \frac{\sqrt{2}}{2}R, \frac{\sqrt{2}}{2}R)$.

证明. Let $F = x^2 + y^2 + z^2 - R^2$. Then a normal vector at P is

$$\nabla F(P) = [2x, 2y, 2z](P) = (0, \sqrt{2}R, \sqrt{2}R).$$

Then the tangent plane is

$$\sqrt{2}R\left(y - \frac{\sqrt{2}}{2}R\right) + \sqrt{2}R\left(z - \frac{\sqrt{2}}{2}R\right) = 0,$$

or $y + z - \sqrt{2}R = 0$. $\square$

Chapter 7

Appendix

7.1 Completeness of the real numbers

Completeness (完备性) is a property of the real numbers that, intuitively, implies that there are no "gaps" or "holes" in the real number line. In the following, we shall briefly discuss how to define the real numbers, as well as its completeness. It is commonly a *mathematical analysis* (数学分析) topic.

Definition 7.1.1 (Cauchy sequence)**.** Let $\{a_n\}$ be a sequence. If for any $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$|a_n - a_m| < \epsilon$$

for any $n, m \geq N$, then we say $\{a_n\}$ is a *Cauchy sequence* (柯西序列).

Example 7.1.2. A monotonic and bounded sequence is a Cauchy sequence.

Roughly speaking, a Cauchy sequence is a sequence that "should be convergent". However, one may easily find a Cauchy sequence of $\mathbb{Q}$ that does not have a limit. This means $\mathbb{Q}$ is not *complete*. To fix this, we need to add "extra points" to the rational numbers so that every Cauchy sequence of $\mathbb{Q}$ has a limit.

Let $\{a_n\}$, $\{b_n\}$ be two Cauchy sequences. We say that $\{a_n\}$, $\{b_n\}$ are *equivalent* (and write $\{a_n\} \sim \{b_n\}$), if

$$\lim_{n\to\infty}(a_n - b_n) = 0.$$

Two equivalent Cauchy sequences mean they "should have the same limit". Note in particular that a rational sequence converges to 0 is not related to the notion of real numbers. Now let $\{a_n\}$ be a Cauchy sequence of $\mathbb{Q}$, and define the *equivalence class* of the Cauchy sequence $\{a_n\}$ by

$$[a_n] \coloneqq \{\{b_n\} : \{b_n\} \text{ is a Cauchy sequence of } \mathbb{Q} \text{ and } \{b_n\} \sim \{a_n\}\}.$$

Then $[a_n]$ includes all possible Cauchy sequences that "should converge to the same limit" as $\{a_n\}$.

Now we are in the position to define the real numbers.

Definition 7.1.3 (Real numbers)**.** The real numbers are defined by

$$\mathbb{R} \coloneqq \{[a_n] : \{a_n\} \text{ is a Cauchy sequence of } \mathbb{Q}\}.$$

Definition 7.1.3 rigorously illustrates the idea that real numbers are the limits of rational sequences. Also, it can be considered as a procedure to "fill the gaps" of the rational numbers (cf. Proposition 1.1.2). This procedure clearly guarantees that every Cauchy sequence of $\mathbb{Q}$ has a limit, and is called a *completion* of $\mathbb{Q}$, sometimes written $\overline{\mathbb{Q}} = \mathbb{R}$. Moreover, after a completion, we no longer need to do a completion again.

Theorem 7.1.4 (Completeness of $\mathbb{R}$, [2, §10.1 定理 1])**.** *A sequence of* $\mathbb{R}$ *is a Cauchy sequence if and only if it is convergent.*

Remark 7.1.5. We say that $\mathbb{R}$ is *complete* because every Cauchy sequence of $\mathbb{R}$ converges. The notion can be extended to other abstract spaces.

Sketch proof of Theorem 7.1.4. Since we have explained the real numbers as Cauchy sequences of $\mathbb{Q}$, one may define the elementary arithmetic, order, absolute value, and the limit of real numbers in a natural way.

Comparing Definitions 1.3.1 and Definition 7.1.1, it is easy to see that convergent sequences of $\mathbb{R}$ are Cauchy sequences. Now assume that $\{A_n\}$ is a Cauchy sequence of $\mathbb{R}$. We write $A_n = \{a_{ni}\}_{i\in\mathbb{N}^*}$ as Cauchy sequences of $\mathbb{Q}$, i.e.

$$\begin{array}{rccccc} A_1 = & a_{11} & a_{12} & \cdots & a_{1n} & \cdots \\ A_2 = & a_{21} & a_{22} & \cdots & a_{2n} & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \\ A_n = & a_{n1} & a_{n2} & \cdots & a_{nn} & \cdots \\ \vdots & \vdots & \vdots & & \vdots & \end{array}$$

Now for $n \in \mathbb{N}^*$, let $b_n = a_{nn}$ and $B = \{b_n\}_{n\in\mathbb{N}^*}$. One may check that $\{b_n\}_{n\in\mathbb{N}^*}$ is a Cauchy sequence of $\mathbb{Q}$, and $A_n \to B$ as $n \to \infty$. $\square$

Next, we turn to focus on the equivalent descriptions of the completeness of $\mathbb{R}$, which will be used to derive properties of continuous functions on closed intervals.

First, by Theorem 7.1.4 and Example 7.1.2, we immediately obtain Proposition 1.1.4. In fact, Proposition 1.1.4 is equivalent to Theorem 7.1.4 (see the proof after Theorem 7.1.8). In other words, Proposition 1.1.4 is an equivalent description of the completeness of $\mathbb{R}$.

Next, we discuss the supremum and infimum:

Definition 7.1.6 (Supremum and infimum)**.** A *lower bound* of a subset $S \subset \mathbb{R}$ is an element $C \in \mathbb{R}$ such that

$$C \leq x$$

for all $x \in S$. A lower bound of S is called an *infimum* (or *greatest lower bound* 下确界), written $\inf(S)$, if for all lower bound C of S, we have

$$C \leq \inf(S).$$

One can similarly define the *upper bound*, and the *supremum* (or *least upper bound* 上确界).

Recall from Proposition 1.1.4 that completeness of $\mathbb{R}$ implies that any monotonic and bounded sequence converges. It implies the existence of infima and suprema:

Theorem 7.1.7 (Least-upper-bound property). *If a nonempty subset $S \subset \mathbb{R}$ has an upper bound, then* $\sup(S)$ *exists in* $\mathbb{R}$*. Similarly, if $S \subset \mathbb{R}$ has an lower bound, then* $\inf(S)$ *exists.*

证明. Suppose that S has an upper bound $C_1 \in \mathbb{R}$. Let $a_1 \in S$. Then by the definition, $a_1 \leq C_1$. If $a_1 = C_1$, then $\sup(S) = C_1$. Thus, assume that $a_1 < C_1$. Consider the average

$$D_1 = \frac{a_1 + C_1}{2}.$$

Then there are two situations:

(1) If $[D, C_1] \cap S = \emptyset$, then write $C_2 = D_1$, $a_2 = a_1$.

(2) If $[D, C_1] \cap S \neq \emptyset$, then let $a_2 \in [D, C_1] \cap S$, $C_2 = C_1$.

In situation (2), if $a_2 = C_2$, then $\sup(S) = C_2$ and we are done. Thus, it completes an inductive step. By induction, we obtain

- an increasing sequence $\{a_n\} \subset S$,
- a decreasing sequence $\{C_n\}$ of upper bounds of S.

In addition, we have

$$0 \leq C_n - a_n \leq 2^{-n}(C_1 - a_1). \tag{7.1.1}$$

Also by Proposition 1.1.4, there are limits $a_n \to \ell$ and $C_n \to L$ as $n \to \infty$. By (7.1.1) and Theorem 1.3.8, we conclude

$$\ell = L$$

and so $\sup(S) = L$. □

The existence of infima and suprema for bounded subsets implies the following:

Theorem 7.1.8 (Bolzano–Weierstrass theorem). *Let $\{a_n\}$ be a bounded sequence. Then there is at least one subsequence $\{a_{n_k}\}$ which converges.*

证明. Let $L_m = \sup\{a_n : n \geq m\}$. Then $\{L_m\}$ is decreasing. Let

$$L = \inf\{L_m : m \geq 1\} = \lim_{m\to\infty} L_m.$$

(We often write $\limsup\limits_{n\to\infty} a_n = L$ and call it the *upper limit* (上极限) of a_n.)

Then for any $k \geq 1$, there exists $a_{n_k} \in \{a_n : n \geq k\}$, such that

$$L - \frac{1}{k} \leq L_k - \frac{1}{k} \leq a_{n_k} \leq L_k.$$

Note that $L_k \to L$, $L - \frac{1}{k} \to L$ as $k \to \infty$. Then by Theorem 1.3.8, we conclude that

$$\lim_{k\to\infty} a_{n_k} = L.$$

This establishes the theorem. $\square$

In the following, we show that the above Theorems are all equivalent.

Proof of Theorem 7.1.4 using Proposition 1.1.4 and Theorem 7.1.7. Let $\{a_n\}$ be a Cauchy sequence. One may easily see that $\{a_n\}$ is bounded.

Assume that $\{a_n\}$ is not convergent. By Theorem 7.1.7, we can define

$$b_m = \inf\{a_n : n \geq m\}, \quad c_m = \sup\{a_n : n \geq m\}. \tag{7.1.2}$$

Clearly, $\{b_n\}$ is increasing and $\{c_n\}$ is decreasing, and

$$b_n \leq a_n \leq c_n \tag{7.1.3}$$

for $n \in \mathbb{N}$. Then by Proposition 1.1.4, $b_n \to b$ and $c_n \to c$ as $n \to \infty$. By Theorem 1.3.13, we have $b \leq c$.

Assume that $b < c$. Let $\epsilon = \frac{c-b}{5}$. Then there is $N \in \mathbb{N}$ so that

$$b_n < b + \epsilon, \quad c - \epsilon < c_n$$

for $n \geq N$. On the other hand, by (7.1.2), there are $m_1, m_2 \geq n$, such that

$$a_{m_1} < b_n + \epsilon, \quad c_n - \epsilon < a_{m_2}.$$

Thus, one calculates

$$a_{m_2} - a_{m_1} > (c_n - \epsilon) - (b_n + \epsilon) > (c - 2\epsilon) - (b + 2\epsilon) = c - b - 4\epsilon = \epsilon.$$

In other words, for any $n > N$, there exist $m_1, m_2 \geq n$ such that

$$|a_{m_2} - a_{m_1}| > \epsilon. \tag{7.1.4}$$

However, by the definition of the Cauchy sequence, there exists $N_1 = N_1(\epsilon) > 0$ such that

$$|a_{m_2} - a_{m_1}| < \epsilon$$

whenever $m_2, m_1 \geq N_1$. This contradicts (7.1.4). Thus, we conclude that $b = c$.

Finally, by Theorem 1.3.8 and (7.1.3), we conclude that $\{a_n\}$ converges. □

Proof of Proposition 1.1.4 using Theorem 7.1.8. Let $\{a_n\}$ be a bounded and increasing sequence of $\mathbb{R}$. Then by Theorem 7.1.8, there is a subsequence $\{a_{n_k}\} \subset \{a_n\}$ that converges to some limit a as $k \to \infty$. Now for $k \in \mathbb{N}$, $n \in [n_k, n_{k+1}) \cap \mathbb{N}$, we define

$$b_n = a_{n_k}, \quad c_n = a_{n_{k+1}}.$$

Then since $\{a_n\}$ is increasing, we have

$$b_n \leq a_n \leq c_n$$

for any $n \geq n_1$. Since $b_n \to a$, $c_n \to a$, as $n \to \infty$. We conclude that $a_n \to a$, as $n \to \infty$ by Theorem 1.3.8. □

Remark 7.1.9. Therefore, Proposition 1.1.4, Theorem 7.1.4, Theorem 7.1.7 and Theorem 7.1.8 are equivalent. They all describe the completeness of $\mathbb{R}$. There are other equivalent descriptions of the completeness of $\mathbb{R}$, such as the *compactness* (紧性) of bounded closed subsets of $\mathbb{R}$. The compactness plays a central role in the theory of topology.

Remark 7.1.10. One may also conclude from the above proofs that there are only three possibilities of a sequence $\{a_n\}$ of $\mathbb{R}$:

(1) $\{a_n\}$ is unbounded.

(2) $\{a_n\}$ is bounded, but has two subsequences that have different limits.

(3) $\{a_n\}$ is convergent.

Next, we discuss the applications of the completeness of $\mathbb{R}$ to the continuous functions on a closed interval.

Definition 7.1.11 (Uniform continuity)**.** Let $X \subset \mathbb{R}$ be a subset, and let $f: X \to \mathbb{R}$ be a function. We say that f is *uniformly continuous* (一致连续) if, for every $\epsilon > 0$, there exists a $\delta > 0$ such that

$$|f(u) - f(v)| < \epsilon$$

for $|u - v| < \delta$.

Theorem 7.1.12 (Heine–Cantor theorem)**.** *Let f be continuous on $[a, b]$. Then f is also uniformly continuous on $[a, b]$.*

证明. Suppose for sake of contradiction that f is not uniformly continuous. Then there exists a $\epsilon > 0$, sequences $\{x_n\}, \{y_n\} \subset [a, b]$ such that

$$|f(x_n) - f(y_n)| \geq \epsilon, \quad |x_n - y_n| < \frac{1}{n}. \tag{7.1.5}$$

Then by Theorem 7.1.8, let $\{x_{n_j}\} \subset \{x_n\}$ be a subsequence so that

$$x_{n_j} \to \ell, \quad (j \to \infty). \tag{7.1.6}$$

By using Theorem 7.1.8 again, let $\{y_{n_{j_k}}\} \subset \{y_{n_j}\}$ be a subsequence so that $\{y_{n_{j_k}}\}$ converges. By (7.1.5), we have

$$x_{n_{j_k}} - \frac{1}{n_{j_k}} \leq y_{n_{j_k}} \leq x_{n_{j_k}} + \frac{1}{n_{j_k}} \tag{7.1.7}$$

for any $k \in \mathbb{N}$. By (7.1.6) and Theorem 1.3.17, we have

$$x_{n_{j_k}} \to \ell \quad (k \to \infty).$$

Then by (7.1.7) and Theorem 1.3.8, we have

$$y_{n_{j_k}} \to \ell \quad (k \to \infty).$$

Now by the continuity of f, we obtain

$$\lim_{k\to\infty} f(x_{n_{j_k}}) = f(\ell) = \lim_{k\to\infty} f(y_{n_{j_k}}).$$

But this contradicts (7.1.5). From this contradiction we conclude that f is in fact uniformly continuous. □

7.2 Linear algebra, continued

7.2.1 Vector space

7.2.1.1 Group and vector space

For a family of sets $\{X_\alpha : \alpha \in A\}$ with an index set A, the *Cartesian product* (笛卡尔积) is defined by

$$\prod_{\alpha \in A} X_\alpha \coloneqq \{(x_\alpha)_{\alpha \in A} : x_\alpha \in X_\alpha\}.$$

In particular, for two sets V, W, the Cartesian product is given by

$$V \times W = \{(v, w) : v \in V,\ w \in W\}.$$

We first discuss the abstraction of addition.

Definition 7.2.1 (Group)**.** A nonempty set V is said to be a *group* (群) if V has an operation $+ : V \times V \to V$ such that

(A.1) (Associativity) For any $u, v, w \in V$, we have $u + (v + w) = (u + v) + w$.

(A.2) (Identity) There is an element $\mathbf{0} \in V$ such that $v + \mathbf{0} = \mathbf{0} + v = v$ for all $v \in V$.

(A.3) (Inverse) For every $v \in V$, there is an element $w \in V$ such that $v + w = w + v = \mathbf{0}$.

We say a group V is an *abelian group* (阿贝尔群/交换群) if the operation $+ : V \times V \to V$ is commutative:

(A.4) (Commutativity) For any $v, w \in V$, we have $v + w = w + v$.

Example 7.2.2.

(1) The set $\mathbb{Z}$ of integers is an abelian group.

(2) $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ with respect to the addition is an abelian group.

(3) The circle $S^1 = \{e^{2\pi i\theta} : \theta \in [0,1]\}$ with respect to the multiplication is an abelian group.

(4) The space $C(\mathbb{R},\mathbb{R})$ of continuous functions $f : \mathbb{R} \to \mathbb{R}$ is an abelian group with respect to the addition

$$(f+g)(x) = f(x) + g(x), \quad x \in \mathbb{R}$$

for any $f, g \in \mathbb{R}$.

Next, we introduce the abstraction of $\mathbb{Q}$:

Definition 7.2.3 (Field)**.** A nonempty set k is said to be a *field* (域) if k has operations $+ : k \times k \to k$, $\cdot : k \times k \to k$ such that

(F.1) $(k, +)$ forms an abelian group.

(F.2) $(k^*, \cdot)$ forms an abelian group, where $k^* = k \setminus \{0\}$ and 0 is the identity element of $(k, +)$.

(F.3) For any $a \in k$, we have $a \cdot 0 = 0$.

(F.4) (Distributivity) For every $a, b, c \in k$, we have

$$a \cdot (b + c) = a \cdot b + a \cdot c.$$

Example 7.2.4. $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are fields.

Now we introduce the abstraction of the Euclidean space $\mathbb{R}^n$.

Definition 7.2.5 (Vector space)**.** A nonempty set V is said to be a *vector space* (or *linear space* 向量空间/线性空间) over a field k if V has two operations:

(A) An *addition* (加法) $+ : V \times V \to V$ so that $(V, +)$ becomes an *abelian group* (阿贝尔群).

(B) A *scalar multiplication* (数乘) $\cdot : k \times V \to V$ such that

(B.1) For any $\lambda \in k$, $v, w \in V$, we have $\lambda \cdot (v + w) = \lambda \cdot v + \lambda \cdot w$.

(B.2) For any $\lambda, \mu \in k$, $v \in V$, we have $(\lambda + \mu) \cdot v = \lambda \cdot v + \mu \cdot v$.

(B.3) For any $\lambda, \mu \in k$, $v \in V$, we have $(\lambda\mu) \cdot v = \lambda \cdot (\mu \cdot v)$.

(B.4) For any $v \in V$, we have $1 \cdot v = v$.

Remark 7.2.6. For the scalar multiplication, we often write $\lambda v = \lambda \cdot v$ for $\lambda \in k$, $v \in V$.

Example 7.2.7.

(1) The *Euclidean space* (欧式空间) $\mathbb{R}^n$ for $n \in \mathbb{N}^*$ is a linear space over $\mathbb{R}$.

(2) The *complex plane* (复数平面) $\mathbb{C}$ is a 1-dimensional linear space over $\mathbb{C}$, and a 2-dimensional linear space over $\mathbb{R}$.

(3) The space $C(\mathbb{R}, \mathbb{R})$ of continuous functions $f : \mathbb{R} \to \mathbb{R}$ is a linear space over $\mathbb{R}$.

(4) The space $C(S^1, \mathbb{C})$ of continuous functions $f : S^1 \to \mathbb{C}$ is a linear space over $\mathbb{C}$.

Next, we discuss what is meant by the dimension of a vector space.

Definition 7.2.8 (Linear independence)**.** Let V be a vector space over k and $v_1, \dots, v_n \in V$ be vectors. We say that $v_1, \dots, v_n$ are *linearly independent* (线性无关) over k, if the **only** elements $\lambda_1, \dots, \lambda_n \in k$ such that

$$\lambda_1 v_1 + \cdots + \lambda_n v_n = 0,$$

are $\lambda_1 = \cdots = \lambda_n = 0$. Otherwise, we say that $v_1, \dots, v_n$ are *linearly dependent* (线性相关) over k.

For a subset $S \subset V$, denote

$$\mathrm{Span}(S) \coloneqq \{\lambda_1 v_1 + \cdots + \lambda_n v_n \in V : \lambda_1, \dots, \lambda_n \in k,\ v_1, \dots, v_n \in S\}.$$

Definition 7.2.9 (Basis)**.** Let V be a vector space over k. A subset S of V is called a *basis* (基) of V if any finite number of elements in S is linearly independent, and $\mathrm{Span}(S) = V$.

Definition 7.2.10 (Dimension)**.** Let V be a vector space over k. Suppose that $S \subset V$ is a basis of V. We say that V is *finite dimensional* (有限维的) if S is a finite subset; Otherwise, we say that V is *infinite dimensional* (无限维的). Moreover, the total number of the elements in S is called the *dimension* (维数) of V.

7.2.1.2 Normed vector space

Now we discuss the abstraction of the *absolute value* (绝对值).

Definition 7.2.11 (Norm)**.** Given a vector space V over a subfield $k \subset \mathbb{C}$ of the complex numbers, a *norm* (范数) on V is a real-valued function $\|\cdot\| : V \to \mathbb{R}$ with the following properties:

(1) (Triangle inequality) For any $\mathbf{x}, \mathbf{y} \in V$, we have $\|\mathbf{x}+\mathbf{y}\| \le \|\mathbf{x}\| + \|\mathbf{y}\|$.

(2) (Absolute homogeneity) For any $\lambda \in k$, $\mathbf{x} \in V$, we have $\|\lambda\mathbf{x}\| = |\lambda|\|\mathbf{x}\|$.

(3) (Positiveness) For every $\mathbf{x} \in V$, if $\|\mathbf{x}\| = 0$, then $\mathbf{x} = \mathbf{0}$.

A vector space equipped with a norm $(V, \|\cdot\|)$ is called a *normed vector space* (赋范向量空间).

Remark 7.2.12. Note that for any $\mathbf{x} \in V$, by (2),

$$\|-\mathbf{x}\| = |-1|\|\mathbf{x}\| = \|\mathbf{x}\|, \quad \|\mathbf{0}\| = |0|\|\mathbf{x}\| = 0.$$

Then by (1), we have

$$0 = \|\mathbf{0}\| = \|\mathbf{x}-\mathbf{x}\| \le \|\mathbf{x}\| + \|-\mathbf{x}\| = 2\|\mathbf{x}\|.$$

Thus, we conclude that for any $\mathbf{x} \in V$, $\|\mathbf{x}\| \ge 0$.

Example 7.2.13.

(1) The absolute value $|\cdot|$ on Euclidean space $\mathbb{R}^n$ for $n \in \mathbb{N}^*$ is a norm.

(2) The uniform norm $\|\cdot\|_\infty$ on $C(\mathbb{R}, \mathbb{R})$ or $C(S^1, \mathbb{C})$ (Definition 4.3.26) is a norm.

(3) The L^2-*norm* $\|f\|_2$ for $f \in C(S^1, \mathbb{C})$ defined by

$$\|f\|_2 := \left(\int_{S^1} |f(z)|^2 dz\right)^{\frac{1}{2}}$$

is a norm on $C(S^1, \mathbb{C})$.

Definition 7.2.14 (Equivalent norms)**.** Two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on a vector space V are said to be *equivalent* if there exists a constant $C > 1$ such that

$$C^{-1}\|x\|_1 \leq \|x\|_2 \leq C\|x\|_1$$

for all $x \in V$.

