Advanced Analysis

Min Yan
Department of Mathematics
Hong Kong University of Science and Technology

September 3, 2009

Contents

Chapter 1

Limit and Continuity

1.1   Limit of Sequence

A *sequence* is an infinite list

$$x_1, \ x_2, \ x_3, \ \ldots, \ x_n, \ x_{n+1}, \ \ldots .$$

The sequence can also be denoted as $\{x_n\}$. The subscript n is called the *index* and does not have to start from 1. For example,

$$x_5, \ x_6, \ x_7, \ \ldots, \ x_n, \ x_{n+1}, \ \ldots,$$

is also a sequence, with the index starting from 5.

In this chapter, the terms x_n of a sequence are assumed to be real numbers and can be plotted on the real number line. A sequence has a limit l if the following implication happens:

$$n \text{ is big } \implies \ x_n \text{ is close to } l.$$

Intuitively, this means that the sequence "accumulates" around l. However, to give a rigorous definition, the meaning of "big", "close" and "implies" has to be more precise. The bigness of a number n is usually measured by $n > N$ for some big N. For example, we say n is "in the thousands" if $N = 1,000$. The closeness between two numbers u and v is usually measured by (the smallness of) the size of $|u - v|$. The implication means that the predetermined smallness of $|x_n - l|$ may be achieved by the bigness of n.

1.1.1   Definition

Definition 1.1.1. A sequence $\{x_n\}$ of real numbers has *limit l* (or *converges to l*), and denoted $\lim_{n \to \infty} x_n = l$, if for any $\epsilon > 0$, there is N, such that

$$n > N \implies |x_n - l| < \epsilon. \tag{1.1.1}$$

A sequence is *convergent* if it has a (finite) limit. Otherwise, the sequence is *divergent*.
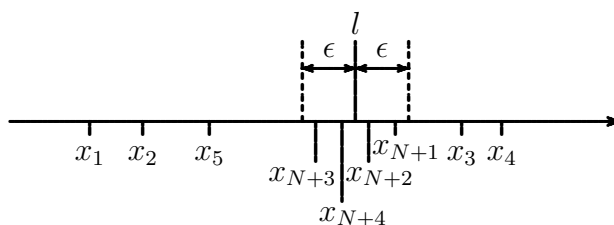
Figure 1.1: for any ϵ, there is N

Note the logical relation between ϵ and N. The predetermined smallness ϵ for $|x_n - l|$ is *arbitrarily* given, while the size N for n is to be found *after* ϵ is given. Thus the choice of N usually depends on ϵ and is often expressed as a function of ϵ.
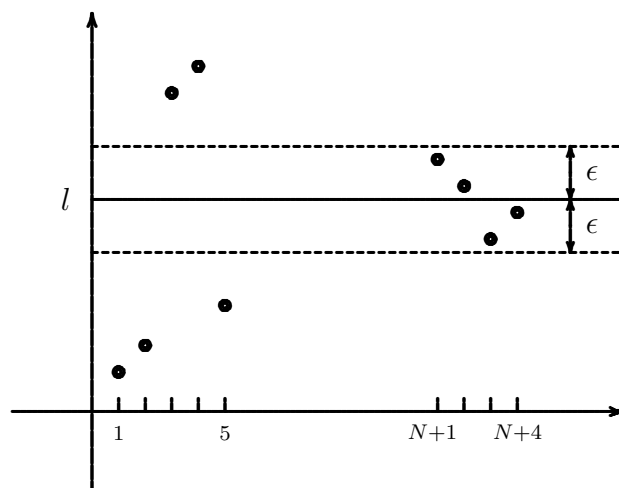
Figure 1.2: another plot of a converging sequence

Since the limit is about the long term behavior of a sequence getting closer to a target, only small ϵ and big N need to be considered establishing a limit. For example, the limit of a sequence is not changed if the first one hundred terms are replaced by other arbitrary numbers. Exercise 1.1.9 contains more examples.

Example 1.1.1. Intuitively, the bigger n is, the smaller $\dfrac{1}{n}$ gets. This suggests $\lim_{n \to \infty} \dfrac{1}{n} = 0$. Rigorously following the definition, for any $\epsilon > 0$, choose $N = \dfrac{1}{\epsilon}$. Then

$$n > N \implies \left| \frac{1}{n} - 0 \right| = \frac{1}{n} < \frac{1}{N} = \epsilon.$$

Thus the implication (1.1.1) is established for the sequence $\left\{ \dfrac{1}{n} \right\}$.

How did we find the suitable N? Our goal is to achieve $\left| \dfrac{1}{n} - 0 \right| < \epsilon$. This is the same as $n > \dfrac{1}{\epsilon}$, which suggests us to take $N = \dfrac{1}{\epsilon}$. Note that our choice of N may not be a natural number.

Example 1.1.2. The sequence $\left\{ \dfrac{n}{n + (-1)^n} \right\}$, with index starting from $n = 2$, is

$$\frac{2}{3}, \ \frac{3}{2}, \ \frac{4}{5}, \ \frac{5}{4}, \ \frac{6}{7}, \ \frac{7}{6}, \ \dots .$$

Plotting the sequence suggests $\lim_{n \to \infty} \dfrac{n}{n + (-1)^n} = 1$.

For the rigorous argument, we observe that

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

In order for the left side to be less than ϵ, it is sufficient to make $\dfrac{1}{n - 1} < \epsilon$, which is the same as $n > \dfrac{1}{\epsilon} + 1$.

Based on the analysis, we have the following formal and rigorous argument for the limit: For any $\epsilon > 0$, choose $N = \dfrac{1}{\epsilon} + 1$. Then

$$n > N \implies \left| \frac{n}{n + (-1)^n} - 1 \right| = \frac{1}{n + (-1)^n} \leq \frac{1}{n - 1} < \frac{1}{N - 1} = \epsilon.$$

Example 1.1.3. Consider the sequence

$$1.4, \ 1.41, \ 1.414, \ 1.4142, \ 1.41421, \ 1.414213, \ 1.4142135, \ 1.41421356, \ \dots$$

of more and more refined decimal approximations of $\sqrt{2}$. The intuition suggests that $\lim_{n \to \infty} x_n = \sqrt{2}$. The rigorous verification means that for any $\epsilon > 0$, we need to find N, such that

$$n > N \implies |x_n - \sqrt{2}| < \epsilon.$$

Since the n-th term x_n is the decimal expansion up to the n-th decimal point, it satisfies $|x_n - \sqrt{2}| < 10^{-n}$. Therefore it suffices to find N such that the following implication holds

$$n > N \implies 10^{-n} < \epsilon.$$

Assume $1 > \epsilon > 0$. Then ϵ has the decimal expansion

$$\epsilon = 0.00 \cdots 0 E_N E_{N+1} E_{N+2} \cdots,$$

with E_N is from $\{1, 2, \dots, 9\}$. In other words, N is the location of the first nonzero digit in the decimal expansion of ϵ. Then for $n > N$, we have

$$\epsilon \geq 0.00 \cdots 0 E_N \geq 0.00 \cdots 01 = 10^{-N} > 10^{-n}.$$

The argument above assumes $1 > \epsilon > 0$. This is not a problem because if we can achieve $|x_n - l| < 0.5$ for $n > N$, then we can certainly achieve $|x_n - l| < \epsilon$ for any $\epsilon \geq 1$ and for the same $n > N$. In other words, we may add the assumption that ϵ is less than a certain fixed number without hurting the overall rigorous argument for the limit. See Exercise 1.1.9 for more freedom we may have in choosing ϵ and N.

Exercise 1.1.1. Rigorously verify the limits.

1. $\lim_{n \to \infty} \dfrac{2n}{n - 2} = 2.$

2. $\lim_{n \to \infty} \dfrac{1}{\sqrt{n}} = 0.$

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

4. $\lim_{n \to \infty} \dfrac{1}{n^{2/3} - n^{1/2}} = 0.$

5. $\lim_{n \to \infty} \dfrac{\cos n}{n} = 0.$

6. $\lim_{n \to \infty} \dfrac{\sqrt{n} - \cos n}{\sqrt{n} + \sin n} = 1.$

Exercise 1.1.2. Let a positive real number $a > 0$ have the decimal expansion

$$a = X.Z_1 Z_2 \cdots Z_n Z_{n+1} \cdots,$$

where X is a non-negative integer, and Z_n is a single digit integer from $\{0, 1, 2, \dots, 9\}$ at the n-th decimal point. Prove the sequence

$$X.Z_1, \ X.Z_1 Z_2, \ X.Z_1 Z_2 Z_3, \ X.Z_1 Z_2 Z_3 Z_4, \ \dots$$

of more and more refined decimal approximations converges to a.

Exercise 1.1.3. Suppose $x_n \le l \le y_n$ and $\lim_{n\to\infty}(x_n - y_n) = 0$. Prove $\lim_{n\to\infty} x_n = \lim_{n\to\infty} y_n = l$.

Exercise 1.1.4. Suppose $|x_n - l| \le y_n$ and $\lim_{n\to\infty} y_n = 0$. Prove $\lim_{n\to\infty} x_n = l$.

Exercise 1.1.5. Suppose $\lim_{n\to\infty} x_n = l$. Prove $\lim_{n\to\infty} |x_n| = |l|$. Is the converse true?

Exercise 1.1.6. Suppose $\lim_{n\to\infty} x_n = l$. Prove $\lim_{n\to\infty} x_{n+3} = l$. Is the converse true?

Exercise 1.1.7. Prove that the limit is not changed if finitely many terms are modified. In other words, if there is N, such that $x_n = y_n$ for $n > N$, then $\lim_{n\to\infty} x_n = l$ if and only if $\lim_{n\to\infty} y_n = l$.

Exercise 1.1.8. Prove the uniqueness of the limit. In other words, if $\lim_{n\to\infty} x_n = l$ and $\lim_{n\to\infty} x_n = l'$, then $l = l'$.

Exercise 1.1.9. Prove the following are equivalent definitions of $\lim_{n\to\infty} x_n = l$.

1. For any $c > \epsilon > 0$, where c is some fixed number, there is N, such that $|x_n - l| < \epsilon$ for all $n > N$.

2. For any $\epsilon > 0$, there is a natural number N, such that $|x_n - l| < \epsilon$ for all $n > N$.

3. For any $\epsilon > 0$, there is N, such that $|x_n - l| \le \epsilon$ for all $n > N$.

4. For any $\epsilon > 0$, there is N, such that $|x_n - l| < \epsilon$ for all $n \ge N$.

5. For any $\epsilon > 0$, there is N, such that $|x_n - l| \le 2\epsilon$ for all $n > N$.

Exercise 1.1.10. Which are equivalent to the definition of $\lim_{n\to\infty} x_n = l$?

1. For $\epsilon = 0.001$, we have $N = 1000$, such that $|x_n - l| < \epsilon$ for all $n > N$.

2. For any $0.001 \ge \epsilon > 0$, there is N, such that $|x_n - l| < \epsilon$ for all $n > N$.

3. For any $\epsilon > 0.001$, there is N, such that $|x_n - l| < \epsilon$ for all $n \ge N$.

4. For any $\epsilon > 0$, there is a natural number N, such that $|x_n - l| \le \epsilon$ for all $n \ge N$.

5. For any $\epsilon > 0$, there is N, such that $|x_n - l| < 2\epsilon^2$ for all $n > N$.

6. For any $\epsilon > 0$, there is N, such that $|x_n - l| < 2\epsilon^2 + 1$ for all $n > N$.

7. For any $\epsilon > 0$, we have $N = 1000$, such that $|x_n - l| < \epsilon$ for all $n > N$.

8. For any $\epsilon > 0$, there are infinitely many n, such that $|x_n - l| < \epsilon$.

9. For infinitely many $\epsilon > 0$, there is N, such that $|x_n - l| < \epsilon$ for all $n > N$.

10. For any $\epsilon > 0$, there is N, such that $l - 2\epsilon < x_n < l + \epsilon$ for all $n > N$.

11. For any natural number K, there is N, such that $|x_n - l| < \dfrac{1}{K}$ for all $n > N$.

The following examples are the most important basic limits. For any given ϵ, the analysis leading to the suitable choice of N will be given. It is left to the reader to write down the rigorous formal argument in line with the definition of limit.

Example 1.1.4. We have
$$\lim_{n\to\infty} \frac{1}{n^a} = 0 \text{ for } a > 0. \tag{1.1.2}$$

The inequality $\left|\dfrac{1}{n^a}\right| < \epsilon$ is the same as $\dfrac{1}{n} < \epsilon^{\frac{1}{a}}$. Thus choosing $N = \epsilon^{-\frac{1}{a}}$ should make the implication (1.1.1) hold.

Example 1.1.5. We have
$$\lim_{n\to\infty} a^n = 0 \text{ for } |a| < 1. \tag{1.1.3}$$

Let $\dfrac{1}{|a|} = 1 + b$. Then $b > 0$ and
$$\frac{1}{|a^n|} = (1+b)^n = 1 + nb + \frac{n(n-1)}{2}b^2 + \cdots > nb.$$

This implies $|a^n| < \dfrac{1}{nb}$. In order to get $|a^n| < \epsilon$, therefore, it suffices to make sure $\dfrac{1}{nb} < \epsilon$. This suggests us to choose $N = \dfrac{1}{b\epsilon}$.

Example 1.1.6. We have
$$\lim_{n\to\infty} \sqrt[n]{n} = 1. \tag{1.1.4}$$

Let $x_n = \sqrt[n]{n} - 1$. Then $x_n > 0$ and
$$n = (1 + x_n)^n = 1 + nx_n + \frac{n(n-1)}{2}x_n^2 + \cdots > \frac{n(n-1)}{2}x_n^2.$$

This implies $x_n^2 < \dfrac{2}{n-1}$. In order to get $|\sqrt[n]{n} - 1| = x_n < \epsilon$, therefore, it suffices to make sure $\dfrac{2}{n-1} < \epsilon^2$. Thus we may choose $N = \dfrac{2}{\epsilon^2} + 1$.

Example 1.1.7. For any a, we have
$$\lim_{n\to\infty} \frac{a^n}{n!} = 0. \tag{1.1.5}$$

Fix an integer $M > |a|$. Then for $n > M$, we have
$$\left|\frac{a^n}{n!}\right| = \frac{|a|^M}{M!}\frac{|a|}{M+1}\frac{|a|}{M+2}\cdots\frac{|a|}{n-1}\frac{|a|}{n} \leq \frac{|a|^M}{M!}\frac{|a|}{n}.$$

Thus in order to get $\left|\dfrac{a^n}{n!}\right| < \epsilon$, we only need to make sure $\dfrac{|a|^M}{M!}\dfrac{|a|}{n} < \epsilon$. This leads to the choice $N = \max\left\{M, \dfrac{|a|^{M+1}}{M!\epsilon}\right\}$.

Exercise 1.1.11. Prove $\dfrac{n!}{n^n} < \dfrac{1}{n}$ and $\dfrac{(n!)^2}{(2n)!} < \dfrac{1}{n+1}$ for $n > 1$. Then use this to prove $\lim_{n\to\infty} \dfrac{n!}{n^n} = \lim_{n\to\infty} \dfrac{(n!)^2}{(2n)!} = 0$.

Exercise 1.1.12. Use the binary expansion of $2^n = (1+1)^n$ to prove $2^n > \dfrac{n(n-1)}{2}$. Then prove $\lim_{n\to\infty} \dfrac{n}{2^n} = 0$.

Exercise 1.1.13. Prove $\lim_{n\to\infty} \sqrt[n]{3} = 1$ and $\lim_{n\to\infty} \sqrt[n]{n^2 + 2n + 3} = 1$.

Exercise 1.1.14. Prove $\sqrt[n]{n!} > \sqrt{\dfrac{n}{2}}$. Then use this to prove $\lim_{n\to\infty} \dfrac{1}{\sqrt[n]{n!}} = 0$.

1.1.2 Property

A sequence is *bounded* if there is a constant B, such that $|x_n| \leq B$ for all n. This is equivalent to the existence of constants B_1 and B_2, such that $B_1 \leq x_n \leq B_2$ for any n. The constants B, B_1, B_2 are respectively called a *bound*, a *lower bound* and an *upper bound*.

Proposition 1.1.2. *Convergent sequences are bounded.*

Proof. Suppose $\lim_{n\to\infty} x_n = l$. For $\epsilon = 1 > 0$, there is N, such that

$$n > N \implies |x_n - l| < 1.$$

Moreover, by taking a bigger natural number if necessary, we may further assume N is a natural number. Then $x_{N+1}, x_{N+2}, \ldots$, have upper bound $l + 1$ and lower bound $l - 1$, and the whole sequence has upper bound $\max\{x_1, x_2, \ldots, x_N, l+1\}$ and lower bound $\min\{x_1, x_2, \ldots, x_N, l-1\}$.
□

Exercise 1.1.15. Prove that if $|x_n| < B$ for $n > N$, then the whole sequence $\{x_n\}$ is bounded. This implies that the boundedness is not changed by modifying finitely many terms in a sequence.

Exercise 1.1.16. Suppose $\lim_{n\to\infty} x_n = 0$ and y_n is bounded. Prove $\lim_{n\to\infty} x_n y_n = 0$.

Proposition 1.1.3 (Arithmetic Rule). *Suppose*

$$\lim_{n\to\infty} x_n = l, \quad \lim_{n\to\infty} y_n = k.$$

Then

$$\lim_{n\to\infty} (x_n + y_n) = l + k, \quad \lim_{n\to\infty} x_n y_n = lk, \quad \lim_{n\to\infty} \frac{x_n}{y_n} = \frac{l}{k},$$

where $y_n \neq 0$ and $k \neq 0$ are assumed in the third equality.

Proof. For any $\epsilon > 0$, there are N_1 and N_2, such that

$$n > N_1 \implies |x_n - l| < \frac{\epsilon}{2},$$
$$n > N_2 \implies |y_n - k| < \frac{\epsilon}{2}.$$

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

$$|(x_n + y_n) - (l + k)| \leq |x_n - l| + |y_n - k| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

This completes the proof that $\lim_{n\to\infty}(x_n + y_n) = l + k$.

By Proposition 1.1.2, we have $|y_n| < B$ for a fixed number B and all n. For any $\epsilon > 0$, there are N_1 and N_2, such that

$$n > N_1 \implies |x_n - l| < \frac{\epsilon}{2B},$$
$$n > N_2 \implies |y_n - k| < \frac{\epsilon}{2|l|}.$$

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

$$|x_n y_n - lk| = |(x_n y_n - ly_n) + (ly_n - lk)|$$
$$\leq |x_n - l||y_n| + |l||y_n - k| < \frac{\epsilon}{2B}B + |l|\frac{\epsilon}{2|l|} = \epsilon.$$

This completes the proof that $\lim_{n\to\infty} x_n y_n = lk$.

Assume $y_n \neq 0$ and $k \neq 0$. We will prove $\lim_{n\to\infty} \frac{1}{y_n} = \frac{1}{k}$. By the product property of the limit, this implies

$$\lim_{n\to\infty} \frac{x_n}{y_n} = \lim_{n\to\infty} x_n \lim_{n\to\infty} \frac{1}{y_n} = l\frac{1}{k} = \frac{l}{k}.$$

For any $\epsilon > 0$, we have $\epsilon' = \min\left\{\frac{\epsilon|k|^2}{2}, \frac{|k|}{2}\right\} > 0$. Then there is N, such that

$$n > N \implies |y_n - k| < \epsilon'$$
$$\iff |y_n - k| < \frac{\epsilon|k|^2}{2}, \quad |y_n - k| < \frac{|k|}{2}$$
$$\implies |y_n - k| < \frac{\epsilon|k|^2}{2}, \quad |y_n| > \frac{|k|}{2}$$
$$\implies \left|\frac{1}{y_n} - \frac{1}{k}\right| = \frac{|y_n - k|}{|y_n k|} < \frac{\frac{\epsilon|k|^2}{2}}{\frac{|k|}{2}|k|} = \epsilon.$$

This completes the proof that $\lim_{n\to\infty} \frac{1}{y_n} = \frac{1}{k}$.

$\square$

Example 1.1.8. By the limit (1.1.2) and the arithmetic rule, we have

$$\lim_{n\to\infty} \frac{2n}{n-2} = \lim_{n\to\infty} \frac{2}{1-\frac{1}{n}} = \frac{\lim_{n\to\infty} 2}{\lim_{n\to\infty} 1 - \lim_{n\to\infty} \frac{1}{n}} = \frac{2}{1-0} = 2.$$

Here is a more complicated example

$$\lim_{n\to\infty} \frac{n^3 + 2n + 2}{2n^3 + 10n^2 + 1} = \lim_{n\to\infty} \frac{1 + 2\frac{1}{n^2} + 2\frac{1}{n^3}}{2 + 10\frac{1}{n} + \frac{1}{n^3}}$$
$$= \frac{1 + 2\left(\lim_{n\to\infty}\frac{1}{n}\right)^2 + 2\left(\lim_{n\to\infty}\frac{1}{n}\right)^3}{2 + 10\left(\lim_{n\to\infty}\frac{1}{n}\right) + \left(\lim_{n\to\infty}\frac{1}{n}\right)^3}$$
$$= \frac{1 + 2\cdot 0^2 + 2\cdot 0^3}{2 + 10\cdot 0 + 0^3} = \frac{1}{2}.$$

The idea can be generalized to obtain

$$\lim_{n\to\infty} \frac{a_p n^p + a_{p-1}n^{p-1} + \cdots + a_1 n + a_0}{b_q n^q + b_{q-1}n^{q-1} + \cdots + b_1 n + b_0} = \begin{cases} 0 & \text{if } p < q \text{ and } b_q \neq 0 \\ \frac{a_p}{b_q} & \text{if } p = q \text{ and } b_q \neq 0 \end{cases}. \quad (1.1.6)$$

Exercise 1.1.17. Suppose $\lim_{n\to\infty} x_n = l$ and $\lim_{n\to\infty} y_n = k$. Prove $\lim_{n\to\infty} \max\{x_n, y_n\} = \max\{l, k\}$ and $\lim_{n\to\infty} \min\{x_n, y_n\} = \min\{l, k\}$. You may use the formula $\max\{x, y\} = \frac{1}{2}(x + y + |x - y|)$ and the similar one for $\min\{x, y\}$.

Proposition 1.1.4 (Order Rule). *Suppose both $\{x_n\}$ and $\{y_n\}$ converge.*

1. *If $x_n \geq y_n$ for big n, then $\lim_{n\to\infty} x_n \geq \lim_{n\to\infty} y_n$.*

2. *If $\lim_{n\to\infty} x_n > \lim_{n\to\infty} y_n$, then $x_n > y_n$ for big n.*

A special case of the property is that $x_n \leq l$ implies $\lim_{n\to\infty} x_n \leq l$, and $\lim_{n\to\infty} x_n < l$ implies $x_n < l$ for sufficiently big n.