Proposition 7.2.15. *Let V be a finite dimensional vector space. Then any two norms on V are equivalent.*

Definition 7.2.16 (Banach space)**.** Let $(V, \|\cdot\|)$ be a normed vector space. We say that $\{\mathbf{x}_n\} \subset V$ is a *Cauchy sequence* if for any $\epsilon > 0$, there is $N = N(\epsilon) > 0$ such that

$$\|\mathbf{x}_n - \mathbf{x}_m\| < \epsilon$$

for any $n, m \geq N$. We say that $(V, \|\cdot\|)$ is a *Banach space* (巴拿赫空间) if every Cauchy sequence $\{\mathbf{x}_n\} \subset V$ has a limit. In other words, there is $\mathbf{y} \in V$ such that for any $\epsilon > 0$, there is $N = N(\epsilon) > 0$ such that

$$\|\mathbf{x}_n - \mathbf{y}\| < \epsilon$$

for any $n \geq N$.

Example 7.2.17. $\mathbb{R}^n$ and $\mathbb{C}^n$ are Banach spaces.

One important consequence on the Banach spaces is the *Banach fixed-point theorem* (巴拿赫不动点定理) or *contraction mapping theorem* (压缩映射定理).

Theorem 7.2.18 (Banach fixed-point theorem)**.** *Let $(V, \|\cdot\|)$ be a Banach space, $\overline{U} \subset V$ be a closed subset. Let $f : \overline{U} \to \overline{U}$ be a contraction, i.e. there is $q \in [0, 1)$ such that*

$$\|f(\mathbf{x}) - f(\mathbf{y})\| \leq q\|\mathbf{x} - \mathbf{y}\|$$

for any $\mathbf{x}, \mathbf{y} \in \overline{U}$. Then f admits a unique fixed point $\mathbf{z} \in \overline{U}$, i.e. $f(\mathbf{z}) = \mathbf{z}$.

证明. Let $\mathbf{x} \in \overline{U}$, and let $\mathbf{x}_n = f^n(\mathbf{x}) = \underbrace{f \circ \cdots \circ f}_{n \text{ copies}}(\mathbf{x})$. Then

$$\begin{aligned}\|\mathbf{x}_n - \mathbf{x}_{n+k}\| = \|f^n(\mathbf{x}) - f^n f^k(\mathbf{x})\| \leq q^n\|\mathbf{x} - f^k(\mathbf{x})\| \\ \leq q^n \sum_{j=1}^{k} \|f^{j-1}(\mathbf{x}) - f^j(\mathbf{x})\| \leq \frac{q^n}{1-q}\|\mathbf{x} - f(\mathbf{x})\|.\end{aligned}$$

It follows that $\{\mathbf{x}_n\} \subset \overline{U}$ is a Cauchy sequence. Let $\mathbf{y} \in V$ be its limit. Since $\overline{U}$ is closed, $\mathbf{y} \in \overline{U}$ by Proposition 6.1.13. Then

$$\begin{aligned}\|f(\mathbf{y}) - \mathbf{y}\| \leq & \|f(\mathbf{y}) - f(\mathbf{x}_n)\| + \|f(\mathbf{x}_n) - \mathbf{x}_n\| + \|\mathbf{x}_n - \mathbf{y}\| \\ \leq & q\|\mathbf{y} - \mathbf{x}_n\| + \|\mathbf{x}_{n+1} - \mathbf{x}_n\| + \|\mathbf{x}_n - \mathbf{y}\|.\end{aligned}$$

Thus, we see that $f(\mathbf{y}) = \mathbf{y}$.

Finally, suppose that there is another $\mathbf{y}' \in V$ so that $f(\mathbf{y}') = \mathbf{y}'$. Then

$$\|\mathbf{y}' - \mathbf{y}\| = \|f(\mathbf{y}') - f(\mathbf{y})\| \leq q\|\mathbf{y}' - \mathbf{y}\|.$$

It forces $\mathbf{y}' = \mathbf{y}$. $\square$

7.2.1.3 Inner product space

Definition 7.2.19 (Inner product)**.** Let V be a vector space over $k = \mathbb{R}$ or $\mathbb{C}$. An *inner product* (内积) is a map $\langle \cdot, \cdot \rangle : V \times V \to k$ that satisfies

(1) For any $v, w \in V$, we have $\langle v, w \rangle = \overline{\langle w, v \rangle}$.

(2) For any $v \in V$, we have $\langle v, v \rangle \geq 0$, and $\langle v, v \rangle = 0$ if and only if $v = \mathbf{0}$.

(3) For any $u, v, w \in V$, $\lambda, \mu \in k$, we have $\langle \lambda u + \mu v, w \rangle = \lambda \langle u, w \rangle + \mu \langle v, w \rangle$.

A vector space V equipped with an inner product $\langle \cdot, \cdot \rangle$ is called an *inner product space* (内积空间).

Remark 7.2.20. Note that for any $u, v, w \in V$, $\lambda, \mu \in k$, by (1) and (3), we have

$$\langle u, \lambda v + \mu w \rangle = \overline{\langle \lambda v + \mu w, u \rangle} = \overline{\lambda \langle v, u \rangle + \mu \langle w, u \rangle} = \bar{\lambda} \langle u, v \rangle + \bar{\mu} \langle u, w \rangle.$$

7.2.2 Finite dimensional linear algebra

For simplicity, we only discuss the field $k = \mathbb{R}$ or $\mathbb{C}$ in this section.

Definition 7.2.21 (Linear map)**.** Given vector spaces V, W over a field k, a *linear map* (or *linear transformation* 线性映射/线性变换) $L : V \to W$ is a map from V to W that satisfies

(1) (Additivity) $L(\mathbf{x}+\mathbf{y}) = L(\mathbf{x}) + L(\mathbf{y})$ for any $\mathbf{x}, \mathbf{y} \in V$.

(2) (Homogeneity) $L(\lambda\mathbf{x}) = \lambda L(\mathbf{x})$ for any $\lambda \in k$, $\mathbf{x} \in V$.

Definition 7.2.22 (Rank)**.** Let $L : V \to W$ be a linear map. Then the *rank* (秩) of L is the dimension of $L(V)$.

Example 7.2.23.

- A linear map $L : \mathbb{R} \to \mathbb{R}$ is of the form
$$L : x \mapsto kx$$
for some $k \in \mathbb{R}$. Its graph is a straight line that passes through the origin.

- A linear map $L : \mathbb{R}^2 \to \mathbb{R}$ is of the form
$$L : (x, y) \mapsto ax + by$$
for some $a, b \in \mathbb{R}$. Its graph is a plane that passes through the origin.

In general, a linear map can be represented by a matrix.

Definition 7.2.24 (Matrix)**.** Let k be a field. A *k-matrix* (or matrix for short) is a rectangular array of real (or complex) numbers, called the *entries* of the matrix. More precisely, given $n, m \in \mathbb{N}^*$, a real $m \times n$ matrix M is of the form

$$M = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix}$$

where $a_{ij} \in k$ for $1 \le i \le m$, $1 \le j \le n$. When $m = n$, we say that M is a *square matrix*.

Example 7.2.25 (Vector)**.** For $n \geq 1$, we usually write the vector $\mathbf{x} \in k^n$ as an $n \times 1$ column matrix (and called the *column vector*):

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

for $x_1, \dots, x_n \in k$.

Example 7.2.26 (Identity)**.** For $n \geq 1$, we define the identity matrix by

$$I = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

Definition 7.2.27 (Product of matrices)**.** Given an $m \times n$ matrix $A = (a_{ij})$, and an $n \times l$ matrix $B = (b_{ij})$, the *product* of A and B is an $m \times l$ matrix $C = (c_{ij})$:

$$C = AB = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1l} \\ b_{21} & b_{22} & \cdots & b_{2l} \\ \vdots & \vdots & & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nl} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1l} \\ c_{21} & c_{22} & \cdots & c_{2l} \\ \vdots & \vdots & & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{ml} \end{bmatrix}$$

where $c_{ij} \in k$ is defined by the formula:

$$c_{ij} := \sum_{r=1}^{n} a_{ir} b_{rj}.$$

Given a square matrix D, we say that D is *invertible* if there is another square matrix, written D^{-1}, such that $DD^{-1} = I$.

Proposition 7.2.28. *For $n, m \in \mathbb{N}^*$, a linear map $L : k^n \to k^m$ can be written as*

$$L : \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \mapsto A\mathbf{x} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

where $A = (a_{ij})$ is a real $m \times n$ matrix. We say that the matrix A represents the linear map L.

Example 7.2.29. For $n \in \mathbb{N}^*$, a linear map $L : k^n \to k$ is of the form

$$L : \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \mapsto a_1x_1 + \cdots + a_nx_n = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

where $a_1, a_2, \ldots, a_n \in k$.

Proposition 7.2.30 (Composition)**.** *Given $m, n, l \in \mathbb{N}^*$, suppose that linear maps $L_1 : k^n \to k^m$, $L_2 : k^l \to k^n$ are represented by the $m \times n$ matrix A and the $n \times l$ matrix B, respectively. Then the composition $L_1 \circ L_2 : k^l \to k^m$ is also linear, and represented by the product matrix AB.*

Proposition 7.2.31. *Suppose that A, B represent the same linear map $L : k^n \to k^n$. Then there is an invertible matrix P such that $A = PBP^{-1}$, and we say A and B are similar (相似).*

In high-dimensional, if $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at x, then the differential df at x is a linear map. It is represented by a matrix, called the *Jacobian matrix* (雅可比矩阵) or *Jacobian* for short. The entries of the Jacobian matrix are the *partial derivatives* (偏导数). See Theorem 6.4.16.

Definition 7.2.32 (Determinant)**.** The *determinant* (行列式) is the unique assignment of a number to a square matrix

$$A \mapsto \det(A)$$

that has the following four properties:

(1) $\det(I) = 1$.

(2) If A_1 is obtained by the exchange of two rows (or of two columns) of A, then $\det(A_1) = -\det(A)$.

(3) If A_2 is obtained by multiplying a row (or a column) of A by a number r, then $\det(A_2) = r \det(A)$.

(4) If A_3 is obtained by adding to a row (or a column) of A a multiple of another row (or a column) of A, then $\det(A_3) = \det(A)$.

Proposition 7.2.33. *If A and B are two $n \times n$ matrices, then*

$$\det(AB) = \det(A) \cdot \det(B).$$

In particular, for an invertible matrix P, we have

$$\det(P^{-1}) = \frac{1}{\det(P)}.$$

Then by Proposition 7.2.31, we can define the *determinant* $\det(L)$ of a linear map $L : k^n \to k^n$ by the determinant of a matrix representing L without causing ambiguity.

Theorem 7.2.34. *Let $L : \mathbb{R}^n \to \mathbb{R}^n$ be a linear map, $S \subset \mathbb{R}^n$ a measurable set (可测集). Then the volumes satisfy:*

$$\mathrm{vol}(L(S)) = |\det(L)| \cdot \mathrm{vol}(S).$$

Definition 7.2.35 (Transpose matrix)**.** Let $A = (a_{ij})$ be an $m \times n$ matrix. Then the *transpose matrix* $A^T = (b_{ij})$ of A is an $n \times m$ matrix defined by

$$b_{ij} = a_{ji}$$

for $1 \leq j \leq m$ and $1 \leq i \leq n$. Similarly, if A is a complex matrix, we define $A^* = (c_{ij})$ of A by

$$c_{ij} = \overline{a_{ji}}$$

for $1 \leq j \leq m$ and $1 \leq i \leq n$.

Proposition 7.2.36. *If A is an $n \times n$ matrix, then*

$$\det(A) = \det(A^T).$$

Definition 7.2.37 (Symmetric and Hermitian matrices)**.** Let M be a $n \times n$ square matrix. We say that M is *symmetric* if

$$M = M^T.$$

If M is a complex matrix, we say that M is *Hermitian* if

$$M = M^*.$$

Definition 7.2.38 (Orthogonal and unitary matrices)**.** Let M be a $n \times n$ square matrix. We say that M is *orthogonal* if

$$M^{-1} = M^T.$$

If M is a complex matrix, we say that M is *unitary* if

$$M^{-1} = M^*.$$

Remark 7.2.39. In particular, for row vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}\mathbf{y}^T$ defines an inner product on $\mathbb{R}^n$. Then for any orthogonal matrix M, we have

$$\langle \mathbf{x}M, \mathbf{y}M \rangle = \mathbf{x}M(\mathbf{y}M)^T = \mathbf{x}MM^T\mathbf{y}^T = \mathbf{x}MM^{-1}\mathbf{y}^T = \mathbf{x}\mathbf{y}^T = \langle \mathbf{x}, \mathbf{y} \rangle.$$

It follows that M preserves the inner product $\langle \cdot, \cdot \rangle$. A similar property holds for the unitary matrices.

Definition 7.2.40 (Diagonal matrix)**.** A $m \times n$ matrix $D = (d_{ij})$ is said to be *diagonal* if the only nonzero entries are d_{ii} for any $1 \leq i \leq \min\{m, n\}$.

Theorem 7.2.41 (Spectral theorem)**.** *Every real symmetric matrix M is orthogonally diagonalizable. In other words, there is an orthogonal matrix P and a real diagonal matrix D such that*

$$M = PDP^T.$$

Moreover, any complex Hermitian matrix N is unitarily diagonalizable. That is, there is an unitary matrix Q and a real diagonal matrix D such that

$$N = QDQ^*.$$

Theorem 7.2.42 (Singular value decomposition)**.** *Every $m \times n$ complex matrix A can be expressed as*

$$A = UDV^*$$

where U and V are $(m \times m)$ and $(n \times n)$ unitary matrices respectively, and D is a $(m \times n)$ diagonal matrix.

Thus, we see that if A is Hermitian, then $A = A^*$ and $UDV^* = VDU^*$, which implies $U = V$. In other words, the spectral decomposition (Theorem 7.2.41) is a special case of the singular value decomposition (Theorem 7.2.42).

7.3 First-order linear differential equation

In general, a differential equation of the form

$$f'(x) + P(x)f(x) = Q(x)$$

is called a *first-order linear differential equation.* To solve this, first write

$$df + fP(x)dx = Q(x)dx. \tag{7.3.1}$$

Then find the antiderivative $k(x)$ of $P(x)$, i.e.

$$k(x) = \int P(x)dx.$$

Then multiply both sides of (7.3.1) by $e^{k(x)}$:

$$e^{k(x)}df + fe^{k(x)}P(x)dx = e^{k(x)}Q(x)dx.$$

By the chain rule, we get

$$e^{k(x)}df + fde^{k(x)} = e^{k(x)}Q(x)dx.$$

By the product rule, we get

$$dfe^{k(x)} = e^{k(x)}Q(x)dx.$$

Integrate both sides

$$fe^{k(x)} = \int e^{k(x)}Q(x)dx.$$

Thus, we conclude

$$f(x) = e^{-k(x)} \int e^{k(x)}Q(x)dx.$$

7.4 Riemann integral and Lebesgue integral

A necessary and sufficient condition for Riemann integrability can be given in terms of *Lebesgue measure*.

Definition 7.4.1 (Set of measure zero)**.** We say a set $E \subset \mathbb{R}$ has *Lebesgue measure zero* (or *measure zero* (零测集) for short) if for every $\epsilon > 0$, there is a collection of open intervals $\{(a_k, b_k) : k \in \mathbb{N}\}$ such that

$$E \subset \bigcup_{k=1}^{\infty} (a_k, b_k), \quad \sum_{k=1}^{\infty} (b_k - a_k) < \epsilon.$$

Example 7.4.2. Finite sets, $\mathbb{N}$, and $\mathbb{Q}$ have measure zero. For $b > a$, $[a, b]$ does not have measure zero.

It has been proven independently by Vitali and by Lebesgue in 1907:

Theorem 7.4.3 (Lebesgue-Vitali theorem)**.** *A function on* $[a, b]$ *is Riemann integrable if and only if it is bounded, and its discontinuities have measure zero.*

Example 7.4.4. The Dirichlet function is not Riemann integrable by Theorem 7.4.3. However, the *Riemann function* (黎曼函数)

$$R(x) := \begin{cases} \dfrac{1}{q} & , \text{ if } x = \dfrac{p}{q} \in \mathbb{Q}, \ (p, q) = 1 \\ 0 & , \text{ if } x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$$

is Riemann integrable, since the discontinuities of $R(x)$ are $\mathbb{Q}$.

Example 7.4.5. A bounded monotone function on $[a, b]$ is Riemann integrable, since any monotone function has at most *countably* many discontinuities. Any *countable set* (可数集) has measure zero.

Lebesgue-Vitali theorem shows some defects of Riemann integral:

- Riemann integrable functions must be **bounded**. We shall see that

  $$\lim_{\epsilon \to 0+0} \int_{\epsilon}^{1} \frac{1}{\sqrt{x}} dx = \lim_{\epsilon \to 0+0} 2 - 2\sqrt{\epsilon} = 2.$$

  However, $\dfrac{1}{\sqrt{x}}$ is not Riemann integrable on $(0, 1]$.

- Riemann integrable functions cannot be defined on an unbounded interval. We shall see that

$$\lim_{n\to+\infty}\int_1^n \frac{1}{x^2}dx = \lim_{n\to+\infty} 1-\frac{1}{n} = 1.$$

However, Riemann integral cannot be defined on $[1,+\infty)$, since we cannot let $\lambda(T)\to 0$ for any finite partition T of $[1,+\infty)$. Then $\frac{1}{x^2}$ cannot be Riemann integrable on $[1,+\infty)$.

To fix these, Lebesgue developed the *Lebesgue integral* (勒贝格积分). In the theory of Lebesgue integral, Riemann integrable functions are **always** Lebesgue integrable. Moreover, the functions $\frac{1}{\sqrt{x}}$ on $(0,1]$, $\frac{1}{x^2}$ on $[1,+\infty)$, and even the Dirichlet function are Lebesgue integrable. The relationship between Riemann and Lebesgue integrable functions is similar to the relationship between $\mathbb{Q}$ and $\mathbb{R}$.

Advantages of Riemann integral:

- The idea is natural and straightforward.
- Most integrals that can be calculated are Riemann integrals.

Advantages of Lebesgue integral:

- The theory is beautiful. Mathematicians are satisfied. (But the idea is sophisticated and difficult to understand.)
- Many more functions are Lebesgue integrable.
- The Lebesgue integral can be defined for functions $f: X\to\mathbb{R}$ on an abstract space X. This is the origin of modern *analysis* (分析学) and *probability theory* (概率论).

7.5 Riemann integrability of continuous functions

Theorem 7.5.1. *Let f be continuous on $[a,b]$. Then f is Riemann integrable.*

证明. Let $\epsilon > 0$ be arbitrary. Since f is continuous, by Theorem 7.1.12, there exists a $\delta > 0$ such that

$$|f(x) - f(y)| < \epsilon \tag{7.5.1}$$

for any $x, y \in [a, b]$ with $|x - y| < \delta$.

Now we shall show the integrability of f by using the *Darboux sum* (cf. [1, P.132, Exercise 7]). Consider a partition $T = \{a = x_0 < x_1 < \cdots < x_n = b\}$ of $[a, b]$ with $\lambda(T) < \delta$. Let M_i and m_i be the maximum and minimum of f on $[x_{i-1}, x_i]$. Then by (7.5.1), we have

$$M_i - m_i < \epsilon.$$

Then the difference of upper and lower Darboux sums is

$$\sum_{i=1}^{n} M_i \Delta x_i - \sum_{i=1}^{n} m_i \Delta x_i = \sum_{i=1}^{n} (M_i - m_i)\Delta x_i \le \sum_{i=1}^{n} \epsilon \Delta x_i < \epsilon(b - a).$$

Now letting $\epsilon \to 0$, we see that the limits of upper and lower Darboux sums are the same, and so f is Riemann integrable. □

7.6 Stirling's formula

Stirling's formula (斯特林公式) is a useful approximation for $n!$ when n is large.

Theorem 7.6.1 (Stirling's formula)**.**

$$\lim_{n\to\infty} \frac{\sqrt{2\pi n} n^n e^{-n}}{n!} = 1.$$

证明. First, recall that we have following identity relative to *Gamma function* Γ:

$$\Gamma(n+1) = \int_0^{+\infty} t^n e^{-t} dt = n!. \tag{7.6.1}$$

So to estimate $n!$, it suffice to estimate the Gamma function. We make the substitution $t = n + s$, obtaining

$$n! = \int_{-n}^{+\infty} (n+s)^n e^{-n-s} ds$$

which we can simplify a little bit as

$$n! = \frac{n^n}{e^n}\int_{-n}^{+\infty}\left(1+\frac{s}{n}\right)^n e^{-s}ds = \frac{n^n}{e^n}\int_{-n}^{+\infty} e^{n\log(1+\frac{s}{n})-s}ds. \tag{7.6.2}$$

Next, we make the substitution $s = \sqrt{n}x$ to the integral (7.6.2), and obtain

$$n! = \frac{\sqrt{n}n^n}{e^n}\int_{-\sqrt{n}}^{+\infty} e^{n\log(1+\frac{x}{\sqrt{n}})-\sqrt{n}x}dx.$$

Note that for fixed x, we have the pointwise convergence

$$\lim_{n\to\infty} e^{n\log(1+\frac{x}{\sqrt{n}})-\sqrt{n}x} = e^{-\frac{x^2}{2}}. \tag{7.6.3}$$

Heuristically, (7.6.3) implies that

$$\lim_{n\to\infty}\int_{-\sqrt{n}}^{+\infty} e^{n\log(1+\frac{x}{\sqrt{n}})-\sqrt{n}x}dx = \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}}dx. \tag{7.6.4}$$

The limit (7.6.4) is true and deduced by the *Lebesgue dominated convergence theorem* (勒贝格控制收敛定理) which will not be discussed here.

Finally, the problem is reduced to calculate

$$\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}}dx = \sqrt{2\pi}$$

which is an important integral, called the *Gaussian integral* (高斯积分), and can be found in [2, P.27, Exercise 26(3)]. So we conclude the Stirling's formula. □

7.7 Partial fraction decomposition

Let k be a *field* (数域). We define the space of polynomials with k coefficients by

$$k[x] = \left\{\sum_{i=0}^{n} a_i x^i : n \in \mathbb{N},\ a_i \in k\right\}.$$

Now let $P(x) = \sum_{i=0}^{n} a_i x^i \in k[x]$ be a polynomial. Then the number n is called the *degree* (次数) of P, and written by $\deg(P)$. In addition, we say r is a *root* (根) of the polynomial $P(x)$ if r is a member of the domain of P and $P(r) = 0$.

Example 7.7.1. Note that roots depend on the domain of P. For instance, $P(x) = x^2 + 1$ does not have roots in $\mathbb{R}$, but $\pm i$ are roots of $P(x)$ in $\mathbb{C}$.