Proof. We prove the second statement first. Suppose $\lim_{n\to\infty} x_n > \lim_{n\to\infty} y_n$. Then by Proposition 1.1.3, $\lim_{n\to\infty}(x_n - y_n) = \lim_{n\to\infty} x_n - \lim_{n\to\infty} y_n > 0$. For $\epsilon = \lim_{n\to\infty}(x_n - y_n) > 0$, there is N, such that

$$n > N \implies |(x_n - y_n) - \epsilon| < \epsilon \implies x_n - y_n > \epsilon - \epsilon = 0 \iff x_n > y_n.$$

By exchanging x_n and y_n in the second statement, we find that

$$\lim_{n\to\infty} x_n < \lim_{n\to\infty} y_n \implies x_n < y_n \text{ for big } n.$$

This further implies that we cannot have $x_n \geq y_n$ for big n. The combined implication

$$\lim_{n\to\infty} x_n < \lim_{n\to\infty} y_n \implies \text{opposite of } (x_n \geq y_n \text{ for big } n)$$

is equivalent to the first statement.

$\square$

In the second part of the proof above, we used the logical fact that "$A \implies B$" is the same as "(not B) $\implies$ (not A)". Moreover, we note that the following two statements are not opposite of each other.

1. There is N, such that $x_n < y_n$ for $n > N$.

2. There is N, such that $x_n \geq y_n$ for $n > N$.

Therefore although the proof above showed that the first statement is a consequence of the second, the two statements are not logically equivalent.

Proposition 1.1.5 (Sandwich Rule). *Suppose*

$$x_n \leq y_n \leq z_n, \quad \lim_{n\to\infty} x_n = \lim_{n\to\infty} z_n = l.$$

Then

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

Proof. For any $\epsilon > 0$, there are N_1 and N_2, such that

$$n > N_1 \implies |x_n - l| < \epsilon,$$
$$n > N_2 \implies |z_n - l| < \epsilon.$$

Then

$$n > \max\{N_1, N_2\} \implies -\epsilon < x_n - l \leq y_n - l \leq z_n - l < \epsilon \implies |y_n - l| < \epsilon.$$

$\square$

Example 1.1.9. To find the limit of the sequence $\left\{\dfrac{\cos n}{n}\right\}$, we compare it with the sequence $\left\{\dfrac{1}{n}\right\}$ in Example 1.1.1. Since $-\dfrac{1}{n} \leq \dfrac{\cos n}{n} \leq \dfrac{1}{n}$ and $\lim_{n\to\infty} \dfrac{1}{n} = \lim_{n\to\infty} -\dfrac{1}{n} = 0$, we get $\lim_{n\to\infty} \dfrac{\cos n}{n} = 0$.

By similar reason, we get $\lim_{n\to\infty} \dfrac{\sin n}{n} = 0$ and $\lim_{n\to\infty} \dfrac{(-1)^n}{n} = 0$. Then by the arithmetic rule,

$$\lim_{n\to\infty} \frac{n - \cos n}{n^2 + (-1)^n \sin n} = \lim_{n\to\infty} \frac{1}{n} \frac{1 - \dfrac{\cos n}{n}}{1 + \dfrac{\sin n}{n} \dfrac{(-1)^n}{n}}$$

$$= \left(\lim_{n\to\infty} \frac{1}{n}\right) \frac{1 - \left(\lim_{n\to\infty} \dfrac{\cos n}{n}\right)}{1 + \left(\lim_{n\to\infty} \dfrac{\sin n}{n}\right)\left(\lim_{n\to\infty} \dfrac{(-1)^n}{n}\right)}$$

$$= 0 \frac{1 - 0}{1 + 0 \cdot 0} = 0.$$

Example 1.1.10. Suppose $\{x_n\}$ is a sequence satisfying $|x_n - l| < \dfrac{1}{n}$. Then we have $l - \dfrac{1}{n} < x_n < l + \dfrac{1}{n}$. Since $\lim_{n\to\infty}\left(l - \dfrac{1}{n}\right) = \lim_{n\to\infty}\left(l + \dfrac{1}{n}\right) = l$, by the sandwich rule, we get $\lim_{n\to\infty} x_n = l$.

Example 1.1.11. For any $a > 1$ and $n > a$, we have $1 < \sqrt[n]{a} < \sqrt[n]{n}$. Thus by the limit (1.1.4) and the sandwich rule, we have $\lim_{n\to\infty} \sqrt[n]{a} = 1$. On the other hand, for $0 < a < 1$, we have $b = \dfrac{1}{a} > 1$ and

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

Combining all the cases, we get $\lim_{n\to\infty} \sqrt[n]{a} = 1$ for any $a > 0$. Furthermore, we have
$$1 < (n + a)^{\frac{1}{n+b}} < (2n)^{\frac{2}{n}}, \quad 1 < (n^2 + an + b)^{\frac{1}{n+c}} < (2n^2)^{\frac{2}{n}}$$

for sufficiently big n. By

$$\lim_{n\to\infty} (2n)^{\frac{2}{n}} = \left(\lim_{n\to\infty} \sqrt[n]{2} \ \lim_{n\to\infty} \sqrt[n]{n}\right)^2 = (1 \cdot 1)^2 = 1,$$
$$\lim_{n\to\infty} (2n^2)^{\frac{2}{n}} = \left(\lim_{n\to\infty} \sqrt[n]{2}\right)^2 \left(\lim_{n\to\infty} \sqrt[n]{n}\right)^4 = 1^2 \cdot 1^4 = 1,$$

and the sandwich rule, we get $\lim_{n\to\infty}(n+a)^{\frac{1}{n+b}} = \lim_{n\to\infty}(n^2 + an + b)^{\frac{1}{n+c}} = 1$. The same idea leads to the limit

$$\lim_{n\to\infty}(a_p n^p + a_{p-1} n^{p-1} + \cdots + a_1 n + a_0)^{\frac{1}{n+a}} = 1. \qquad (1.1.7)$$

Example 1.1.12. Consider $\lim_{n\to\infty}\dfrac{n^p}{a^n}$ for $|a| > 1$ and any p. The special case $p = 0$ is the limit (1.1.3), and the special case $p = 1$, $a = 2$ is Exercise 1.1.12.

Let $|a| = 1 + b$. Since $|a| > 1$, we have $b > 0$. Fix a natural number $P \geq p$. Then for $n > P$,

$$|a|^n = 1 + nb + \frac{n(n-1)}{2}b^2 + \cdots + \frac{n(n-1)\cdots(n-P)}{(P+1)!}b^{P+1} + \cdots$$
$$> \frac{n(n-1)\cdots(n-P)}{(P+1)!}b^{P+1}.$$

Thus

$$0 < \left|\frac{n^p}{a^n}\right| \leq \frac{n^P}{|a|^n} < \frac{(P+1)!}{n(n-1)\cdots(n-P)}\frac{n^P}{b^{P+1}}$$
$$= \frac{1}{n}\frac{n}{(n-1)}\frac{n}{(n-2)}\cdots\frac{n}{(n-P)}\frac{(P+1)!}{b^{P+1}}.$$

By $\lim_{n\to\infty}\dfrac{1}{n} = 0$, $\lim_{n\to\infty}\dfrac{n}{n-k} = 1$, the fact that P and b are fixed constants, and the arithmetic rule, the right side has limit 0 as $n \to \infty$. By the sandwich rule, we conclude that

$$\lim_{n\to\infty}\frac{n^p}{a^n} = 0 \text{ for } |a| > 1 \text{ and any } p. \qquad (1.1.8)$$

Exercise 1.1.18. Redo Exercise 1.1.4 by using the sandwich rule.

Exercise 1.1.19. Let $a > 0$ be a constant. Then $\dfrac{1}{n} < a < n$ for big n. Use this and the limit (1.1.4) to prove $\lim_{n\to\infty}\sqrt[n]{a} = 1$.

Exercise 1.1.20. Compute the limits.

1. $\lim_{n\to\infty}\dfrac{2n^{7/4} - 3n^{3/2}}{(3n^{3/4} - n^{1/2} + 1)(n+2)}$.

2. $\lim_{n\to\infty}(\sqrt{n^2 + n} - n)$.

3. $\lim_{n\to\infty}\dfrac{3^{n-1} - 8\cdot 7^n + (-1)^{n+1}}{8^{n-1} + 2(n+1)(-5)^n}$.

4. $\lim_{n\to\infty}\dfrac{(-1)^n n + 1}{\sqrt{n^3} + (-1)^n}$.

5. $\lim_{n\to\infty}\dfrac{n! + 10^n}{n^{10} + n^n}$.

6. $\lim_{n\to\infty}\dfrac{(n^2 + n + 3)(n!2^n + 5^n)}{(n+2)!2^n + 5^n}$.

7. $\lim_{n\to\infty}\sqrt[n]{\sqrt{n} + \cos n} + \sin n$.

8. $\lim_{n\to\infty}\sqrt[n]{11\cdot 2^n + 2\cdot 5^n + 5\cdot 11^n}$.

9. $\lim_{n\to\infty}\sqrt[n]{11^n n - 5^n(n^2 + 1)}$.

10. $\lim_{n\to\infty}\sqrt[n]{1 + \sqrt[n]{2 + \sqrt[n]{3 + n}}}$.

11. $\lim_{n\to\infty}\sqrt[n]{\dfrac{1\cdot 3\cdot 5\cdots(2n-1)}{2\cdot 4\cdot 5\cdots 2n}}$.

Exercise 1.1.21. Compute the limits.

1. $\lim_{n\to\infty}\dfrac{a^n}{a^n + 1}$, where $a \neq -1$.

2. $\lim_{n\to\infty}(n + a)^{\frac{n}{n^2+bn+c}}$.

3. $\lim_{n\to\infty} \sqrt[n]{a^n + b^n + c^n}$, where $a, b, c > 0$.

4. $\lim_{n\to\infty} \left(\dfrac{1}{\sqrt{1 + n^2}} + \dfrac{1}{\sqrt{2 + n^2}} + \cdots + \dfrac{1}{\sqrt{n + n^2}} \right)$.

1.1.3   Infinity and Infinitesimal

A changing numerical quantity is an *infinity* if it tends to get arbitrarily big. For sequences, this means the following.

Definition 1.1.6. A sequence $\{x_n\}$ diverges to *infinity*, denoted $\lim_{n\to\infty} x_n = \infty$, if for any b, there is N, such that

$$n > N \implies |x_n| > b. \tag{1.1.9}$$

It diverges to *positive infinity*, denoted $\lim_{n\to\infty} x_n = +\infty$, if for any b, there is N, such that

$$n > N \implies x_n > b. \tag{1.1.10}$$

It diverges to *negative infinity*, denoted $\lim_{n\to\infty} x_n = -\infty$, if for any b, there is N, such that

$$n > N \implies x_n < b. \tag{1.1.11}$$

Example 1.1.13. We rigorously verify $\lim_{n\to\infty} \dfrac{n^2}{n + (-1)^n} = +\infty$. For any $b > 0$, choose $N = 2b$. Then

$$n > N \implies \frac{n^2}{n + (-1)^n} \geq \frac{n^2}{n + 1} > \frac{n^2}{2n} > \frac{N}{2} = b.$$

Exercise 1.1.22. Rigorously verify the divergence to infinity.

1. $\lim_{n\to\infty}(100 + 10n - n^2) = -\infty$.

2. $\lim_{n\to\infty} \dfrac{(-1)^n \sqrt{n}}{2 + \sin n} = \infty$.

3. $\lim_{n\to\infty} \sqrt{n + \sqrt{(n - 1) + \cdots + \sqrt{2 + \sqrt{1}}}} = +\infty$.

Exercise 1.1.23. Infinities must be unbounded. Is the converse true?

Exercise 1.1.24. Suppose $\lim_{n\to\infty} x_n = +\infty$ and $\lim_{n\to\infty} y_n = +\infty$. Prove $\lim_{n\to\infty}(x_n + y_n) = +\infty$ and $\lim_{n\to\infty} x_n y_n = +\infty$.

Exercise 1.1.25. Suppose $\lim_{n\to\infty} x_n = \infty$ and $|x_n - x_{n+1}| < c$ for some constant c.

1. Prove that either $\lim_{n\to\infty} x_n = +\infty$ or $\lim_{n\to\infty} x_n = -\infty$.

2. If we further know $\lim_{n\to\infty} x_n = +\infty$, prove that for any $a > x_1$, some term x_n lies in the interval $(a, a + c)$.

A changing numerical quantity is an *infinitesimal* if it tends to get arbitrarily small. For sequences, this means that for any $\epsilon > 0$, there is N, such that

$$n > N \implies |x_n| < \epsilon. \tag{1.1.12}$$

This simply means $\lim_{n \to \infty} x_n = 0$.

Note that the implications (1.1.9) and (1.1.12) are equivalent by changing x_n to $\dfrac{1}{x_n}$ and taking $\epsilon = \dfrac{1}{b}$. Therefore we have

$$\{x_n\} \text{ is an infinity} \iff \left\{\frac{1}{x_n}\right\} \text{ is an infinitesimal.}$$

For example, the infinitesimals (1.1.2), (1.1.3), (1.1.5), (1.1.8) tell us that $\{n^a\}$ (for $a > 0$), $\{a^n\}$ (for $|a| > 1$), $\left\{\dfrac{n!}{a^n}\right\}$, and $\left\{\dfrac{a^n}{n^p}\right\}$ (for $|a| > 1$) are infinities. Moreover, the first case in the limit (1.1.6) tells us

$$\lim_{n \to \infty} \frac{a_p n^p + a_{p-1} n^{p-1} + \cdots + a_1 n + a_0}{b_q n^q + b_{q-1} n^{q-1} + \cdots + b_1 n + b_0} = \infty \quad \text{if } p > q \text{ and } a_p \neq 0.$$

On the other hand, since $\lim_{n \to \infty} x_n = l$ is equivalent to $\lim_{n \to \infty}(x_n - l) = 0$, we have

$$\{x_n\} \text{ converges to } l \iff \{x_n - l\} \text{ is an infinitesimal.}$$

For example, the limit (1.1.4) tells us that $\{\sqrt[n]{n} - 1\}$ is an infinitesimal.

Exercise 1.1.26. How to characterize a positive infinity $\{x_n\}$ in terms of the infinitesimal $\left\{\dfrac{1}{x_n}\right\}$?

Exercise 1.1.27. Explain the infinities.

1. $\lim_{n \to \infty} \dfrac{n!}{a^n} = \infty$ for any $a \neq 0$.

2. $\lim_{n \to \infty} \dfrac{n!}{a^n + b^n} = \infty$ if $a + b \neq 0$.

3. $\lim_{n \to \infty} \dfrac{1}{\sqrt[n]{n} - 1} = +\infty$.

4. $\lim_{n \to \infty} \dfrac{1}{\sqrt[n]{n} - \sqrt[n]{2n}} = -\infty$.

Some properties of finite limits can be extended to infinities and infinitesimals. For example, if $\lim_{n \to \infty} x_n = +\infty$ and $\lim_{n \to \infty} y_n = +\infty$, then $\lim_{n \to \infty}(x_n + y_n) = +\infty$. The property can be denoted as the arithmetic rule $(+\infty) + (+\infty) = +\infty$. Moreover, if $\lim_{n \to \infty} x_n = 1$, $\lim_{n \to \infty} y_n = 0$, and $y_n < 0$ for big n, then $\lim_{n \to \infty} \dfrac{x_n}{y_n} = -\infty$. Thus we have another arithmetic rule $\dfrac{1}{0^-} = -\infty$. Common sense suggests more arithmetic rules such as

$$c + \infty = \infty, \quad c \cdot \infty = \infty (\text{for } c \neq 0), \quad \infty \cdot \infty = \infty, \quad \frac{c}{0} = \infty (\text{for } c \neq 0), \quad \frac{c}{\infty} = 0,$$

where c is a finite number and represents a sequence convergent to c.

We must be careful in applying arithmetic rules involving infinities and infinitesimals. For example, we have

$$\lim_{n \to \infty} n^{-1} = 0, \quad \lim_{n \to \infty} 2n^{-1} = 0, \quad \lim_{n \to \infty} n^{-2} = 0,$$

$$\lim_{n \to \infty} \frac{n^{-1}}{2n^{-1}} = \frac{1}{2}, \quad \lim_{n \to \infty} \frac{n^{-1}}{n^{-2}} = +\infty, \quad \lim_{n \to \infty} \frac{n^{-2}}{2n^{-1}} = 0.$$

This shows that $\dfrac{0}{0}$ has no definite value.

Example 1.1.14. By the (extended) arithmetic rule, we have

$$\lim_{n \to \infty} (n^2 + 3)(-2)^n = \lim_{n \to \infty} (n^2 + 3) \lim_{n \to \infty} (-2)^n = \infty \cdot \infty = \infty.$$

$$\lim_{n \to \infty} \left(n + \frac{1}{n} \right) = \lim_{n \to \infty} n + \lim_{n \to \infty} \frac{1}{n} = (+\infty) + 0 = +\infty.$$

$$\lim_{n \to \infty} \frac{\sqrt[n]{n}}{\sqrt[n]{n} - 1} = \frac{\lim_{n \to \infty} \sqrt[n]{n}}{\lim_{n \to \infty}(\sqrt[n]{n} - 1)} = \lim_{n \to \infty} \frac{1}{0^+} = +\infty.$$

Exercise 1.1.28. Prove the properties of infinities.

1. (bounded)$+\infty = \infty$: If $\{x_n\}$ is bounded and $\lim_{n \to \infty} y_n = \infty$, then $\lim_{n \to \infty}(x_n + y_n) = \infty$.

2. $\min\{+\infty, +\infty\} = +\infty$: If $\lim_{n \to \infty} x_n = \lim_{n \to \infty} y_n = +\infty$, then $\lim_{n \to \infty} \min\{x_n, y_n\} = +\infty$.

3. Sandwich rule: If $x_n \ge y_n$ and $\lim_{n \to \infty} y_n = +\infty$, then $\lim_{n \to \infty} x_n = +\infty$.

4. $(> c > 0) \cdot (+\infty) = +\infty$: If $x_n > c$ for some constant $c > 0$ and $\lim_{n \to \infty} y_n = +\infty$, then $\lim_{n \to \infty} x_n y_n = +\infty$.

Exercise 1.1.29. Show that it is not necessarily true that $\infty + \infty = \infty$ by constructing examples of sequences $\{x_n\}$ and $\{y_n\}$ that diverge to ∞ but one of the following holds.

1. $\lim_{n \to \infty}(x_n + y_n) = 2$.

2. $\lim_{n \to \infty}(x_n + y_n) = +\infty$.

3. $\{x_n + y_n\}$ is bounded and divergent.

Exercise 1.1.30. Show that one cannot make a definite conclusion on $0 \cdot \infty$ by constructing examples of sequences $\{x_n\}$ and $\{y_n\}$, such that $\lim_{n \to \infty} x_n = 0$ and $\lim_{n \to \infty} y_n = \infty$ but one of the following holds.

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

2. $\lim_{n \to \infty} x_n y_n = 0$.

3. $\lim_{n \to \infty} x_n y_n = \infty$.

4. $\{x_n y_n\}$ is bounded and divergent.

Exercise 1.1.31. Provide counterexamples to the wrong extensions of the arithmetic rules.

$$\frac{+\infty}{+\infty} = 1, \quad (+\infty) - (+\infty) = 0, \quad 0 \cdot \infty = 0, \quad 0 \cdot \infty = \infty, \quad 0 \cdot \infty = 1.$$

1.1.4   Additional Exercise

Ratio Rule

Exercise 1.1.32. Suppose $\left|\dfrac{x_{n+1}}{x_n}\right| \leq \left|\dfrac{y_{n+1}}{y_n}\right|$.

1. Prove that $|x_n| \leq c|y_n|$ for some constant c.

2. Prove that $\lim_{n\to\infty} y_n = 0$ implies $\lim_{n\to\infty} x_n = 0$.

3. Prove that $\lim_{n\to\infty} x_n = \infty$ implies $\lim_{n\to\infty} y_n = \infty$.

Exercise 1.1.33. Suppose $\lim_{n\to\infty} \dfrac{x_{n+1}}{x_n} = l$. What can you say about $\lim_{n\to\infty} x_n$ by looking at the value of l?

Exercise 1.1.34. Use the ratio rule to study the limits (1.1.8) and $\lim_{n\to\infty} \dfrac{(n!)^2 a^n}{(2n)!}$.

Power Rule

Exercise 1.1.35. Suppose $\lim_{n\to\infty} x_n = l > 0$. Prove $\lim_{n\to\infty} x_n^\alpha = l^\alpha$ by the following steps.

1. Assume $x_n \geq 1$ and $l = 1$. By using the sandwich rule and the fact that $\lim_{n\to\infty} x_n^a = 1$ for any integer a, prove that $\lim_{n\to\infty} x_n^\alpha = 1$ for any number α. The same argument applies to the case $x_n \leq 1$.

2. Use $\min\{x_n, 1\} \leq x_n \leq \max\{x_n, 1\}$, Exercise 1.1.17 and the sandwich rule to remove the assumption $x_n \geq 1$ in the first part.

3. Use the arithmetic rule to prove the limit for general l.

Average Rule

Exercise 1.1.36. Suppose $\lim_{n\to\infty} x_n = l$. Let $y_n = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$.

1. Prove that if $|x_n - l| < \epsilon$ for $n > N$, where N is a natural number, then

$$n > N \implies |y_n - l| < \frac{|x_1| + |x_2| + \cdots + |x_N| + N|l|}{n} + \epsilon.$$

2. Use the first part and Proposition 1.1.2 to prove $\lim_{n\to\infty} y_n = l$.

3. What happens if $l = +\infty$ or ∞?

Exercise 1.1.37. Find suitable condition on a sequence $\{a_n\}$ of positive numbers, such that $\lim_{n\to\infty} x_n = l$ implies $\lim_{n\to\infty} \dfrac{a_1 x_1 + a_2 x_2 + \cdots + a_n x_n}{a_1 + a_2 + \cdots + a_n} = l$.

1.2   Convergence of Sequence Limit

The discussion of convergent sequences in Section 1.1 is based on the explicit value of the limit. However, there are many cases that a sequence must be convergent, but the value of the limit is not known. The limit of the world record in 100 meter dash is one such example. In such cases, the existence of the limit cannot be established by using the definition alone. A more fundamental theory is needed.

1.2.1   Necessary Condition

A *subsequence* of a sequence $\{x_n\}$ is obtained by selecting some terms from the sequence. The indices of the selected terms can be arranged as a *strictly increasing* sequence $n_1 < n_2 < \cdots < n_k < \cdots$, and the subsequence can be denoted as $\{x_{n_k}\}$. The following are two examples of subsequences.

$$\{x_{3k}\}: x_3,\ x_6,\ x_9,\ x_{12},\ x_{15},\ x_{18},\ \dots$$
$$\{x_{2^k}\}: x_2,\ x_4,\ x_8,\ x_{16},\ x_{32},\ x_{64},\ \dots$$

Note that if $\{x_n\}$ starts from $n = 1$, then $n_k \geq k$. Thus by reindexing the terms if necessary, we will always assume $n_k \geq k$ in subsequent proofs.

Proposition 1.2.1. *Suppose a sequence converges to l. Then all its subsequences converge to l.*

Proof. Suppose $\lim_{n\to\infty} x_n = l$. For any $\epsilon > 0$, there is N, such that $n > N$ implies $|x_n - l| < \epsilon$. Then

$$k > N \implies n_k \geq k > N \implies |x_{n_k} - l| < \epsilon.$$

$\square$

Example 1.2.1. By $\lim_{n\to\infty} \sqrt[n]{n} = 1$, we have $\lim_{k\to\infty} \sqrt[2k]{2k} = 1$. Taking the square of the limit, we get $\lim_{n\to\infty} \sqrt[n]{2n} = (\lim_{k\to\infty} \sqrt[2k]{2k})^2 = 1$.

Example 1.2.2. The sequence $\{(-1)^n\}$ has subsequences $\{(-1)^{2k}\} = \{1\}$ and $\{(-1)^{2k+1}\} = \{-1\}$. Since the two subsequences have different limits, the original sequence $\{(-1)^n\}$ diverges.

Exercise 1.2.1. Prove the sequences diverge.

1. $\dfrac{(-1)^n 2n + 1}{n + 2}$.

2. $\dfrac{(-1)^n 2n(n+1)}{(\sqrt{n}+2)^3}$.

3. $\sqrt{n}\left(\sqrt{n + (-1)^n} - \sqrt{n - (-1)^n}\right)$.

4. $\dfrac{n \sin \dfrac{n\pi}{3}}{n \cos \dfrac{n\pi}{2} + 2}$.

5. $\dfrac{(-1)^n n}{n + 1}$.

6. $x_{2n} = \dfrac{1}{n}$, $x_{2n+1} = \sqrt[n]{n}$.

7. $\sqrt[n]{2^n + 3^{(-1)^n n}}$.

8. $1 + n \sin \dfrac{n\pi}{2}$.

9. $\cos^n \dfrac{2n\pi}{3}$.

Exercise 1.2.2. Prove that $\lim_{n\to\infty} x_n = l$ if and only if $\lim_{k\to\infty} x_{2k} = l$ and $\lim_{k\to\infty} x_{2k+1} = l$.

Exercise 1.2.3. What is wrong with the following application of Propositions 1.1.3 and 1.2.1: The sequence $x_n = (-1)^n$ satisfies $x_{n+1} = -x_n$. Therefore $\lim_{n\to\infty} x_n = \lim_{n\to\infty} x_{n+1} = -\lim_{n\to\infty} x_n$, and we get $\lim_{n\to\infty} x_n = 0$.

Exercise 1.2.4. Prove that if a sequence diverges to infinity, then all its subsequences diverge to infinity.

Theorem 1.2.2 (Cauchy[1] Criterion). *Suppose a sequence $\{x_n\}$ converges. Then for any $\epsilon > 0$, there is N, such that*

$$m, n > N \implies |x_m - x_n| < \epsilon.$$

Proof. Suppose $\lim_{n \to \infty} x_n = l$. For any $\epsilon > 0$, there is N, such that $n > N$ implies $|x_n - l| < \dfrac{\epsilon}{2}$. Then $m, n > N$ implies

$$|x_m - x_n| = |(x_m - l) - (x_n - l)| \leq |x_m - l| + |x_n - l| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

$\square$

Cauchy criterion plays a critical role in analysis. Therefore the criterion is called a *theorem* instead of just a proposition. Moreover, the property described in the criterion is given a special name.

Definition 1.2.3. A sequence $\{x_n\}$ is called a *Cauchy sequence* if for any $\epsilon > 0$, there is N, such that

$$m, n > N \implies |x_m - x_n| < \epsilon. \tag{1.2.1}$$

Theorem 1.2.2 says that convergent sequences must be Cauchy sequences. The converse that any Cauchy sequence is convergent is also true and is one of the most fundamental results in analysis.

Example 1.2.3. Consider the sequence $\{(-1)^n\}$. For $\epsilon = 1 > 0$ and any N, we can find an even $n > N$. Then $m = n + 1 > N$ is odd and $|x_m - x_n| = 2 > \epsilon$. Therefore the Cauchy criterion fails and the sequence diverges.

Example 1.2.4 (Oresme[2]). The *harmonic sequence*

$$x_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}$$

satisfies

$$x_{2n} - x_n = \frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{2n} \geq \frac{1}{2n} + \frac{1}{2n} + \cdots + \frac{1}{2n} = \frac{n}{2n} = \frac{1}{2}.$$

Thus for $\epsilon = \dfrac{1}{2}$ and any N, we have $|x_m - x_n| > \dfrac{1}{2}$ by taking any natural number $n > N$ and $m = 2n$. Therefore the Cauchy criterion fails and the harmonic sequence diverges.

---

[1] Augustin Louis Cauchy, born 1789 in Paris (France), died 1857 in Sceaux (France). His contributions to mathematics can be seem by the numerous mathematical terms bearing his name, including Cauchy integral theorem (complex functions), Cauchy-Kovalevskaya theorem (differential equations), Cauchy-Riemann equations, Cauchy sequences. He produced 789 mathematics papers and his collected works were published in 27 volumes.

[2] Nicole Oresme, born 1323 in Allemagne (France), died 1382 in Lisieux (France). Oresme is best known as an economist, mathematician, and a physicist. He was one of the most famous and influential philosophers of the later Middle Ages. His contributions to mathematics were mainly contained in his manuscript *Tractatus de configuratione qualitatum et motuum* (Treatise on the Configuration of Qualities and Motions).

Example 1.2.5. We show the sequence $\{\sin n\}$ diverges. The real line is divided by $k\pi \pm \dfrac{\pi}{4}$ into intervals of length $\dfrac{\pi}{2} > 1$. Therefore for any integer k, there are integers m and n satisfying

$$2k\pi + \frac{\pi}{4} < m < 2k\pi + \frac{3\pi}{4}, \;\; 2k\pi - \frac{\pi}{4} > n > 2k\pi - \frac{3\pi}{4}.$$

Moreover, by taking k to be a big positive number, m and n can be as big as we wish. Then $\sin m > \dfrac{1}{\sqrt{2}}$, $\sin n < -\dfrac{1}{\sqrt{2}}$, and we have $|\sin m - \sin n| > \sqrt{2}$. Thus the sequence $\{\sin n\}$ is not Cauchy and must diverge.

Exercise 1.2.5. Prove the sequences diverge.

1. $x_n = 1 + \dfrac{1}{\sqrt{2}} + \dfrac{1}{\sqrt{3}} + \cdots + \dfrac{1}{\sqrt{n}}.$ 2. $x_n = \dfrac{1}{2} + \dfrac{2}{5} + \dfrac{3}{10} + \cdots + \dfrac{n}{n^2+1}.$

Exercise 1.2.6. Prove that if $\lim_{n\to\infty} x_n = +\infty$ and $|x_n - x_{n+1}| < c$ for some constant $c < \pi$, then $\{\sin x_n\}$ diverges. Exercise 1.1.25 might be helpful here.

1.2.2   Supremum and Infimum

To discuss the converse of Theorem 1.2.2, we have to consider the difference between rational and real numbers. Specifically, consider the converse of Cauchy criterion stated for the real and the rational number systems:

1. Real number Cauchy sequences always have *real* number limits.

2. Rational number Cauchy sequences always have *rational* number limits.

The key distinction here is that a sequence of rational numbers may have an irrational number as the limit. For example, the rational number sequence of the decimal approximations of $\sqrt{2}$ in Example 1.1.3 is a Cauchy sequence but has no rational number limit. This shows that the second statement is wrong.

   Therefore the truthfulness of the first statement is closely related to the fundamental question of the definition of real numbers. In other words, the establishment of the first property must also point to the key difference between the rational and the real number systems. One solution to the fundamental question is to simply use the converse of Cauchy criterion as the way of constructing real numbers from rational numbers (by requiring that all Cauchy sequences converge). This is the *topological* approach and can be dealt with in the larger context of the *completion of metric spaces*. Alternatively, real numbers can be constructed by considering the order among the numbers. The subsequent discussion will be based on this more intuitive approach, which is called the *Dedekind*[3] *cut*.

   The order relation between real numbers enables us to introduce the following concept.

---

[3]Julius Wilhelm Richard Dedekind, born 1831 and died 1916 in Braunschweig (Germany). Dedekind came up with the idea of the cut on November 24 of 1858 while thinking how to teach calculus. He made important contributions to algebraic number theory. His work introduced a new style of mathematics that influenced generations of mathematicians.

Definition 1.2.4. Let X be a nonempty set of numbers. An *upper bound* of X is a number B such that $x \leq B$ for any $x \in X$. The *supremum* of X is the least upper bound of the set and is denoted $\sup X$.

The supremum $\lambda = \sup X$ is characterized by the following properties.

1. λ is an upper bound: For any $x \in X$, we have $x \leq \lambda$.

2. Any number smaller than λ is not an upper bound: For any $\epsilon > 0$, there is $x \in X$, such that $x > \lambda - \epsilon$.

The *lower bound* and the *infimum* $\inf X$ can be similarly defined and characterized.

Example 1.2.6. Both the set $\{1, 2\}$ and the interval $[0, 2]$ have 2 as the supremum. In general, the *maximum* of a set X is a number $\xi \in X$ such that $\xi \geq x$ for any $x \in X$, and the maximum (if exists) is always the supremum.

On the other hand, the interval $(0, 2)$ has no maximum but still has 2 as the supremum. The similar discussion on *minimum* can be made.

Example 1.2.7. $\sqrt{2}$ is the supremum of the set

$$\{1.4, \ 1.41, \ 1.414, \ 1.4142, \ 1.41421, \ 1.414213, \ 1.4142135, \ 1.41421356, \ \dots\}$$

of its decimal expansions. It is also the supremum of the set

$$\left\{\frac{m}{n} : m \text{ and } n \text{ are natural numbers satisfying } m^2 < 2n^2\right\}$$

of positive rational numbers whose squares are less than 2.

Example 1.2.8. Let L_n be the length of an edge of the inscribed regular n-gon in a circle of radius 1. Then 2π is the supremum of the set $\{3L_3, 4L_4, 5L_5, \dots\}$ of the circumferences of the inscribed regular n-gons.

Exercise 1.2.7. Find the suprema and the infima.

1. $\{a + b : a, b \text{ are rational. } a^2 < 3, |2b + 1| < 5\}$.

2. $\left\{\dfrac{n}{n+1} : n \text{ is a natural number}\right\}$.

3. $\left\{\dfrac{(-1)^n n}{n+1} : n \text{ is a natural number}\right\}$.

4. $\left\{\dfrac{m}{n} : m \text{ and } n \text{ are natural numbers satisfying } m^2 > 3n^2\right\}$.

5. $\left\{\dfrac{1}{2^m} + \dfrac{1}{3^n} : m \text{ and } n \text{ are natural numbers}\right\}$.

6. $\{nR_n : n \geq 3 \text{ is a natural number}\}$, where R_n is the length of an edge of the circumscribed regular n-gon around a circle of radius 1.

Exercise 1.2.8. Prove the supremum is unique.

Exercise 1.2.9. Suppose X is a nonempty bounded set of numbers. Prove that $\lambda = \sup X$ is characterized by the following two properties.

1. λ is an upper bound: For any $x \in X$, we have $x \le \lambda$.

2. λ is the limit of a sequence in X: There are $x_n \in X$, such that $\lambda = \lim_{n \to \infty} x_n$.

The following are some properties of the supremum and infimum.

Proposition 1.2.5. *Suppose X and Y are nonempty bounded sets of numbers.*

1. *If $x \le y$ for any $x \in X$ and $y \in Y$, then $\sup X \le \inf Y$.*

2. *If $|x - y| \le c$ for any $x \in X$ and $y \in Y$, then $|\sup X - \sup Y| \le c$ and $|\sup X - \inf Y| \le c$.*

3. *If $X + Y = \{x + y \colon x \in X, y \in Y\}$, then $\sup(X + Y) = \sup X + \sup Y$ and $\inf(X + Y) = \inf X + \inf Y$.*

4. *If $cX = \{cx \colon x \in X\}$, then $\sup(cX) = c \sup X$ when $c > 0$ and $\sup(cX) = c \inf X$ when $c < 0$.*

5. *If $XY = \{xy \colon x \in X, y \in Y\}$ and all numbers in X, Y are positive, then $\sup(XY) = \sup X \sup Y$ and $\inf(XY) = \inf X \inf Y$.*

6. *If $X^{-1} = \{x^{-1} \colon x \in X\}$ and all numbers in X are positive, then $\sup X^{-1} = (\inf X)^{-1}$.*

Proof. In the first property, fix any $y \in Y$. Then y is an upper bound of X. Therefore $\sup X \le y$. Since $\sup X \le y$ for any $y \in Y$, $\sup X$ is a lower bound of Y. Therefore $\sup X \le \inf Y$.

Now consider the second property. For any $\epsilon > 0$, there are $x \in X$ and $y \in Y$, such that $\sup X - \epsilon < x \le \sup X$ and $\sup Y - \epsilon < y \le \sup Y$. Then

$$(\sup X - \epsilon) - \sup Y < x - y < \sup X - (\sup Y - \epsilon),$$

which means exactly $|(x - y) - (\sup X - \sup Y)| < \epsilon$. This further implies

$$|\sup X - \sup Y| < |x - y| + \epsilon \le c + \epsilon.$$

Since this holds for any $\epsilon > 0$, we conclude that $|\sup X - \sup Y| \le c$. The inequality $|\sup X - \inf Y| \le c$ can be similarly proved.

Next we prove the third property. For any $x \in X$ and $y \in Y$, we have $x + y \le \sup X + \sup Y$, so that $\sup X + \sup Y$ is an upper bound of $X + Y$. On the other hand, for any $\epsilon > 0$, there are $x, y \in X, Y$, such that $x > \sup X - \epsilon$ and $y > \sup Y - \epsilon$. Then $x + y \in X + Y$ satisfies $x + y > \sup X + \sup Y - 2\epsilon$. Since $\epsilon > 0$ is arbitrary, this shows that any number smaller than $\sup X + \sup Y$ is not an upper bound of $X + Y$. Therefore $\sup X + \sup Y$ is the supremum of $X + Y$.

The proof of the rest are left as exercises.

$\square$

Exercise 1.2.10. Finish the proof of Proposition 1.2.5.

Exercise 1.2.11. Suppose X_i are nonempty sets of numbers. Let $X = \cup_i X_i$ and $l_i = \sup X_i$. Prove that $\sup X = \sup_i l_i$.

The existence of the supremum is what distinguishes the real numbers from the rational numbers.

Definition 1.2.6. Real numbers is a set with the usual arithmetic operations and the order satisfying the usual properties, and the additional property that any bounded set of real numbers has the supremum.

The arithmetic operations are addition, subtraction, multiplication and division. An *order* on a set S is a relation $x < y$ defined for pairs $x, y \in S$, satisfying the following properties:

- *transitivity*: $x < y$ and $y < z \implies x < z$.

- *exclusivity*: If $x < y$, then $x \neq y$.

The two properties imply that $x < y$ and $y < x$ cannot happen at the same time. The following are some (but not all) of the usual arithmetic and order properties.

- *commutativity*: $a + b = b + a$, $ab = ba$.

- *distributivity*: $a(b + c) = ab + ac$.

- *unit*: There is a special number 1 such that $1a = a$.

- *order compatible with addition*: $a < b \implies a + c < b + c$.

- *order compatible with multiplication*: $a < b, 0 < c \implies ac < bc$.

Because of these properties, the real numbers form an *ordered field*. Since the rational numbers also has the arithmetic operations and the order satisfying these usual properties, the rational numbers also form an ordered field. Thus the key distinction between the real and rational numbers is the existence of the supremum. A bounded set of rational numbers may not have rational number supremum. A bounded set of real numbers always has real number supremum. Due to the existence of supremum, the real numbers form a *complete ordered field*.

Finally, we remark that the real numbers is the logical foundation of mathematical analysis. In this course, all the proof are logically derived from the properties (or axioms) about the arithmetic operations and the order (including the existence of the supremum). In fact, we will also use the exponential a^b for any positive real number a and any real number b. This means the existence and the usual properties of such exponentials. Some (but not all) properties are listed below.

- *zero*: $a^0 = 1$.

- *unit*: $a^1 = a$, $1^a = 1$.

- *exponential compatible with addition*: $a^{b+c} = a^b a^c$.

- *exponential compatible with multiplication*: $a^{bc} = (a^b)^c$, $(ab)^c = a^c b^c$.

- *exponential compatible with order*: $a > b$, $c > 0 \implies a^c > b^c$; and $a > 1$, $b > c \implies a^b > a^c$.

Note that the exponential and the associated properties are not assumptions added to the real numbers. They can be derived from the existing arithmetic operations and order relation.

1.2.3 Monotone Sequence

A sequence $\{x_n\}$ is *increasing* if $x_{n+1} \geq x_n$. It is *strictly increasing* if $x_{n+1} > x_n$. The concepts of *decreasing* and *strictly decreasing* sequences can be similarly defined. A sequence is *monotone* if it is either increasing or decreasing.

Proposition 1.2.7. *Bounded monotone sequences of real numbers are convergent. Unbounded monotone sequences of real numbers diverge to infinity.*

Since an increasing sequence $\{x_n\}$ satisfies $x_n \geq x_1$, the sequence has x_1 as a lower bound. Therefore it is bounded if and only if it has an upper bound, and the proposition says that an increasing sequence with upper bound must be convergent. Similar remarks can be made for decreasing sequences.

Proof. Let $\{x_n\}$ be a bounded increasing sequence. The sequence has a real number supremum $l = \sup\{x_n\}$. For any $\epsilon > 0$, by the second property characterizing the supremum, there is N, such that $x_N > l - \epsilon$. Then because the sequence is increasing, $n > N$ implies $x_n \geq x_N > l - \epsilon$. We also have $x_n \leq l$ because l is an upper bound. Thus we conclude that

$$n > N \implies l - \epsilon < x_n \leq l \implies |x_n - l| < \epsilon.$$

This proves that the sequence converges to l.

Let $\{x_n\}$ be an unbounded increasing sequence. Then it has no upper bound. In other words, for any b, there is N, such that $x_N > b$. Since the sequence is increasing, we have

$$n > N \implies x_n > x_N > b.$$

Thus we conclude that $\{x_n\}$ diverges to $+\infty$.

The proof for the decreasing sequences is similar.

$\square$

Example 1.2.9. The sequence in Example 1.2.4 is clearly increasing. Since it is divergent, the sequence has no upper bound. In fact, the proof of Proposition 1.2.7 tells us

$$\lim_{n \to \infty} \left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} \right) = +\infty.$$

On the other hand, the increasing sequence

$$x_n = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2}$$

satisfies

$$x_n < 1 + \frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \cdots + \frac{1}{(n-1)n}$$
$$= 1 + \left(\frac{1}{1} - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \cdots + \left(\frac{1}{n-1} - \frac{1}{n}\right)$$
$$= 1 + \frac{1}{1} - \frac{1}{n} < 2,$$

and must be convergent.

Much later on, we will see that the sequence

$$x_n = 1 + \frac{1}{2^p} + \frac{1}{3^p} + \cdots + \frac{1}{n^p}$$

converges if and only if $p > 1$.

Example 1.2.10. Suppose a sequence is given inductively by

$$x_1 = 1, \quad x_{n+1} = \sqrt{2 + x_n}.$$

We claim that the sequence is increasing. It is easy to see that $x_{n+1} > x_n$ is equivalent to $x_n^2 - x_n - 2 = (x_n - 2)(x_n + 1) < 0$. Since the sequence is clearly positive, the problem becomes $x_n < 2$. First $x_1 < 2$. Second if $x_n < 2$, then $x_{n+1} < \sqrt{2 + 2} = 2$. Thus the induction proves the inequality $x_n < 2$.

Since the sequence is increasing and has upper bound 2, it is convergent. The limit l satisfies

$$l^2 = \lim_{n \to \infty} x_{n+1}^2 = 2 + \lim_{n \to \infty} x_n = 2 + l.$$

Solving the equation for $l \geq 0$, we conclude that the limit is $l = 2$.

Example 1.2.11. Let $x_n = \left(1 + \frac{1}{n}\right)^n$. The binomial expansion tells us

$$x_n = 1 + n\left(\frac{1}{n}\right) + \frac{n(n-1)}{2!}\left(\frac{1}{n}\right)^2 + \frac{n(n-1)(n-2)}{3!}\left(\frac{1}{n}\right)^3$$
$$+ \cdots + \frac{n(n-1)\cdots 1}{n!}\left(\frac{1}{n}\right)^n$$
$$= 1 + \frac{1}{1!} + \frac{1}{2!}\left(1 - \frac{1}{n}\right) + \frac{1}{3!}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)$$
$$+ \cdots + \frac{1}{n!}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\cdots\left(1 - \frac{n-1}{n}\right).$$

By comparing the similar formula for x_{n+1}, we find the sequence is strictly increasing. The formula also tells us (see Example 1.2.9)

$$x_n < 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{n!} < 1 + 1 + \frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \cdots + \frac{1}{(n-1)n} < 3.$$

Therefore the sequence converges. The limit has a special notation

$$e = \lim_{n \to \infty}\left(1 + \frac{1}{n}\right)^n = 2.71828182845904\cdots$$

and is a fundamental constant of the nature.

Exercise 1.2.12. Prove the sequences converge.

1. $x_n = 1 + \dfrac{1}{2^3} + \dfrac{1}{3^3} + \cdots + \dfrac{1}{n^3}.$ 3. $x_n = 1 + \dfrac{1}{2^2} + \dfrac{1}{3^3} + \cdots + \dfrac{1}{n^n}.$

2. $x_n = 1 + \dfrac{1}{2!} + \dfrac{1}{3!} + \cdots + \dfrac{1}{n!}.$ 4. $x_n = 1 + \dfrac{1}{2} + \dfrac{1}{2^2} + \cdots + \dfrac{1}{2^n}.$

Exercise 1.2.13. Prove that $\lim_{n\to\infty} a^n = 0$ for $|a| < 1$ in the following steps.

1. Prove that for $0 \le a < 1$, the sequence $\{a^n\}$ converges.

2. Prove that the limit of $\{a^n\}$ must be 0.

3. Prove that $\lim_{n\to\infty} a^n = 0$ for $-1 < a \le 0$.

Exercise 1.2.14. Consider the sequence $x_n = 1 + \dfrac{1}{\sqrt{2}} + \dfrac{1}{\sqrt{3}} + \cdots + \dfrac{1}{\sqrt{n}} - 2\sqrt{n + \alpha}.$

1. Prove that if $\alpha \le \dfrac{1}{2}$, then the sequence is strictly decreasing.

2. Prove that if $\alpha > \dfrac{1}{2}$, then the sequence is strictly increasing for $n > \dfrac{1}{16\alpha - 8}.$

3. Prove the sequence converges to the same limit for all α.

Exercise 1.2.15. For any $a > 0$, let $x_n = \sqrt{a + \sqrt{a + \sqrt{a + \cdots + \sqrt{a}}}}$, where the square root appears n times. Prove $x_n < x_{n+1} < \dfrac{1 + \sqrt{4a + 1}}{2}$ and find the limit of the sequence.