We say that a field k is *algebraically closed* if every non-constant polynomial in $k[x]$ has a root in k. The seminal *fundamental theorem of algebra* (代数基本定理) states that the field $\mathbb{C}$ of complex numbers is algebraically closed:

Theorem 7.7.2 (Fundamental theorem of algebra). *Every polynomial $P(x) \in \mathbb{C}[x]$ of degree $\deg(P) \geq 1$ has a root in $\mathbb{C}$.*

Example 7.7.3 (Polynomial with real coefficients). Let $P(x) \in \mathbb{R}[x]$ be a polynomial with real coefficients. Although $P(x)$ does not always have real roots, the fundamental theorem of algebra implies that $P(x)$ must have a complex root $z_0 \in \mathbb{C}$.

Moreover, if $z_0 \in \mathbb{C} \setminus \mathbb{R}$ is a root of $P(x)$, then the complex conjugate $\bar{z}_0$ is also a root of $P(x)$. In fact, it follows from the fact that

$$\overline{z_1 + z_2} = \bar{z}_1 + \bar{z}_2, \quad \overline{z_1 \cdot z_2} = \bar{z}_1 \cdot \bar{z}_2$$

for any $z_1, z_2 \in \mathbb{C}$. Thus,

$$\begin{aligned} 0 = \overline{P(z_0)} = \overline{a_0 + a_1 z_0 + \cdots + a_n z_0^n} &= \bar{a}_0 + \overline{a_1 z_0} + \cdots + \overline{a_n z_0^n} \\ &= a_0 + a_1 \bar{z}_0 + \cdots + a_n \bar{z}_0^n = P(\bar{z}_0). \end{aligned}$$

The following is a direct corollary of the fundamental theorem of algebra:

Corollary 7.7.4. *Every polynomial $P(x) = \sum_{i=0}^n a_i x^i \in \mathbb{C}[x]$ of degree $n \geq 1$ admits a representation in the form*

$$P(x) = a_n (x - z_1) \cdots (x - z_n)$$

where $z_1, \ldots, z_n \in \mathbb{C}$ (and the numbers $z_1, \ldots, z_n$ are not necessarily all distinct). This representation is unique up to the order of the factors.

证明. Let $Q(x) \in \mathbb{C}[x]$ be another polynomial with $\deg(Q) \leq \deg(P)$. Then the *polynomial long division* (多项式除法) implies that

$$P(x) = q(x)Q(x) + r(x)$$

where $q(x), r(x) \in \mathbb{C}[x]$ with $\deg(r) < \deg(Q)$. Thus, if $\deg(Q) = 1$, then $\deg(r) = 0$ and so $r(x) \equiv C$ is a constant.

Let z_1 be a root of $P(x)$. Then $P(x) = q(x)(x - z_1) + r$. Since $P(z_1) = r$, we see that $r = 0$. Thus, we have the representation $P(x) = q(x)(x - z_1)$. Note that the degree $\deg(q) = \deg(P) - 1$. Thus, the consequence follows from induction. $\square$

Corollary 7.7.5. *Every polynomial $P(x) \in \mathbb{R}[x]$ can be expressed as a product of linear and quadratic polynomials (二次多项式) with real coefficients. In other words,*

$$P(x) = P_1(x)P_2(x)\cdots P_m(x) \tag{7.7.1}$$

for $P_i(x) \in \mathbb{R}[x]$ with $\deg(P_i) \le 2$ $(i = 1, \ldots, m)$.

证明. By Corollary 7.7.4, we can write

$$P(x) = a_n(x - z_1)\cdots(x - z_n)$$

where $a_n \in \mathbb{R}$, $z_1, \ldots, z_n \in \mathbb{C}$. If $z_1 \in \mathbb{R}$, then we write $P_1(x) = x - z_1$, and complete an inductive step. If $z_1 \in \mathbb{C} \setminus \mathbb{R}$, then Example 7.7.3 implies that $\bar{z}_1$ is also a root of $P(x)$, i.e. $\bar{z}_1 = z_i$ for some $i = 1, \ldots, m$. Then we write

$$P_1(x) = (x - z_1)(x - z_i) = (x - z_1)(x - \bar{z}_1) = x^2 - (z_1 + \bar{z}_1)x + |z_1|^2.$$

Thus, the consequence follows from induction. $\square$

Let $P, Q \in k[x]$ be two polynomials. Then the *greatest common divisor* (最大公因式) of P and Q is a polynomial $G(x) \in k[x]$ of the highest possible degree, so that

$$P(x) = q_1(x)G(x), \quad Q(x) = q_2(x)G(x),$$

for some $q_1(x), q_2(x) \in k[x]$. It is also written by $\gcd(P, Q)$. Usually, we can use the *Euclidean algorithm* (欧几里得算法) to find the greatest common divisors.

Example 7.7.6. By multiplying out all the identical factors in (7.7.1), we can rewrite that product

$$P(x) = (P_1(x))^{d_1}\cdots(P_k(x))^{d_k}.$$

Then one easily sees that $\gcd((P_i(x))^{d_i}, (P_j(x))^{d_j}) = 1$ for any $i \ne j$.

Theorem 7.7.7 (Bézout's identity (裴蜀定理))**.** *For any $P, Q \in k[x]$, there exist $a, b \in k[x]$ such that*

$$\gcd(P, Q) = aP + bQ.$$

Theorem 7.7.8 (Partial fraction decomposition (部分分式分解))**.** *Let $P(x), Q(x) \in \mathbb{R}[x]$ and*

$$Q(x) = c\prod_{i=1}^{l}(x - x_i)^{k_i} \cdot \prod_{i=1}^{n}(x^2 + p_i x + q_i)^{m_i}.$$

Then there exists a unique partial fraction decomposition of the rational fraction $\frac{P(x)}{Q(x)}$:

$$\frac{P(x)}{Q(x)} = p(x) + \sum_{i=1}^{l}\sum_{j=1}^{k_i}\frac{A_{ij}}{(x - x_i)^j} + \sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{B_{ij}x + D_{ij}}{(x^2 + p_i x + q_i)^j}$$

where $p(x) \in \mathbb{R}[x]$ and $A_{ij}, B_{ij}, D_{ij} \in \mathbb{R}$.

Remark 7.7.9. The partial fraction decomposition can also be obtained for other fields k.

Proof of Proposition 7.7.8. First, if $\deg(P) \geq \deg(Q)$, then by the polynomial long division, we have

$$P(x) = q(x)Q(x) + r(x)$$

where $q(x), r(x) \in \mathbb{C}[x]$ with $\deg(r) < \deg(Q)$. It follows that

$$\frac{P(x)}{Q(x)} = q(x) + \frac{r(x)}{Q(x)}.$$

Thus, by replacing $P(x)$ by $r(x)$ if necessary, we may assume that $\deg(P) < \deg(Q)$.

Next, we claim that

Claim 7.7.10. *Suppose that $Q = Q_1Q_2$ for $Q_i \in \mathbb{R}[x]$ with $\gcd(Q_1, Q_2) = 1$. Then there exist $P_1, P_2 \in \mathbb{R}[x]$ such that*

$$\frac{P(x)}{Q(x)} = \frac{P_1(x)}{Q_1(x)} + \frac{P_2(x)}{Q_2(x)}$$

and $\deg(P_i) < \deg(Q_i)$ for $i = 1, 2$.

Proof of Claim 7.7.10. Bézout's identity (Theorem 7.7.7) asserts the existence of $C, D \in \mathbb{R}[x]$ such that

$$CQ_1 + DQ_2 = 1.$$

By the polynomial long division, we have

$$D(x)P(x) = q(x)Q_1(x) + P_1(x)$$

where $q(x), P_1(x) \in \mathbb{C}[x]$ with $\deg(P_1) < \deg(Q_1)$. Let

$$P_2(x) = C(x)P(x) + q(x)Q_2(x).$$

One calculates

$$\begin{aligned}\frac{P(x)}{Q(x)} &= \frac{P \cdot (CQ_1 + DQ_2)}{Q_1Q_2} \\ &= \frac{PD}{Q_1} + \frac{PC}{Q_2} \\ &= \frac{qQ_1 + P_1}{Q_1} + \frac{P_2 - qQ_2}{Q_2} = \frac{P_1}{Q_1} + \frac{P_2}{Q_2}.\end{aligned}$$

In particular, we have $P = P_1Q_2 + P_2Q_1$. To see $\deg(P_2) < \deg(Q_2)$, one observes

$$\begin{aligned}\deg(P_2) &= \deg(P - P_1Q_2) - \deg(Q_1) \\ &\leq \max\{\deg(P), \deg(P_1Q_2)\} - \deg(Q_1) \\ &< \max\{\deg(Q), \deg(Q_1Q_2)\} - \deg(Q_1) = \deg(Q_2).\end{aligned}$$

Thus, we establish the claim. $\square$

Therefore, by Example 7.7.6 and Claim 7.7.10, it reduces to show that for $P, Q \in \mathbb{R}[x]$ with $\deg(P) < \deg(Q^k)$, for $k > 1$, we have

$$\frac{P}{Q^k} = \frac{P_1}{Q} + \frac{P_2}{Q^2} + \cdots + \frac{P_k}{Q^k}$$

for $P_i \in \mathbb{R}[x]$ with $\deg(P_i) < \deg(Q)$.

This indeed follows from the polynomial long division again: If $\deg(P) \geq \deg(Q)$, then we have

$$P(x) = q(x)Q(x) + P_k(x)$$

where $q(x), P_k(x) \in \mathbb{C}[x]$ with $\deg(P_k) < \deg(Q)$. It follows that

$$\frac{P}{Q^k} = \frac{q}{Q^{k-1}} + \frac{P_k}{Q^k},$$

and

$$\begin{aligned}\deg(q) &= \deg(P - P_k) - \deg(Q) \\ &\leq \deg(P) - \deg(Q) < \deg(Q^k) - \deg(Q) = \deg(Q^{k-1}).\end{aligned}$$

The consequence now follows from induction. □

7.8 Real analytic functions

7.8.1 Power series, continued

Definition 7.8.1 (Absolute convergence)**.** For a series $\sum_{k=m}^{\infty} a_k$, if $\sum_{k=m}^{\infty} |a_k|$ is convergent, then we say the series $\sum_{k=m}^{\infty} a_k$ is *absolutely convergent* (绝对收敛).

We first discuss the convergence of a power series $\sum_{k=0}^{\infty} c_k(x - x_0)^k$. In general, the closer x gets to x_0, the easier it is for this series to converge.

Definition 7.8.2 (Radius of convergence)**.** Let $\sum_{k=0}^{\infty} c_k(x - x_0)^k$ be a formal power series. We define the *radius of convergence* (收敛半径) of this series to be the quantity

$$R := \frac{1}{\limsup_{n\to\infty} |c_n|^{\frac{1}{n}}} = \frac{1}{\lim_{m\to\infty} \sup\{|c_n|^{\frac{1}{n}} : n \geq m\}} \in [0, +\infty].$$

Theorem 7.8.3. *Let $\sum_{k=0}^{\infty} c_k(x - x_0)^k$ be a formal power series, and let R be its radius of convergence.*

(1) (Divergence outside of R) If $x \in \mathbb{R}$ is such that $|x - x_0| > R$, then the power series $\sum_{k=0}^{\infty} c_k(x - x_0)^k$ is divergent.

(2) (Convergence inside of R) If $x \in \mathbb{R}$ is such that $|x - x_0| < R$, then the power series $\sum_{k=0}^{\infty} c_k(x - x_0)^k$ is absolutely convergent.

证明. (1) Write $X = x - x_0$. For $n \in \mathbb{N}$, let

$$S_n(x) = \sum_{k=1}^{n} c_k(x - x_0)^k = \sum_{k=1}^{n} c_k X^k$$

be a partial sum. If the series converges, then we have

$$\lim_{n\to\infty} c_{n+1}X^{n+1} = \lim_{n\to\infty} S_{n+1}(x) - S_n(x) = 0. \tag{7.8.1}$$

On the other hand, by the assumption, we have $|X| > R(1+\epsilon)$ for some $\epsilon > 0$. It means

$$|c_{n+1}X^{n+1}| \geq |c_{n+1}|R^{n+1}(1+\epsilon)^{n+1} \nrightarrow 0, \quad (n \to \infty).$$

It contradicts (7.8.1). Hence, the series is divergent.

(2) Now by the assumption we have $|X| < R(1-\epsilon)$ for some $\epsilon > 0$. Then the series

$$\sum_{k=1}^{\infty} |c_k||X^k| \leq \sum_{k=1}^{\infty} |c_k|R^k(1-\epsilon)^k \lesssim \sum_{k=1}^{\infty}(1-\epsilon)^k < +\infty.$$

Therefore, we conclude the series is absolutely convergent. □

Theorem 7.8.4. *Let $\sum_{k=0}^{\infty} c_k(x - x_0)^k$ be a power series, and let R be its radius of convergence. Then $f(x) = \sum_{k=0}^{\infty} c_k(x - x_0)^k$ is well defined on $(x_0 - R, x_0 + R)$. Moreover, the following properties hold:*

(1) (Uniform convergence on bounded intervals) For any $r \in (0, R)$, the series $\sum_{k=0}^{\infty} c_k(x - x_0)^k$ converges uniformly to f on the interval $[x_0 - r, x_0 + r]$. In particular, f is continuous on $(a - R, a + R)$.

(2) (Differentiation of power series) The function f is differentiable on $(x_0 - R, x_0 + R)$. Moreover, For any $r \in (0, R)$, the series

$$\sum_{k=1}^{\infty} c_k k(x - x_0)^{k-1}$$

converges uniformly to f' on the interval $[x_0 - r, x_0 + r]$. In other words, the derivative operator and sum operator commute:

$$\frac{d}{dx}\left[\sum_{k=0}^{\infty} c_k(x-x_0)^k\right] = \sum_{k=0}^{\infty} \frac{d}{dx}\left[c_k(x-x_0)^k\right].$$

(3) (Integration of power series) For any closed interval $[u, v] \subset (x_0 - R, x_0 + R)$, we have

$$\int_u^v f(t)dt = \sum_{k=0}^{\infty} \frac{c_k}{k+1}[(u-x_0)^{k+1} - (v-x_0)^{k+1}].$$

In other words, the integral operator and sum operator commute:

$$\int_u^v \left[\sum_{k=0}^{\infty} c_k(t-x_0)^k\right] dt = \sum_{k=0}^{\infty} \left[\int_u^v c_k(t-x_0)^k dt\right].$$

证明. (1) Write $X = x - x_0$. For $n \in \mathbb{N}$, let

$$S_n(x) = \sum_{k=1}^{n} c_k(x-x_0)^k = \sum_{k=1}^{n} c_k X^k$$

be a partial sum with $|X| \le r$. Then $r < R - 2\epsilon$ for some $\epsilon > 0$, and

$$\|S_n - f\|_\infty \le \sup_{|X|\le r} \sum_{k=n+1}^{\infty} |c_k X^k| \le \sum_{k=n+1}^{\infty} |c_k| r^k \le \sum_{k=n+1}^{\infty} \frac{r^k}{(R-\epsilon)^k} \to 0, \quad n \to \infty.$$

Then for any $x, y \in [x_0 - r, x_0 + r]$, we have

$$\begin{aligned} |f(x) - f(y)| \le & |f(x) - S_n(x)| + |S_n(x) - S_n(y)| + |S_n(y) - f(y)| \\ \le & |S_n(x) - S_n(y)| + 2\|S_n - f\|_\infty. \end{aligned}$$

This leads to the continuity of f on $[x_0 - r, x_0 + r]$ for any $r \in (0, R)$.

(2) We use the notation similar to (1). For the derivative series, we calculate the radius of convergence:

$$R' = \frac{1}{\limsup_{n\to\infty} |c_n n|^{\frac{1}{n}}} = \frac{1}{\limsup_{n\to\infty} |c_n|^{\frac{1}{n}}} = R.$$

Then by Theorem 7.8.3(2) and Theorem 7.8.4(1), we conclude that the derivative series

$$S'(y) := \sum_{k=0}^{\infty} c_{k+1}(k+1)(y-x_0)^k$$

is absolutely and uniformly convergent on $[x_0 - r, x_0 + r]$. In particular, S' is continuous on $(x_0 - R, x_0 + R)$.

Now for $y \in [x_0 - r, x_0 + r]$, letting $n \to \infty$, we have

$$\begin{aligned}\left|S_n(y) - f(x_0) - \int_{x_0}^{y} S'(t)dt\right| &= \left|\int_{x_0}^{y} S_n(t) - S'(t)dt\right| \\ &\leq \int_{x_0}^{y} |S_n(t) - S'(t)|\, dt \\ &\leq \int_{x_0}^{y} \|S_n - S'\|_\infty dt = (y - x_0)\|S_n - S'\|_\infty \to 0.\end{aligned}$$

Since $S_n(y) \to f(y)$ as $n \to \infty$, by the first fundamental theorem of calculus (Theorem 2.9.3), we conclude that

$$f'(y) = S'(y)$$

for any $y \in [x_0 - r, x_0 + r]$ and any $r \in (0, R)$.

(3) Similar to (2), for the integral series, we calculate the radius of convergence:

$$R_I = \frac{1}{\limsup\limits_{n\to\infty} |c_n(n+1)^{-1}|^{\frac{1}{n}}} = \frac{1}{\limsup\limits_{n\to\infty} |c_n|^{\frac{1}{n}}} = R.$$

Then by Theorem 7.8.3(2) and Theorem 7.8.4(1)(2), we conclude that the integral series

$$S_I(y) := \sum_{k=0}^{\infty} \frac{c_k}{k+1}(y - x_0)^{k+1}$$

is absolutely and uniformly convergent on $[x_0 - r, x_0 + r]$. In particular, S_I is derivative on $(x_0 - R, x_0 + R)$.

Now by Theorem 7.8.4(1)(2), for $y \in (x_0 - R, x_0 + R)$ we have

$$\frac{d}{dy}S_I(y) = \sum_{k=0}^{\infty} \frac{d}{dy}\left[\frac{c_k}{k+1}(y - x_0)^{k+1}\right] = \sum_{k=0}^{\infty} c_k(y - x_0)^k = f(y).$$

The consequence follows from Newton-Leibniz formula. □

7.8.2 Real analytic functions

A function $f(x)$ which is lucky enough to be representable as a power series has a special name; it is a real *analytic function* (解析函数).

Definition 7.8.5 (Real analytic functions). Let $x_0 \in (A, B)$, $f : (A, B) \to \mathbb{R}$ be a function. We say that f is *real analytic* at x_0 if there is a neighborhood $U_r(x_0)$ of x_0, and a power series $\sum_{k=0}^{\infty} c_k(x - x_0)^k$ such that

$$f(x) = \sum_{k=0}^{\infty} c_k(x - x_0)^k$$

for $x \in U_r(x_0)$. If f is real analytic at every point x_0 of (A, B), we say that f is *real analytic* on (A, B).

From Theorem 7.8.4(2), we see that if f is real analytic at x_0, then f is differentiable on a neighborhood $U_R(x_0)$ of x_0, and more importantly, the derivative f' is again real analytic at x_0. We then conclude that a real analytic function f at x_0 must be infinitely differentiable on a neighborhood $U_R(x_0)$ of x_0. Moreover, the coefficients of the power series are determined by the Taylor's formula:

Proposition 7.8.6 (Taylor series). *Let $x_0 \in (A, B)$, $f : (A, B) \to \mathbb{R}$ be a real analytic function at x_0 and has the power series expansion:*

$$f(x) = \sum_{k=0}^{\infty} c_k(x - x_0)^k$$

for $x \in U_r(x_0)$, where $U_r(x_0)$ is some neighborhood of x_0. Then for any $n \in \mathbb{N}$, the partial sum

$$S_n = \sum_{k=0}^{n} c_k(x - x_0)^k$$

of the power series is the Taylor polynomial of order n of $f(x)$ at x_0. In particular, for any $k \in \mathbb{N}$, we have

$$c_k = \frac{f^{(k)}(x_0)}{k!}.$$

证明. Since f is analytic at x_0, we have

$$\frac{f(x) - S_n}{(x - x_0)^n} = \frac{1}{(x - x_0)^n} \sum_{k=n+1}^{\infty} c_k(x - x_0)^k = \sum_{k=1}^{\infty} c_{k+n}(x - x_0)^k.$$

Now note that

$$\sum_{k=1}^{\infty} |c_{k+n}(x - x_0)^k| \leq \sum_{k=1}^{\infty} |c_{k+n}(k + n) \cdots (k + 1)(x - x_0)^k|.$$

Then by Theorem 7.8.4(2), the series $\sum_{k=1}^{\infty} c_{k+n}(x-x_0)^k$ is absolutely convergent, and it determines a function $G(x)$ that is real analytic at x_0. In particular, $G(x)$ is continuous at x_0. Therefore, we conclude that

$$\lim_{x\to x_0} \frac{f(x)-S_n}{(x-x_0)^n} = 0.$$

Then by Theorem 4.3.14, we establish the proposition. □

Thus, by the Taylor's theorem (Theorem 4.3.12), we have

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k$$

for $x \in U_R(x_0)$. Also, the uniqueness of Taylor's polynomials (Theorem 4.3.14) implies that the coefficients of the series are uniquely determined by f, similar to the decimal representation of irrationals. In particular, we have shown that the assumption in Definition 4.3.21 that partial sums are required to be Taylor polynomials is redundant.

7.9 Uniform approximation by polynomials

In this section, we give a proof of the *Weierstrass approximation theorem* (Theorem 4.3.29).

Definition 7.9.1 (Compactly supported functions)**.** A function $f:\mathbb{R}\to\mathbb{R}$ is said to be *compactly supported* if it is supported on some closed interval $[a,b]$, i.e. $f(x)=0$ for any $x\notin[a,b]$.

Definition 7.9.2 (Convolution)**.** Let $f:\mathbb{R}\to\mathbb{R}$ and $g:\mathbb{R}\to\mathbb{R}$ be continuous, compactly supported functions. We define the *convolution* $f*g:\mathbb{R}\to\mathbb{R}$ of f and g to be the function

$$(f*g)(t) := \int_{-\infty}^{\infty} f(y)g(t-y)dy.$$

Convolutions play an important role in Fourier analysis and in partial differential equations (PDE). One motivation of this notion is the identity

$$\int_{-\infty}^{\infty} F(t)[(f*g)(t)dt] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(x+y)[f(x)dx][g(y)dy].$$

Proposition 7.9.3 (Basic properties of convolution)**.** *Let $f, g, h : \mathbb{R} \to \mathbb{R}$ be continuous, compactly supported functions. Then we have:*

(a) The convolution $f * g$ is also continuous, compactly supported.

(b) $f * g = g * f$.

(c) $f * (g + h) = f * g + f * h$.

(d) $f * (rg) = (rf) * g = r(f * g)$ for any $r \in \mathbb{R}$.

A key fact of convolutions is that $f * g$ takes good properties of f and g.