The result can be extended to a sequence defined by $x_1 \le b$, $x_{n+1} = f(x_n)$, where $f(x)$ be a function satisfying $b \ge f(x) \ge x$ for $x \le b$. See Figure 1.3.
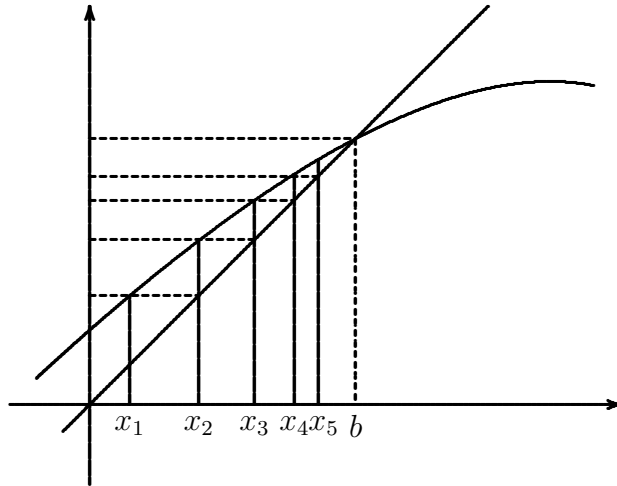
Figure 1.3: recursively defined convergent sequence

Exercise 1.2.16. Prove the inductively defined sequences converge and find the limits.

1. $x_1 = 1,\ x_{n+1} = 1 + \dfrac{2}{x_n}.$ 2. $x_1 = 1,\ x_{n+1} = \dfrac{x_n^2 + 2}{2x_n}.$

Exercise 1.2.17. Discuss the convergence of the sequence defined by $x_1 = \alpha$, $x_{n+1} = \beta + \gamma x_n$. Do the same to the sequence defined by $x_1 = \alpha$, $x_{n+1} = \beta + \dfrac{\gamma}{x_n}$.

Exercise 1.2.18. Let $a, b > 0$. Define sequences by $a_1 = a$, $b_1 = b$, $a_n = \dfrac{a_{n-1} + b_{n-1}}{2}$, $b_n = \dfrac{2a_{n-1}b_{n-1}}{a_{n-1} + b_{n-1}}$. Use $\dfrac{a + b}{2} \geq \dfrac{2ab}{a + b}$ to prove the sequences converge. Moreover, find the limits.

Exercise 1.2.19. Let $x_n = \left(1 + \dfrac{1}{n}\right)^n$ and $y_n = \left(1 + \dfrac{1}{n}\right)^{n+1}$.

1. Use induction to prove $(1+x)^n \geq 1+nx$ for $x > -1$ and any natural number n.

2. By showing $\dfrac{x_{n+1}}{x_n} > 1$ and $\dfrac{y_{n-1}}{y_n} > 1$, prove $\{x_n\}$ is increasing and $\{y_n\}$ is decreasing.

3. Prove $\{x_n\}$ and $\{y_n\}$ converge to the same limit e.

4. Prove $e - x_n < \dfrac{e}{n}$.

5. Prove $\lim_{n\to\infty} \left(1 - \dfrac{1}{n}\right)^{-n} = e$.

Exercise 1.2.20. Prove $e \geq 1 + \dfrac{1}{1!} + \dfrac{1}{2!} + \cdots + \dfrac{1}{n!}$ for any n.

1.2.4 Convergent Subsequence

By Proposition 1.1.2, any convergent sequence is bounded. While bounded sequences may not converge (see Example 1.2.2), the following still holds.

Theorem 1.2.8 (Bolzano[4]-Weierstrass[5] Theorem)**.** *A bounded sequence of real numbers has a convergent subsequence.*

Recall that by Proposition 1.2.1, any subsequence of a convergent sequence is convergent. Theorem 1.2.8 shows that if the original sequence is only assumed to be bounded, then "any subsequence" should be changed to "some subsequence".

Proof. Let $\{x_n\}$ be a bounded sequence. Then all x_n lie in a bounded interval $I = [a, b]$.

---

[4]Bernard Placidus Johann Nepomuk Bolzano, born 1781 and died 1848 in Prague (Bohemia, now Czech). Bolzano insisted that many results which were thought "obvious" required rigorous proof and made fundamental contributions to the foundation of mathematics. He understood the need to redefine and enrich the concept of number itself and define the Cauchy sequence four years before Cauchy's work appeared.

[5]Karl Theodor Wilhelm Weierstrass, born 1815 in Ostenfelde (Germany), died 1848 in Berlin (Germany). In 1864, he found a continuous but nowhere differentiable function. His lectures on analytic functions, elliptic functions, abelian functions and calculus of variations influenced many generations of mathematicians, and his approach still dominates the teaching of analysis today.

Divide I into two equal halves $I' = \left[a, \dfrac{a+b}{2}\right]$ and $I'' = \left[\dfrac{a+b}{2}, b\right]$. Then either I' or I'' must contain infinitely many x_n. We denote this interval by $I_1 = [a_1, b_1]$ and find $x_{n_1} \in I_1$.

Further divide I_1 into two equal halves $I'_1 = \left[a_1, \dfrac{a_1 + b_1}{2}\right]$ and $I''_1 = \left[\dfrac{a_1 + b_1}{2}, b_1\right]$. Then either I'_1 or I''_1 must contain infinitely many x_n. We denote this interval by $I_2 = [a_2, b_2]$. Because I_2 contains infinitely many x_n, we can find $x_{n_2} \in I_2$ with $n_2 > n_1$.

Keep going, we get a sequence of intervals

$$I = [a, b] \supset I_1 = [a_1, b_1] \supset I_2 = [a_2, b_2] \supset \cdots \supset I_k = [a_k, b_k] \supset \cdots$$

with the length of I_k being $b_k - a_k = \dfrac{b - a}{2^k}$. Moreover, we have a subsequence $\{x_{n_k}\}$ satisfying $x_{n_k} \in I_k$.

The inclusion relation between the intervals tells us

$$a \le a_1 \le a_2 \le \cdots \le a_k \le \cdots \le b_k \le \cdots \le b_2 \le b_1 \le b.$$

Thus $\{a_k\}$ and $\{b_k\}$ are bounded and monotone sequences. By Proposition 1.2.7, both sequences converge. Moreover, the length of I_k and the limit (1.1.3) tell us $\lim_{k \to \infty}(b_k - a_k) = 0$. Therefore the two sequences have the same limit. Denote $l = \lim_{k \to \infty} a_k = \lim_{k \to \infty} b_k$.

The property $x_{n_k} \in I_k$ means $a_k \le x_{n_k} \le b_k$. By the sandwich rule, we get $\lim_{k \to \infty} x_{n_k} = l$. Thus we find a convergent subsequence $\{x_{n_k}\}$.
$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Example 1.2.12. Let $\{x_n\}$, $\{y_n\}$, $\{z_n\}$ be sequences converging to l_1, l_2, l_3, respectively. Then for the sequence

$$x_1, \ y_1, \ z_1, \ x_2, \ y_2, \ z_2, \ x_3, \ y_3, \ z_3, \ \ldots, \ x_n, \ y_n, \ z_n, \ \ldots$$

the limits of convergent subsequences are l_1, l_2, l_3.

We need to explain that any $l \ne l_1, l_2, l_3$ is not the limit of any subsequence. There is $\epsilon > 0$, such that (take $\epsilon = \dfrac{1}{2} \min\{|l - l_1|, |l - l_2|, |l - l_3|\}$, for example)

$$|l - l_1| \ge 2\epsilon, \ \ |l - l_2| \ge 2\epsilon, \ \ |l - l_3| \ge 2\epsilon.$$

Then there are N_1, N_2, N_3, such that

$$\begin{aligned}
n > N_1 &\implies |x_n - l_1| < \epsilon, \\
n > N_2 &\implies |y_n - l_2| < \epsilon, \\
n > N_3 &\implies |z_n - l_3| < \epsilon.
\end{aligned}$$

Since $|l - l_1| \ge 2\epsilon$ and $|x_n - l_1| < \epsilon$ imply $|x_n - l| > \epsilon$, we have

$$n > \max\{N_1, N_2, N_3\} \implies |x_n - l| > \epsilon, \ \ |y_n - l| > \epsilon, \ \ |z_n - l| > \epsilon.$$

This implies that l cannot be the limit of any convergent subsequence.

A more direct argument about the limits of convergent subsequences is the following. Let l be the limit of a convergent subsequence $\{w_m\}$ of the combined sequence. The subsequence $\{w_m\}$ must contain infinitely many terms from at least one of the three sequences. If $\{w_m\}$ contains infinitely many terms from $\{x_n\}$, then it contains a subsequence $\{x_{n_k}\}$ of $\{x_n\}$. By Proposition 1.2.1, we get

$$l = \lim w_m = \lim x_{n_k} = \lim x_n = l_1.$$

The second equality is due to the fact that $\{x_{n_k}\}$ is a subsequence of $\{w_m\}$. The third equality is due to $\{x_{n_k}\}$ being a subsequence of $\{x_n\}$.

Exercise 1.2.21. For sequences in Exercise 1.2.1, find all the limits of convergent subsequences.

Exercise 1.2.22. Any real number is the limit of a sequence of the form $\dfrac{n_1}{10}$, $\dfrac{n_2}{100}$, $\dfrac{n_3}{1000}$, ..., where n_k are integers. Based on this observation, construct a sequence so that the limits of convergent subsequences are all the numbers between 0 and 1.

Exercise 1.2.23. Prove that a number is the limit of a convergent subsequence of $\{x_n\}$ if and only if it is the limit of a convergent subsequence of $\{x_n \sqrt[n]{n}\}$.

Exercise 1.2.24. Suppose $\{x_n\}$ and $\{y_n\}$ are two bounded sequences. Prove that there are n_k, such that both subsequences $\{x_{n_k}\}$ and $\{y_{n_k}\}$ converge.

The following technical result provides a criterion for a number to be the limit of a subsequence.

Proposition 1.2.9. *l is the limit of a convergent subsequence of $\{x_n\}$ if and only if for any $\epsilon > 0$ and N, there is $n > N$, such that $|x_n - l| < \epsilon$.*

Proof. Let l be the limit of a convergent subsequence $\{x_{n_k}\}$. Let $\epsilon > 0$ and N be given. Then there is K, such that $k > K$ implies $|x_{n_k} - l| < \epsilon$. It is easy to find $k > K$, such that $n_k > N$ (take $k = \max\{K, N\} + 1$, for example). Then for $n = n_k$, we have $n > N$ and $|x_n - l| < \epsilon$.

Conversely, suppose for any $\epsilon > 0$ and N, there is $n > N$, such that $|x_n - l| < \epsilon$. Then for $\epsilon = 1$, there is n_1 such that $|x_{n_1} - l| < 1$. Next, for $\epsilon = \dfrac{1}{2}$ and $N = n_1$, there is $n_2 > n_1$ such that $|x_{n_2} - l| < \dfrac{1}{2}$. Keep going, by taking $\epsilon = \dfrac{1}{k+1}$ and $N = n_k$ to find $x_{n_{k+1}}$, we construct a subsequence $\{x_{n_k}\}$ satisfying $|x_{n_k} - l| < \dfrac{1}{k}$. The inequality implies $\lim_{k \to \infty} x_{n_k} = l$. $\qquad\square$

Here is a remark that is very useful for the discussion of subsequences. Suppose P is a property about terms in a sequence ($x_n > l$ or $x_n > x_{n+1}$, for examples). Then the following statements are equivalent:

1. For any N, there is $n > N$, such that x_n has property P.

2. There are infinitely many x_n with property P.

3. There is a subsequence $\{x_{n_k}\}$ such that each term x_{n_k} has property P.

In particular, the criterion for l to be the limit of a convergent subsequence of $\{x_n\}$ is that for any $\epsilon > 0$, there are infinitely many x_n satisfying $|x_n - l| < \epsilon$.

Exercise 1.2.25. Let $\{x_n\}$ be a sequence. Suppose $\lim_{k\to\infty} l_k = l$ and for each k, l_k is the limit of a convergent subsequence of $\{x_n\}$. Prove that l is also the limit of a convergent subsequence.

Let $\{x_n\}$ be a bounded sequence. Then the set $\mathrm{LIM}\{x_n\}$ of all the limits of convergent subsequences of $\{x_n\}$ is also bounded. The supremum of $\mathrm{LIM}\{x_n\}$ is called the *upper limit* and denoted $\overline{\lim}_{n\to\infty} x_n$. The infimum of $\mathrm{LIM}\{x_n\}$ is called the *lower limit* and denoted $\underline{\lim}_{n\to\infty} x_n$. For example, for the sequence in Example 1.1.14, the upper limit is $\max\{l_1, l_2, l_3\}$ and the lower limit is $\min\{l_1, l_2, l_3\}$.

The following characterizes the upper limit. The lower limit can be similarly characterized.

Proposition 1.2.10. *Suppose $\{x_n\}$ is a bounded sequence and l is a number.*

1. *If $l < \overline{\lim}_{n\to\infty} x_n$, then there are infinitely many $x_n > l$.*

2. *If $l > \overline{\lim}_{n\to\infty} x_n$, then there are only finitely many $x_n > l$.*

The characterization gives us a clear picture of the upper limit. Let us start with a big l and move downwards. As we decrease l, the number of terms $x_n > l$ will increase. If l is an upper bound of $\{x_n\}$, then this number is zero. When l is lowered and is no longer an upper bound, some terms in the sequence will be higher than l. The number might be finite at the beginning. But there will be a threshold, such that if l is below the threshold, then the number of $x_n > l$ becomes infinite. The reason for the existence of such a threshold is that when l is so low to become a lower bound, then all (in particular, infinitely may) $x_n > l$.

This threshold is the upper limit.

Proof. If $l < \overline{\lim}_{n\to\infty} x_n$, then by the definition of the upper limit, some $l' > l$ is the limit of a subsequence. By Proposition 1.1.4, all the terms in the subsequence except finitely many will be bigger than l. Thus we find infinitely many $x_n > l$.

The second statement is the same as the following: If there are infinitely many $x_n > l$, then $l \leq \overline{\lim}_{n\to\infty} x_n$. We will prove this equivalent statement.

Since there are infinitely many $x_n > l$, there is a subsequence $\{x_{n_k}\}$ satisfying $x_{n_k} > l$. By Theorem 1.2.8, the subsequence has a further convergent subsequence $\{x_{n_{k_p}}\}$. Then $x_{n_k} > l$ implies $\lim x_{n_{k_p}} \geq l$. This gives a number in $\mathrm{LIM}\{x_n\}$ that is no less than l, so that $\overline{\lim}_{n\to\infty} x_n = \sup \mathrm{LIM}\{x_n\} \geq l$. $\square$

Proposition 1.2.11. *The upper and lower limits are limits of convergent subsequences. Moreover, the sequence converges if and only if the upper and lower limits are equal.*

The first conclusion is $\overline{\lim}_{n\to\infty} x_n, \underline{\lim}_{n\to\infty} x_n \in \mathrm{LIM}\{x_n\}$. In the second conclusion, the equality $\overline{\lim}_{n\to\infty} x_n = \underline{\lim}_{n\to\infty} x_n = l$ means $\mathrm{LIM}\{x_n\} = \{l\}$, which basically says that all convergent subsequences have the same limit.

Proof. Denote $l = \overline{\lim} x_n$. For any $\epsilon > 0$, we have $l + \epsilon > \overline{\lim} x_n$ and $l - \epsilon < \overline{\lim} x_n$. Applying Proposition 1.2.10, we know there are infinitely many $x_n > l - \epsilon$ and only finitely many $x_n > l + \epsilon$. Therefore there are infinitely many x_n satisfying $l + \epsilon \geq x_n > l - \epsilon$. Thus we have proved that for any $\epsilon > 0$, there are infinitely many x_n satisfying $|x_n - l| \leq \epsilon$. By Proposition 1.2.9 (and the remark after the proof), this shows that l is the limit of a convergent subsequence.

For the second part, Proposition 1.2.1 says that if $\{x_n\}$ converges to l, then $\mathrm{LIM}\{x_n\} = \{l\}$, so that $\overline{\lim} x_n = \underline{\lim} x_n = l$. Conversely, suppose $\overline{\lim} x_n = \underline{\lim} x_n = l$. Then for any $\epsilon > 0$, applying the second part of Proposition 1.2.10 to $l + \epsilon > \overline{\lim} x_n$, we find only finitely many $x_n > l + \epsilon$. Applying the similar property for the lower limit to $l - \epsilon < \underline{\lim} x_n$, we also find only finitely many $x_n < l - \epsilon$. Thus $|x_n - l| \leq \epsilon$ holds for all but finitely many x_n. If N is the biggest index for those x_n that do not satisfy $|x_n - l| \leq \epsilon$, then we get $|x_n - l| \leq \epsilon$ for all $n > N$. This proves that $\{x_n\}$ converges to l. □

Exercise 1.2.26. Find all the upper and lower limits of bounded sequences in Exercise 1.2.1.

Exercise 1.2.27. Prove the properties of upper and lower limits.

1. $\overline{\lim}_{n\to\infty}(-x_n) = -\underline{\lim}_{n\to\infty} x_n$.

2. $\overline{\lim}_{n\to\infty} x_n + \overline{\lim}_{n\to\infty} y_n \geq \overline{\lim}_{n\to\infty}(x_n + y_n) \geq \underline{\lim}_{n\to\infty} x_n + \overline{\lim}_{n\to\infty} y_n$.

3. If $x_n > 0$, then $\overline{\lim}_{n\to\infty} \dfrac{1}{x_n} = \dfrac{1}{\underline{\lim}_{n\to\infty} x_n}$.

4. If $x_n \geq 0$ and $y_n \geq 0$, then $\overline{\lim}_{n\to\infty} x_n \cdot \overline{\lim}_{n\to\infty} y_n \geq \overline{\lim}_{n\to\infty}(x_n y_n) \geq \underline{\lim}_{n\to\infty} x_n \cdot \overline{\lim}_{n\to\infty} y_n$.

Exercise 1.2.28. Prove that if $\overline{\lim}_{n\to\infty} \left| \dfrac{x_{n+1}}{x_n} \right| < 1$, then $\lim_{n\to\infty} x_n = 0$. Prove that if $\underline{\lim}_{n\to\infty} \left| \dfrac{x_{n+1}}{x_n} \right| > 1$, then $\lim_{n\to\infty} x_n = \infty$.

Exercise 1.2.29. Prove that the upper limit l of a bounded sequence $\{x_n\}$ is characterized by the following two properties.

1. l is the limit of a convergent subsequence.

2. For any $\epsilon > 0$, there is N, such that $x_n < l + \epsilon$ for any $n > N$.

The characterization may be compared with the one for the supremum in Exercise 1.2.9.

Exercise 1.2.30. Let $\{x_n\}$ be a bounded sequence. Let $y_n = \sup\{x_n, x_{n+1}, x_{n+2}, \dots\}$. Then $\{y_n\}$ is a bounded decreasing sequence. Prove that $\lim_{n\to\infty} y_n = \overline{\lim}_{n\to\infty} x_n$. Find the similar formula for $\underline{\lim}_{n\to\infty} x_n$.

1.2.5   Convergence of Cauchy Sequence

Now we are ready to prove the converse of Theorem 1.2.2.

Theorem 1.2.12. *Any Cauchy sequence of real numbers is convergent.*

Proof. Let $\{x_n\}$ be a Cauchy sequence. We claim the sequence is bounded. For $\epsilon = 1 > 0$, there is N, such that $m, n > N$ implies $|x_m - x_n| < 1$. Taking $m = N + 1$, we find $n > N$ implies $x_{N+1} - 1 < x_n < x_{N+1} + 1$. Therefore $\max\{x_1, x_2, \ldots, x_N, x_{N+1} + 1\}$ is an upper bound for the sequence, and $\min\{x_1, x_2, \ldots, x_N, x_{N+1} - 1\}$ is a lower bound.

By Theorem 1.2.8, there is a subsequence $\{x_{n_k}\}$ converging to a limit l. Thus for any $\epsilon > 0$, there is K, such that

$$k > K \implies |x_{n_k} - l| < \frac{\epsilon}{2}.$$

On the other hand, since $\{x_n\}$ is a Cauchy sequence, there is N, such that

$$m, n > N \implies |x_m - x_n| < \frac{\epsilon}{2}.$$

Now for any $n > N$, we can easily find some $k > K$, such that $n_k > N$ ($k = \max\{K, N\} + 1$, for example). Then we have both $|x_{n_k} - l| < \frac{\epsilon}{2}$ and $|x_{n_k} - x_n| < \frac{\epsilon}{2}$. The inequalities imply $|x_n - l| < \epsilon$. Thus we established the implication

$$n > N \implies |x_n - l| < \epsilon.$$

$\square$

The proof of Theorem 1.2.12 does not use the upper and lower limits. Alternatively, we observe that for any $\epsilon > 0$, the Cauchy condition implies that any two subsequences will be within ϵ of each other after finitely many terms. This implies that the difference between the limits of any two convergent subsequences cannot be more than ϵ. Since ϵ can be arbitrarily small, this implies the upper and the lower limits must be the same. The Proposition 1.2.11 then shows that the whole sequence converges.

Example 1.2.13. For the sequence

$$x_n = 1 - \frac{1}{2^2} + \frac{1}{3^2} - \cdots + \frac{(-1)^n}{n^2},$$

and $m > n$, we have

$$
\begin{aligned}
|x_m - x_n| &= \left| \frac{(-1)^{n+1}}{(n+1)^2} + \frac{(-1)^{n+2}}{(n+2)^2} + \cdots + \frac{(-1)^m}{m^2} \right| \\
&< \frac{1}{n(n+1)} + \frac{1}{(n+1)(n+2)} + \cdots + \frac{1}{(m-1)m} \\
&= \left(\frac{1}{n} - \frac{1}{n+1} \right) + \left(\frac{1}{n+1} - \frac{1}{n+2} \right) + \cdots + \left(\frac{1}{m-1} - \frac{1}{m} \right) \\
&= \frac{1}{n} - \frac{1}{m} < \frac{1}{n}.
\end{aligned}
$$

For any $\epsilon > 0$, let $N = \dfrac{1}{\epsilon}$. Then for $m > n > N$, we have

$$0 < x_m - x_n < \frac{1}{n} < \epsilon.$$

Thus $\{x_n\}$ is a Cauchy sequence and must converge.

Exercise 1.2.31. Prove the sequences converge.

1. $x_n = 1 + \dfrac{1}{2!} + \dfrac{1}{3!} + \cdots + \dfrac{1}{n!}$.

2. $x_n = 1 + \dfrac{1}{2^2} + \dfrac{1}{3^3} + \cdots + \dfrac{1}{n^n}$.

3. $x_n = 1 - \dfrac{1}{2^3} + \dfrac{1}{3^3} + \cdots + (-1)^n \dfrac{1}{n^3}$.

4. $x_n = \sin 1 + \dfrac{\sin 2}{2^3} + \dfrac{\sin 3}{3^3} + \cdots + \dfrac{\sin n}{n^3}$.

Exercise 1.2.32. Let $|a| \le 1$. Define a sequence by $x_1 = a$, $x_{n+1} = \dfrac{1}{4}(1 + x_n^2)$.

1. Prove $|x_n| \le 1$ and $|x_{n+1} - x_n| < \dfrac{1}{2^{n-2}}$.

2. Prove $\dfrac{1}{2} + \dfrac{1}{2^2} + \dfrac{1}{2^3} + \cdots + \dfrac{1}{2^n} = 1 - \dfrac{1}{2^n}$.

3. Prove the sequence converges and find the limit.

1.2.6 Open Cover

A set X of numbers is *closed* if $x_n \in X$ and $\lim_{n\to\infty} x_n = l$ implies $l \in X$. Intuitively, this means that one cannot escape X by taking limits. For example, the order rule on the limit tells us that closed intervals $[a, b]$ are closed sets. The proof of Bolzano-Weierstrass Theorem also gives us the following result, which in modern topological language says that bounded and closed sets of numbers are *compact*.

Theorem 1.2.13 (Heine[6]-Borel[7] Theorem). *Suppose X is a bounded and closed set of numbers. Suppose $\{(a_i, b_i)\}$ is a collection of open intervals such that $X \subset \cup(a_i, b_i)$. Then $X \subset (a_{i_1}, b_{i_1}) \cup (a_{i_2}, b_{i_2}) \cup \cdots \cup (a_{i_n}, b_{i_n})$ for finitely many intervals in the collection.*

When $X \subset \cup(a_i, b_i)$ happens, we say $\mathcal{U} = \{(a_i, b_i)\}$ is an *open cover* of X. The theorem says that if X is bounded and closed, then any cover by open intervals has a finite *subcover*.

---

[6]Heinrich Eduard Heine, born 1821 in Berlin (Germany), died 1881 in Halle (Germany). In addition to the Heine-Borel theorem, he introduced the idea of uniform continuity.

[7]Félix Edouard Justin Émile Borel, born 1871 in Aveyron (France), died 1881 in Paris (France). Borel' measure theory was the beginning of the modern theory of functions of a real variable. He was French Minister of the Navy from 1925 to 1940.

Proof. Suppose $X \subset I = [\alpha, \beta]$ for a bounded and closed interval I. Suppose X cannot be covered by finitely many open intervals in $\mathcal{U} = \{(a_i, b_i)\}$.

Similar to the proof of Bolzano-Weierstrass Theorem, we divide the interval into two equal halves $I' = \left[\alpha, \dfrac{\alpha + \beta}{2}\right]$ and $I'' = \left[\dfrac{\alpha + \beta}{2}, \beta\right]$. Then either $X' = X \cap I'$ or $X'' = X \cap I''$ cannot be covered by finitely many open intervals in $\mathcal{U}$. We denote the corresponding interval by $I_1 = [\alpha_1, \beta_1]$ and denote $X_1 = X \cap I_1$.

Further divide I_1 into two equal halves $I_1' = \left[\alpha_1, \dfrac{\alpha_1 + \beta_1}{2}\right]$ and $I_1'' = \left[\dfrac{\alpha_1 + \beta_1}{2}, \beta_1\right]$. Then either $X_1' = X_1 \cap I_1'$ or $X_1'' = X_1 \cap I_1''$ cannot be covered by finitely many open intervals in $\mathcal{U}$. We denote the corresponding interval by $I_2 = [\alpha_2, \beta_2]$ and denote $X_2 = X \cap I_2$.

Keep going, we get a sequence of intervals

$$I = [\alpha, \beta] \supset I_1 = [\alpha_1, \beta_1] \supset I_2 = [\alpha_2, \beta_2] \supset \cdots \supset I_k = [\alpha_k, \beta_k] \supset \cdots$$

with the length of I_k being $\beta_k - \alpha_k = \dfrac{\beta - \alpha}{2^k}$, and $X_k = X \cap I_k$ cannot be covered by finitely many open intervals in $\mathcal{U}$.

As argued in the proof before, we have converging limit $l = \lim_{k \to \infty} \alpha_k = \lim_{k \to \infty} \beta_k$. Moreover, picking $x_k \in X_k$, we have $\alpha_k \le x_k \le \beta_k$. By the sandwich rule, we get $l = \lim_{k \to \infty} x_k$. Then by the assumption that X is closed, we get $l \in X$.

Since $X \subset \cup(a_i, b_i)$, we have $l \in (a_{i_0}, b_{i_0})$ for some interval $(a_{i_0}, b_{i_0}) \in \mathcal{U}$. Then by $l = \lim_{k \to \infty} \alpha_k = \lim_{k \to \infty} \beta_k$, we have $X_k \subset I_k = [\alpha_k, \beta_k] \subset (a_{i_0}, b_{i_0})$ for sufficiently big k. In particular, X_k can be covered by one open interval in $\mathcal{U}$. The contradiction shows that X must be covered by finitely many open intervals from $\mathcal{U}$.

$\square$

Exercise 1.2.33. Find a collection $\mathcal{U} = \{(a_i, b_i)\}$ that covers $(0, 1]$, but $(0, 1]$ cannot be covered by finitely many intervals in $\mathcal{U}$. Find similar counterexample for $[0, +\infty)$ in place of $(0, 1]$.

Exercise 1.2.34 (Lebesgue[8]). Suppose $[\alpha, \beta]$ is covered by a collection $\mathcal{U} = \{(a_i, b_i)\}$. Denote

$$X = \{x \in [\alpha, \beta] \colon [\alpha, x] \text{ is covered by finitely many intervals in } \mathcal{U}\}.$$

1. Prove that $\sup X \in X$.

2. Prove that if $x \in X$ and $x < \beta$, then $x + \delta \in X$ for some $\delta > 0$.

3. Prove that $\sup X = \beta$.

---

[8]Henri Léon Lebesgue, born 1875 in Beauvais (France), died 1941 in Paris (France). His 1901 paper "Sur une généralisation de l'intégrale définie" introduced the concept of measure and revolutionized the integral calculus. He also made major contributions in other areas of mathematics, including topology, potential theory, the Dirichlet problem, the calculus of variations, set theory, the theory of surface area and dimension theory.

This proves Heine-Borel Theorem for bounded and closed intervals.

Exercise 1.2.35. Prove Heine-Borel Theorem for a bounded and closed set X in the following way. Suppose X is covered by a collection $\mathcal{U} = \{(a_i, b_i)\}$.

1. Prove that there is $\delta > 0$, such that for any $x \in X$, $(x - \delta, x + \delta) \subset (a_i, b_i)$ for some $(a_i, b_i) \in \mathcal{U}$.

2. Use the boundedness of X to find finitely many numbers $c_1, c_2, \ldots, c_n$, such that $X \subset (c_1, c_1 + \delta) \cup (c_2, c_2 + \delta) \cup \cdots \cup (c_n, c_n + \delta)$.

3. Prove that if $X \cap (c_j, c_j + \delta) \neq \emptyset$, then $(c_j, c_j + \delta) \subset (a_i, b_i)$ for some $(a_i, b_i) \in \mathcal{U}$.

4. Prove that X is covered by no more than n open intervals in $\mathcal{U}$.

1.2.7 Additional Exercise

Extended Supremum and Extended Upper Limit

Exercise 1.2.36. Extend the number system by including the "infinite numbers" $+\infty$, $-\infty$ and introduce the order $-\infty < x < +\infty$ for any real number x. Then for any nonempty set X of real numbers and possibly $+\infty$ or $-\infty$, we have $\sup X$ and $\inf X$ similarly defined. Prove that there are exactly three possibilities for $\sup X$.

1. If X has no upper bound or $+\infty \in X$, then $\sup X = +\infty$.

2. If X has a finite number as an upper bound and contains at least one finite real number, then $\sup X$ is a finite real number.

3. If $X = \{-\infty\}$, then $\sup X = -\infty$.

Write down the similar statements for $\inf X$.

Exercise 1.2.37. For a not necessarily bounded sequence $\{x_n\}$, extend the definition of $\mathrm{LIM}\{x_n\}$ by adding $+\infty$ if there is a subsequence diverging to $+\infty$, and adding $-\infty$ if there is a subsequence diverging to $-\infty$. Define the upper and lower limits as the supremum and infimum of $\mathrm{LIM}\{x_n\}$, using the extension of the concepts in Exercise 1.2.36. Prove the following extensions of Proposition 1.2.11.

1. A sequence with no upper bound must have a subsequence diverging to $+\infty$. This means $\overline{\lim}_{n \to \infty} x_n = +\infty$.

2. If there is no subsequence with finite limit and no subsequence diverging to $-\infty$, then the whole sequence diverges to $+\infty$.

Supremum and Infimum in Ordered Set

Recall that an order on a set is a relation $x < y$ between pairs of elements satisfying the transitivity and the exclusivity. The concepts of upper bound, lower bound, supremum and infimum can be defined for subsets of an ordered set in a way similar to numbers.

Exercise 1.2.38. Provide a characterization of the supremum similar to numbers.

Exercise 1.2.39. Prove that the supremum, if exists, must be unique.

Exercise 1.2.40. An order is defined for all subsets of the plane $\mathbb{R}^2$ by $A \leq B$ if A is contained in B. Let R be the set of all rectangles centered at the origin and with circumference 1. Find the supremum and infimum of R.

Alternative Proof of Bolzano-Weierstrass Theorem

We say a term x_n in a sequence has property P if there is M, such that $m > M$ implies $x_m > x_n$. The property means that $x_m > x_n$ for sufficiently big m.

Exercise 1.2.41. Suppose there are infinitely many terms in a sequence $\{x_n\}$ with property P. Prove that for any term x_n with property P, there is $m > n$, such that $x_m > x_n$ and x_m also has property P. Then construct an increasing subsequence $\{x_{n_k}\}$ in which each x_{n_k} has property P.

Exercise 1.2.42. Suppose there are only finitely many terms in a sequence $\{x_n\}$ with property P. Prove that there is N, such that for any $n > N$, there is a term $x_m \leq x_n$ with $m > n$. Then use this to construct a decreasing subsequence $\{x_{n_k}\}$.

Exercise 1.2.43. Suppose the sequence $\{x_n\}$ is bounded. Prove that the increasing subsequence constructed in Exercise 1.2.41 converges to $\underline{\lim} x_n$, and the decreasing subsequence constructed in Exercise 1.2.42 converges to $\overline{\lim} x_n$.

Set Version of Bolzano-Weierstrass Theorem

A number l is an *accumulation point* of a set X of numbers if for any $\epsilon > 0$, there is $x \in X$ satisfying $0 < |x - l| < \epsilon$. The set version of Bolzano-Weierstrass Theorem says that any infinite, bounded and closed set of numbers has an accumulation point. Theorem 1.2.8 may be considered as the sequence version of Bolzano-Weierstrass Theorem.

Exercise 1.2.44. What does it mean for a point l not to be an accumulation point? Use your answer and Heine-Borel Theorem to prove the set version of Bolzano-Weierstrass Theorem.

Exercise 1.2.45. Use the set version of Bolzano-Weierstrass Theorem to prove the sequence version.

1.3   Limit of Function

The limit process is not restricted to sequences only. For a function $f(x)$ defined near (but not necessarily at) a, we may also consider its tendency as x approaches a. This leads to the definition of the limit of functions.

Many properties of the limit of sequences can be extended to functions. The two types of limits are also closely related.

1.3.1   Definition

Definition 1.3.1. A function $f(x)$ defined near a has *limit l* at a, denoted $\lim_{x \to a} f(x) = l$, if for any $\epsilon > 0$, there is $\delta > 0$, such that

$$0 < |x - a| < \delta \implies |f(x) - l| < \epsilon. \tag{1.3.1}$$
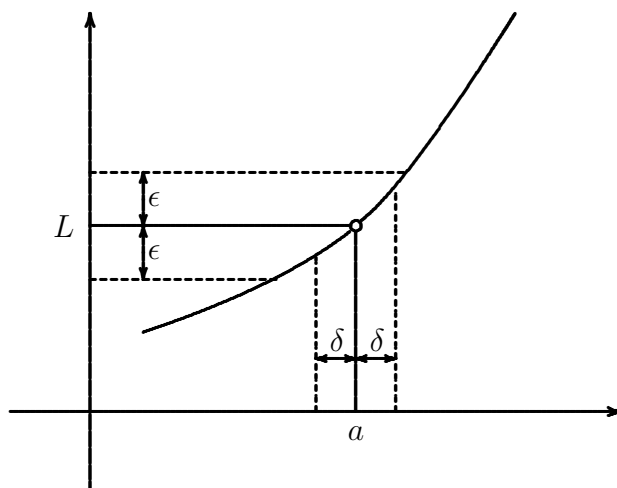
Figure 1.4: for any ϵ, there is δ

Similar to the limit of sequences, the predetermined smallness ϵ for $|f(x)-l|$ is *arbitrarily* given, while the size δ for $|x-a|$ is to be found *after* ϵ is given. Thus the choice of δ usually depends on ϵ and is often expressed as a function of ϵ. Moreover, since the limit is about what happens when numbers are close, only small ϵ and δ need to be considered.

Example 1.3.1. The graph for the function $f(x) = x^2$ suggests $\lim_{x \to 2} x^2 = 4$. Rigorously following the definition, for any $\epsilon > 0$, choose $\delta = \min \left\{ 1, \dfrac{\epsilon}{5} \right\}$. Then

$$0 < |x - 2| < \delta \implies |x - 2| < 1, \ \ |x - 2| < \frac{\epsilon}{5}$$
$$\implies |x + 2| < 5, \ \ |x - 2| < \frac{\epsilon}{5}$$
$$\implies |x^2 - 4| = |x + 2||x - 2| < 5\frac{\epsilon}{5} = \epsilon.$$

How did we choose δ? We try to achieve $|x^2 - 4| = |x + 2||x - 2| < \epsilon$ by requiring $0 < |x - 2| < \delta$. Note that when x is close to 2, $|x + 2|$ is close to 4 and $|x^2 - 4|$ is close to 4δ. More precisely, if we give some room to x by choosing $\delta \le 1$, so that $|x + 2|$ is not more than 5, then we end up with the requirement $5\delta \le \epsilon$. Combining $\delta \le 1$ and $5\delta \le \epsilon$ together yields our choice for δ.

Example 1.3.2. The graph for the function $f(x) = \sqrt{x}$ suggests that $\lim_{x \to 4} \sqrt{x} = 2$. Rigorously following the definition, for any $\epsilon > 0$, choose $\delta = \epsilon$. Then

$$0 < |x - 4| < \delta \implies |\sqrt{x} - 2| = \frac{|\sqrt{x} - 2||\sqrt{x} + 2|}{|\sqrt{x} + 2|} = \frac{|x - 4|}{|\sqrt{x} + 2|} < \frac{\delta}{|\sqrt{x} + 2|} < \delta = \epsilon.$$

Example 1.3.3. The graph for the function $f(x) = \dfrac{1}{x}$ suggests $\lim_{x \to 1} \dfrac{1}{x} = 1$. To rigorously argue by the definition, we estimate how small

$$\left| \frac{1}{x} - 1 \right| = \frac{|x - 1|}{|x|}$$

can be when $0 < |x - 1| < \delta$. Note that in the quotient, the numerator $|x - 1| < \delta$ and the denominator $|x|$ should be close to 1 for small δ. Specifically, if $\delta < \dfrac{1}{2}$,

then $|x - 1| < \delta$ will imply $|x| > \dfrac{1}{2}$, and we get $\dfrac{|x - 1|}{|x|} < 2\delta$. We conclude that

$0 < |x - 1| < \delta = \min\left\{\dfrac{1}{2}, \dfrac{\epsilon}{2}\right\}$ implies $\left|\dfrac{1}{x} - 1\right| < \epsilon$.

Example 1.3.4. Note that the limit of $f(x)$ at a does not depend on $f(a)$. In fact, the function f does not even need to be defined at a. For example, the function

$$f(x) = \begin{cases} \dfrac{1}{x} & \text{if } x \neq 1 \\ 2 & \text{if } x = 1 \end{cases}$$

has limit $\lim_{x \to 1} f(x) = \lim_{x \to 1} \dfrac{1}{x} = 1$. In fact, the argument for the limit in Example 1.1.3 can be used here without any change.

Exercise 1.3.1. Rigorously verify the limits.

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

2. $\lim_{x \to 2} 3x^3 = 24$.

3. $\lim_{x \to 1} \sqrt{x} = 1$.

4. $\lim_{x \to -2} \dfrac{1}{x} = -\dfrac{1}{2}$.

5. $\lim_{x \to -2} \dfrac{2}{x^2} = \dfrac{1}{2}$.

6. $\lim_{x \to a} |x| = |a|$.

Exercise 1.3.2. Prove that $\lim_{x \to a} f(x) = l$ implies $\lim_{x \to a} |f(x)| = |l|$. Prove the converse is true if $l = 0$. What about the converse in case $l \neq 0$?

Exercise 1.3.3. Prove the following are equivalent definitions of $\lim_{x \to a} f(x) = l$.

1. For any $\epsilon > 0$, there is $\delta > 0$, such that $0 < |x - a| \leq \delta$ implies $|f(x) - l| < \epsilon$.

2. For any $c > \epsilon > 0$, where c is some fixed number, there is $\delta > 0$, such that $0 < |x - a| < \delta$ implies $|f(x) - l| \leq \epsilon$.

3. For any natural number n, there is $\delta > 0$, such that $0 < |x - a| < \delta$ implies $|f(x) - l| \leq \dfrac{1}{n}$.

4. For any $1 > \epsilon > 0$, there is $\delta > 0$, such that $0 < |x - a| < \delta$ implies $|f(x) - l| < \dfrac{\epsilon}{1 - \epsilon}$.

Exercise 1.3.4. Which are equivalent to the definition of $\lim_{x \to a} f(x) = l$?

1. For $\epsilon = 0.001$, we have $\delta = 0.01$, such that $0 < |x - a| \leq \delta$ implies $|f(x) - l| < \epsilon$.

2. For any $\epsilon > 0$, there is $\delta > 0$, such that $|x - a| < \delta$ implies $|f(x) - l| < \epsilon$.

3. For any $\epsilon > 0$, there is $\delta > 0$, such that $0 < |x - a| < \delta$ implies $0 < |f(x) - l| < \epsilon$.

4. For any $0.001 \geq \epsilon > 0$, there is $\delta > 0$, such that $0 < |x - a| \leq 2\delta$ implies $|f(x) - l| \leq \epsilon$.

5. For any $\epsilon > 0.001$, there is $\delta > 0$, such that $0 < |x - a| \leq 2\delta$ implies $|f(x) - l| \leq \epsilon$.

6. For any $\epsilon > 0$, there is a rational number $\delta > 0$, such that $0 < |x - a| < \delta$ implies $|f(x) - l| < \epsilon$.

7. For any $\epsilon > 0$, there is a natural number N, such that $0 < |x - a| < \dfrac{1}{N}$ implies $|f(x) - l| < \epsilon$.

8. For any $\epsilon > 0$, there is $\delta > 0$, such that $0 < |x - a| < \delta$ implies $|f(x) - l| < \epsilon^2$.

9. For any $\epsilon > 0$, there is $\delta > 0$, such that $0 < |x - a| < \delta$ implies $|f(x) - l| < \epsilon^2 + 1$.

10. For any $\epsilon > 0$, there is $\delta > 0$, such that $0 < |x - a| < \delta + 1$ implies $|f(x) - l| < \epsilon^2$.

1.3.2 Variation

In the definition of $\lim_{x \to a} f(x)$, x may approach a from right (i.e., $x > a$) or from left (i.e., $x < a$). The two approaches may be treated separately, leading to *one side limits*.

Definition 1.3.2. A function $f(x)$ defined for $x > a$ and near a has *right limit* l at a, denoted $\lim_{x \to a^+} f(x) = l$, if for any $\epsilon > 0$, there is $\delta > 0$, such that

$$0 < x - a < \delta \implies |f(x) - l| < \epsilon. \tag{1.3.2}$$

A function $f(x)$ defined for $x < a$ and near a has *left limit* l at a, denoted $\lim_{x \to a^-} f(x) = l$, if for any $\epsilon > 0$, there is $\delta > 0$, such that

$$-\delta < x - a < 0 \implies |f(x) - l| < \epsilon. \tag{1.3.3}$$

The one side limits are often denoted as $f(a^+) = \lim_{x \to a^+} f(x)$ and $f(a^-) = \lim_{x \to a^-} f(x)$. Moreover, the one side limits and the usual (two side) limit are related as follows. The proof is left as an exercise.

Proposition 1.3.3. $\lim_{x \to a} f(x) = l$ *if and only if* $\lim_{x \to a^+} f(x) = \lim_{x \to a^-} f(x) = l$.

Example 1.3.5. For the function

$$f(x) = \begin{cases} \dfrac{1}{x} & \text{if } x > 1 \\ 3x - 2 & \text{if } x < 1 \end{cases},$$

we have

$$\lim_{x \to 1^+} f(x) = \lim_{x \to 1} \frac{1}{x} = 1, \quad \lim_{x \to 1^-} f(x) = \lim_{x \to 1} (3x - 2) = 1.$$

Therefore $\lim_{x \to 1} f(x) = 1$. On the other hand, for the function

$$g(x) = \begin{cases} \dfrac{1}{x} & \text{if } x > 1 \\ 3x & \text{if } x < 1 \end{cases},$$

we have

$$\lim_{x \to 1^+} g(x) = \lim_{x \to 1} \frac{1}{x} = 1, \quad \lim_{x \to 1^-} g(x) = \lim_{x \to 1} 3x = 3.$$

Since the left and right limits are not equal, $g(x)$ diverges at 1.

Example 1.3.6. If $\alpha > 0$, then for any $\epsilon > 0$, we have

$$0 < x < \delta = \epsilon^{\frac{1}{\alpha}} \implies 0 < x^{\alpha} < \delta^{\alpha} = \epsilon.$$

This shows $\lim_{x \to 0^+} x^{\alpha} = 0$.

Exercise 1.3.5. Let $[x]$ be the biggest integer n satisfying $n \le x$. For example, $[1.1] = 1$, $[0.99] = 0$, $[-3.2] = -4$. Compute the limits.

1. $\lim_{x \to 2^+} [x]$.   3. $\lim_{x \to \sqrt{2}^+} [x]$.   5. $\lim_{x \to 4^+} [\sqrt{x}]$.

2. $\lim_{x \to 2^-} [x]$.   4. $\lim_{x \to \sqrt{2}^-} [x]$.   6. $\lim_{x \to 4^-} [\sqrt{x}]$.

Exercise 1.3.6. Determine whether limits exist.

1. $\lim_{x \to 0} \begin{cases} \sqrt{x} & \text{if } x > 0 \\ -\sqrt{-x} & \text{if } x < 0 \end{cases}$.   3. $\lim_{x \to 1} \begin{cases} 1 & \text{if } x > 1 \\ x & \text{if } x < 1 \end{cases}$.

2. $\lim_{x \to 1} \begin{cases} 2 & \text{if } x > 1 \\ x & \text{if } x < 1 \end{cases}$.   4. $\lim_{x \to 1} \begin{cases} 1 & \text{if } x > 2 \\ x & \text{if } x < 2 \end{cases}$.

   Another variation of the function limit is to consider what happens when x gets very big.

Definition 1.3.4. A function $f(x)$ has limit l at ∞, denoted $\lim_{x \to \infty} f(x) = l$, if for any $\epsilon > 0$, there is N, such that

$$|x| > N \implies |f(x) - l| < \epsilon. \tag{1.3.4}$$

   The limit at the infinity can also be split into the limits $\lim_{x \to +\infty} f(x)$, $\lim_{x \to -\infty} f(x)$ at positive and negative infinities. Proposition 1.3.3 still holds for the limit at infinity.

   The index n in sequences are usually taken as positive numbers only, so that the sequence limit $\lim_{n \to \infty} x_n$ really means $\lim_{n \to +\infty} x_n$. For the function limit, if only ∞ is used without sign, then both negative and positive infinities are included.