Lemma 7.9.4. *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function supported on $[0, 1]$, $g : \mathbb{R} \to \mathbb{R}$ be a polynomial supported on $[-1, 1]$. Then $f * g$ is a polynomial on $[0, 1]$.*

证明. Since g is a polynomial on $[-1, 1]$, we may find an integer $n \geq 0$ and $r_0, \dots, r_n$ such that

$$g(x) = r_0 + r_1 x + \cdots + r_n x^n$$

for all $x \in [-1, 1]$. It follows that

$$\begin{aligned} f * g(t) &= \int_0^1 f(y) \sum_{j=0}^{n} r_j (t - y)^j dy \\ &= \int_0^1 f(y) \sum_{j=0}^{n} \sum_{k=0}^{j} r_j \frac{j!}{k!(j-k)!} t^k (-y)^{j-k} dy \\ &= \int_0^1 f(y) \sum_{k=0}^{n} \sum_{j=k}^{n} r_j \frac{j!}{k!(j-k)!} t^k (-y)^{j-k} dy \\ &= \sum_{k=0}^{n} t^k \left(\int_0^1 f(y) \sum_{j=k}^{n} r_j \frac{j!}{k!(j-k)!} (-y)^{j-k} dy \right). \end{aligned}$$

Thus, $f * g$ is a polynomial on $[0, 1]$. □

Definition 7.9.5 (Approximation to the identity)**.** Let $\epsilon > 0$. A function $g_\epsilon : \mathbb{R} \to \mathbb{R}$ is said to be an *ϵ-approximation to the identity* (or a *nascent delta function*) if it obeys the following three properties:

(a) g_ϵ is supported on $[-1, 1]$, and $g_\epsilon(x) \geq 0$ for all $x \in [-1, 1]$.

(b) g_ϵ is continuous with $\int_{-\infty}^{\infty} g_\epsilon = 1$.

(c) $g_\epsilon(x) < \epsilon$ for all $\epsilon \le |x| \le 1$.

Remark 7.9.6. The *Dirac delta function* (狄拉克 δ 函数) is a *generalized function* 广义函数 (or a *measure* 测度) that can be viewed as the limit of a sequence of ϵ-approximations to the identity:

$$\delta(x) = \lim_{\epsilon \to 0+0} g_\epsilon(x).$$

The limit is meant in a weak sense:

$$\lim_{\epsilon \to 0+0} \int_{-\infty}^{\infty} f(x) g_\epsilon(x) dx = g(0)$$

for all continuous functions f having compact support.

The following lemma explains the terminology:

Lemma 7.9.7. *Let $\|\cdot\|_\infty$ be the uniform norm on $[0,1]$ (Definition 4.3.26). Let $f: \mathbb{R} \to \mathbb{R}$ be continuous, compactly supported on $[0,1]$. For $\epsilon > 0$, let $g_\epsilon : \mathbb{R} \to \mathbb{R}$ be an ϵ-approximation to the identity. Then we have*

$$\lim_{\epsilon \to 0+0} \|f * g_\epsilon - f\|_\infty = 0.$$

证明. By Theorem 7.1.12, for any $\epsilon > 0$, there exists a $\delta = \delta(\epsilon) > 0$ such that

$$|f(x) - f(y)| < \epsilon \tag{7.9.1}$$

for any $|x - y| < \delta$. We may further assume that $\delta < \epsilon$.

On the other hand, one easily deduces that

$$1 - 2\delta \le \int_{-\delta}^{\delta} g_\delta dx \le 1.$$

Now for any $\epsilon > 0$, $t \in [0,1]$, we let $\delta = \delta(\epsilon)$ be as in (7.9.1). Then we have

$$\begin{aligned}
&|f * g_\delta(t) - f(t)| \\
=&\left|\int_{-\infty}^{\infty} f(t-y)g_\delta(y)dy - f(t)\right| \\
\leq&\left|\int_{-\delta}^{\delta} f(t-y)g_\delta(y)dy - f(t)\right| + \left|\int_{\delta}^{1} f(t-y)g_\delta(y)dy\right| + \left|\int_{-1}^{-\delta} f(t-y)g_\delta(y)dy\right| \\
\leq&\left|\int_{-\delta}^{\delta} f(t-y)g_\delta(y)dy - f(t)\right| + 2\|f\|_\infty\delta \\
\leq&\left|\int_{-\delta}^{\delta} f(t)g_\delta(y)dy - f(t)\right| + \left|\int_{-\delta}^{\delta} \epsilon g_\delta(y)dy\right| + 2\|f\|_\infty\delta \\
\leq& 2\delta\|f\|_\infty + \epsilon + 2\|f\|_\infty\delta \leq (1+4\|f\|_\infty)\epsilon.
\end{aligned}$$

That is, $\|f * g_\delta - f\|_\infty \leq (1+4\|f\|_\infty)\epsilon$. $\square$

Another key fact is that polynomials can be approximations to the identity:

Lemma 7.9.8. *For any $\epsilon > 0$, there exists an ϵ-approximation to the identity which is a polynomial P_ϵ on $[-1,1]$.*

证明. Note first that for any $n \in \mathbb{N}$, $y \in [0,1]$, we have

$$c_n := \int_{-1}^{1}(1-x^2)^n dx \geq n^{-\frac{1}{2}}.$$

Then for any $\epsilon > 0$, one may find $n(\epsilon) > 0$ such that

$$(1-x^2)^{n(\epsilon)} < \epsilon$$

for any $\epsilon \leq |x| \leq 1$.

Now we define

$$P_\epsilon(x) := \begin{cases} c_{n(\epsilon)}^{-1}(1-x^2)^{n(\epsilon)} & \text{, if } x \in [-1,1] \\ 0 & \text{, if } x \notin [-1,1] \end{cases}.$$

Then $P_\epsilon(x)$ meets the requirement. $\square$

Proof of Theorem 4.3.29. Note that if P is a polynomial on $[0,1]$, then $Q(t) = P\left(\frac{t-A}{B-A}\right)$ is a polynomial of $t \in [A,B]$. Thus, by replacing f by $x \mapsto f(A+(B-A)x)$ if necessary, we may assume that $[A,B] = [0,1]$. Now the theorem follows from Lemmas 7.9.4, 7.9.7, 7.9.8. $\square$

In a similar spirit, one may obtain a much more general theorem known as the *Stone-Weierstrass theorem.*

7.10 Fourier series, continued

7.10.1 L^2 structure

We will need a couple more operations on the functions on S^1, before finding the Fourier series of them.

Definition 7.10.1 (Inner product)**.** If $f, g \in C(S^1, \mathbb{C})$ are continuous functions on S^1, we define the *inner product* $\langle f, g\rangle$ to be the quantity

$$\langle f, g\rangle = \int_{S^1} f(z)\overline{g(z)}dz = \int_0^1 f(e^{2\pi i\theta})\overline{g(e^{2\pi i\theta})}d\theta.$$

(Note that $f(z)\overline{g(z)}$ will be Riemann integrable since both functions are continuous.)

Example 7.10.2. $\langle z^n, z^n\rangle = 1$ for any $n \in \mathbb{Z}$, and $\langle z^n, z^m\rangle = 0$ for any two different $m, n \in \mathbb{Z}$. Moreover, $\langle f, f\rangle \geq 0$ for any continuous $f : S^1 \to \mathbb{C}$.

Remark 7.10.3. In general, the inner product $\langle f, g\rangle$ will be a complex number.

Roughly speaking, the inner product $\langle f, g\rangle$ is to the space $C(S^1, \mathbb{C})$ what the inner product $x \cdot y$ is to Euclidean space $\mathbb{R}^n$.

Lemma 7.10.4 (Properties of the inner product)**.** *Let $f, g, h \in C(S^1, \mathbb{C})$ be continuous functions on S^1.*

(1) (Hermitian property) We have $\langle g, f\rangle = \overline{\langle f, g\rangle}$.

(2) (Positivity) We have $\langle f, f\rangle \geq 0$. Moreover, we have $\langle f, f\rangle = 0$ if and only if $f \equiv 0$.

(3) (Bilinearity) We have

$$\langle f + g, h\rangle = \langle f, h\rangle + \langle g, h\rangle, \quad \langle f, g + h\rangle = \langle f, g\rangle + \langle f, h\rangle.$$

For any $c \in \mathbb{C}$, we have

$$\langle cf, g\rangle = c\langle f, g\rangle, \quad \langle f, cg\rangle = \bar{c}\langle f, g\rangle.$$

Definition 7.10.5 (L^2-norm). For a continuous $f : S^1 \to \mathbb{C}$, the *L^2-norm* (L^2 范数) $\|f\|_2$ of f is defined by

$$\|f\|_2 = \sqrt{\langle f, f\rangle} = \left(\int_{S^1} f(z)\overline{f(z)}dz\right)^{\frac{1}{2}} = \left(\int_{S^1} |f(z)|^2 dz\right)^{\frac{1}{2}}.$$

Lemma 7.10.6 (Properties of the L^2-norm). *Let $f, g \in C(S^1, \mathbb{C})$ be continuous functions on S^1.*

(1) (Non-degeneracy) We have $\|f\|_2 = 0$ if and only if $f = 0$.

(2) (Cauchy-Schwarz inequality) We have

$$|\langle f, g\rangle| \le \|f\|_2 + \|g\|_2.$$

(3) (Triangle inequality) We have

$$\|f + g\|_2 \le \|f\|_2 + \|g\|_2.$$

(4) (Pythagoras theorem 勾股定理) If $\langle f, g\rangle = 0$, then

$$\|f + g\|_2^2 \le \|f\|_2^2 + \|g\|_2^2.$$

(5) (Homogeneity) For any $c \in \mathbb{C}$, we have

$$\|cf\|_2 = |c| \cdot \|f\|_2.$$

In particular, since we have the triangle inequality, we conclude that $\|f - g\|_2$ measures the distance between f and g. Hence, we say that a sequence $f_n \in C(S^1, \mathbb{C})$ *converges in the L^2-norm* to a function $f \in C(S^1, \mathbb{C})$, denoted $f_n \xrightarrow{L^2} f$ as $n \to \infty$, if we have

$$\lim_{n\to\infty} \|f_n - f\|_2 = 0.$$

Now we are in the position to discuss the Fourier series (Definition 4.3.30) of a function $f \in C(S^1, \mathbb{C})$. For $n \in \mathbb{Z}$, the Fourier transform (Definition 4.3.32) of f is given by

$$\hat{f}(n) := \langle f, z^n\rangle = \int_{S_1} f(z)z^{-n}dz.$$

Then clearly, if $f(z) = \sum_{n=-N}^{N} c_n z^n$ is a trigonometric polynomial, then we have the *Fourier inversion formula*:

$$f(z) = \sum_{n=-N}^{N} \hat{f}(n) z^n = \sum_{n=-N}^{N} \langle f, z^n \rangle z^n = \sum_{n=-\infty}^{\infty} \langle f, z^n \rangle z^n.$$

Then, by the *Pythagoras theorem* (勾股定理), we have the *Plancherel formula*:

$$\|f\|_2^2 = \sum_{n=-N}^{N} |\hat{f}(n)|^2 = \sum_{n=-\infty}^{\infty} |\hat{f}(n)|^2. \tag{7.10.1}$$

Thus, we prove Theorems 4.3.34 and 4.3.35 in the case when f is a trigonometric polynomial.

In the following, we shall extend the Fourier inversion and Plancherel formula to general functions in $C(S^1, \mathbb{C})$. To do this, we will need a version of the *Weierstrass approximation theorem*.

7.10.2 Weierstrass approximation theorem for trigonometric polynomials

In this section, we discuss the *Weierstrass approximation theorem* for trigonometric polynomials. It is again a special case of the *Stone-Weierstrass theorem*. The latter is beyond the scope of this text. So we only present the proof of this special case. It is very similar to the proof of Theorem 4.3.29 in Section 7.9.

Theorem 7.10.7 (Weierstrass approximation theorem)**.** *Let $f \in C(S^1, \mathbb{C})$. Then for any $\epsilon > 0$, there is a trigonometric polynomial $P : S^1 \to \mathbb{C}$ such that*

$$\|f - P\|_\infty < \epsilon.$$

Definition 7.10.8 (Convolution)**.** Let $f, g \in C(S^1, \mathbb{C})$. We define the *convolution* $f * g : S^1 \to \mathbb{C}$ of f and g to be the function

$$(f * g)(z) := \int_{S^1} f(y) g(zy^{-1}) dy = \int_0^1 f(e^{2\pi i x}) g(e^{2\pi i (t-x)}) dx$$

for $z = e^{2\pi i t} \in S^1$.

Similar to Proposition 7.9.3, we have:

Proposition 7.10.9 (Basic properties of convolution)**.** *Let $f, g, h \in C(S^1, \mathbb{C})$. Then we have:*

(a) The convolution $f * g$ is continuous, i.e. $f * g \in C(S^1, \mathbb{C})$.

(b) $f * g = g * f$.

(c) $f * (g + h) = f * g + f * h$.

(d) $f * (cg) = (cf) * g = c(f * g)$ for any $c \in \mathbb{R}$.

The following is analogous to Lemma 7.9.4.

Lemma 7.10.10. *Let $f \in C(S^1, \mathbb{C})$. Let $P(z) = \sum_{n=-N}^{N} c_n z^n$ be a trigonometric polynomial. Then we have:*

$$f * P(z) = \sum_{n=-N}^{N} c_n \hat{f}(n) z^n.$$

Again, we need the notion of approximation to the identity.

Definition 7.10.11 (Approximation to the identity)**.** Let $\epsilon > 0$. A function $g_\epsilon \in C(S^1, \mathbb{C})$ is said to be an ϵ-*approximation to the identity* (or a *nascent delta function*) if it obeys the following three properties:

(a) $g_\epsilon(z) \geq 0$ for all $z \in S^1$.

(b) $\int_{-\infty}^{\infty} g_\epsilon = 1$.

(c) $g_\epsilon(e^{2\pi i t}) < \epsilon$ for all $\epsilon \leq t \leq 1 - \epsilon$.

Lemma 7.10.12. *Let $\| \cdot \|_\infty$ be the uniform norm on S^1 (Definition 4.3.26). Let $f \in C(S^1, \mathbb{C})$. For $\epsilon > 0$, let $g_\epsilon : S^1 \to \mathbb{C}$ be an ϵ-approximation to the identity. Then we have*

$$\lim_{\epsilon \to 0+0} \|f * g_\epsilon - f\|_\infty = 0.$$

证明. By Theorem 7.1.12, for any $\epsilon > 0$, there exists a $\delta = \delta(\epsilon) > 0$ such that

$$|f(e^{2\pi it}) - f(e^{2\pi is})| < \epsilon \tag{7.10.2}$$

for any $|t-s| < \delta$. We may further assume that $\delta < \epsilon$.

On the other hand, one easily deduces that

$$1 - 2\delta \leq \int_{-\delta}^{\delta} g_\delta(e^{2\pi ix})dx \leq 1.$$

Now for any $\epsilon > 0$, $z = e^{2\pi it}, y = e^{2\pi is} \in S^1$, we let $\delta = \delta(\epsilon)$ be as in (7.10.2). Then we have

$$\begin{aligned}
&|f * g_\delta(z) - f(z)| \\
=&\left|\int_{S^1} f(zy^{-1})g_\delta(y)dy - f(z)\right| \\
\leq&\left|\int_{-\delta}^{\delta} f(e^{2\pi i(t-s)})g_\delta(e^{2\pi is})ds - f(e^{2\pi it})\right| + \left|\int_{\delta}^{1-\delta} f(e^{2\pi i(t-s)})g_\delta(e^{2\pi is})ds\right| \\
\leq&\left|\int_{-\delta}^{\delta} f(e^{2\pi i(t-s)})g_\delta(e^{2\pi is})ds - f(e^{2\pi it})\right| + 2\|f\|_\infty\delta \\
\leq&\epsilon\left|\int_{-\delta}^{\delta} g_\delta(e^{2\pi is})ds\right| + \left|f(e^{2\pi it})\int_{-\delta}^{\delta} g_\delta(e^{2\pi is})ds - f(e^{2\pi it})\right| + 2\|f\|_\infty\delta \\
\leq&\epsilon + 2\delta\|f\|_\infty + 2\|f\|_\infty\delta \leq (1 + 4\|f\|_\infty)\epsilon.
\end{aligned}$$

That is, $\|f * g_\delta - f\|_\infty \leq (1 + 4\|f\|_\infty)\epsilon$. □

Lemma 7.10.13. *For any $\epsilon > 0$, there exists an ϵ-approximation to the identity which is a trigonometric polynomial P_ϵ on S^1.*

证明. Note first that for any $N \in \mathbb{N}^*$, $z \in S^1$, we define the *Fejér kernel*

$$F_N(z) := \sum_{n=-N}^{N} \left(1 - \frac{|n|}{N}\right) z^n.$$

One then deduces that

$$F_N(z) = \frac{1}{N}\left|\sum_{n=0}^{N-1} z^n\right|^2 = \begin{cases} \frac{\sin^2(\pi Nx)}{N\sin^2(\pi x)} & , \text{ if } z = e^{2\pi ix} \neq 1 \\ N & , \text{ if } z = 1 \end{cases}.$$

Thus, $F(z) \geq 0$ for any $z \in S^1$. Also, we have

$$\int_{S^1} F_N(z)dz = \sum_{n=-N}^{N} \left(1 - \frac{|n|}{N}\right) \int_{S^1} z^n dz = 1.$$

Finally, since $\sin^2(\pi N x) \leq 1$, we have

$$F_N(e^{2\pi i x}) \leq \frac{1}{N\sin^2(\pi x)} \leq \frac{1}{N\sin^2(\pi \epsilon)}$$

for $\epsilon < x < 1-\epsilon$. Thus, by choosing N large enough, we can make $F_N(e^{2\pi i x}) \leq \epsilon$ for all $\epsilon < x < 1-\epsilon$. □

Therefore, combining Lemmas 7.10.10, 7.10.12, 7.10.13, we conclude Theorem 7.10.7.

7.10.3 Fourier and Plancherel theorems

In this section, we can deduce the Fourier inversion theorem (Theorem 4.3.34) and Plancherel theorem (Theorem 4.3.35) using Weierstrass approximation theorem (Theorem 7.10.7).

Theorem 7.10.14 (Weierstrass M-test)**.** *Let (X, d) be a metric space. For $n \in \mathbb{N}$, $f_n : X \to \mathbb{C}$ be a sequence of bounded continuous functions on X such that the series*

$$\sum_{n=1}^{\infty} \|f_n\|_\infty < +\infty.$$

Then the series $\sum_{n=1}^{\infty} f_n$ converges uniformly to some continuous function $f : X \to \mathbb{C}$.

证明. For a fixed $x \in X$, consider the partial sum

$$S_i = \sum_{n=1}^{i} f_n(x).$$

Then for any $k \geq 1$,

$$|S_i - S_{i+k}| = \left|\sum_{n=i+1}^{i+k} f_n(x)\right| \leq \sum_{n=i+1}^{\infty} \|f_n\|_\infty \to 0, \quad i \to \infty. \tag{7.10.3}$$

Thus, $\{S_i\}$ is a Cauchy sequence in $\mathbb{C}$. By Theorem 7.1.4, there is a number $S = S(x) \in \mathbb{C}$ such that

$$S_i \to S, \quad i \to \infty. \tag{7.10.4}$$

To show the uniform convergence, fix $i \in \mathbb{N}$. Then by (7.10.4), for $x \in X$, $\epsilon > 0$, there exists $k(x, \epsilon) > i$ such that

$$|S(x) - S_{k(x,\epsilon)}(x)| \le \epsilon.$$

Then by (7.10.3), we have

$$|S(x) - S_i(x)| \le |S(x) - S_{k(x,\epsilon)}(x)| + |S_{k(x,\epsilon)}(x) - S_i(x)| \le \epsilon + \sum_{n=i+1}^{\infty} \|f_n\|_\infty.$$

Since $x \in X$ and $\epsilon > 0$ are chosen arbitrarily, we conclude that

$$\|S - S_i\|_\infty \le \sum_{n=i+1}^{\infty} \|f_n\|_\infty.$$

It leads to the uniform convergence.

Finally, for $x_n \to x \in X$, we have

$$\begin{aligned} |S(x) - S(x_n)| \le & |S(x) - S_i(x)| + |S_i(x) - S_i(x_n)| + |S_i(x_n) - S(x_n)| \\ \le & |S_i(x) - S_i(x_n)| + 2\|S - S_i\|_\infty. \end{aligned}$$

This leads to the continuity of S. □

Proof of Theorem 4.3.34. By Theorem 7.10.14, we see that $\sum_{n\in\mathbb{Z}} \hat{f}(n) z^n$ uniformly converges to some continuous function S. Thus, we have

$$\lim_{N\to\infty} \|S - F_N\|_\infty = 0, \quad F_N = \sum_{n=-N}^{N} \hat{f}(n) z^n.$$

In particular, since $\|\cdot\|_2 \le \|\cdot\|_\infty$, we have

$$\lim_{N\to\infty} \|S - F_N\|_2 = 0. \tag{7.10.5}$$

We shall show that $S = f$. First, by Theorem 7.10.7, for $\epsilon > 0$, there exists a trigonometric polynomial P of order $N_0 = N_0(f, \epsilon)$ such that

$$\|f - P\|_2 \le \|f - P\|_\infty \le \epsilon.$$

Now note that for $|n| \leq N_0$

$$\langle f - F_{N_0}, z^n \rangle = \langle f, z^n \rangle - \langle F_{N_0}, z^n \rangle = \hat{f}(n) - \hat{f}(n) = 0.$$

Then since $F_{N_0} - P$ is a trigonometric polynomial of order N_0, we have

$$\langle f - F_{N_0}, F_{N_0} - P \rangle = 0.$$

Thus, by the Pythagoras theorem, we have

$$\|f - F_{N_0}\|_2 \leq \|f - F_{N_0}\|_2 + \|F_{N_0} - P\|_2 = \|f - P\|_2 < \epsilon.$$

It follows that

$$\lim_{N \to \infty} \|f - F_N\|_2 = 0. \tag{7.10.6}$$

Comparing (7.10.5)(7.10.6), we conclude that $S = f$. □

Proof of Theorem 4.3.35. Let $\epsilon > 0$ and

$$F_N = \sum_{n=-N}^{N} \hat{f}(n) z^n.$$

By Fourier inversion theorem (7.10.6), there exists an $N = N(\epsilon) > 0$ such that

$$\|f - F_M\|_2 < \epsilon$$

for all $M \geq N$. It follows that

$$\|f\|_2 - \epsilon \leq \|F_M\|_2 \leq \|f\|_2 + \epsilon.$$

On the other hand, by (7.10.1), we have

$$S_M := \sum_{n=-M}^{M} |\hat{f}(n)|^2 = \|F_M\|_2^2.$$

Then one deduces

$$(\|f\|_2 - \epsilon)^2 \leq S_M \leq (\|f\|_2 + \epsilon)^2 \tag{7.10.7}$$

In particular, we get

$$|S_M - S_N| \leq (\|f\|_2 + \epsilon)^2 - (\|f\|_2 - \epsilon)^2 = 4 \|f\|_2 \epsilon$$

for all $M \geq N$. This means $\{S_M\}$ is a Cauchy sequence. Suppose that $S_M \to S$ as $M \to \infty$. Then by Theorem 1.3.1 and (7.10.7), we obtain

$$\|f\|_2 - \epsilon \leq \sum_{n=-\infty}^{\infty} |\hat{f}(n)|^2 \leq \|f\|_2 + \epsilon.$$

Letting $\epsilon \to 0$, we establish the Plancherel theorem. $\square$

Chapter 8

Problems Sets

8.1 Homework 1

- 9.10:

  [1] Exercises 1.1 (P.17): 4(1)(2).