Example 1.3.7. We verify that $\lim_{x \to \infty} \dfrac{2x + 1}{x - 1} = 2$. For any $\epsilon > 0$, take $N = \dfrac{3}{\epsilon} + 1$. Then

$$|x| > N \implies |x - 1| \ge |x| - 1 > \frac{3}{\epsilon} \implies \left| \frac{2x + 1}{x - 1} - 2 \right| = \frac{3}{|x - 1|} < \epsilon.$$

Example 1.3.8. By treating n in Example 1.1.1 as a real number instead of an integer, the same argument leads to $\lim_{x \to \infty} \dfrac{1}{x} = 0$. More generally, the argument for the limit (1.1.2) also works and gives us

$$\lim_{x \to +\infty} \frac{1}{x^{\alpha}} = 0 \ \text{ for } \ \alpha > 0. \tag{1.3.5}$$

The limit can be rephrased as

$$\lim_{x \to +\infty} x^{\alpha} = 0 \ \text{ for } \ \alpha < 0. \tag{1.3.6}$$

Note that only $+\infty$ is considered here because x^α may not be defined for negative x and non-integer α. Of course in the special case n is a natural number, the same argument leads to

$$\lim_{x \to \infty} \frac{1}{x^n} = 0 \text{ for natural number } n. \tag{1.3.7}$$

Example 1.3.9. The analogue of the limit (1.1.3) is

$$\lim_{x \to +\infty} \alpha^x = 0 \text{ for } 0 < \alpha < 1. \tag{1.3.8}$$

The proof, however, needs to be modified. Again let $\dfrac{1}{\alpha} = 1 + \beta$. Then $\beta > 0$. For any $\epsilon > 0$, take $N = \dfrac{1}{\beta\epsilon} + 1$. Then

$$x > N \implies x > n \text{ for some integer } n \text{ satisfying } x > n > \frac{1}{\beta\epsilon}$$

$$\implies \frac{1}{\alpha^x} > \frac{1}{\alpha^n} > n\beta > \frac{1}{\epsilon}$$

$$\implies 0 < \alpha^x < \epsilon,$$

where the inequality $\dfrac{1}{\alpha^n} > n\beta$ was proved in Example 1.2.12.

Exercise 1.3.7. Rigorously verify the limits.

1. $\lim_{x \to \infty} \dfrac{x}{x^2 + 1} = 0.$

2. $\lim_{x \to \infty} \dfrac{\sin x}{x} = 0.$

3. $\lim_{x \to \infty} \dfrac{x^2 - 1}{x^2 + 1} = 1.$

4. $\lim_{x \to \infty} \dfrac{x^2 + x - 1}{x^2 + 1} = 1.$

Exercise 1.3.8. Prove

$$\lim_{x \to -\infty} \alpha^x = 0 \text{ for } \alpha > 1. \tag{1.3.9}$$

Exercise 1.3.9. Prove the extension

$$\lim_{x \to +\infty} x^\beta \alpha^x = 0 \text{ for } 0 < \alpha < 1 \tag{1.3.10}$$

of the limit (1.3.8).

The final variation is the divergence to infinity.

Definition 1.3.5. A function $f(x)$ diverges to *infinity* at a, denoted $\lim_{x \to a} f(x) = \infty$, if for any b, there is $\delta > 0$, such that

$$0 < |x - a| < \delta \implies |f(x)| > b. \tag{1.3.11}$$

The divergence to positive and negative infinities, denoted $\lim_{x \to a} f(x) = +\infty$ and $\lim_{x \to a} f(x) = -\infty$ respectively, can be similarly defined. Moreover, the divergence to infinity at the left of a, the right of a, or when a is various kind of infinities, can also be similarly defined.

Similar to the sequences diverging to infinity, we know $f(x)$ is an infinity if and only if $\lim_{x \to a} \dfrac{1}{f(x)} = 0$, i.e., the reciprocal is an *infinitesimal*.

Example 1.3.10. We verify that $\lim_{x\to 1^-} \dfrac{x}{x^2 - 1} = -\infty$. For any $b > 0$, choose $\delta = \min\left\{\dfrac{1}{2}, \dfrac{1}{b}\right\}$. Then

$$-\delta < x - 1 < 0 \implies -\delta < x - 1 < 0, \ \ x > 1 - \delta \geq \frac{1}{2}, \ \ x + 1 < 2$$

$$\implies \frac{x}{x^2 - 1} = \frac{x}{(x+1)(x-1)} < \frac{\frac{1}{2}}{2(-\delta)} < -\frac{b}{4}.$$

Since b can be arbitrarily big, $\dfrac{b}{4}$ can also be arbitrarily big. Therefore $\dfrac{x}{x^2 - 1}$ is a negative infinity at the left of 1.

Example 1.3.11. Taking the reciprocals of the infinitesimals in the Examples 1.3.6 and 1.3.8, we get some infinities. By considering the sign, we actually get positive or negative infinities. Combining the results with earlier examples, we get the limits of the power functions

$$\lim_{x\to 0^+} x^\alpha = \begin{cases} 0 & \text{if } \alpha > 0 \\ 1 & \text{if } \alpha = 0 \\ +\infty & \text{if } \alpha < 0 \end{cases}, \quad \lim_{x\to +\infty} x^\alpha = \begin{cases} 0 & \text{if } \alpha < 0 \\ 1 & \text{if } \alpha = 0 \\ +\infty & \text{if } \alpha > 0 \end{cases}. \qquad (1.3.12)$$

Taking the reciprocal of the limits of the exponential function in Example 1.3.9 and in (1.3.9) give us

$$\lim_{x\to +\infty} \alpha^x = \begin{cases} 0 & \text{if } 0 < \alpha < 1 \\ 1 & \text{if } \alpha = 1 \\ +\infty & \text{if } \alpha > 1 \end{cases}, \quad \lim_{x\to -\infty} \alpha^x = \begin{cases} 0 & \text{if } \alpha > 1 \\ 1 & \text{if } \alpha = 1 \\ +\infty & \text{if } 0 < \alpha < 1 \end{cases}. \qquad (1.3.13)$$

Exercise 1.3.10. Rigorously verify the limits.

1. $\lim_{x\to 0} \dfrac{1-x}{x^2(1+x)} = +\infty$.

2. $\lim_{x\to 0^+} \dfrac{1-x}{x} = +\infty$.

3. $\lim_{x\to -\infty} \dfrac{x^2 + 1}{x + 1} = -\infty$.

4. $\lim_{x\to +\infty} \dfrac{x^2 + 1}{2x - 5} = +\infty$.

1.3.3   Property

The properties of the limit of sequences can be extended to functions.

Proposition 1.3.6. *The limit of functions have the following properties.*

1. *Boundedness: A function convergent at a is bounded near a.*

2. *Arithmetic: Suppose $\lim_{x\to a} f(x) = l$ and $\lim_{x\to a} g(x) = k$. Then*

$$\lim_{x\to a}(f(x) + g(x)) = l + k, \ \ \lim_{x\to a} f(x)g(x) = lk, \ \ \lim_{x\to a} \frac{f(x)}{g(x)} = \frac{l}{k},$$

where $g(x) \neq 0$ and $k \neq 0$ are assumed in the third equality.

3. *Order: Suppose* $\lim_{x \to a} f(x) = l$ *and* $\lim_{x \to a} g(x) = k$. *If* $f(x) \geq g(x)$ *for x close to a, then $l \geq k$. Conversely, if $l > k$, then $f(x) > g(x)$ for x close to a.*

4. *Sandwich: Suppose* $f(x) \leq g(x) \leq h(x)$ *and* $\lim_{x \to a} f(x) = \lim_{x \to a} h(x) = l$. *Then* $\lim_{x \to a} g(x) = l$.

5. *Composition: Suppose* $\lim_{x \to a} g(x) = b$, $\lim_{y \to b} f(y) = c$. *If* $g(x) \neq b$ *for x near a, or $f(b) = c$, then $\lim_{y \to a} f(g(x)) = c$.*

When something happens near a, we mean that there is $\delta > 0$, such that it happens when $0 < |x - a| < \delta$. For example, $f(x)$ is bounded near a if there is $\delta > 0$ and B, such that $0 < |x - a| < \delta$ implies $|f(x)| < B$.

The first four properties are parallel to the Propositions 1.1.2, 1.1.3, 1.1.4, 1.1.5 for the limit of sequences, and can be proved in the similar way. The fifth property means that for $y = g(x)$ and $z = f(y)$, we have

$$\lim_{x \to a} y = b, \ \lim_{y \to b} z = c \implies \lim_{x \to a} z = c.$$

Its analogue for the sequence is Proposition 1.2.1. The property can be proved by combining the following two implications together

$$0 < |x - a| < \delta \implies |g(x) - b| < \mu,$$
$$0 < |y - b| < \mu \implies |f(y) - c| < \epsilon,$$

where μ is found for the given ϵ, and then δ is found for the given μ. However, there is the technical problem that the right side of the first implication does not quite match the left side of the second implication. To make them match, the implications should be modified either as

$$0 < |x - a| < \delta \implies 0 < |g(x) - b| < \mu,$$
$$0 < |y - b| < \mu \implies |f(y) - c| < \epsilon,$$

or as

$$0 < |x - a| < \delta \implies |g(x) - b| < \mu,$$
$$|y - b| < \mu \implies |f(y) - c| < \epsilon.$$

The first modification simply additionally requires that $g(x) \neq b$ for x near a. The second modification simply additionally requires that $f(b) = c$.

Example 1.3.12. A *polynomial* is a function of the form

$$p(x) = c_n x^n + c_{n-1} x^{n-1} + \cdots + c_1 x + c_0$$

By repeatedly applying the arithmetic rule to $\lim_{x \to a} c = c$ (c is a constant) and $\lim_{x \to a} x = a$, we get

$$\lim_{x \to a} p(x) = c_n a^n + c_{n-1} a^{n-1} + \cdots + c_1 a + c_0 = p(a). \tag{1.3.14}$$

More generally, a *rational function*

$$r(x) = \frac{c_n x^n + c_{n-1} x^{n-1} + \cdots + c_1 x + c_0}{d_m x^m + d_{m-1} x^{m-1} + \cdots + d_1 x + d_0}$$

is a quotient of two polynomials. By applying the arithmetic rule to (1.3.14), we get

$$\lim_{x \to a} r(x) = \frac{c_n a^n + c_{n-1} a^{n-1} + \cdots + c_1 a + c_0}{d_m a^m + d_{m-1} a^{m-1} + \cdots + d_1 a + d_0} = r(a), \qquad (1.3.15)$$

as long as the denominator is nonzero.

Example 1.3.13. We have $\lim_{x \to 1}(x^2 + x + 2) = 4$ by Example 1.3.12 and $\lim_{x \to 4} \sqrt{x} = 2$ by Example 1.3.2. Then we get $\lim_{x \to 1} \sqrt{x^2 + x + 2} = 2$ by the composition rule.

Proposition 1.3.6 was stated for the two side limit $\lim_{x \to a} f(x) = l$ with finite a and l only. The properties also hold when a is replaced by a^+, a^-, ∞, $+\infty$ and $-\infty$. Some properties still hold when l is replaced by various infinities. Here are some examples.

1. The arithmetic rule $(+\infty) + (+\infty) = +\infty$ for sequences can be extended to functions. For example, if $\lim_{x \to a^+} f(x) = +\infty$ and $\lim_{x \to a^+} g(x) = +\infty$, then $\lim_{x \to a^+}(f(x) + g(x)) = +\infty$. In general, all the valid arithmetic rules for sequences that involve infinities and infinitesimals are still valid for functions. However, as in the sequence case, the same care needs to be taken in applying the arithmetic rules to infinities and infinitesimals.

2. In the composition rule, a, b, c can be basically any symbols. For example, if $\lim_{x \to \infty} g(x) = b$, $g(x) > b$ and $\lim_{y \to b^+} f(y) = c$, then $\lim_{x \to \infty} f(g(x)) = c$.

3. If $f(x) \geq g(x)$ and $\lim_{x \to a} g(x) = +\infty$, then $\lim_{x \to a} f(x) = +\infty$. This extends the sandwich rule to $+\infty$. There is similar sandwich rule for $-\infty$ but no sandwich rule for ∞.

Example 1.3.14. By repeatedly applying the arithmetic rule to $\lim_{x \to \infty} c = c$ (c is a constant) and $\lim_{x \to \infty} \dfrac{1}{x} = 0$, we get

$$\lim_{x \to \infty} \frac{2x^5 + 10}{-3x + 1} = \lim_{x \to \infty} x^4 \frac{2 + 10\dfrac{1}{x}}{-3 + \dfrac{1}{x}} = (+\infty)\frac{2 + 10 \cdot 0}{-3 + 0} = -\infty,$$

$$\lim_{x \to \infty} \frac{x^3 - 2x^2 + 1}{-x^3 + 1} = \lim_{x \to \infty} \frac{1 - 2\dfrac{1}{x} + \dfrac{1}{x^3}}{-1 + \dfrac{1}{x^3}} = \frac{2 - 2 \cdot 0 + 0^3}{-1 + 0} = -2.$$

In general, we have the limit

$$\lim_{x \to \infty} \frac{c_n x^n + c_{n-1} x^{n-1} + \cdots + c_1 x + c_0}{d_m x^m + d_{m-1} x^{m-1} + \cdots + d_1 x + d_0} = \begin{cases} 0 & \text{if } m > n, d_m \neq 0 \\ \dfrac{c_n}{d_m} & \text{if } m = n, d_m \neq 0 \\ \infty & \text{if } m < n, c_n \neq 0 \end{cases} \qquad (1.3.16)$$

of a rational function at ∞.

Example 1.3.15. The power function x^α is defined for any α and $x > 0$. In Example 1.3.11, we obtained the limits at 0^+ and $+\infty$. Now we may consider the limit at any finite $a > 0$.

Fix a natural number $n > |\alpha|$. Then for $x > 1$, we have $\dfrac{1}{x^n} < x^\alpha < x^n$. By $\lim_{x\to 1} x^n = 1$, $\lim_{x\to 1} \dfrac{1}{x^n} = 1$ from (1.3.15) and the sandwich rule, we get $\lim_{x\to 1+} x^\alpha = 1$. Note that the limit is taken from the right because the sandwich inequality holds for $x > 1$ only. Similarly, from the inequality $\dfrac{1}{x^n} > x^\alpha > x^n$ for $0 < x < 1$, we get $\lim_{x\to 1-} x^\alpha = 1$. Thus we conclude that $\lim_{x\to 1} x^\alpha = 1$. For the limit of the power function at any $a > 0$, the composition rule gives us

$$\lim_{x\to a} x^\alpha = \lim_{x\to 1} (ax)^\alpha = a^\alpha \lim_{x\to 1} x^\alpha = a^\alpha. \qquad (1.3.17)$$

Example 1.3.16. Suppose $\lim_{x\to 0+} f(x) = l$. By $\lim_{x\to 0} x^2 = 0$, $x^2 > 0$ for $x \neq 0$, and the composition rule, we have $\lim_{x\to 0} f(x^2) = l$.

Conversely, suppose $\lim_{x\to 0} f(x^2) = l$. Then by the composition rule and $\lim_{x\to 0+} \sqrt{x} = 0$ in (1.3.12), we get $\lim_{x\to 0+} f(x) = \lim_{x\to 0+} f((\sqrt{x})^2) = l$.

Thus we conclude that $\lim_{x\to 0} f(x^2) = \lim_{x\to 0+} f(x)$. The equality means that one limit exists if and only if the other also exists. Moreover, the two limits are equal when they exist.

By the extension of the composition rule, the equality $\lim_{x\to 0} f(x^2) = \lim_{x\to 0+} f(x)$ also holds even if the limits diverge to infinity.

Exercise 1.3.11. Rewrite the limits as $\lim_{x\to a} f(x)$ for suitable a.

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

2. $\lim_{x\to 0} f\left(\dfrac{1}{x}\right)$.

3. $\lim_{x\to 0} f((x+1)^3)$.

4. $\lim_{x\to 2+} f\left(\dfrac{1}{x}\right)$.

Exercise 1.3.12. Compute the limits.

1. $\lim_{x\to 1} \dfrac{x^2 + 3\sqrt{x} + 1}{(\sqrt{x} + 1)(\sqrt{x} + 2)}$.

2. $\lim_{x\to -3} \sqrt[3]{x - 5}$.

3. $\lim_{x\to 0} \dfrac{\sqrt{2x+1} - \sqrt{x+1}}{\sqrt{x+2} - \sqrt{2x+2}}$.

4. $\lim_{x\to +\infty} \dfrac{\sqrt{2x+1} - \sqrt{x+1}}{\sqrt{x+2} - \sqrt{2x+2}}$.

5. $\lim_{x\to 1} \dfrac{\sqrt[3]{x} - 1}{x - 1}$.

6. $\lim_{x\to 0} \dfrac{x}{\sqrt[3]{2x+1} - \sqrt[3]{x+1}}$.

7. $\lim_{x\to +\infty} \dfrac{3x^{\frac{8}{3}} + 5x^{\frac{5}{3}} + 2}{x^{\frac{5}{2}} - 3x^{\frac{3}{2}} + 2x^{\frac{1}{2}}}$.

8. $\lim_{x\to 1+} \dfrac{3x^{\frac{8}{3}} + 5x^{\frac{5}{3}} + 2}{x^{\frac{5}{2}} - 3x^{\frac{3}{2}} + 2x^{\frac{1}{2}}}$.

9. $\lim_{x\to +\infty} x\left(\sqrt{\dfrac{x+1}{x-1}} - 1\right)$.

10. $\lim_{x\to 1+} x\left(\sqrt{\dfrac{x+1}{x-1}} - 1\right)$.

11. $\lim_{x\to +\infty} \dfrac{x + 1 - \sqrt{x^2 + 3}}{x - 1}$.

12. $\lim_{x\to 1+} \dfrac{x + 1 - \sqrt{x^2 + 3}}{x - 1}$.

Exercise 1.3.13. Prove the properties of limit.

1. If $\lim_{x\to a} f(x) = l$ and $\lim_{x\to a} g(x) = k$, then $\lim_{x\to a} \max\{f(x), g(x)\} = \max\{l, k\}$ and $\lim_{x\to a} \min\{f(x), g(x)\} = \min\{l, k\}$.

2. If $\lim_{x\to a^+} f(x) = \infty$ and there are $c > 0$ and $\delta > 0$, such that $0 < x - a < \delta$ implies $g(x) > c$, then $\lim_{x\to a^+} f(x)g(x) = \infty$.

3. If $\lim_{x\to a} g(x) = +\infty$ and $\lim_{y\to+\infty} f(y) = c$, then $\lim_{x\to a} f(g(x)) = c$.

4. If $f(x) \le g(x)$ and $\lim_{x\to+\infty} g(x) = -\infty$, then $\lim_{x\to+\infty} f(x) = -\infty$.

1.3.4   Limit of Trignometric Function

We already know the limits (1.3.12) and (1.3.17) of the power function x^α. We also know the limits (1.3.15) and (1.3.16) of the polynomials and rational functions. In this section, we will study the limits of trigonometric functions.

We begin by establishing an important inequality. In Figure 1.5 are a circle of radius 1 and an angle $0 < x < \dfrac{\pi}{2}$. The arc PB has length x, and the area of the fan OBP is $\dfrac{1}{2}x$. Both the triangles OBP and OBQ have the base $OB = 1$ and respective heights $AP = \sin x$, $BQ = \tan x$. Since the fan OBP is sandwiched between the triangles OBP and OBQ, we have the inequality between the areas

$$\frac{1}{2}\sin x < \frac{1}{2}x < \frac{1}{2}\tan x \quad \text{for } 0 < x < \frac{\pi}{2}.$$

Therefore we get

$$0 < \sin x < x \quad \text{for } 0 < x < \frac{\pi}{2}, \tag{1.3.18}$$

and

$$\cos x < \frac{\sin x}{x} < 1 \quad \text{for } 0 < x < \frac{\pi}{2}. \tag{1.3.19}$$
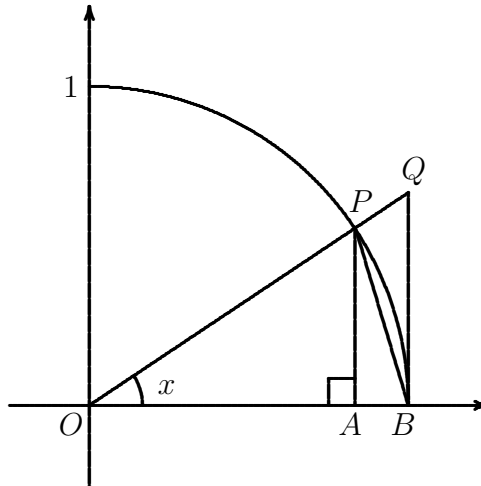
Figure 1.5: trignometric function

Applying the sandwich rule to (1.3.18), we get $\lim_{x\to 0^+} \sin x = 0$. By the composition rule, $\lim_{x\to 0^-} \sin x = \lim_{-x\to 0^-} \sin(-x) = -\lim_{x\to 0^+} \sin x = 0$. Thus by Proposition 1.3.3,

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

This further implies

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

Then by the addition formulae for the sine and the cosine functions, we have

$$\lim_{x\to a}\sin x = \lim_{x\to 0}\sin(a+x) = \lim_{x\to 0}(\sin a \cos x + \cos a \sin x)$$
$$= \sin a \cdot 1 + \cos a \cdot 0 = \sin a, \tag{1.3.20}$$
$$\lim_{x\to a}\cos x = \lim_{x\to 0}\cos(a+x) = \lim_{x\to 0}(\cos a \cos x + \sin a \sin x)$$
$$= \cos a \cdot 1 + \sin a \cdot 0 = \cos a. \tag{1.3.21}$$

Finally, by the arithmetic rule, we get

$$\lim_{x\to a}\tan x = \frac{\lim_{x\to a}\sin x}{\lim_{x\to a}\cos x} = \frac{\sin a}{\cos a} = \tan a, \tag{1.3.22}$$

and similarly for the limits of the other trignometric functions.

Applying $\lim_{x\to 0}\cos x = 1$ and the sandwich rule to (1.3.19), we get

$$\lim_{x\to 0}\frac{\sin x}{x} = 1. \tag{1.3.23}$$

This implies

$$\lim_{x\to 0}\frac{\tan x}{x} = \frac{\lim_{x\to 0}\dfrac{\sin x}{x}}{\lim_{x\to 0}\cos x} = 1. \tag{1.3.24}$$

$$\lim_{x\to 0}\frac{1-\cos x}{x^2} = \lim_{x\to 0}\frac{2\sin^2\frac{x}{2}}{x^2} = \lim_{x\to 0}\frac{2\sin^2 x}{(2x)^2} = \frac{1}{2}\left(\lim_{x\to 0}\frac{\sin x}{x}\right)^2 = \frac{1}{2}. \tag{1.3.25}$$

Example 1.3.17. In the limits (1.3.23) and (1.3.24), we substitute x by $x - \dfrac{\pi}{2}$ and get (the composition rule is used here)

$$\lim_{x\to\frac{\pi}{2}}\frac{\cos x}{x - \dfrac{\pi}{2}} = -1, \quad \lim_{x\to\frac{\pi}{2}}\frac{1}{\left(x - \dfrac{\pi}{2}\right)\tan x} = -1.$$

Taking the reciprocal of the second limit, we get

$$\lim_{x\to\frac{\pi}{2}}\left(x - \frac{\pi}{2}\right)\tan x = -1.$$

Example 1.3.18. Since $\lim_{x\to+\infty}(\sqrt{x+1} - \sqrt{x}) = \lim_{x\to+\infty}\dfrac{1}{\sqrt{x+1}+\sqrt{x}} = 0.$ By the composition rule, we have

$$\lim_{x\to+\infty}\sin\frac{\sqrt{x+1}-\sqrt{x}}{2} = 0, \quad \lim_{x\to+\infty}\frac{\sin(\sqrt{x+1}-\sqrt{x})}{\sqrt{x+1}-\sqrt{x}} = 1.$$

Then by the fact that cos is bounded, we have

$$\lim_{x\to+\infty}(\sin\sqrt{x+1} - \sin\sqrt{x}) = \lim_{x\to+\infty}2\sin\frac{\sqrt{x+1}-\sqrt{x}}{2}\cos\frac{\sqrt{x+1}+\sqrt{x}}{2} = 0.$$

Moreover,

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

$$= \lim_{x \to +\infty} \frac{\sqrt{x}}{\sqrt{x+1} + \sqrt{x}} = \lim_{x \to +\infty} \frac{1}{\sqrt{1 + \dfrac{1}{x}} + 1} = \frac{1}{2}.$$

Exercise 1.3.14. Compute the limits.

1. $\lim_{x \to \pi} \dfrac{\sin x}{x - \pi}$.

2. $\lim_{x \to \infty} \dfrac{\sin x}{x - \pi}$.

3. $\lim_{x \to 0} \dfrac{\sin x}{x - \pi}$.

4. $\lim_{x \to -\frac{3\pi}{2}} \dfrac{x + \dfrac{3\pi}{2}}{\cos x}$.

5. $\lim_{x \to \pi} \dfrac{1 + \cos x}{(x - \pi)^2}$.

6. $\lim_{x \to 0} \dfrac{\tan 2x}{\tan 3x}$.

7. $\lim_{x \to 0} \dfrac{\sin(\tan x)}{x}$.

8. $\lim_{x \to 0} \dfrac{1 - \cos x}{\sin x}$.

9. $\lim_{x \to 0} \dfrac{1 - \cos x}{\sin x} \sin \dfrac{1}{x^2}$.

10. $\lim_{x \to 0} \dfrac{\tan x - \sin x}{x^3}$.

11. $\lim_{x \to 0} \dfrac{\sin x^2}{(\sin x)^2}$.

12. $\lim_{x \to 0} \dfrac{1 - \cos x^2}{(1 - \cos x)^2}$.

13. $\lim_{x \to 0} \dfrac{\cos x - \cos 2x}{x^2}$.

14. $\lim_{x \to 0} \dfrac{\sqrt{\cos x} - \sqrt[3]{\cos x}}{\sin^2 x}$.

1.3.5   Limit of Exponential Function

For $a > 0$, we know $\lim_{n \to \infty} a^{\frac{1}{n}} = 1$. This suggests that

$$\lim_{x \to 0} a^x = 1. \tag{1.3.26}$$

For the case $a \geq 1$ and $x > 0$, we will prove this by comparing with the known limit of $a^{\frac{1}{n}}$. In other words, we compare x with the reciprocals of natural numbers.

For $0 < x < 1$, we have $\dfrac{1}{n + 1} \leq x < \dfrac{1}{n}$ for some natural number n. Moreover, the smaller x is, the bigger n will be (the precise estimation is given in the formal argument below). This implies

$$1 \leq a^{\frac{1}{n+1}} \leq a^x \leq a^{\frac{1}{n}}.$$

Repeating the proof of the sandwich rule, by the limit $\lim_{n \to \infty} a^{\frac{1}{n}} = 1$, for any $\epsilon > 0$, there is N, such that $n > N$ implies $|a^{\frac{1}{n}} - 1| < \epsilon$. Then

$$0 < x < \delta = \frac{1}{N + 1} \implies \frac{1}{n + 1} \leq x < \frac{1}{n} \text{ for some natural number } n > N$$

$$\implies |a^x - 1| \leq |a^{\frac{1}{n}} - 1| < \epsilon.$$

This completes the proof of $\lim_{x \to 0^+} a^x = 1$ in case $a \geq 1$.

For the case $0 < a \leq 1$, we have $\dfrac{1}{a} \geq 1$, and the arithmetic rule gives us

$$\lim_{x \to 0^+} a^x = \frac{1}{\lim_{x \to 0^+} \left(\frac{1}{a}\right)^x} = 1.$$

Further by the composition rule, we have

$$\lim_{x \to 0^-} a^x = \lim_{x \to 0^+} a^{-x} = \frac{1}{\lim_{x \to 0^+} a^x} = 1.$$

This completes the proof of the limit (1.3.26) in all cases.

The limit (1.3.26) is a special case of $\lim_{x \to b} a^x$, which we expect to be a^b. The equality $\lim_{x \to b} a^x = a^b$ can be derived from the following general result.

Proposition 1.3.7 (Exponential Rule). *Suppose $\lim_{x \to a} f(x) = l > 0$ and $\lim_{x \to a} g(x) = k$. Then $\lim_{x \to a} f(x)^{g(x)} = l^k$.*

Proof. We prove the case a is finite. The proof for the other cases is similar.

By the limit (1.3.17) and the composition rule, we have $\lim_{x \to a} f(x)^k = l^k$. On the other hand, choose $A > 1$ satisfying $A^{-1} < l < A$. Then $\lim_{x \to a} f(x) = l$ tells us that there is $\delta > 0$, such that $0 < |x - a| < \delta$ implies $A^{-1} < f(x) < A$. This further implies

$$A^{-|g(x)-k|} < f(x)^{g(x)-k} < A^{|g(x)-k|}.$$

The assumption $\lim_{x \to a} g(x) = k$ implies $\lim_{x \to a} |g(x) - k| = 0$. By the limit (1.3.26) and the composition rule, we have

$$\lim_{x \to a} A^{-|g(x)-k|} = \lim_{x \to a} A^{|g(x)-k|} = 1.$$

Then by the sandwich rule, we get $\lim_{x \to a} f(x)^{g(x)-k} = 1$. Thus we conclude that

$$\lim_{x \to a} f(x)^{g(x)} = \lim_{x \to a} f(x)^k \lim_{x \to a} f(x)^{g(x)-k} = l^k \cdot 1 = l^k.$$

$\square$

Next we establish the limit

$$\lim_{x \to 0^+} x^x = 1, \tag{1.3.27}$$

which is not covered by the exponential rule above. Note that for the special case $x = \dfrac{1}{n}$, we have $x^x = \dfrac{1}{\sqrt[n]{n}}$. So the limit is closely related to the limit (1.1.4). This suggests us to prove (1.3.27) by comparing with (1.1.4). We have

$$\frac{1}{n+1} \leq x < \frac{1}{n} \implies 1 > x^x > \left(\frac{1}{n+1}\right)^{\frac{1}{n}} = \frac{1}{(n+1)^{\frac{1}{n}}}$$

$$\implies |x^x - 1| < \left| \frac{1}{(n+1)^{\frac{1}{n}}} - 1 \right|.$$

Based on the estimation and the known limit $\lim_{n\to\infty} \dfrac{1}{(n+1)^{\frac{1}{n}}} = 1$ (see (1.1.7)), an argument for the limit (1.3.27) can be made similar to the proof of (1.3.26).

Other forms of the exponential rule can be found in Exercise 1.3.19.

Example 1.3.19. By the limit (1.3.27), we have $\lim_{x\to 0^-} |x|^x = \lim_{x\to 0^+} x^{-x} = \dfrac{1}{\lim_{x\to 0^+} x^x} = 1$. Thus $\lim_{x\to 0} |x|^x = 1$.

Example 1.3.20. By the limit (1.3.27) and the exponential rule, we have

$$\lim_{x\to 0} (5 - 3x + 4x^2)^x = 5^0 = 1.$$

$$\lim_{x\to 0} (2x^2 - x^3)^x = \lim_{x\to 0} (|x|^x)^2 (2 - x)^x = 1^2 \cdot 2^0 = 1.$$

$$\lim_{x\to 0^+} (2x + x^3)^{\sqrt{x}} = \lim_{x\to 0^+} (2x^2 + x^6)^x = \lim_{x\to 0} (|x|^x)^2 (2 + x^4)^x = 1^2 \cdot 2^0 = 1.$$

Example 1.3.21. By the limits (1.3.23) and the exponential rule, we have

$$\lim_{x\to 0^+} (\sin x)^x = \lim_{x\to 0^+} \left(\frac{\sin x}{x}\right)^x x^x = 1^0 \cdot 1 = 1.$$

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

Exercise 1.3.15. Compute the limits.

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

2. $\lim_{x\to 0^+} x^{(x^x)}$.

3. $\lim_{x\to 0^+} (\tan x)^x$.

4. $\lim_{x\to 0} |x|^{\tan x}$.

5. $\lim_{x\to 0} (\cos x)^x$.

6. $\lim_{x\to +\infty} x^{-x}$.

7. $\lim_{x\to +\infty} x^{\frac{1}{x}}$.

8. $\lim_{x\to +\infty} (2^x + 3^x)^{\frac{1}{x}}$.

9. $\lim_{x\to +\infty} (2^x + 3^x)^{\frac{x}{x^2+1}}$.

Exercise 1.3.16. Prove $\lim_{x\to 0} |p(x)|^x = 1$ for any nonzero polynomial $p(x)$.

Exercise 1.3.17. Finish the proof of the limit (1.3.27).

Exercise 1.3.18. For any $a > 0$, we have $x < a < \dfrac{1}{x}$ for sufficiently small $x > 0$. Use this to derive the limit (1.3.26) from the limit (1.3.27).

Exercise 1.3.19. Prove the following exponential rules.

1. $l^{+\infty} = +\infty$ for $l > 1$: If $\lim_{x\to a} f(x) = l > 1$ and $\lim_{x\to a} g(x) = +\infty$, then $\lim_{x\to a} f(x)^{g(x)} = +\infty$.

2. $(0^+)^k = 0$ for $k > 0$: If $f(x) > 0$, $\lim_{x\to a} f(x) = 0$ and $\lim_{x\to a} g(x) = k > 0$, then $\lim_{x\to a} f(x)^{g(x)} = 0$.

From the two rules, further derive the following exponential rules.

1. $l^{+\infty} = 0$ for $0 < l < 1$.

2. $l^{-\infty} = 0$ for $l > 1$.

3. $(0^+)^k = +\infty$ for $k < 0$.

4. $(+\infty)^k = 0$ for $k > 0$.

Exercise 1.3.20. Provide counterexamples to the wrong exponential rules.

$$(+\infty)^0 = 1, \quad 1^{+\infty} = 1, \quad 0^0 = 1, \quad 0^0 = 0.$$

Finally we consider $\lim_{x \to \infty} \left(1 + \dfrac{1}{x}\right)^x$. Similar to the way the limit (1.3.27) is derived, we compare with the natural number case, which is the definition of e in Example 1.2.11.

For $x > 1$, we have $n \le x \le n + 1$ for some natural number n. This implies

$$\left(1 + \frac{1}{n+1}\right)^n \le \left(1 + \frac{1}{x}\right)^x \le \left(1 + \frac{1}{n}\right)^{n+1}. \qquad (1.3.28)$$

From Example 1.2.11, we know

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

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

Therefore for any $\epsilon > 0$, there is $N > 0$, such that

$$n > N \implies \left|\left(1 + \frac{1}{n}\right)^{n+1} - e\right| < \epsilon, \quad \left|\left(1 + \frac{1}{n+1}\right)^n - e\right| < \epsilon, \qquad (1.3.29)$$

Then for $x > N + 1$, we have $n \le x \le n + 1$ for some natural number $n > N$. By the inequalities (1.3.28) and (1.3.29), this further implies

$$-\epsilon < \left(1 + \frac{1}{n+1}\right)^n - e \le \left(1 + \frac{1}{x}\right)^x - e \le \left(1 + \frac{1}{n}\right)^{n+1} - e < \epsilon.$$

This proves that $\lim_{x \to +\infty} \left(1 + \dfrac{1}{x}\right)^x = e$, which further implies

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

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

Thus we conclude that

$$\lim_{x \to \infty} \left(1 + \frac{1}{x}\right)^x = e. \qquad (1.3.30)$$

Example 1.3.22. By the composition rule, we have

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

$$\lim_{x \to 0} (1 - x)^{\frac{1}{x}} = \lim_{x \to \infty} \left(1 + \frac{1}{x}\right)^{-x} = \left(\lim_{x \to \infty} \left(1 + \frac{1}{x}\right)^x\right)^{-1} = e^{-1}.$$

By the exponential rule and the composition rule, we have

$$\lim_{x \to -\infty} \left(1 + \frac{1}{x}\right)^{x^2} = \lim_{x \to -\infty} \left(\left(1 + \frac{1}{x}\right)^x\right)^x = e^{-\infty} = 0.$$

$$\lim_{x \to 0}(1 + x + x^2)^{\frac{1}{x}} = \lim_{x \to 0}\left((1 + x + x^2)^{\frac{1}{x+x^2}}\right)^{\frac{x+x^2}{x}}$$

$$= \left(\lim_{x \to 0}(1 + x + x^2)^{\frac{1}{x+x^2}}\right)^{\lim_{x \to 0}\frac{x+x^2}{x}}$$

$$= \left(\lim_{y \to \infty}\left(1 + \frac{1}{y}\right)^y\right)^{\lim_{x \to 0}(1+x)} = e^1 = e.$$

The examples also show that there is no exponential rule for $1^{+\infty}$ or $1^{-\infty}$.

Exercise 1.3.21. Compute the limits.

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

2. $\lim_{x \to \infty}\left(1 + \frac{1}{x^2}\right)^x$.

3. $\lim_{x \to 0}\left(\frac{2+x}{2+3x}\right)^{\frac{1}{x}}$.

4. $\lim_{x \to \infty}\left(\frac{2+x}{2+3x}\right)^{\frac{1}{x}}$.

5. $\lim_{x \to 0}\left(\frac{2+x+x^2}{2-3x+2x^2}\right)^{\frac{1}{x^2+x}}$.

6. $\lim_{x \to \infty}\left(\frac{2+x+x^2}{2-3x+2x^2}\right)^{\frac{1}{x^2+x}}$.

Exercise 1.3.22. Prove that $n \le x \le n+1$ implies $\left(1 - \frac{1}{n+1}\right)^{-n} \le \left(1 - \frac{1}{x}\right)^{-x} \le \left(1 - \frac{1}{n}\right)^{-(n+1)}$. Then use this and Exercise 1.2.19 to prove $\lim_{x \to -\infty}\left(1 + \frac{1}{x}\right)^x = e$.

1.3.6   More Property

In Section 1.3.5, we saw the close relation between the sequence limit and the function limit. In fact, we used the sequence limit to derive the function limit.

Proposition 1.3.8. *For a function $f(x)$, $\lim_{x \to a} f(x) = l$ if and only if $\lim_{n \to \infty} f(x_n) = l$ for any sequence $\{x_n\}$ satisfying $x_n \ne a$ and $\lim_{n \to \infty} x_n = a$.*

Proof. We prove the case a and l are finite. The proof for the other cases is similar.

Suppose $\lim_{x \to a} f(x) = l$. Suppose $\{x_n\}$ is a sequence satisfying $x_n \ne a$ and $\lim_{n \to \infty} x_n = a$. Then for any $\epsilon > 0$, there are $\delta > 0$ and N, such that

$$0 < |x - a| < \delta \implies |f(x) - l| < \epsilon.$$
$$n > N \implies |x_n - a| < \delta.$$

The assumption $x_n \ne a$ further implies $|x_n - a| > 0$. Thus combining the two implications together leads to

$$n > N \implies |f(x_n) - l| < \epsilon,$$

and we conclude $\lim_{n \to \infty} x_n = a$.

Conversely, assume $\lim_{x \to a} f(x) \neq l$ (which means either the limit does not exist, or the limit exists but is not equal to l). Then there is $\epsilon > 0$, such that for any $\delta > 0$, there is x satisfying $0 < |x - a| < \delta$ and $|f(x) - l| \geq \epsilon$. Specifically, by choosing $\delta = \dfrac{1}{n}$ for natural numbers n, we find a sequence $\{x_n\}$ satisfying $0 < |x_n - a| < \dfrac{1}{n}$ and $|f(x_n) - l| \geq \epsilon$. The first inequality implies $x_n \neq a$ and $\lim_{n \to \infty} x_n = a$. The second inequality implies $\lim_{n \to \infty} f(x_n) \neq l$. In conclusion, we find a sequence satisfying $x_n \neq a$ and $\lim_{n \to \infty} x_n = a$ but $\lim_{n \to \infty} f(x_n) \neq l$.

$\square$

Example 1.3.23. Consider $\lim_{x \to +\infty} \cos x$. The sequences $x_n = 2n\pi$, $y_n = (2n+1)\pi$ diverge to $+\infty$. However, the sequences $\{\cos x_n\} = \{1\}$ and $\{\cos y_n\} = \{-1\}$ converge to different limits. By Proposition 1.3.8, $\cos x$ diverges at $+\infty$.

Example 1.3.24. The *Dirichlet*[9] *function* is

$$D(x) = \begin{cases} 1 & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational} \end{cases}.$$

For any a, we can find a sequence $\{x_n\}$ of rational numbers and a sequence $\{y_n\}$ of irrational numbers convergent to but not equal to a (for $a = 0$, take $x_n = \dfrac{1}{n}$ and $y_n = \dfrac{\sqrt{2}}{n}$, for example). Then $f(x_n) = 1$ and $f(y_n) = 0$, so that $\lim_{n \to \infty} f(x_n) = 1$ and $\lim_{n \to \infty} f(y_n) = 0$. This implies that the Dirichlet function diverges everywhere.

Example 1.3.25. The restriction of the limits (1.3.23), (1.3.27), (1.3.30) to $\left\{\dfrac{1}{n}\right\}$, $\left\{\dfrac{1}{\sqrt{n}}\right\}$, $\{\sqrt{n}\}$ give

$$\lim_{n \to +\infty} n \sin \frac{1}{n} = 1.$$

$$\lim_{n \to \infty} n^{\frac{1}{\sqrt{n}}} = \left(\lim_{n \to \infty} (\sqrt{n})^{\frac{1}{\sqrt{n}}}\right)^2 = 1.$$

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

Exercise 1.3.23. Determine convergence.

1. $\lim_{x \to \frac{\pi}{2}} \tan x$.

2. $\lim_{x \to \frac{\pi}{2}} \left(x - \dfrac{\pi}{2}\right) \tan x$.

3. $\lim_{x \to \frac{\pi}{2}} \dfrac{\tan x + 1}{\tan x - 1}$.

4. $\lim_{x \to \pi} \tan x$.

---

[9] Johann Peter Gustav Lejeune Dirichlet, born 1805 in Düren (French Empire, now Germany), died in Göttingen (Germany). He proved the famous Fermat's Last Theorem for the case $n = 5$ in 1825. He made fundamental contributions to the analytic number theory, partial differential equation, and Fourier series. He introduced his famous function in 1829.

5. $\lim_{x \to 1} \sin \dfrac{x}{x^2 - 1}$.   6. $\lim_{x \to 1} (\sqrt{1 + x} - \sqrt{2x}) \sin \dfrac{x}{x^2 - 1}$.

Exercise 1.3.24. Compute the limits.

1. $\lim_{n \to \infty} \left(\dfrac{n^2 - 1}{2n + 1} \right)^{\sin \frac{1}{\sqrt{n} - 1}}$.

4. $\lim_{n \to \infty} (\sin \sqrt{n + 1} - \sin \sqrt{n})$.

2. $\lim_{n \to \infty} \left(\sin \dfrac{1}{n} \right)^{\tan \frac{1}{n}}$.

5. $\lim_{n \to \infty} \left(1 + \sin \dfrac{1}{n} \right)^n$.

3. $\lim_{n \to \infty} \left(\sin \dfrac{1}{n} \right)^{\frac{1}{\sqrt{n}}}$.

6. $\lim_{n \to \infty} \left(1 + \dfrac{(-1)^n}{n^2} \right)^n$.

Exercise 1.3.25. Prove that $\lim_{x \to a} f(x)$ converges if and only if $\lim_{n \to \infty} f(x_n)$ converges for any sequence $\{x_n\}$ satisfying $x_n \neq a$ and $\lim_{n \to \infty} x_n = a$.

Exercise 1.3.26. Prove that $\lim_{x \to a^+} f(x) = l$ if and only if $\lim_{n \to \infty} f(x_n) = l$ for any *strictly decreasing* sequence $\{x_n\}$ converging to a. Moreover, state the similar criterion for $\lim_{x \to +\infty} f(x) = l$.

Exercise 1.3.27. Prove the exponential rule for the sequence limits similar to Proposition 1.3.7: If $\lim_{n \to \infty} x_n = l > 0$ and $\lim_{n \to \infty} y_n = k$, then $\lim_{n \to \infty} x_n^{y_n} = l^k$. Moreover, extend the exponential rule similar to Exercise 1.3.19.

A function is *increasing* if

$$x > y \implies f(x) \geq f(y).$$

It is *strictly increasing* if

$$x > y \implies f(x) > f(y).$$

The concepts of *decreasing* and *strictly decreasing* functions can be similarly defined. A function is *monotone* if it is either increasing or decreasing.

Proposition 1.2.7 on the convergence of monotone sequences can be extended to the one side limits and the limits at signed infinities of monotone functions. The following result is stated for the right limit, and can be proved in similar way.

Proposition 1.3.9. *Suppose $f(x)$ is a monotone function defined and bounded for $x > a$ and x near a. Then $\lim_{x \to a^+} f(x)$ converges.*

We will use derivatives to determine whether certain functions are monotone. The method can be combined with the proposition to show the convergence of limits.

Exercise 1.3.28. Suppose $f(x)$ is an increasing function defined and *un*bounded for $x > a$ and x near a. Prove that $\lim_{x \to a^+} f(x) = -\infty$.

Finally, the Cauchy criterion can also be applied to the limit of functions.

Proposition 1.3.10. *A function $f(x)$ has limit at a if and only if for any $\epsilon > 0$, there is $\delta > 0$, such that*

$$0 < |x - a| < \delta, \ \ 0 < |y - a| < \delta \implies |f(x) - f(y)| < \epsilon.$$

Similar statements can be made for the one side limits and the limits at infinities.

Proof. For the convergence implying the Cauchy condition, the proof of Theorem 1.2.2 can be adopted without much change.

Conversely, assume $f(x)$ satisfies the Cauchy condition. Choose a sequence $\{x_n\}$ satisfying $x_n \neq a$ and $\lim_{n \to \infty} x_n = a$. We prove that $\{f(x_n)\}$ is a Cauchy sequence. For any $\epsilon > 0$, we have $\delta > 0$ as given by the Cauchy condition for $f(x)$. Then for $\delta > 0$, there is N, such that

$$n > N \implies |x_n - a| < \delta.$$

The assumption $x_n \neq a$ also implies $0 < |x_n - a| < \delta$. Combined with the Cauchy condition for $f(x)$, we get

$$m, n > N \implies 0 < |x_m - a| < \delta, \ 0 < |x_n - a| < \delta$$
$$\implies |f(x_m) - f(x_n)| < \epsilon.$$

Therefore $\{f(x_n)\}$ is a Cauchy sequence and must converge to a limit l.