  [1] Exercises 1.2 (P.27): 1(2), 2(4), 3(2), 5, 7, 10(2).

- 9.12:

  [1] Exercises 1.2 (P.29): 12(6).

  [1] Exercises 1.3 (P.46): 3(1)(2), 4(5), 5, 6, 7(3)(4), 8(1) (This is the famous *Basel problem* 巴塞尔问题, and you should search the history of it).

- 9.19:

  [1] Exercises 1.4 (P.61): 1(3), 2, 3(7)(8)(10)(14), 4(3)(4)(7).

  [1] Exercises of Chapter 1 (P.74): 9, 13.

8.2 Homework 2

- 9.24:

  [1] Exercises 1.5 (P.69): 3, 4(2), 5(4), 6, 7(3)(4).

[1] Exercises 1.6 (P.73): 2, 5.

[1] Exercises of Chapter 1 (P.74): 15, 20, 21, 22, 23(3)(4), 24.

- 9.26:

  [1] Exercises 1.6 (P.73): 1.

  [1] Exercises 2.1 (P.88): 4, 9, 13, 14.

  [1] Exercises 2.2 (P.99): 4(8)(14)(16).

  [1] Exercises of Chapter 2 (P.145): 3, 4, 8, 11.

8.3 Homework 3

- 10.8:

  [1] Exercises 2.5 (P.113): 8(3), 10(2)(3).

  [1] Exercises 2.6 (P.118): 5, 8.

  Extra Problem 8.3.1 (2020 Calculus B, Midterm)**.** Let $f(x) = (\arcsin x)^2$, find $f^{(n)}(0)$ for $n \in \mathbb{N}^*$.

  Extra Problem 8.3.2 (2023 Calculus A, Midterm)**.** Find the second derivative $\dfrac{d^2 f}{dx^2}$ of the function

$$f(x) = \begin{cases} x^4 \sin \dfrac{1}{x} & , \text{ if } x \neq 0 \\ 0 & , \text{ if } x = 0 \end{cases} .$$

  Extra Problem 8.3.3 (2023 Calculus A, Midterm)**.** Let $f(x) = x^n(1-x)^n$, and

$$F(x) = f(x) - f''(x) + f^{(4)}(x) - \cdots + (-1)^n f^{(2n)}(x).$$

  Find $\dfrac{d}{dx}(F'(x)\sin x - F(x)\cos x)$.

- 10.10:

  [1] Exercises 2.7 (P.124): 14, 18.

  [1] Exercises 2.8 (P.132): 7 (cf. *Darboux sum* 达布和).

  [1] Exercises of Chapter 2 (P.146): 10 (cf. *Cauchy-Schwarz inequality* 柯西-施瓦茨不等式), 14(1)(2)(3).

  [1] Exercises of Chapter 3 (P.207): 16(2)(3), 30, 31(1).

  Extra Problem 8.3.4 (2023 Calculus A, Midterm)**.** Let $f(x)$ be a Riemann integrable function on $[0, 1]$. Find the limit

  $$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} (-1)^{k-1} f\left(\frac{k}{n}\right).$$

  Extra Problem 8.3.5 (2022 Calculus A, Midterm)**.** Solve the differential equation

  $$g'(x) + g(x) = 1 + e^{-x}. \tag{8.3.1}$$

  In other words, find $g(x)$ that satisfies the equation (8.3.1).

8.4 Homework 4

- 10.15:

  [1] Exercises 2.9 (P.138): 1(4), 2, 5, 6.

  [1] Exercises 2.10 (P.144): 2, 3(2)(3), 4(3), 5 (a weaker form of the *Lagrange's mean value theorem* 拉格朗日中值定理 (Theorem 4.1.6)).

  [1] Exercises of Chapter 3 (P.205): 2, 5, 11, 12, 19, 31(2) (in most literature, it is called *Young's inequality* 杨氏不等式).

  Extra Problem 8.4.1 (2023 Calculus A, Midterm)**.** Find the derivative $\dfrac{df}{dx}$ of the function

  $$f(x) = \int_{\cot x}^{\tan x} \sqrt{1 + t^2}dt.$$

- 10.17:

  [1] Exercises 3.1 (P.155): 15, 24, 35.

  [1] Exercises 3.2 (P.162): 10, 16.

  Extra Problem 8.4.2 (2023 Calculus A, Midterm)**.** Solve

  $$\int \frac{\arctan e^x}{e^x + e^{-x}} dx.$$

  Extra Problem 8.4.3 (2021 Calculus B, Midterm)**.** Show that for any continuous function $f(x)$ on $[0, 2\pi]$, we have

  $$\lim_{n\to\infty} \int_0^{2\pi} f(x) \sin nx dx = 0.$$

  (If f is replaced by an integrable function, it is called the *Riemann-Lebesgue lemma* 黎曼-勒贝格引理.)

  Extra Problem 8.4.4 (2023 Calculus A, Midterm)**.** Let $y = y(x)$ be a function determined by the equation $y^2(x - y) = x^2$. Solve

  $$\int \frac{1}{y^2} dx.$$

  Extra Problem 8.4.5 (Gamma function)**.** For $x \in (0, +\infty)$, the *Gamma function* (伽玛函数) is defined by

  $$\Gamma(x) = \lim_{r\to+\infty} \int_0^r t^{x-1} e^{-t} dt = \int_0^{+\infty} t^{x-1} e^{-t} dt.$$

  Show that

  (1) $\Gamma(1) = 1$.

  (2) For $n \in \mathbb{N}^*$, $\Gamma(n + 1) = n \cdot \Gamma(n)$. Then conclude that

  $$\Gamma(n + 1) = n!.$$

  See [2, P.367] for more details.

Extra Problem 8.4.6 (Heisenberg uncertainty principle 海森堡不确定性原理)**.** Let f be a function on $[0,1]$ with a continuous derivative function f'. Suppose further that $f(0) = f(1) = 0$ and $\int_0^1 f^2(x)dx = 1$. Show that

$$\left(\int_0^1 (f'(x))^2 dx\right)\left(\int_0^1 (xf(x))^2 dx\right) \geq \frac{1}{4}. \tag{8.4.1}$$

Remark 8.4.7. A *complex-valued function* $f : U \to \mathbb{C}$ is a function so that its codomain is the *complex numbers* (复数). In the Fourier analysis, the *Fourier transform* (傅里叶变换, see Definition 4.3.32) is an assignment of a complex-valued function $\hat{f}$ to a complex-valued function f:

$$f(x) \mapsto \hat{f}(\xi)$$

that has the following properties:

- It is *unitary*, in the sense that

  $$\int_{-\infty}^{+\infty} |\hat{f}(\xi)|^2 d\xi = \int_{-\infty}^{+\infty} |f(x)|^2 dx. \tag{8.4.2}$$

  It is also called the *Plancherel theorem* (普朗歇尔定理), which is considered as a generalization of *Parseval's identity* (帕塞瓦尔恒等式).

- The derivative operator transforms to an algebraic operator:

  $$\widehat{\frac{d}{dx}f}(\xi) = i\xi \cdot \hat{f}(\xi). \tag{8.4.3}$$

  See [2, P.411] for more details.

Then, combining (8.4.2), (8.4.3) with the inequality (8.4.1), we obtain the following beautiful inequality, namely the *Heisenberg uncertainty principle*:

$$\left(\int_{-\infty}^{+\infty} |\xi\hat{f}(\xi)|^2 d\xi\right)\left(\int_{-\infty}^{+\infty} |xf(x)|^2 dx\right) \geq \frac{1}{4}.$$

8.5 Homework 5

8.5.1 Homework from the Textbook

- [1] Exercises 3.3 (P.174): 6, 19, 32.

  [1] Exercises 3.4 (P.185): 21, 22, 23, 24(1)(2), 25.

  [1] Exercises 3.5 (P.198): 14, 16, 24.

  [1] Exercises of Chapter 3 (P.206): 8(3)(4), 13(1)(3), 24, 32.

8.5.2 Practice Exam

Problem 8.5.1 (2022 Calculus A, Midterm (5pts))**.** Is the irrational power of an irrational number always irrational? That is, is it true that $x^y \in \mathbb{Q}^c$ for any $x, y \in \mathbb{Q}^c$?

Problem 8.5.2 (2022 Calculus A, Midterm (15pts))**.**

(1) $\lim\limits_{x \to +\infty} \dfrac{\sqrt{x + 2\sqrt{x + 2\sqrt{x}}}}{\sqrt{x+4}}$.

(2) $\lim\limits_{x \to 1} \dfrac{1}{x-1} - \dfrac{2}{x^2-1} + \dfrac{3}{x^3-1} - \dfrac{4}{x^4-1}$.

(3) $\lim\limits_{x \to +\infty} 2021\sqrt{x+2021} + 2023\sqrt{x+2023} - 2 \cdot 2022\sqrt{x+2022}$.

Problem 8.5.3 (2021 Calculus B, Midterm (15pts))**.**

(1) Find the derivative $\dfrac{df}{dx}$ of the function $f(x) = x^{\arcsin x}$ $(x \in (0,1))$.

(2) Find the derivative $\dfrac{df}{dx}$ of the function $f(x) = \displaystyle\int_e^{e^x} \frac{dt}{1 + \ln t}$ $(x \in (1, +\infty))$.

(3) Find $\dfrac{d^3 f}{dx^3}(0)$ of the function $f(x) = \arctan x$.

Problem 8.5.4 (2023 Calculus A, Midterm (10pts))**.**

(1) Find $\displaystyle\int \sqrt{1+x^2}dx$.

(2) Find $\int_0^{\pi} f(x)dx$, where

$$f(x) = \int_0^x \frac{\sin t}{\pi - t} dt.$$

Problem 8.5.5 (2020 Calculus B, Midterm (20pts))**.**

(1) Find $\int_0^1 \ln(x + \sqrt{x^2+1})dx$.

(2) Find $\int \frac{4x^3 + 2x^2 + 3x + 1}{x(x+1)(x^2+1)} dx$.

(3) Find $\int_0^1 x^4\sqrt{1-x^2}dx$.

(4) Find $\int_{-1}^1 (x^4 + 2x^2 + 1)\sin^3 x dx$.

Problem 8.5.6 (2021 Calculus B, Midterm (10pts))**.** For $x > 0$, let

$$p(x) = \int_0^x \frac{dt}{\sqrt{t^3 + 2021}}.$$

Prove that the equation

$$p(x+1) = p(x) + \sin x$$

has infinitely many different positive real solutions.

Problem 8.5.7 (2022 Calculus B, Midterm (15pts))**.** Let

$$f(x) = \int_0^{2\pi} \ln(1 - 2x\cos t + x^2)dt.$$

(1) (5pts) Prove that $2f(x) = f(x^2)$ for any $x \in (-1, 1)$.

(2) (5pts) Prove that $f(x)$ is bounded on $(-\frac{1}{2}, \frac{1}{2})$.

(3) (5pts) Find the explicit formula of $f(x)$ for $x \in (-1, 1)$.

Problem 8.5.8 (2021 Calculus C, Midterm (10pts))**.** Suppose that $x_1 > 1$ and

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

for any $n \in \mathbb{N}^*$. Show that the limit $\lim_{n\to\infty} x_n$ exists, and find the limit.

8.6 Homework 6

8.6.1 Practice Exam

Problem 8.6.1 (2022 Calculus A, Midterm (10pts)). Find a sequence $\{a_n\} \subset \mathbb{R}$ such that both of the following hold:

(1) $\lim_{n\to\infty}(a_n - e^n) = 0$;

(2) $\lim_{n\to\infty}(f(a_n) - f(e^n)) \neq 0$ where $f(x) = x\ln x$, for $x \in (0, +\infty)$.

Problem 8.6.2 (2023 Calculus A, Midterm (10pts)). Determine the number a, b so that the function

$$f(x) = \lim_{n\to\infty} \frac{x^{2n-1} + ax^2 + bx}{x^{2n} + 1}$$

is continuous on $\mathbb{R}$.

Problem 8.6.3 (2022 Calculus A, Midterm (5pts)). Is there a sequence $\{a_n\}$ of real numbers so that both of the following hold?

(1) $\lim_{n\to\infty} a_n = 1$.

(2) $\lim_{n\to\infty} a_n^n = 1.001$.

Problem 8.6.4 (2022 Calculus A, Midterm (5pts)). Find the n-th derivative $y^{(n)}$ of

$$y = (x^2 + 2x + 2)e^{-x}.$$

Problem 8.6.5 (2023 Calculus C, Midterm (10pts)).

(1) Let $f(x) = \begin{cases} x^2 \ln(2 + \cos\frac{1}{x}) & , \text{ if } x \neq 0 \\ 0 & , \text{ if } x = 0 \end{cases}$. Find the derivative $f'(x)$. Is there any discontinuity of $f'(x)$? What kind of discontinuity is it?

(2) Suppose that $f(x)$ is differentiable at 0, and satisfies

$$|f(x)| \leq \ln(1 + |\arcsin x|)$$

for $|x| < 1$. Prove that $|f'(0)| \leq 1$.

Problem 8.6.6 (2021 Calculus B, Midterm (15pts)).

(1) Find $\int_0^{\frac{\pi}{2}} \frac{\sin x \cos x}{1+\sin^2 x} dx$.

(2) Find the arc length of the parabola $y = \frac{1}{2}x^2$ from $(0,0)$ to $(1, \frac{1}{2})$.

(3) For odd $n = 2k+1$ ($k \in \mathbb{N}^*$), find the area of the rose with n-petals:

$$r(\theta) = \sin(n\theta)$$

for $\theta \in [0, 2\pi]$.

Problem 8.6.7 (2022 Calculus A, Midterm (10pts)). Let $f(x)$ be continuous on $[a,b]$, and

$$F(x) = \int_a^x f(t)dt$$

for $x \in [a,b]$. Prove that

(1) $F(x)$ is continuous on $[a,b]$.

(2) $F(x)$ is differentiable on (a,b) and $F'(x) = f(x)$ for $x \in (a,b)$.

Problem 8.6.8 (2020 Calculus B, Midterm (5pts)). Find the derivative

$$\frac{d}{dx} \int_{x^3+1}^{2^x} \frac{\sin t}{t^4+2} dt.$$

Problem 8.6.9 (2022 Calculus B, Midterm (5pts)). Find the integral

$$\int \frac{dx}{\sqrt[3]{(x+1)(x-1)^5}}.$$

Problem 8.6.10 (2022 Calculus C, Final (5pts)). Find the integral

$$\int_{-1}^{1} (1+x^7)(1-x^2)^{\frac{3}{2}} dx.$$

Problem 8.6.11 (2022 Calculus C, Final (10pts)). Suppose that $f(x)$ is continuous on $[0,T]$. Show that

$$\int_0^T f(t) \left(\int_0^t f(s)ds \right) dt = \frac{1}{2} \left(\int_0^T f(t)dt \right)^2.$$

Problem 8.6.12 (2021 Calculus B, Midterm (10pts)). Let $a_1 > 0$. Suppose that a sequence $\{a_n\}$ satisfies

$$a_{n+1} = \frac{1}{2}\left(a_n + \frac{1}{a_n}\right)$$

for any $n \in \mathbb{N}^*$. Show that the limit $\lim_{n\to\infty} a_n$ exists, and find the limit.

8.6.2 Practice Exam

Problem 8.6.13 (2023 Calculus A, Midterm (20pts)).

(1) Find $\lim_{n\to\infty} \frac{3^n}{n!}$.

(2) Find $\lim_{n\to\infty} \left[\frac{1}{(n+1)^3} + \frac{2}{(n+2)^3} + \cdots + \frac{n}{(2n)^3}\right]$.

(3) Find $\lim_{x\to+\infty} \sin\left(\left(\sqrt{x^2+x} - \sqrt{x^2-x}\right)\pi\right)$.

(4) Find $\lim_{n\to\infty} \left[\frac{1}{n^2}\sum_{k=1}^{n} k\ln(n+k) - \frac{n+1}{2n}\ln n\right]$.

Problem 8.6.14 (2022 Calculus B, Midterm (10pts)).

(1) Find $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \sin\left(\frac{k}{n} - \frac{1}{2(n^k)}\right)$.

(2) Find $\lim_{x\to 0}(1+\tan^2 x)^{\frac{1}{\sin^2 x}}$.

Problem 8.6.15 (2022 Calculus B, Midterm (15pts)).

(1) Find the derivative $f'(x)$ of the function

$$f(x) = x^{\sqrt{x}}$$

for $x > 0$.

(2) Find the derivative $f'(x)$ of the function

$$f(x) = \int_0^{\sin x} \frac{dt}{\sqrt{1-t^3}}$$

for $x < 1$.

(3) Find the 4th derivative $f^{(4)}(x)$ of the function

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

for $x \neq \pm 1$.

Problem 8.6.16 (2023 Calculus A, Midterm (15pts)).

(1) Find $\int_0^1 \frac{\sqrt{x}}{1+\sqrt{x}} dx$.

(2) Find $\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{\sin^2 x}{1+e^x} dx$.

(3) Find $\int \sqrt{1+x^2} dx$.

Problem 8.6.17 (2020 Calculus C, Final (10pts)). Let f be continuous on $[a, b]$. Show that

$$\lim_{h \to 0} \int_a^b \frac{f(x+h) - f(x)}{h} dx = f(b) - f(a).$$

Problem 8.6.18 (2020 Calculus B, Midterm (10pts)). Suppose that a function $f(x)$ has a continuous derivative. Show that for any $x \in [0, 1]$, we have

$$|f(x)| \leq \int_0^1 |f(t)| dt + \int_0^1 |f'(t)| dt.$$

Also, find all possible functions $f(x)$ so that the above equality holds for all $x \in [0, 1]$.

Problem 8.6.19 (2020 Calculus B, Midterm (10pts)). Suppose $f : [a, b] \to [a, b]$ satisfies

$$|f(x) - f(y)| \leq |x - y|$$

for any $x, y \in [a, b]$. Let $x_1 \in [a, b]$, and

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

Show that the limit $\lim_{n \to \infty} x_n$ exists.

Problem 8.6.20 (2020 Calculus B, Midterm (10pts)). Let

$$f(x) = \begin{cases} x^m \sin \frac{1}{x} & , \text{ if } x \neq 0 \\ 0 & , \text{ if } x = 0 \end{cases}$$

for $m \in \mathbb{N}^*$. Find $f'(x)$ and $f''(x)$ for $x \neq 0$. Determine m so that $f(x)$ has a continuous second derivative.

8.6.3 Extra Problems

Problem 8.6.21 (2023 Calculus B, Midterm)**.**

(1) Prove that for $x \in (-\frac{\pi}{2}, \frac{\pi}{2})$, we have

$$-1 < \frac{4\sin x}{3+\sin^2 x} < 1.$$

(2) For $x \in (-\frac{\pi}{2}, \frac{\pi}{2})$, find the derivative $f'(x)$ of the function

$$f(x) = \arcsin\left(\frac{4\sin x}{3+\sin^2 x}\right).$$

(3) Prove that

$$\int_0^{\frac{\pi}{2}} \frac{dx}{\sqrt{4\cos^2 x + \sin^2 x}} = \int_0^{\frac{\pi}{2}} \frac{dx}{\sqrt{\frac{9}{4}\cos^2 x + 2\sin^2 x}}.$$

Problem 8.6.22 (2023 Calculus B, Midterm)**.** Suppose that $f(x)$ and $g(x)$ are continuous on $[0,1]$, and satisfy

$$f(0) = g(0), \quad \sin f(1) = \sin g(1), \quad \cos f(1) = \cos g(1),$$

and

$$(\cos f(x) + \cos g(x))^2 + (\sin f(x) + \sin g(x))^2 \neq 0$$

for any $x \in [0,1]$. Prove that $f(1) = g(1)$.

Problem 8.6.23 (2020 Calculus A, Midterm)**.**

(1) Find $\lim\limits_{x\to 0} x^x$.

(2) Find $\lim\limits_{x\to 0} \dfrac{\sqrt[3]{1+x}-1}{x+\tan x}$.

(3) Find $\lim\limits_{n\to\infty} \cos\dfrac{a}{2}\cos\dfrac{a}{2^2}\cdots\cos\dfrac{a}{2^n}$ for $a \in (0,1)$.

(4) Find $\lim\limits_{n\to\infty} \dfrac{2^n n!}{n^n}$.

Remark 8.6.24. For Problem 8.6.23(4), note that $\ln x$ ($x \in (0,1]$) is unbounded, and so is not Riemann integrable, but it is Lebesgue integrable. So if you want to turn the problem into an integral, it still works. (But there are many other methods. For instance, you can use the *Stirling's formula* (Appendix 7.6) if you prefer.)

Problem 8.6.25 (2020 Calculus A, Midterm)**.**

(1) Suppose that for any $k \in \mathbb{N}^*$, $\lim_{n\to\infty}(x_{n+k} - x_n) = 0$. Does $\lim_{n\to\infty} x_n$ exist?

(2) Suppose that $\lim_{n\to\infty} x_n = l$. Show that

$$\lim_{n\to\infty} \frac{1}{n^2}(x_1 + 2x_2 + \cdots + nx_n) = \frac{l}{2}.$$

Problem 8.6.26 (2020 Calculus A, Midterm)**.**

(1) Let $a > 0$, and $f(x)$ be continuous on $[0, 2a]$ with $f(0) = f(2a)$. Show that there exists a point $x_0 \in [0, 2a]$, such that $f(x_0) = f(x_0 + a)$.

(2) Is there a function $f : \mathbb{R} \to \mathbb{R}$ that is *locally unbounded*? In other words, can we find a function f so that the following holds?

For any $x \in \mathbb{R}$, $\epsilon > 0$, and any $M > 0$, there exists $y \in \mathbb{R}$ so that $|y - x| < \epsilon$ and $|f(y)| > M$.

Problem 8.6.27 (2020 Calculus A, Midterm)**.** Let $[a, b]$ be some interval, $f : [a, b] \to [a, b]$ be a continuous function, $x_1 \in [a, b]$, and $x_{n+1} = f(x_n)$. Suppose that $\lim_{n\to\infty}(x_{n+1} - x_n) = 0$. Does $\lim_{n\to\infty} x_n$ exist? (This is a difficult analysis problem.)

8.7 Midterm Exams

Problem 8.7.1 (2024 Calculus B, Midterm)**.** Find

$$\lim_{x\to 0} \left(\frac{1 + 2\sin^2 x}{\cos(2x)}\right)^{\csc^2 x}.$$

Problem 8.7.2 (2024 Calculus B, Midterm)**.** Let $f:(-1,1)\to\mathbb{R}$ be defined by

$$f(x)=\int_0^{\arcsin x}\frac{dt}{\sqrt{1+(\sin t)^2}}.$$

Find the second derivative $f''(x)$.

Problem 8.7.3 (2024 Calculus B, Midterm)**.** Find

$$\int\frac{4x^2+4x-11}{(2x-1)(2x+3)(2x-5)}dx.$$

Problem 8.7.4 (2024 Calculus B, Midterm)**.** Let $T\subset\mathbb{R}^2$ be the region enclosed by the curve

$$y=\frac{\ln x}{\sqrt{\pi}},\quad 1\le x\le 2,$$

and the lines $x=2$, $y=0$. Let $A\subset\mathbb{R}^3$ be the solid generated by revolving the region T about the x-axis. Find the volume of the solid A.

Problem 8.7.5 (2024 Calculus B, Midterm)**.** Show that there are exactly two solutions in $\mathbb{R}$ to the equation

$$x^{18}+x^{12}-\cos x=0.$$

Problem 8.7.6 (2024 Calculus B, Midterm)**.** Let $A:[0,1]\to[0,1]$, $B:[0,1]\to\mathbb{R}$ be continuous functions. For any continuous $f:[0,1]\to\mathbb{R}$, define the continuous function $Tf:[0,1]\to\mathbb{R}$ by

$$(Tf)(x)=B(x)+\int_0^x A(t)f(t)dt.$$

Prove that there is at most one continuous function f such that $Tf=f$. In other words, show that if f,g are continuous so that $Tf=f$ and $Tg=g$, then $f=g$ on $[0,1]$.

Problem 8.7.7 (2024 Calculus A, Midterm)**.** Let $f:\mathbb{R}\to\mathbb{R}$ be defined by

$$f(x)=(x^2-3x+2)^{100}\cos\frac{\pi x^2}{4}.$$

(i) Calculate $f^{(n)}(1)$ for $n=1,2,\ldots,100$.

(ii) Calculate $f^{(101)}(2)$.

Problem 8.7.8 (2024 Calculus A, Midterm)**.**

(i) Suppose that $f(x)$ is differentiable at $x = a$. Show that

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

(ii) Show that even if $f(x)$ is continuous at $x = a$ and $\lim_{h\to 0}\frac{f(a+h)-f(a-h)}{h}$ exists, $f'(a)$ may not exist.

(iii) Suppose that $f(x)$ is differentiable at $x = a$. Show that for any $k \in \mathbb{R}\backslash\{0,1\}$,

$$\lim_{h\to 0}\frac{f(a+kh)-f(a+h)}{h} = (k-1)f'(a).$$

(iv) Find a function f so that $\lim_{h\to 0}\frac{f(a+kh)-f(a+h)}{h}$ exists for all $k \in \mathbb{R}\setminus\{0,1\}$, and $f'(a)$ does not exist.

Problem 8.7.9 (2024 Calculus A, Midterm)**.**

(i) Let $x_n > 0$, $\lim\limits_{n\to+\infty} x_n = a > 0$. Prove that

$$\lim_{n\to+\infty}\sqrt[n]{x_1x_2\cdots x_n} = a.$$

(ii) Suppose that $\lim\limits_{n\to+\infty}\frac{1}{n}\sum\limits_{k=1}^{n} x_n = a$ and $\lim\limits_{n\to+\infty} n(x_n - x_{n-1}) = 0$. Prove that

$$\lim_{n\to+\infty} x_n = a.$$

8.8 Homework 7

- 10.31:

  [1] Exercises 4.1 (P.216): 6, 7 (The determinant is called *Wrońskian* 朗斯基行列式), 11, 12.

  [1] Exercises 4.2 (P.226): 16, 18.

  [1] Exercises of Chapter 4 (P.259): 5, 9, 10, 18.

Extra Problem 8.8.1 (2023 Calculus C, Midterm, cf. Gronwall's inequality)**.** Let f be continuous on $[a, b]$ and differentiable on (a, b). Suppose that $f(a) = 0$ and

$$|f'(x)| \leq \frac{1}{2(b-a)}|f(x)|$$

for $x \in (a, b)$. Prove that $f(x) = 0$ for all $x \in [a, b]$.

- 11.5:

  [1] Exercises 4.4 (P.239): 1(1)(4), 3(1).

8.9 Homework 8

- 11.12:

  [1] Exercises of Chapter 4 (P.259): 8, 11, 13, 15(1)(2)(3), 17, 21, 23, 25, 29, 31(1)(2), 32.

Extra Problem 8.9.1 (2020 Calculus C, Final & 2023 Calculus A, Final)**.** Find the limit

$$\lim_{x\to 0} \frac{1}{\sin x}\left[(1+x)^{\frac{1}{x}} - e\right].$$

Extra Problem 8.9.2 (2023 Calculus A, Final)**.** Suppose that $n \in \mathbb{N}$, $a \in \mathbb{R}$, $f(x)$ is $n+1$ times differentiable at $x = 0$, and satisfies

$$f(0) = f'(0) = \cdots = f^{(n-1)}(0) = 0, \quad f^{(n)}(0) = a.$$

Find the limit

$$\lim_{x\to 0} \frac{f(e^x - 1) - f(x)}{x^{n+1}}.$$

Extra Problem 8.9.3 (2022 Calculus A, Final)**.** Suppose that a continuous function $f(x) : [0, 2] \to \mathbb{R}$ has a continuous derivative, and $f(0) = f(2) = 0$. Let $M = \max_{x\in[0,2]} |f(x)|$.

(1) Prove that there is a $\xi \in (0, 2)$ such that $|f'(\xi)| \geq M$.

(2) Prove that if $|f'(\xi)| \leq M$ for any $x \in (0, 2)$, then $f \equiv 0$.

- 11.14:

  [1] Exercises 4.5 (P.237): 3, 5.

  [1] Exercises 4.6 (P.254): 2(3), 3.

  [1] Exercises of Chapter 4 (P.260): 20, 27, 28.

  Extra Problem 8.9.4 (2020 Calculus B, Final)**.** Let $f : [-1,1] \to \mathbb{R}$ be defined by the formula

$$f(x) = x^{\frac{2}{3}} - (x^2-1)^{\frac{1}{3}}, \quad x \in [-1,1].$$

  Find all the local minima of f.

  Extra Problem 8.9.5 (Basel problem, enjoy!)**.** Let $f \in C(S^1, \mathbb{C})$ be the function defined by $f(e^{2\pi i x}) = (1-2x)^2$ for $x \in [0,1)$.

  (1) Show that for $n \in \mathbb{Z}$, the Fourier transform of f is given by

$$\hat{f}(n) = \begin{cases} \dfrac{2}{\pi^2 n^2} & , \text{ if } n \in \mathbb{Z} \setminus \{0\} \\ \\ \dfrac{1}{3} & , \text{ if } n = 0 \end{cases}.$$

  (2) Using the Fourier inversion theorem, conclude that

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

  (3) Using the Plancherel theorem, conclude that

$$\sum_{n=1}^{\infty} \frac{1}{n^4} = \frac{\pi^4}{90}.$$

8.10 Homework 9

- You need to have basic knowledge of *Linear Algebra*. See Appendix 7.2.2 if your are not familiar with it.

- 11.19:

  [1] Exercises 5.1 (P.267): 7(2), 8, 10 (This is called the *parallelogram law* 平行四边形恒等式), 11.

  [1] Exercises 5.2 (P.274): 9(3), 11, 12, 15.

  [1] Exercises 6.1 (P.304): 4(1)(2).

  [1] Exercises of Chapter 5 (P.295): 8, 9, 10(1)(2).

  Extra Problem 8.10.1 (2022 Analysis I, Final, rapidly decreasing functions 速降函数)**.** Suppose that f is infinitely differentiable on $\mathbb{R}$, and for any $k \in \mathbb{N}$, there is $M_k > 0$ such that

  $$|x|^k|f(x)| + |f^{(k)}(x)| < M_k$$

  for any $x \in \mathbb{R}$. Prove that for any $k, l \in \mathbb{N}$, there is $M_{k,l} > 0$ such that

  $$|x|^k|f^{(l)}(x)| < M_{k,l}$$

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

  Extra Problem 8.10.2 (2023 Calculus A, Final)**.** Let $f : \mathbb{R} \to \mathbb{R}$ be an infinitely differentiable and periodic function on $\mathbb{R}$. Prove that for any $k \in \mathbb{N}^*$, there exists $\xi \in \mathbb{R}$ such that

  $$f^{(k)}(\xi) = 0.$$

  Extra Problem 8.10.3 (2022 Analysis I, Final)**.** Suppose that f is twice differentiable on $[-2, 2]$, $|f(x)| \leq 1$ and

  $$[f(0)]^2 + [f'(0)]^2 = 4.$$

  Prove that there exists $\xi \in (-2, 2)$ such that

  $$f(\xi) + f''(\xi) = 0.$$

  Extra Problem 8.10.4 (2021 Calculus B, Final)**.** Let $n \geq 3$ be an integer, S^1 be a circle of radius 1. Find the maximum of the areas:

  $$\max(\text{Area}(P_n))$$

where P_n is any *cyclic n-gon* (内接 n 边形) of S^1, i.e. the *vertices* (端点) of P_n are all in S^1.

- 11.21:

  [1] Exercises 5.3 (P.283): 9, 15, 19, 20.

  [1] Exercises of Chapter 5 (P.295): 11(1)(2).

  Extra Problem 8.10.5 (2022 Calculus A, Final)**.** Let $P_1(a_1, b_1, c_1)$ and $P_2(a_2, b_2, c_2)$ be two different points on the sphere

  $$S^2 = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}.$$

  Let $O = (0, 0, 0)$ be the origin. Calculate

  $$(\overrightarrow{OP_1} \times \overrightarrow{OP_2})^2 + (\overrightarrow{OP_1} \cdot \overrightarrow{OP_2})^2$$

  where $(\vec{\mathbf{r}})^2 = \vec{\mathbf{r}} \cdot \vec{\mathbf{r}}$ is the inner product of a vector $\vec{\mathbf{r}}$ with itself.

  Extra Problem 8.10.6 (2023 Calculus A, Final & 2022 Calculus B, Final)**.** Suppose that the plane $P = \{(x, y, z) \in \mathbb{R}^3 : x + 3y + 2z = 6\}$ intersects the x-axis, y-axis, and z-axis at A, B, and C, respectively.

  (1) Find the area of the triangle ΔABC.

  (2) Find the equation of the sphere that passes through A, B, C and the identity $(0, 0, 0)$.

8.11 Homework 10

- 11.26:

  [1] Exercises 5.4 (P.289): 1, 3(4)(5).

  [1] Exercises 5.5 (P.294): 3, 4.

  [1] Exercises of Chapter 5 (P.295): 13(1)(2), 15, 16.

  Extra Problem 8.11.1 (2022 Calculus A, Final)**.**

(1) Let L be given by

$$\begin{cases} x - 2y + z = 0 \\ 5x + 2y - 5z = -6 \end{cases}.$$

Prove that $(1, 2, 3) \in L$. And find the standard form of L.

(2) Find the normal plane of the curve

$$\begin{cases} x = 7t - 14 \\ y = 4t^2 \\ z = 3t^3 \end{cases}$$

at the time $t = 1$.

Extra Problem 8.11.2 (2022 Calculus A, Final)**.** Suppose that f is a twice differentiable function on $[a, b]$, and

$$f(a) = f(b) = 0, \quad f\left(\frac{a+b}{2}\right) > 0.$$

Prove that there exists a $\xi \in (a, b)$ such that $f''(\xi) < 0$.

Extra Problem 8.11.3 (2022 Calculus B, Final)**.** Find the tangent line of the curve in $\mathbb{R}^2$:

$$e^{xy} + xy + y^2 = 2$$

at the point $(0, 1)$.

- 11.28:

  [1] Exercises 6.1 (P.304): 1(3)(6), 2(1)(2)(3)(4), 3.

  [1] Exercises 6.2 (P.313): 3(2)(3)(4).

  [1] Exercises of Chapter 6 (P.389): 4, 5, 10(2)(3).

Extra Problem 8.11.4 (2022 Calculus A, Final)**.** Determine whether the following limits exist or not. Find the limit if it exists.

(1) $\lim\limits_{x \to 0} \dfrac{\int_0^{x^3} \sin^3 2t dt}{\int_0^{x^2} \tan t^5 dt}$;

(2) $\lim_{(x,y)\to(0,0)} \frac{x^2y}{x^4+y^2}$;

(3) $\lim_{n\to\infty} \sum_{k=1}^{n} \frac{n}{n^2+k^2}$.

Extra Problem 8.11.5 (2021 Calculus B, Final)**.**

(1) Find the limit $\lim_{x\to 0} \frac{\tan^4 x}{(1-\frac{1}{2}x\sin x)^{\frac{1}{2}} - (\cos x)^{\frac{1}{2}}}$.

(2) Let $n \geq 1$ be an integer, $a_1, \ldots, a_n$ be a finite sequence of positive real numbers. Find

$$\lim_{x\to 0} \left(\frac{1}{n} \sum_{k=1}^{n} a_k^x \right)^{\frac{1}{x}}.$$

Extra Problem 8.11.6 (2022 Calculus B, Final)**.** Determine whether the following limits exist or not. Find the limit if it exists.

(1) $\lim_{(x,y)\to(0,0)} \frac{24\cos\sqrt{x^2+y^2} - 24 + 12(x^2+y^2)}{(\tan\sqrt{x^2+y^2})^4}$;

(2) $\lim_{(x,y)\to(0,0)} (x+\ln(1+y))\cos\left(\frac{1}{x^2+y^2}\right)$;

(3) $\lim_{(x,y)\to(0,0)} \frac{x\sin y}{(\sin x)^2 + (\sin y)^2}$.

Extra Problem 8.11.7 (2023 Analysis III, Midterm)**.** Determine whether the following limit exists or not:

$$\lim_{(x,y,z)\to(0,0,0)} \frac{\sin(xyz)}{\sqrt{x^2+y^2+z^2}}.$$

Find the limit if it exists.

Extra Problem 8.11.8. Determine whether the following limit exists or not:

$$\lim_{(x,y,z)\to(0,0,0)} \left(\frac{|xyz|}{x^2+y^2+z^2} \right)^{|x|+|y|+|z|}.$$

Find the limit if it exists.

8.12 Homework 11

- 12.3:

  [1] Exercises 6.3 (P.319): 2, 4.

  [1] Exercises 6.4 (P.332): 8 (cf. *Cauchy-Riemann equations* 柯西-黎曼方程), 12, 14, 15 (cf. *Green's Theorem* 格林公式 [2, §8.3 定理 3]), 16(1)(2)(3)(4).

  Extra Problem 8.12.1 (2023 Calculus A, Final)**.** Let $D \subset \mathbb{R}^2$ be a region containing the origin $(0,0)$. Let $f : D \to \mathbb{R}$ be continuous at $(0,0)$. Suppose that the limit

  $$\lim_{(x,y)\to(0,0)} \frac{f(x,y)}{x^2+y^2}$$

  exists. Prove that f is differentiable at $(0,0)$.

  Extra Problem 8.12.2 (2021 Calculus B, Final)**.** Let $f : \mathbb{R}^3 \to \mathbb{R}$ be defined by

  $$f(x,y,z) = \begin{cases} \dfrac{xyz}{x^2+y^2+z^2} & , \text{ if } (x,y,z) \neq (0,0,0) \\ \\ 0 & , \text{ if } (x,y,z) = (0,0,0) \end{cases}.$$

  (1) Calculate the partial derivatives $f_x(0,0,0)$, $f_y(0,0,0)$, $f_z(0,0,0)$.

  (2) Is f differentiable at $(0,0,0)$? Prove your judgment.

- 12.5:

  [1] Exercises 6.5 (P.342): 6, 9.

  [1] Exercises of Chapter 6 (P.390): 8, 12(1)(2), 15, 16 (cf. *cylindrical coordinate system* 柱坐标 [2, §7.3.2]), 17 (cf. *spherical coordinate system* 球坐标 [2, §7.3.3]), 24.

  Extra Problem 8.12.3 (2023 Calculus A, Final)**.** Let D be a region containing the unit circle:

  $$S^1 = \{(u,v) : u^2+v^2 = 1\}.$$

Suppose that there is a function $f(u,v) \in C^1(D)$ so that

$$f(x, 1-x) = 1 \tag{8.12.1}$$

for $(x, 1-x) \in D$. Prove that there are at least two points (u_1, v_1), $(u_2, v_2) \in S^1$ that satisfy

$$v f_u(u,v) = u f_v(u,v).$$

Remark 8.12.4. In fact, the condition (8.12.1) is redundant if we are familiar with the *compactness* (紧集) of S^1.

Extra Problem 8.12.5 (2021 Calculus B, Final & 2022 Calculus B, Final)**.** Let $f, g \in C^2(\mathbb{R})$. For $x, y \in \mathbb{R}$ with $x \neq 0$, define

$$h(x,y) = x f\left(\frac{y}{x}\right) + g\left(\frac{y}{x}\right).$$

Find

$$x^2 h_{xx}(x,y) + 2xy h_{yx}(x,y) + y^2 h_{yy}(x,y).$$

Extra Problem 8.12.6 (2022 Calculus B, Final)**.** Let $n \geq 3$ be an integer. For $\mathbf{0} \neq (x_1, \ldots, x_n) \in \mathbb{R}^n$, define

$$u(x_1, \ldots, x_n) = \left(\sum_{k=1}^{n} x_k^2\right)^{\frac{2-n}{2}}.$$

Find $\Delta u = \sum_{k=1}^{n} u_{x_k x_k}$.

Extra Problem 8.12.7 (2023 Analysis III, Midterm)**.** Let

$$f(x,y,z) = \left(\frac{x}{y}\right)^z.$$

Find $\dfrac{\partial f}{\partial x}$, $\dfrac{\partial f}{\partial y}$, $\dfrac{\partial f}{\partial z}$.

Extra Problem 8.12.8 (2022 Analysis III, Midterm)**.** Let $m, n \in \mathbb{N}^*$. We say a function $f : \mathbb{R}^n \to \mathbb{R}$ is *homogeneous of degree m* (m 次齐次函数) if

$$f(t\mathbf{x}) = t^m f(\mathbf{x})$$

for any $\mathbf{x} \in \mathbb{R}^n$ and $t \in \mathbb{R}$. Prove that a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is homogeneous of degree m if and only if

$$\sum_{i=1}^{n} x_i f_{x_i}(\mathbf{x}) = mf(\mathbf{x})$$

for any $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$.

Extra Problem 8.12.9 (2022 Analysis III, Midterm)**.** Let $n \geq 2$ be an integer, and let $V : \mathbb{R}^n \to \mathbb{R}$ be the *Vandermonde determinant* (范德蒙行列式), i.e. for $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$,

$$V(\mathbf{x}) = \det \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ 1 & x_3 & x_3^2 & \cdots & x_3^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{bmatrix}.$$

Prove that:

(1) $\displaystyle\sum_{i=1}^{n} x_i V_{x_i}(\mathbf{x}) = \frac{n(n-1)}{2} V(\mathbf{x})$, for $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$.

(2) $\displaystyle\sum_{i=1}^{n} V_{x_i}(\mathbf{x}) = 0$, for $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$.

8.13 Homework 12

- 12.10:

  [1] Exercises 6.6 (P.349): 7, 8, 9.

  [1] Exercises 6.7 (P.355): 3, 5.

  Extra Problem 8.13.1 (2020 Calculus B, Final & 2022 Calculus B, Final)**.** Let $a, b \in \mathbb{R}$ with $b \neq 0$. Let

  $$f(x, y) = \arctan \frac{x}{y}.$$

  Find the Taylor's polynomial of order 2 of $f(x, y)$ at (a, b).

- 12.12:

  [1] Exercises 6.8 (P.369): 5, 8, 10.

  [1] Exercises of Chapter 6 (P.391): 13, 21, 22(1)(2), 25, 26, 28(1)(2), 29.

Extra Problem 8.13.2 (2023 Calculus A, Final)**.** Let $F(u,v) \in C^2(\mathbb{R}^2)$, and let $z = z(x,y)$ be determined by

$$F(x-z, y-z) = 0.$$

Calculate

$$\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial x \partial y} + \frac{\partial^2 z}{\partial y \partial x} + \frac{\partial^2 z}{\partial y^2}.$$

Extra Problem 8.13.3 (2023 Calculus A, Final)**.** Consider the system of equations

$$\begin{cases} xy + yz^2 + 4 = 0 \\ x^2y + yz - z^2 + 5 = 0 \end{cases}. \tag{8.13.1}$$

Try to discuss which implicit functions can be determined by the system (8.13.1) near the point $P_0(1,-2,1)$. Fine the derivatives of the implicit functions at P_0.

Extra Problem 8.13.4 (2021 Calculus B, Final)**.** Let

$$F(x,y,z) = x^3 + (y^2-1)z^3 - xyz.$$

(1) Prove that there is a neighborhood of $(1,1)$ in $\mathbb{R}^2$, and an implicit function $z(x,y)$ such that $z(1,1) = 1$ and

$$F(x,y,z(x,y)) = 0$$

for $(x,y) \in D$.

(2) Find the direction/unit vector $\mathbf{v}$ in which the function $z(x,y)$ decreases most quickly from $(1,1)$.

(3) Suppose that $\mathbf{n} \in \mathbb{R}^3$ is the normal vector of the plane $x+2y-2z = 1$, and the z-coordinate of $\mathbf{n}$ is positive. Find the cosine of the angle between $\mathbf{n}$ and $(\mathbf{v}, 0)$: $\cos(\mathbf{n}, (\mathbf{v}, 0))$.

Extra Problem 8.13.5 (2022 Calculus B, Final)**.** Let $r > 0$, $D = \{(x, y) \in \mathbb{R}^2 : \sqrt{x^2 + y^2} < r\}$, $f \in C^3(D)$. Suppose that $f(0, 0) = 0$, $df = 0$ at $(0, 0)$, and

$$d^2 f = Edx^2 + 2Fdxdy + Gdy^2$$

at $(0, 0)$ for some $E, F, G \in \mathbb{R}$.

(1) Prove that there exist two functions $a : D \to \mathbb{R}$, $b : D \to \mathbb{R}$ such that $a(0, 0) = 0$, $b(0, 0) = 0$, and

$$f(x, y) = xa(x, y) + yb(x, y)$$

for any $(x, y) \in D$.

(2) Suppose that $E > 0$, $EG - F^2 < 0$. Which type of quadric is approximated by surface $z = f(x, y)$ near $(0, 0, 0) \in \mathbb{R}^3$? (cf. [2, §8.4.2].)

(3) Determine whether there is an injective map $g : D_1 \to \mathbb{R}^2$

$$g : (u, v) \mapsto (x(u, v), y(u, v))$$

on a sufficiently small neighborhood $D_1 \subset \mathbb{R}^2$ of $(0, 0)$, such that $x(u, v)$, $y(u, v) \in C^1(D_1)$, and

$$f(g(u, v)) = f(x(u, v), y(u, v)) = u^2 - v^2.$$

Extra Problem 8.13.6 (2023 Analysis III, Midterm)**.** Consider the equation

$$x_1 x_2 + u + \mathrm{sh}[(|\mathbf{x}|^2 + 1)u] = 0$$

for $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$, $u \in \mathbb{R}$. Prove that the equation determines a unique function $u(\mathbf{x}) \in C^1(\mathbb{R}^n)$.

Extra Problem 8.13.7 (2022 Analysis III, Midterm)**.** Let $z = z(x, y)$ be given by

$$x + e^{yz} + z^2 = 0.$$

Find $\dfrac{\partial^2 z}{\partial x^2}$, $\dfrac{\partial^2 z}{\partial x \partial y}$, $\dfrac{\partial^2 z}{\partial y^2}$ at $(-2, 0, 1)$.

Extra Problem 8.13.8 (2017 Analysis III, Midterm)**.** Let $f : \mathbb{R} \to \mathbb{R}$ be a strictly increasing function with $f(0) = 0$, and let $D \subset \mathbb{R}^2$ be a bounded region. Suppose $z(x, y)$ is continuous on $\overline{D}$ and has partial derivatives z_x and z_y on D such that

(1) $z(x, y) = 0$ for $(x, y) \in \partial D$,

(2) $z_x(x, y) + z_y(x, y) = f(z(x, y))$ for $(x, y) \in D$.

Prove that $z(x, y) = 0$ for any $(x, y) \in D$.

8.14 Homework 13

8.14.1 Homework from the Textbook

- [1] Exercises 6.9 (P.381): 1(3)(4), 4, 8, 10.

  [1] Exercises of Chapter 6 (P.393): 30, 33.

8.14.2 Practice Exam

Problem 8.14.1 (2022 Calculus B, Final, 15pts)**.** Determine whether the following limits exist or not. Find the limit if it exists.

(1) $\displaystyle\lim_{x\to 0} \frac{\sqrt[3]{1+x^2+x^4}-1}{(\ln(1+x))^2}$;

(2) $\displaystyle\lim_{(x,y)\to(0,0)} (e^y - 1)\frac{xy}{x^2+y^2}$;

(3) $\displaystyle\lim_{(x,y)\to(0,0)} \frac{xy}{x^2+(e^y-1)^2}$.

Problem 8.14.2 (2022 Calculus B, Final, 5pts)**.** Let $a \in \mathbb{R}$, $f, g \in C^2(\mathbb{R})$, and $h(x, t) = f(x + at) + g(x - at)$. Find $h_{tt} - a^2 h_{xx}$.

Problem 8.14.3 (2018 Analysis III, Midterm, 10pts)**.** Let E be the curve determined by

$$\begin{cases} x^2 + 2y^2 + 2z^2 = 2 \\ x + y + z = 0 \end{cases}.$$

Let $Q(1,0,0) \in \mathbb{R}^3$. Find the points $P_1, P_2 \in E$ so that

$$|P_1Q| = \min_{P \in E} |PQ|, \quad |P_2Q| = \max_{P \in E} |PQ|.$$

Problem 8.14.4 (2023 Calculus A, Final, 10pts)**.** Find the extrema of the function

$$f(x,y) = (y - x^2)(y - x^3).$$

Problem 8.14.5 (2022 Calculus A, Final, 10pts)**.**

(1) Let $z = \arctan \dfrac{(x-3)y + (x^2+x-1)y^2}{(x-2)y + (x-3)^2 y^4}$. Find $\dfrac{\partial z}{\partial y}\Big|_{(3,0)}$.

(2) Let $z = z(x,y)$ be determined by

$$m\left(x + \frac{z}{y}\right)^n + n\left(y + \frac{z}{x}\right)^m = 1$$

where $m, n \in \mathbb{N}^*$. Find $x\dfrac{\partial z}{\partial x} + y\dfrac{\partial z}{\partial y} + xy$.

Problem 8.14.6 (2020 Calculus B, Final & 2022 Calculus B, Final, 10pts)**.** Let $A, B, C \in \mathbb{R}^3$ be the intersections of the plane

$$K: \ x + 2y + 3z = 6$$

and x-axis, y-axis, z-axis, respectively. Suppose there is a point $H \in \mathbb{R}^3$ so that the distance between H and K is 1. Let $M \in K$ be the orthogonal projection of H on K, i.e. $\|H - M\| = 1$. Suppose further that $M \in \Delta ABC$. Let p, q, r be the distance of M and BC, CA, AB, respectively.

(1) Find the area of the triangle ΔABC.

(2) Find the area $S(p,q,r)$ of the surface of the tetrahedron (四面体) $ABCH$ in terms of p, q, r.

(3) Find the constraints of p, q, r.

(4) Find the extrema of $S(p,q,r)$ with constraint.

Problem 8.14.7 (2018 Analysis III, Midterm, 10pts)**.** Let $D \subset \mathbb{R}^3$ be a region, $P_0(x_0, y_0, z_0) \in D$, $F \in C^2(D)$ with $F_z(P_0) \neq 0$. Suppose that $z = f(x, y)$ is the function determined by the equation

$$F(x, y, z) = 0, \quad z_0 = f(x_0, y_0).$$

Show that:

(1) If (x_0, y_0) is a local extremum of $f(x, y)$, then $F_x(P_0) = F_y(P_0) = 0$.

(2) If $F_x(P_0) = F_y(P_0) = 0$ and the matrix

$$H(P_0) = \begin{bmatrix} \dfrac{\partial^2 F}{\partial x^2}(P_0) & \dfrac{\partial^2 F}{\partial x \partial y}(P_0) \\ \dfrac{\partial^2 F}{\partial x \partial y}(P_0) & \dfrac{\partial^2 F}{\partial y^2}(P_0) \end{bmatrix}$$

is positive definite, then (x_0, y_0) is a local extremum of $f(x, y)$.

Problem 8.14.8 (2022 Calculus B, Final, 10pts)**.** Let L be the intersection of the plane $2x + y - 3z = 0$ and the plane $x + 2y - z - 2 = 0$.

(1) Find the standard form of the line L.

(2) Find the sphere centered at the origin $(0, 0, 0)$ tangent to the line L.

Problem 8.14.9 (2020 Calculus B, Final & 2022 Calculus B, Final, 10pts)**.** Suppose that the potential of a charge $P \in \mathbb{R}^3 \setminus \{(x, y, 0) \in \mathbb{R}^3\}$ is given by

$$V(P) = \left(\frac{2y}{z}\right)^x.$$

Find the unit vector in which the potential V decreases most quickly from $(1, \frac{1}{2}, 1)$.

Problem 8.14.10 (2020 Analysis I, Final, 10pts)**.** Let f be continuous on $[0, 1]$ and differentiable on $(0, 1)$. Prove that there exists $\xi \in (0, 1)$ such that

$$f(1) - f(0) = (\ln \sqrt{3})(1 + 2\xi) f'(\xi).$$

8.14.3 Practice Exam

Problem 8.14.11 (2022 Analysis I, Final, 10pts). Find

$$\lim_{x\to 0} \frac{[1+\ln(1+x)]^{\frac{1}{\tan x}} - e(1-x)}{x^2}.$$

Problem 8.14.12 (2022 Calculus B, Final, 10pts). In $\mathbb{R}^2$, find the tangent line of the curve

$$x^2(y+1)^2 + \int_0^{xy} \frac{dt}{\sqrt{t^3+1}} = 1$$

at $(1,0)$.

Problem 8.14.13 (2018 Analysis III, Midterm, 10pts). Let

$$f(x,y) = \begin{cases} \dfrac{xy^3}{x^2+y^4} & , \text{ if } (x,y) \neq (0,0) \\ \\ 0 & , \text{ if } (x,y) = (0,0) \end{cases}.$$

Prove that $f(x,y)$ is continuous at $(0,0)$, but not differentiable at $(0,0)$.

Problem 8.14.14 (2022 Calculus B & 2024 Calculus B, Final, 10pts). Find the Taylor's formula of the function

$$f(x,y) = x^{\sqrt{y}}$$

of order 2 at $(1,1)$.

Problem 8.14.15 (2022 Calculus B, Final, 10pts). Let $p \in \mathbb{R}$ with $p > 4$. Let

$$f(x) = (x^4)^{\frac{1}{p}} + (1-x^4)^{\frac{1}{p}}$$

for $x \in [-1,1]$. Find all the local minima of f.

Problem 8.14.16 (2023 Analysis III, Midterm, 10pts). Let

$$f(x,y) = \frac{ax+by+c}{\sqrt{x^2+y^2+1}}$$

where at least one of a, b, c is nonzero. Find the local extrema of f.

Problem 8.14.17 (2018 Analysis III, Midterm, 10pts)**.** Find a function $f(x, y)$ so that obey **all** of the following properties:

(a) f_x and f_y exist at $(0, 0)$;

(b) The directional derivative of f in any direction exists at $(0, 0)$;

(c) f is not differentiable at $(0, 0)$.

Problem 8.14.18 (2022 Calculus B, Final, 10pts)**.** For any $p \in \mathbb{R}$, consider the equation

$$x^p - 3x^2y + 3xy^2 - y^p = 0.$$

Prove that there are neighborhoods U, W of $1 \in \mathbb{R}$ such that the equation determines a unique function $y = f(x) : U \to W$.

Problem 8.14.19 (2020 Analysis I, Final, 10pts)**.** For $x \in (0, +\infty)$, define

$$f(x) = x \ln(2^{\frac{1}{x}} + 3^{\frac{1}{x}}).$$

(1) Prove that f is strictly increasing.

(2) Find the asymptote of f.

(3) Prove that f is strictly convex.

Problem 8.14.20 (2021 Analysis I, Final, 10pts)**.** Let f be continuous on $[-1, 1]$ and three times differentiable on $(-1, 1)$. Suppose that $f(0) = f'(0) = f(-1) = 0$, $f(1) = 1$. Prove that there is a $\xi \in (-1, 1)$ such that $f'''(\xi) = 3$.

8.14.4 Practice Exam

Problem 8.14.21 (2020 Analysis I, Final, 10pts)**.** Suppose that f, g are two functions such that

$$f''(x) < 0, \quad g''(x) > 0, \quad f(x) \neq g(x)$$

for any $x \in \mathbb{R}$. Prove that there are constants k and b such that

$$f(x) < kx + b < g(x)$$

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

Problem 8.14.22 (2023 Calculus A, Final, 10pts). The intersection E of the plane $x+y+z=1$ and the cylinder $x^2+y^2=1$ is an ellipse. Let $O(0,0,0)$ be the origin. Find the points $P_1, P_2 \in E$ so that

$$|P_1O| = \min_{P\in E} |PO|, \quad |P_2O| = \max_{P\in E} |PO|.$$

Problem 8.14.23 (2018 Analysis III, Midterm, 10pts). Let $D \subset \mathbb{R}^2$ be a region, $f(\mathbf{t}) \in C^1(D)$. Suppose that the tangent plane

$$T_{\mathbf{x}} = \{(\mathbf{h}, df(\mathbf{t},\mathbf{h})) \in \mathbb{R}^3 : \mathbf{h} \in \mathbb{R}^2\}$$

is parallel to the plane $x+y+z=0$ for any $\mathbf{t} \in \mathbb{R}^2$. Show that there is a constant $c \in \mathbb{R}$ such that $f(x,y) = c - x - y$.

Problem 8.14.24 (2022 Analysis III, Midterm, 10pts). Determine whether the following limit exists or not:

$$\lim_{(x,y,z)\to(0,0)} \frac{x^3}{x^2+y^3}.$$

Find the limit if it exists.

Problem 8.14.25 (2022 Calculus B, Final, 10pts). Let $z(x,y)$ be determined by

$$z^3 + xz - 2y = 0.$$

Find the maximum of the directional derivatives of $z(x,y)$ at $(1,1)$.

Problem 8.14.26 (2023 Analysis III, Midterm, 10pts). Let

$$f(x,y,z) = (x^3 - 2y^3)e^{-zx^2} + (2 + z\cos y)x^2 - 3y^2 + x.$$

(1) Fix $z=0$. Find the local minima of the function $(x,y) \mapsto f(x,y,0)$.

(2) Prove that there is a $\delta > 0$ such that for any $z \in (-\delta, \delta)$, the function $(x,y) \mapsto f(x,y,z)$ has at least one local minimum.

Problem 8.14.27 (2023 Analysis III, Midterm, 10pts). Find all the 2023rd order partial derivatives of the function

$$f(x,y) = e^{-x^2y} \cdot \sin x$$

at $(0,0)$.

Problem 8.14.28 (2021 Calculus B, Final, 10pts)**.** Let $n \geq 2$ be an integer. Find the Taylor's formula of the function

$$f(x) = \frac{1 - 2x + 5x^2}{(1 - 2x)(1 + x^2)}$$

of order $(2n + 1)$ at $x = 0$.

Problem 8.14.29 (2022 Calculus B, Final, 10pts)**.** Let $f(x) = (\sin x)^{\frac{2}{3}} + (\cos x)^{\frac{2}{3}}$ for $x \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. Find all the local minima of f.

Problem 8.14.30 (2020 Analysis I, Final, 10pts)**.** Find functions $f, g \in C^\infty(\mathbb{R})$ so that both of the following properties hold:

(a) $f^{(n)}(0) = g^{(n)}(0) = 0$ for all $n \in \mathbb{N}^*$,

(b) $x = 0$ is an extremum of f, but not an extremum of g.

8.14.5 Extra Problems

Problem 8.14.31 (2022 Calculus A, Final)**.**

(1) Calculate $\displaystyle\int_{-1}^{1} \left(\frac{\sin^2 x}{1 + e^x} + \frac{\cos^2 x}{1 + e^{-x}} \right) dx$.

(2) Calculate $\displaystyle\int_0^{\frac{\pi}{4}} \ln \frac{\sin(x + \frac{\pi}{4})}{\cos x} dx$.

Problem 8.14.32 (2023 Calculus B, Final)**.**

(1) Prove that for any $x \in \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$, we have

$$2\int_0^x \frac{dt}{\sqrt{1 - t^2}} = \int_0^{2x\sqrt{1-x^2}} \frac{dt}{\sqrt{1 - t^2}}.$$

(2) Prove that for any $x \in \left(-\frac{1}{\sqrt[4]{6}}, \frac{1}{\sqrt[4]{6}}\right)$, we have

$$2\int_0^x \frac{dt}{\sqrt{1 - t^4}} = \int_0^{\frac{2x\sqrt{1-x^4}}{1+x^4}} \frac{dt}{\sqrt{1 - t^4}}.$$

Problem 8.14.33 (2020 Analysis I, Final)**.** Find the limits:

(1) $\lim\limits_{x\to\frac{\pi}{4}-0}(\tan x)^{\tan 2x}$;

(2) $\lim\limits_{x\to 0}\dfrac{\tan x-\sin x}{\sin x-\cos x}$;

(3) $\lim\limits_{x\to+\infty}\dfrac{\int_0^x e^{t^2}dt}{e^{x^2}}$.

Problem 8.14.34 (2020 Analysis I, Final)**.** Let $f:[a,b]\to\mathbb{R}$ be continuous on $[a,b]$ and twice differentiable on (a,b). Suppose that $f(a)=f(b)=0$, and there are $B:(a,b)\to\mathbb{R}$ and $C:(a,b)\to(-\infty,0)$ such that

$$f''(x)+B(x)f'(x)+C(x)f(x)=0$$

for $x\in(a,b)$. Prove that $f(x)=0$ for any $x\in[a,b]$.

Problem 8.14.35 (2017 Analysis III, Midterm)**.** Find the tangent line and the normal plane of the following curve at $(1,-1,1)$:

$$\begin{cases} 3x^2y+y^2z+2=0 \\ 2xz-x^2y-3=0 \end{cases}.$$

Problem 8.14.36 (2022 Analysis III, Midterm)**.** For $n>1$, find the distance between the point $P(p_1,\ldots,p_n)$ and the hyperplane $a_1x_1+\cdots+a_nx_n+b=0$, using the strategy for finding the extrema with constraint.

Problem 8.14.37 (2018 Analysis III, Midterm)**.** Let

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

Let $E=\{(x,y)\in\mathbb{R}^2 : y>f(x)\}$. Find all possible $a\in\mathbb{R}$ so that E is open in $\mathbb{R}^2$.

Problem 8.14.38 (2023 Calculus B, Final)**.** Let $P(x)$ be a continuous function on $[0,1]$ with $P(0)=0$ and $P(1)=1$. Suppose that $P(x)$ is differentiable on $(0,1)$, and $P'(x)>0$ for any $x\in(0,1)$. Prove that for any $A,B\in(0,+\infty)$, $n\in\mathbb{N}^*$, there are $\theta_0,\theta_1,\ldots,\theta_n\in(0,1)$ such that

$$0<\theta_0<\theta_1<\cdots<\theta_n<1,$$

and

$$(A+B)^n = \sum_{k=0}^{n} \frac{1}{P'(\theta_k)} \frac{n!}{k!(n-k)!} A^{n-k} B^k.$$

Problem 8.14.39 (2017 Analysis III, Midterm)**.** Let $z = f(x, y)$ be an infinitely differentiable function, and let

$$g(t, \theta) = f(t\cos\theta, t\sin\theta).$$

(1) Suppose that for any $\theta \in \mathbb{R}$, we have $g_t(0, \theta) = 0$, $g_{tt}(0, \theta) > 0$. Prove that f attains a local minimum at $(0, 0)$.

(2) Suppose that for any $\theta \in \mathbb{R}$, $g(\cdot, \theta) : t \mapsto f(t\cos\theta, t\sin\theta)$ attains a local minimum at $t = 0$. Does f attain a local minimum at $(0, 0)$?

Problem 8.14.40 (2022 Analysis III, Midterm)**.** Let $f(x, y)$ be continuous on $[a, b] \times [a, b]$. For $t \in [a, b]$, let

$$\varphi(t) = \max_{x \in [a,t]} \max_{y \in [a,x]} f(x, y).$$

Prove that $\varphi(t)$ is continuous on $[a, b]$.

Problem 8.14.41 (2023 Analysis III, Midterm)**.** For $n \geq 2$, let

$$S^{n-1} = \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : x_1^2 + \cdots + x_n^2 = 1\}.$$

(1) For $\mathbf{x} \in S^{n-1}$, let

$$f(\mathbf{x}) = \sum_{i,j=1}^{n} a_{ij} x_i x_j.$$

where $a_{ij} = a_{ji}$ for any i, j. Find the local extrema of f.

(2) Given k points $\{P_i(a_i, b_i, c_i) \in \mathbb{R}^3 : 1 \leq i \leq k\}$, find the point $Q \in S^2$ so that the following distance function attains the minimum:

$$F(Q) = |Q - P_1|^2 + |Q - P_2|^2 + \cdots + |Q - P_k|^2.$$

Problem 8.14.42 (2018 Analysis III, Midterm)**.** Let

$$f(x, y, z) = 2x^2 + y^4 + z^4$$

with the constraint $xyz = 1$. Can f attain a global minimum? Find it if possible.

8.15 Final Exams

Problem 8.15.1 (2024 Calculus B, Final)**.** Determine whether the following limits exist or not. Find the limit if it exists.

(1) $\lim\limits_{(x,y)\to(0,0)} \dfrac{xy}{x^2+(\tan y)^2}$;

(2) $\lim\limits_{(x,y)\to(0,0)} \left(\dfrac{xy}{e^x-1}+\sin y\right)\sin\dfrac{1}{x^2+y^2}$.

Problem 8.15.2 (2024 Calculus B, Final)**.** Suppose that $z(x,y)$ is determined by the equation

$$z^3+x^2z-2y^3=0.$$

Find the maximum directional derivative of $z(x,y)$ at $(1,1)$.

Problem 8.15.3 (2024 Calculus B, Final)**.** Let $f:\mathbb{R}^2\to\mathbb{R}$ be defined by

$$f(x,y)=x^2+2xy\sin(x+y)-y^2.$$

Show that there is a neighborhood D of $(0,0)$, and a C^1 invertible map $(u,v)\mapsto(x(u,v),y(u,v))$ for $(u,v)\in D$ such that $x(0,0)=0$, $y(0,0)=0$ and

$$f(u,v)=u^2-v^2$$

for all $(u,v)\in D$.

Problem 8.15.4 (2024 Calculus B, Final)**.** Find the minimum distance between $O(0,0,0)$ and the surface

$$(x-y)^2-z^2=4.$$

Problem 8.15.5 (2024 Calculus B, Final)**.** Let $f:[-1,1]\to\mathbb{R}$ be a Riemann integrable function on $[-1,1]$. Suppose that $\lim\limits_{x\to0}f(x)=A$ for some A. Prove that

$$\lim_{n\to+\infty}\int_{-1}^{1}\frac{nf(x)}{1+n^2x^2}dx=\pi A.$$

Problem 8.15.6 (2024 Calculus A, Final)**.** Find the limit

$$\lim_{(x,y)\to(0,0)}\left(x^2+y^2\right)^{x^2}.$$

Problem 8.15.7 (2024 Calculus A, Final)**.** Find the limit

$$\lim_{x \to +\infty} \left(\frac{\pi}{2} - \arctan x \right)^{\frac{1}{\ln x}}.$$

Problem 8.15.8 (2024 Calculus A, Final)**.** Let f be twice differentiable on $[a, b]$. Suppose that $f(a) = f(b) = 0$, and there is a point $c \in (a, b)$ such that

$$f(c) > 0.$$

Prove that there exists a point $\xi \in (a, b)$ such that $f''(\xi) < 0$.

Problem 8.15.9 (2024 Calculus A, Final)**.**

(1) Let Σ be a plane containing P_0, with a normal vector $\mathbf{n}$. Let $P_1 \in \mathbb{R}^3 \setminus \Sigma$. Use $\overrightarrow{P_0P_1}$ and $\mathbf{n}$ to express the distance between P_1 and Σ.

(2) Let L be a line containing P_0, with a directional vector $\mathbf{v}$. Let $P_1 \in \mathbb{R}^3 \setminus L$. Use $\overrightarrow{P_0P_1}$ and $\mathbf{v}$ to express the distance between P_1 and L.

(3) Let L_1, L_2 be *skew lines* (异面直线) containing P_1, P_2, with directional vectors $\mathbf{v}_1$, $\mathbf{v}_2$, respectively. Use $\overrightarrow{P_1P_2}$ and $\mathbf{v}_1$, $\mathbf{v}_2$ to express the distance between L_1 and L_2.

Problem 8.15.10 (2024 Calculus A, Final)**.** Let

$$f(x, y) = \begin{cases} y \arctan \frac{1}{\sqrt{x^2+y^2}} & , \text{ if } (x, y) \neq (0, 0) \\ 0 & , \text{ if } (x, y) = (0, 0) \end{cases}.$$

Discuss whether f is differentiable at $(0, 0)$.

Problem 8.15.11 (2024 Calculus A, Final)**.** Let

$$f(x, y) = \begin{cases} \frac{2xy^3}{x^2+y^4} & , \text{ if } x^2 + y^2 \neq 0 \\ 0 & , \text{ if } x^2 + y^2 = 0 \end{cases}.$$

Calculate the directional derivative $\dfrac{\partial f}{\partial \mathbf{v}}(0, 0)$, where $\mathbf{v} = (\cos \alpha, \sin \alpha)$, $\alpha \in [0, 2\pi)$ is a unit vector.

Problem 8.15.12 (2024 Calculus A, Final)**.** Let $g(x,y)$ be a function on $\mathbb{R}^2$.

(1) Suppose that g attains a minimum at (x_0,y_0). Does $h_\theta(t)=g(x_0+t\cos\theta, y_0+t\sin\theta)$ attain a local minimum at $t=0$ for any $\theta\in[0,2\pi)$?

(2) Suppose that for any $\theta\in\mathbb{R}$, $h_\theta(t)=g(x_0+t\cos\theta,y_0+t\sin\theta)$ attains a local minimum at $t=0$. Does g attain a local minimum at (x_0,y_0)?

Problem 8.15.13 (2024 Calculus A, Final)**.** Let $z=z(x,y)$ be a function determined by the equation:

$$(x^2+y^2)z+\ln z+2(x+y+1)=0.$$

Find the local extrema of $z(x,y)$.

Problem 8.15.14 (2024 Calculus A, Final)**.** Let $f(x)$ be a twice differentiable function on $[a,b]$. Suppose that $f(a)=f(b)=f'(a)=f'(b)=0$, and

$$|f''(x)|\leq M$$

for $x\in[a,b]$. Prove that

$$|f(x)|\leq\frac{M}{16}(b-a)^2$$

for $x\in[a,b]$.

8.16 Hints

8.4.6: Use the Cauchy-Schwarz inequality.
8.6.26: Consider the Riemann function.

Sketch proof of Problem 8.6.27. Let $g(x)=f(x)-x$. Then by the definition, g is continuous on $[a,b]$ and $g(a)\geq 0$ and $g(b)\leq 0$. Suppose that

$$\lim_{n\to\infty}(x_{n+1}-x_n)=\lim_{n\to\infty}f(x_n)-x_n=\lim_{n\to\infty}g(x_n)=0. \tag{8.16.1}$$

If $g(x_n)=0$ for some n, then we are done. Now assume that

$$g(x_n)\neq 0 \tag{8.16.2}$$

for any $n \in \mathbb{N}$. Then by (8.16.1), x_n tends to the zeros of g.

Now suppose that there are two subsequences $y_i = x_{n_i} \to y$, and $z_j = x_{n_j} \to z$ with $y < z$. By the continuity of g, we have $g(y) = g(z) = 0$. Clearly, $g([y, z]) \neq \{0\}$ by (8.16.2). Choose $w \in (y, z)$ so that $|g(w)| = A > 0$. Then by the continuity of g, there is an interval $(w - \epsilon, w + \epsilon) \subset [y, z]$ so that

$$|g(x)| \geq A/2$$

for $x \in (w - \epsilon, w + \epsilon)$.

Now let N be sufficiently large so that

$$|g(x_n)| = |x_{n+1} - x_n| < \min\{A/100, \epsilon/100\} \tag{8.16.3}$$

for any $n \geq N$. Then by the assumption, since $\{x_n\}$ has subsequences tending to y and z, there are points x_n in between. Moreover, by (8.16.3), one may show that there are points $x_n \in (w - \epsilon, w + \epsilon)$. This leads to a contradiction. $\square$

8.7.9: Use Proposition 1.3.21.

8.9.5: (1) Evaluate the Fourier series of f at $x = 0$. (2) Use Euler's formula and calculate the integral.

8.10.1: Induct on l and apply the Taylor's formula.

8.10.3: Consider $F(x) = (f(x))^2 + (f'(x))^2$ for $x \in [-2, 2]$. Or consider $F(x) = f(x)\sin x + f'(x)\cos x$ for $x \in [-2, 2]$.

8.12.8: Fix $\mathbf{x}$ and consider $F(t) = f(t\mathbf{x})$.

8.12.9: $V(\mathbf{x}) = \prod\limits_{1 \leq i < j \leq n} (x_j - x_i)$.

Sketch proof of Extra Problem 8.13.6. Let $F(x, y, u) = xy + u + \mathrm{sh}[(x^2 + y^2 + 1)u] = 0$. Then

$$F_u(x, y, u) = 1 + (x^2 + y^2 + 1)\,\mathrm{ch}[(x^2 + y^2 + 1)u] > 0$$

for any $(x, y, u) \in \mathbb{R}^3$. In particular, for fixed $(x_0, y_0) \in \mathbb{R}^2$, $G(u) = F(x_0, y_0, u)$ is strictly increasing with respect to u. Moreover,

$$\lim_{u \to -\infty} G(u) = -\infty, \quad \lim_{u \to +\infty} G(u) = +\infty.$$

Thus, there is a unique $u_0 = u_0(x_0, y_0)$ such that $0 = G(u_0) = F(x_0, y_0, u_0)$. This defines a function u_0 on $\mathbb{R}^2$. $\square$

Proof of Problem 8.14.7. (1) Write $p_0 = (x_0, y_0)$. By the chain rule, we have

$$0 = F_x + F_z f_x, \quad 0 = F_y + F_z f_y. \tag{8.16.4}$$

Since f attains a local extremum at p_0 (and f is of class C^1 in a neighborhood of p_0), we have

$$\mathbf{0} = [f_x(p_0), f_y(p_0)] = \left[-\frac{F_x(P_0)}{F_z(P_0)}, -\frac{F_y(P_0)}{F_z(P_0)}\right].$$

Thus, $0 = F_x(P_0) = F_y(P_0)$.

(2) Since $0 = F_x(P_0) = F_y(P_0)$, p_0 is a stationary point of f. Differentiate (8.16.4) again:

$$0 = (F_x + F_z f_x)_x = F_{xx} + F_{xz} f_x + (F_{zx} + F_{zz} f_x) f_x + F_z f_{xx}.$$

In particular, we have

$$0 = F_{xx}(P_0) + F_z(P_0) f_{xx}(p_0).$$

Similarly, one may calculate

$$\begin{aligned} H(P_0) &= \begin{bmatrix} \dfrac{\partial^2 F}{\partial x^2}(P_0) & \dfrac{\partial^2 F}{\partial x \partial y}(P_0) \\ \dfrac{\partial^2 F}{\partial x \partial y}(P_0) & \dfrac{\partial^2 F}{\partial y^2}(P_0) \end{bmatrix} \\ &= \begin{bmatrix} -F_z(P_0) f_{xx}(p_0) & -F_z(P_0) f_{xy}(p_0) \\ -F_z(P_0) f_{xy}(p_0) & -F_z(P_0) f_{yy}(p_0) \end{bmatrix} = -F_z(P_0)\mathbf{H}_f(p_0). \end{aligned}$$

This means $\mathbf{H}_f(p_0)$ is positive or negative definite. Thus, f attains a local extremum at p_0. $\square$

Proof of Problem 8.14.10. Let $D = \frac{f(1)-f(0)}{\ln\sqrt{3}}$. Consider the differential equation $\frac{D}{1+2x} dx = df$. Then we have

$$\frac{D}{2}\ln(1+2x) = f + C.$$

Then let $F(x) = f(x) - \frac{D}{2}\ln(1+2x)$. Then $F(0) = f(0)$, $F(1) = f(0)$. Then use Rolle's theorem. $\square$

Proof of Problem 8.14.19. (1) Write $f(x) = x\ln(1+(\frac{3}{2})^{\frac{1}{x}}) + \ln 2$, $a = \frac{3}{2} > 1$. Then

$$\begin{aligned} f'(x) =& \ln(1+a^{\frac{1}{x}}) + \frac{x}{1+a^{\frac{1}{x}}} a^{\frac{1}{x}} \ln a \cdot \left(-\frac{1}{x^2}\right) \\ =& \frac{(1+a^{\frac{1}{x}})\ln(1+a^{\frac{1}{x}}) - a^{\frac{1}{x}} \ln a^{\frac{1}{x}}}{1+a^{\frac{1}{x}}} > 0. \end{aligned}$$

(3) We shall show that if $x \nearrow$, then $f'(x) \nearrow$. Let $y = a^{\frac{1}{x}} > 1$. Then as $x \nearrow$, $\frac{1}{x} \searrow$, $y \searrow$. Thus, we need to show that

$$g(y) = \ln(1+y) - \frac{y}{1+y} \ln y$$

is strictly decreasing. Now

$$g'(y) = \frac{1}{1+y} - \frac{(\ln y + 1)(1+y) - y \ln y}{(1+y)^2} = \frac{-\ln y}{(1+y)^2} < 0.$$

□

8.14.20: Use Taylor's formula.

Proof of Problem 8.14.21. By the assumptions, $g(x) - f(x) > 0$. Moreover,

$$g''(x) - f''(x) > 0.$$

The latter means $g'(x) - f'(x)$ is strictly increasing. Suppose that $g'(x_0) - f'(x_0) = 0$ for some $x_0 \in \mathbb{R}$. Pick the line

$$y = g'(x_0)(x - x_0) + g(x_0)$$

and we are done.

Thus, assume that $g'(x) - f'(x) > 0$ for all $x \in \mathbb{R}$ (the case with reverse inequality is similar). Then by the bounded convergence,

$$\lim_{x \to -\infty} g'(x) - f'(x) \geq 0.$$

Also, since $g'(x) \searrow$, $f'(x) \nearrow$ as $x \searrow -\infty$, for any $x_1 \in \mathbb{R}$,

$$g'(x) \geq f'(x) \geq f'(x_1)$$

for any $x \leq x_1$. In particular, by the bounded convergence,

$$\lim_{x \to -\infty} g'(x) = G,$$

for some $G \in \mathbb{R}$. Moreover,

$$f'(x) \leq G \leq g'(x)$$

for $x \in \mathbb{R}$. Thus, for any $x_2 \in \mathbb{R}$, we have

$$g(x) - Gx \geq f(x) - Gx$$

for $x \leq x_2$. Since $g(x) - Gx \searrow$ and $f(x) - Gx \nearrow$ as $x \searrow -\infty$, by the bounded convergence,

$$\lim_{x \to -\infty} g(x) - Gx = A.$$

Then we can pick the line

$$y = Gx + A.$$

$\square$

Proof of Problem 8.14.26. (1) One calculates

$$f_x(x, y, 0) = 3x^2 + 4x + 1, \quad f_y(x, y, 0) = -6y^2 - 6y.$$

Then the stationary points are $(-1, 0)$, $(-1, -1)$, $(-\frac{1}{3}, 0)$, $(-\frac{1}{3}, -1)$. Further,

$$\mathbf{H}_f(x, y, 0) = \begin{bmatrix} 6x + 4 & 0 \\ 0 & -12y - 6 \end{bmatrix}.$$

Then only $\mathbf{H}_f(-\frac{1}{3}, -1, 0)$ is positive definite. Thus, $f(x, y, 0)$ attains the minimum at $(-\frac{1}{3}, -1)$.

(2) Let $g^z : (x, y) \mapsto f(x, y, z)$. If

$$\mathbf{0} = (\partial_x g^z, \partial_y g^z) = (f_x(x, y, z), f_y(x, y, z)) =: G(x, y, z),$$

$$\mathbf{H}_{g^z}(x, y) = \begin{bmatrix} g^z_{xx} & g^z_{yx} \\ g^z_{xy} & g^z_{yy} \end{bmatrix} = \begin{bmatrix} f_{xx}(x, y, z) & f_{yx}(x, y, z) \\ f_{xy}(x, y, z) & f_{yy}(x, y, z) \end{bmatrix} \text{ is positive definite,}$$

then g^z attains a minimum at (x, y). Now $G(-\frac{1}{3}, -1, 0) = \mathbf{0}$. Moreover,

$$\det G_{(x,y)}\left(-\frac{1}{3}, -1, 0\right) = \det \begin{bmatrix} g^0_{xx} & g^0_{xy} \\ g^0_{yx} & g^0_{yy} \end{bmatrix} > 0.$$

Then by the implicit function theorem, there is a map $z \mapsto (x(z), y(z))$ on a neighborhood $(-\delta, \delta)$ such that $0 \mapsto (-\frac{1}{3}, -1)$ and

$$G(x(z), y(z), z) = \mathbf{0}, \quad G_{(x,y)}(x(z), y(z), z) = \begin{bmatrix} g^z_{xx} & g^z_{xy} \\ g^z_{yx} & g^z_{yy} \end{bmatrix}.$$

By the continuity of $G_{(x,y)}$ on $(-\delta, \delta)$, (by shrinking the neighborhood if necessary), $\det G_{(x,y)} > 0$ and $g^z_{xx} > 0$ on $(-\delta, \delta)$. Therefore, $\mathbf{H}_{g^z}(x(z), y(z))$ is positive definite, and so g^z attains a minimum at $(x(z), y(z))$ for $z \in (-\delta, \delta)$. $\square$

8.14.30: Consider [1, P.238-239].

8.14.39(2): A noncontinuous counterexample:

$$f(x, y) = \begin{cases} x & , \text{ if } y = x^2 \\ 0 & , \text{ if } y \neq x^2 \end{cases}.$$

A C^∞ counterexample:

$$f(x, y) = \begin{cases} \left(x - \left|\frac{y}{2x}\right|\right)\left(x - \left|\frac{y}{x}\right|\right) e^{-\frac{1}{x^2 - y^2}} & , \text{ if } |x| > |y| \\ \left(y - \left|\frac{x}{2y}\right|\right)\left(y - \left|\frac{x}{y}\right|\right) e^{-\frac{1}{y^2 - x^2}} & , \text{ if } |x| < |y| \\ 0 & , \text{ if } |x| = |y| \end{cases}.$$

Proof of Problem 8.14.41. (1) Let

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}.$$

Then $f(\mathbf{x}) = \mathbf{x}A\mathbf{x}^T$ where $\mathbf{x} = (x_1, \ldots, x_n) \in S^{n-1}$ is a row vector.

By the spectral theorem, there is an orthogonal matrix P and a real diagonal matrix D such that

$$A = PDP^T, \quad D = \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{bmatrix}.$$

Note that $\mathbf{x}P^{-1} \in S^1$ if and only if $\mathbf{x} \in S^1$ via Remark 7.2.39. Let $g(\mathbf{x}) = f(\mathbf{x}P^{-1}) = d_1x_1^2 + \cdots + d_nx_n^2$. Let $d_M = \max_i d_i$, $d_m = \min_i d_i$. Then

$$d_m = d_m(x_1^2 + \cdots + x_n^2) \leq g(\mathbf{x}) \leq d_M(x_1^2 + \cdots + x_n^2) = d_M.$$

Also g attains the maximum d_M at $x_M = 1$, and the minimum d_m at $x_m = 1$. Therefore, f attains the maximum d_M at the eigenvector associated to d_M, and the minimum d_m at the associated to d_m.

(2) Write $Q = (x, y, z) \in S^2$. Then

$$F(Q) = F(x, y, z) = \sum_{i=1}^{k}(x - a_i)^2 + (y - b_i)^2 + (z - c_i)^2.$$

Note that since $F(x, y, z)$ is continuous on S^2, it must attain a minimum. Now consider the Lagrange's function

$$L(x, y, z, \lambda) = F(x, y, z) + \lambda(x^2 + y^2 + z^2 - 1).$$

Then one calculates

$$\begin{cases} L_x = 2kx - (a_1 + \cdots + a_k) + 2\lambda x = 0 \\ L_y = 2ky - (b_1 + \cdots + b_k) + 2\lambda y = 0 \\ L_z = 2kz - (c_1 + \cdots + c_k) + 2\lambda z = 0 \\ L_\lambda = x^2 + y^2 + z^2 - 1 = 0 \end{cases}. \tag{8.16.5}$$

Let $A = a_1 + \cdots + a_k$, $B = b_1 + \cdots + b_k$, $C = c_1 + \cdots + c_k$, $T = \sqrt{A^2 + B^2 + C^2}$. If $T = 0$, then (8.16.5) implies that any point $(x, y, z) \in S^2$ is stationary. Moreover, one deduces that

$$\begin{aligned} F(x, y, z) =& \sum_{i=1}^{k}(x^2 - 2a_ix + a_i^2) + (y^2 - 2b_iy + b_i^2) + (z^2 - 2c_iz + c_i^2) \\ =& k + \sum_{i=1}^{k} a_i^2 + b_i^2 + c_i^2 \end{aligned}$$

is a constant function for $(x, y, z) \in S^2$. If $T > 0$, then (8.16.5) implies two stationary points

$$(x, y, z) = \left(\frac{A}{T}, \frac{B}{T}, \frac{C}{T}\right) \text{ or } \left(-\frac{A}{T}, -\frac{B}{T}, -\frac{C}{T}\right).$$

Now one calculates

$$\begin{aligned}
&F\left(\frac{A}{T},\frac{B}{T},\frac{C}{T}\right)-F\left(-\frac{A}{T},-\frac{B}{T},-\frac{C}{T}\right)\\
=&\sum_{i=1}^{k}\left(\frac{A}{T}-a_i\right)^2+\left(\frac{B}{T}-b_i\right)^2+\left(\frac{C}{T}-c_i\right)^2\\
&-\sum_{i=1}^{k}\left(-\frac{A}{T}-a_i\right)^2+\left(-\frac{B}{T}-b_i\right)^2+\left(-\frac{C}{T}-c_i\right)^2\\
=&\sum_{i=1}^{k}\left(\frac{A}{T}-a_i-\frac{A}{T}-a_i\right)\left(\frac{A}{T}-a_i+\frac{A}{T}+a_i\right)\\
&+\sum_{i=1}^{k}\left(\frac{B}{T}-b_i-\frac{B}{T}-b_i\right)\left(\frac{B}{T}-b_i+\frac{B}{T}+b_i\right)\\
&+\sum_{i=1}^{k}\left(\frac{C}{T}-c_i-\frac{C}{T}-c_i\right)\left(\frac{C}{T}-c_i+\frac{C}{T}+c_i\right)\\
=&\sum_{i=1}^{k}-2a_i\frac{2A}{T}+\sum_{i=1}^{k}-2b_i\frac{2B}{T}+\sum_{i=1}^{k}-2c_i\frac{2C}{T}\\
=&-\frac{4(A^2+B^2+C^2)}{T}=-4T<0.
\end{aligned}$$

Therefore, $F(x,y,z)$ attains the minimum at $Q=\left(\frac{A}{T},\frac{B}{T},\frac{C}{T}\right)$. $\square$

Proof of Problem 8.15.3. Let $O=(0,0)$, and

$$\begin{cases} u(x,y)=x+y\sin(x+y) \\ v(x,y)=y\sqrt{1+\sin(x+y)} \end{cases}. \tag{8.16.6}$$

Then $f(x,y)=u^2(x,y)-v^2(x,y)$, $(u(O),v(O))=O$. One calculates

$$\mathbf{J}_{(u,v)}(O)=\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Then by the inverse function theorem, we conclude that there is a neighborhood D of O, and an inverse function $(u,v)\mapsto(x(u,v),y(u,v))$ on D. It follows that

$$(x(O),y(O))=O,\quad f(x,y)=f(x(u,v),y(u,v))=u^2-v^2$$

for $(u,v)\in D$. $\square$

Proof of Problem 8.15.5. Since f is Riemann integrable, f is bounded, i.e. $|f| < M$ for some M. We rewrite the integral

$$\int_{-1}^{1} \frac{nf(x)}{1+n^2x^2}dx = \int_{-n}^{n} \frac{f(\frac{y}{n}) - A}{1+y^2}dy + \int_{-n}^{n} \frac{A}{1+y^2}dy. \tag{8.16.7}$$

(For those who are familiar with the *real analysis* (实分析), the consequence follows immediately from the *Lebesgue dominated convergence theorem* (勒贝格控制收敛定理).) For the second part of the right hand side of (8.16.7), one calculates

$$\lim_{n\to+\infty} \int_{-n}^{n} \frac{A}{1+y^2}dy = A \lim_{n\to+\infty} \arctan(n) - \arctan(-n) = A\pi.$$

For the first part, we first unwrap the condition $\lim\limits_{z\to 0} f(z) = A$: for any $\epsilon > 0$, there is a $\delta = \delta(\epsilon) > 0$ such that

$$|f(z) - A| < \epsilon$$

for any $|z| < \delta$. Letting $z = \frac{y}{n}$, we have

$$\begin{aligned}
&\left|\int_{-n}^{n} \frac{f(\frac{y}{n}) - A}{1+y^2}dy\right| \\
\leq &\left|\int_{-n\delta}^{n\delta} \frac{f(\frac{y}{n}) - A}{1+y^2}dy\right| + \left|\int_{n\delta}^{n} \frac{f(\frac{y}{n}) - A}{1+y^2}dy\right| + \left|\int_{-n}^{-n\delta} \frac{f(\frac{y}{n}) - A}{1+y^2}dy\right| \\
\leq &\int_{-n\delta}^{n\delta} \frac{|f(\frac{y}{n}) - A|}{1+y^2}dy + \frac{2(M+A)n}{1+(\delta n)^2} \\
\leq &\epsilon\pi + \frac{2(M+A)n}{1+(\delta n)^2}.
\end{aligned}$$

Then one can find $N = N(\epsilon, \delta(\epsilon))$ such that

$$\frac{2(M+A)n}{1+(\delta n)^2} < \epsilon\pi, \quad \text{and so} \quad \left|\int_{-n}^{n} \frac{f(\frac{y}{n}) - A}{1+y^2}dy\right| < 2\epsilon\pi$$

for any $n \geq N$. This implies

$$\lim_{n\to+\infty} \int_{-n}^{n} \frac{f(\frac{y}{n}) - A}{1+y^2}dy = 0.$$

□

8.15.14: Suppose that f attains a maximum at $\xi \in [a, b]$. Then either $\xi \leq \frac{a+b}{2}$ or $\xi > \frac{a+b}{2}$. Now consider $f|_{[a,\xi]}$ or $f|_{[\xi,b]}$.

参考文献

[1] 李忠, 周建莹. 高等数学 (上册). 第三版. 北京大学出版社, 2023.

[2] 李忠, 周建莹. 高等数学 (下册). 第三版. 北京大学出版社, 2023.

索引