Next we prove that l is also the limit of the function at a. For any $\epsilon > 0$, we have $\delta > 0$ as given by the Cauchy condition for $f(x)$. Then we can find x_n satisfying $0 < |x_n - a| < \delta$ and $|f(x_n) - l| < \epsilon$. Moreover, for x satisfying $0 < |x - a| < \delta$, we may apply the Cauchy condition to x and x_n to get $|f(x) - f(x_n)| < \epsilon$. Therefore

$$|f(x) - l| \leq |f(x_n) - l| + |f(x) - f(x_n)| < 2\epsilon.$$

This completes the proof that $\lim_{x \to a} f(x) = l$.

$\square$

Example 1.3.26. Consider $\lim_{x \to +\infty} \cos x$. For any N, we can find a natural number n such that $x = 2n\pi, y = (2n+1)\pi > N$, but $|\cos x - \cos y| = 2$. Thus the Cauchy criterion fails for $\epsilon = 2$, and $\lim_{x \to +\infty} \cos x$ diverges.

Exercise 1.3.29. Use Proposition 1.3.10 to show the divergence of limits.

1. $\lim_{x \to 1} \dfrac{1 + x}{1 - x}$.

2. $\lim_{x \to 0} \sin \dfrac{1}{x}$.

3. $\lim_{x \to \infty} \dfrac{2 + \sin x}{2 - \sin x}$.

1.3.7 Order of Infinity and Infinitesimal

The limit $\lim_{x \to a} f(x) = l$ means $f(x) - l$ is an infinitesimal. The limit $\lim_{x \to a} f(x) = \infty$ means $f(x)$ is an infinity. Thus the concept of limit is really a matter of infinitesimals and infinities.

The infinities can be compared. For example, as $x \to \infty$, the infinity x^2 gets bigger much faster than the infinity x does. In general, bigger infinities are considered to have higher *order*. The following are more specific ways of comparing infinities $f(x)$ and $g(x)$ at a. The comparisons can also be applied to sequences.

If $\lim_{x \to a} \dfrac{f(x)}{g(x)} = 0$ (or equivalently, $\lim_{x \to a} \dfrac{g(x)}{f(x)} = \infty$), then we denote $f(x) = o(g(x))$ and call $g(x)$ an *infinity of higher order* than $f(x)$. For example,

$$x^\alpha = o(x^\beta) \text{ at } +\infty \text{ for } \beta > \alpha > 0,$$
$$x^\alpha = o(x^\beta) \text{ at } 0^+ \text{ for } \beta < \alpha < 0,$$
$$a^x = o(b^x) \text{ at } +\infty \text{ for } b > a > 1,$$
$$x^\alpha = o(a^x) \text{ at } +\infty \text{ for } \alpha > 0, a > 1,$$
$$n^\alpha = o(n^\beta) \text{ for } \beta > \alpha > 0,$$
$$n^\alpha = o(a^x) \text{ for } \alpha > 0, a > 1,$$
$$a^n = o(n!),$$
$$n! = o(n^n).$$

If $|f(x)| \leq c|g(x)|$ for a constant c and x close to a, then we denote $f(x) = O(g(x))$ and say the order of $f(x)$ is no higher than that of $g(x)$. For example, $x(1 + \sin x) = O(x)$ at ∞, $n^{2+(-1)^n} = O(n^3)$. We also denote $f(x) = O(1)$ to mean that $f(x)$ is bounded, although the constant 1 is not an infinity. For example, $\sin x = O(1)$ and $\cos x = O(1)$ at ∞.

If $c_1|g(x)| \leq |f(x)| \leq c_2|g(x)|$ for constants $c_2 > c_1 > 0$ and x close to a, that is, $f(x) = O(g(x))$ and $g(x) = O(f(x))$, then $f(x)$ and $g(x)$ are *infinities of the same order*. For example, a polynomial of degree n has the same order as x^n at ∞, $x(3 + 2\sin x)$ and x have the same order at ∞.

If $\lim_{x \to a} \dfrac{f(x)}{g(x)} = 1$, then we denote $f(x) \sim g(x)$ and call $f(x)$ and $g(x)$ *equivalent infinities*. For example, $x^3 + 2x - 5 \sim x^3$ and $2^x + x^4 \sim 2^x$ at ∞, $\dfrac{3}{2(x-1)} \sim \dfrac{x+2}{x^2 - 1}$ at 1, $3^n - 2^{n+2} \sim 3^n$.

Example 1.3.27. By

$$\frac{\sqrt{x + \sqrt{x + \sqrt{x}}}}{\sqrt{x}} = \sqrt{1 + \frac{\sqrt{x + \sqrt{x}}}{x}} = \sqrt{1 + \sqrt{\frac{1}{x} + \frac{1}{x^{\frac{3}{2}}}}},$$

we get $\lim_{x \to +\infty} \dfrac{\sqrt{x + \sqrt{x + \sqrt{x}}}}{\sqrt{x}} = 1$ and conclude $\sqrt{x + \sqrt{x + \sqrt{x}}} \sim \sqrt{x}$ at $+\infty$.

Exercise 1.3.30. Compare the infinities at $+\infty$.

$$\sqrt{x + 1} - 1, \quad \sqrt[3]{\sqrt{x} + 1} - 1, \quad \sqrt{1 + \sqrt{1 + \sqrt{1 + x}}}, \quad x^2(2 + \cos x), \quad x^x, \quad 2^x.$$

Exercise 1.3.31. Find α, so that the infinities at ∞ have the same order as x^α.

$$x(2 + \cos x) + x^2(2 - \sin x), \quad \sqrt{6x^3 + 5x^5}, \quad \sqrt{x + \sqrt{x^2 + \sqrt{x^3}}}.$$

Exercise 1.3.32. Is it true that $f_1(x) = o(g(x))$ and $f_2(x) = o(g(x)) \implies f_1(x) + f_2(x) = o(g(x))$? Discuss the similar properties for product, composition, etc. Discuss the similar properties for other types of comparisons.

Exercise 1.3.33. Given a sequence of functions $f_1(x), f_2(x), \ldots, f_n(x), \ldots$ that are infinities as $x \to a$. Construct a function $f(x)$ such that $f_n(x) = o(f(x))$ for all n. In other words, $f(x)$ diverges to infinity much faster than any $f_n(x)$.

Similar to the infinities, the infinitesimals can also be compared. In general, smaller infinitesimals are considered to have higher *order*. The following comparisons are defined for infinitesimals $f(x)$ and $g(x)$ at a.

If $\lim_{x \to a} \dfrac{f(x)}{g(x)} = 0$, then we denote $f(x) = o(g(x))$ and call $f(x)$ an *infinitesimal of higher order* than $g(x)$. For example,

$$x^\beta = o(x^\alpha) \text{ at } 0^+ \text{ for } \beta > \alpha > 0,$$
$$b^x = o(a^x) \text{ at } -\infty \text{ for } b > a > 1,$$
$$n^\beta = o(n^\alpha) \text{ for } \beta < \alpha < 0,$$
$$a^n = o(n^\alpha) \text{ for } \alpha < 0, 0 < a < 1.$$

If $|f(x)| \le c|g(x)|$ for a constant c and x close to a, then we denote $f(x) = O(g(x))$ and say the order of $g(x)$ is no higher than that of $f(x)$. For example, $\sin x \sin \dfrac{1}{x} = O(x)$ at 0, $\dfrac{\cos x}{x} = O\left(\dfrac{1}{x}\right)$ at ∞.

If $c_1|g(x)| \le |f(x)| \le c_2|g(x)|$ for constants $c_2 > c_1 > 0$ and x close to a, that is, $f(x) = O(g(x))$ and $g(x) = O(f(x))$, then $f(x)$ and $g(x)$ are *infinitesimals of the same order*. For example, $x^2 + 2x - 3$ and $x - 1$ have the same order at 1, $1 - \cos x$ and x^2 have the same order at 0.

If $\lim_{x \to a} \dfrac{f(x)}{g(x)} = 1$, then we denote $f(x) \sim g(x)$ and call $f(x)$ and $g(x)$ *equivalent infinitesimals*. For example, $\sin x \sim x$ and $\tan x \sim x$ at 0, $x^2 + 2x - 3 \sim 4(x - 1)$ at 1.

Example 1.3.28. By

$$\frac{\sqrt{x + \sqrt{x + \sqrt{x}}}}{x^{\frac{1}{8}}} = \sqrt{x^{\frac{3}{4}} + \frac{\sqrt{x + \sqrt{x}}}{x^{\frac{1}{4}}}} = \sqrt{x^{\frac{3}{4}} + \sqrt{x^{\frac{1}{2}} + 1}},$$

we get $\lim_{x \to 0^+} \dfrac{\sqrt{x + \sqrt{x + \sqrt{x}}}}{x^{\frac{1}{8}}} = 1$ and conclude $\sqrt{x + \sqrt{x + \sqrt{x}}} \sim x^{\frac{1}{8}}$ at 0^+.

Example 1.3.29. The signs $o(g(x))$ and $O(g(x))$ are often used as a function $f(x)$ satisfying the indicated comparison with $g(x)$. For example, we may write

$$\sin x = x + o(x), \quad \cos x = 1 - \frac{1}{2}x^2 + o(x^2)$$

to mean that

$$\sin x = x + f_1(x), \quad \cos x = 1 - \frac{1}{2}x^2 + f_2(x),$$

with

$$\lim_{x \to 0} \frac{f_1(x)}{x} = \lim_{x \to 0} \frac{f_2(x)}{x^2} = 0.$$

Exercise 1.3.34. Compare the infinitesimals at 0.

$$\sqrt{x + 1} - 1, \quad \sqrt[3]{\sqrt{x + 1}} - 1, \quad \tan x - x, \quad \sin x + \cos x - 1, \quad \sin x \sin \frac{1}{x}.$$

Exercise 1.3.35. Find α, so that the infinitesimals at 0 have the same order as x^α.

$$\sin 2x - 2\sin x, \quad \sqrt[2]{6x^3 + 5x^5}, \quad \sqrt{x + \sqrt{x^2 + \sqrt{x^3}}}.$$

Exercise 1.3.36. Do Exercise 1.3.32 again for the infinitesimals.

Exercise 1.3.37. How do the comparisons change when we use the reciprocal $f(x) \leftrightarrow \dfrac{1}{f(x)}$ to convert between infinities and infinitesimals?

1.3.8   Additional Exercise

Extended Supremum and Extended Upper Limit

If $\lim_{x \to 0} \dfrac{f(x)}{g(x)} = 1$, then we expect $f(x_1) + f(x_2) + \cdots + f(x_n)$ and $g(x_1) + g(x_2) + \cdots + g(x_n)$ to be very close to each other when $x_1, x_2, \ldots, x_n$ are very small. The following exercises indicate some cases our expectation is fulfilled.

Exercise 1.3.38. Suppose $f(x)$ is a function on $(0, 1]$ satisfying $\lim_{x \to 0^+} \dfrac{f(x)}{x} = 1$. Prove

$$\lim_{n \to \infty} \left(f\left(\frac{1}{n^2}\right) + f\left(\frac{2}{n^2}\right) + f\left(\frac{3}{n^2}\right) + \cdots + f\left(\frac{n}{n^2}\right) \right)$$
$$= \lim_{n \to \infty} \left(\frac{1}{n^2} + \frac{2}{n^2} + \frac{3}{n^2} + \cdots + \frac{n}{n^2} \right) = \lim_{n \to \infty} \frac{n+1}{2n} = \frac{1}{2}.$$

Exercise 1.3.39. Suppose $g(x) > 0$ and $\lim_{x \to 0} \dfrac{f(x)}{g(x)} = 1$. Suppose for each natural number n, there are nonzero numbers $x_{n,1}, x_{n,2}, \ldots, x_{n,k_n}$, so that $\{x_{n,k}\}$ *uniformly converges* to 0: For any $\epsilon > 0$, there is N, such that $n > N$ implies $|x_{n,k}| < \epsilon$. Prove that if
$$\lim_{n \to \infty} (g(x_{n,1}) + g(x_{n,2}) + \cdots + g(x_{n,k_n})) = a,$$
then
$$\lim_{n \to \infty} (f(x_{n,1}) + f(x_{n,2}) + \cdots + f(x_{n,k_n})) = a.$$

Upper and Lower Limits of Functions

Suppose $f(x)$ is defined near (but not necessarily at) a. Let

$$\mathrm{LIM}_a f = \left\{ \lim_{n \to \infty} f(x_n) \colon x_n \neq a, \ \lim_{n \to \infty} x_n = a, \ \lim_{n \to \infty} f(x_n) \text{ converges} \right\}.$$

Define
$$\overline{\lim_{x \to a}} f(x) = \sup \mathrm{LIM}_a f, \quad \underline{\lim_{x \to a}} f(x) = \inf \mathrm{LIM}_a f.$$

Similar definitions can be made when a is replaced by a^+, a^-, ∞, $+\infty$ and $-\infty$.

Exercise 1.3.40. Prove the analogue of Proposition 1.2.9: $l \in \mathrm{LIM}_a f$ if and only if for any $\epsilon > 0$ and $\delta > 0$, there is x satisfying $0 < |x - a| < \delta$ and $|f(x) - l| < \epsilon$.

Exercise 1.3.41. Prove the analogue of Exercise 1.2.25: If $\lim_{k \to \infty} l_k = l$ and $l_k \in \mathrm{LIM}_a f$ for each k, then $l \in \mathrm{LIM}_a f$.

Exercise 1.3.42. Prove the analogue of Proposition 1.2.10:

1. If $l < \overline{\lim}_{x \to a} f(x)$, then for any $\delta > 0$, there is x satisfying $0 < |x - a| < \delta$ and $f(x) > l$.

2. If $l > \overline{\lim}_{x \to a} f(x)$, then there is $\delta > 0$, such that $0 < |x - a| < \delta$ implies $f(x) < l$.

The two properties completely characterize the upper limit.

Exercise 1.3.43. Prove the analogue of Proposition 1.2.11: The upper and the lower limits of $f(x)$ at a belong to $\mathrm{LIM}_a f$, and $f(x)$ converges at a if and only if the upper and the lower limits at a are equal.

Exercise 1.3.44. Prove the extension of Proposition 1.3.3:

$$\mathrm{LIM}_a f = \mathrm{LIM}_{a^+} f \cup \mathrm{LIM}_{a^-} f.$$

In particular, we have

$$\overline{\lim_{x \to a}} f(x) = \max \left\{ \overline{\lim_{x \to a^+}} f(x), \ \overline{\lim_{x \to a^-}} f(x) \right\},$$

$$\underline{\lim_{x \to a}} f(x) = \min \left\{ \underline{\lim_{x \to a^+}} f(x), \ \underline{\lim_{x \to a^-}} f(x) \right\}.$$

Exercise 1.3.45. Extend the arithmetic and order properties of upper and lower limits in Exercise 1.2.27.

Exercise 1.3.46. Prove the analogue of Exercise 1.2.30:

$$\overline{\lim_{x \to a}} f(x) = \lim_{\delta \to 0^+} \sup\{f(x) \colon 0 < |x - a| < \delta\}.$$

1.4 Continuous Function

Changing quantities are often described by functions. Most of the changes in the real world are smooth, gradual and well behaved. For example, people do not often press the brake when driving a car, and the weather do not suddenly change from summer to winter. The functions describing such well behaved changes are at least continuous.

1.4.1 Definition

A function is continuous if its graph "does not break". The graph of a function $f(x)$ may break at a for various reasons. But the breaks are always one of the two types: Either $\lim_{x \to a} f(x)$ diverges or the limit converges but the limit value is not $f(a)$.

In Figure 1.6, the function is continuous at a_4 and is not continuous at the other four points. Specifically, $\lim_{x \to a_1} f(x)$ exists but is not equal to $f(a_1)$. $\lim_{x \to a_2} f(x)$ does not exist because the left and the right limits are not the same. $\lim_{x \to a_3} f(x)$ does not exist because the function is not bounded near a_3. $\lim_{x \to a_5} f(x)$ does not exist because the left limit does not exist.

The observation leads to the following definition.
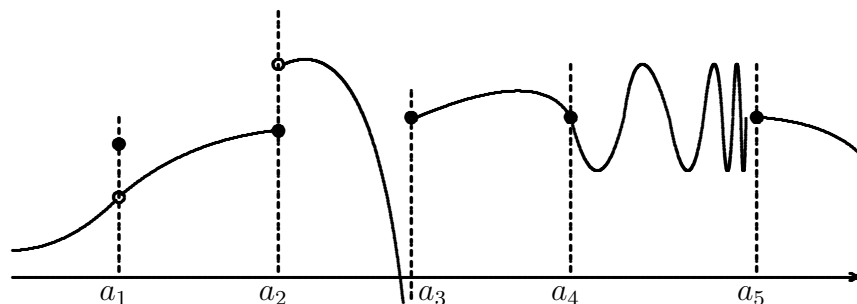
Figure 1.6: continuity and discontinuity

Definition 1.4.1. A function $f(x)$ defined near (and include) a is *continuous* at a if $\lim_{x \to a} f(x) = f(a)$.

Using the ϵ-δ language, the continuity of $f(x)$ at a means that for any $\epsilon > 0$, there is $\delta > 0$, such that

$$|x - a| < \delta \implies |f(x) - f(a)| < \epsilon. \tag{1.4.1}$$

A function is *right continuous* at a if $\lim_{x \to a^+} f(x) = f(a)$, and is *left continuous* if $\lim_{x \to a^-} f(x) = f(a)$. For example, the function in Figure 1.6 is left continuous at a_2 and right continuous at a_5. A function is continuous at a if and only if it is both left and right continuous at a.

A function defined on an open interval (a, b) is continuous if it is continuous at every point on the interval. A function defined on a closed interval $[a, b]$ is continuous if it is continuous at every point on (a, b), is right continuous at a, and is left continuous at b. Continuities for functions on other kinds of intervals can be defined similarly.

Most basic functions are continuous. The limits (1.3.15), (1.3.17), (1.3.20), (1.3.21), (1.3.22), (1.3.26) show that polynomials, rational functions, trignometric functions, power functions, and exponential functions are continuous (at the places where the functions are defined). Then the arithmetic rule and the composition rule in Proposition 1.3.6 imply the following.

Proposition 1.4.2. *The arithmetic combinations and the compositions of continuous functions are still continuous.*

By the proposition, functions such as $\sin^2 x + \tan x^2$, $\dfrac{x}{\sqrt{x}+1}$, $2^x \cos \dfrac{\tan x}{x^2+1}$ are continuous. Thus examples of discontinuity can only be found among more exotic functions.

Proposition 1.3.8 implies the following criterion for the continuity in terms of sequences.

Proposition 1.4.3. *A function $f(x)$ is continuous at a if and only if $\lim_{n \to \infty} f(x_n) = f(a)$ for any sequence $\{x_n\}$ converging to a.*

The conclusion of the proposition can be written as

$$\lim_{n \to \infty} f(x_n) = f\left(\lim_{n \to \infty} x_n\right). \tag{1.4.2}$$

Moreover, by the composition rule for the limit of functions, the continuity of $f(x)$ at $a = \lim_{y \to b} g(y)$ also means

$$\lim_{y \to b} f(g(y)) = f\left(\lim_{y \to b} g(y)\right). \tag{1.4.3}$$

Therefore the function and the limit can be interchanged *if the function is continuous.*
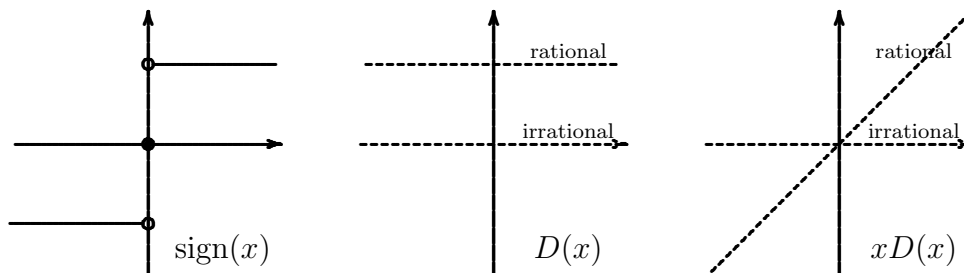
Figure 1.7: discontinuous functions

Example 1.4.1. The sign function

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

is continuous everywhere except at 0. The Dirichlet function in Example 1.3.24 is not continuous everywhere. Multiplying x to the Dirichlet function produces a function

$$xD(x) = \begin{cases} x & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational} \end{cases}$$

that is continuous only at 0.

Example 1.4.2 (Thomae[10]). For a rational number $x = \dfrac{p}{q}$, where p is an integer, q is a natural number, and p, q are coprime, define $R(x) = \dfrac{1}{q}$. For an irrational number x, define $R(x) = 0$. Finally define $R(0) = 1$. We will show that $R(x)$ is continuous at precisely all the irrational numbers.

By the way $R(x)$ is defined, for any integer N, the only numbers x satisfying $R(x) \geq \dfrac{1}{N}$ are the rational numbers $\dfrac{p}{q}$ with $q \leq N$. Therefore on any bounded interval, we have $R(x) < \dfrac{1}{N}$ for all except finitely many rational numbers. For any irrational a, let $\epsilon > 0$ be the smallest distance between a and these rational numbers. Then we have $|R(x) - R(a)| = R(x) < \dfrac{1}{N}$ for all x satisfying $|x - a| < \epsilon$. This proves that $f(x)$ is continuous at a. On the other hand, for rational $a = \dfrac{p}{q}$,

---

[10]Karl Johannes Thomae, born 1840 in Laucha (Germany), died 1921 in Jena (Germany). Thomae made important contributions to the function theory. In 1870 he showed the continuity in each variable does not imply the joint continuity. He constructed the example here in 1875.

we can find irrational numbers x arbitrarily close to a, but with $|R(x) - R(a)| = R(a) = \dfrac{1}{q}$. Since q does not depend on x, we do not have $\lim_{x \to a} R(x) = R(a)$, and the function is not continuous at the rational point.

Exercise 1.4.1. Which functions are not continuous? Where do the discontinuities happen, and for what reason?

1. $|x|$.

2. $\dfrac{x^2 \sin x}{x^2 + \sin x}$.

3. $[x] \sin x$.

4. $[x] \sin \pi x$.

5. $\begin{cases} \dfrac{|x|}{x} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$.

6. $\begin{cases} \dfrac{|x|}{x} \sin x & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$.

7. $\begin{cases} \dfrac{|x|}{x} \sin x & \text{if } x \neq 0 \\ 1 & \text{if } x = 0 \end{cases}$.

8. $\begin{cases} \sin \dfrac{1}{x} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$.

9. $\begin{cases} x \sin \dfrac{1}{x} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$.

10. $\begin{cases} x^x & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$.

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

12. $\begin{cases} x & \text{if } x \text{ is rational} \\ 1 - x & \text{if } x \text{ is irrational} \end{cases}$.

Exercise 1.4.2. Construct functions on $(0, 2)$ satisfying the requirements.

1. $f(x)$ is not continuous at $\dfrac{1}{2}$, 1 and $\dfrac{3}{2}$ and is continuous everywhere else.

2. $f(x)$ is continuous at $\dfrac{1}{2}$, 1 and $\dfrac{3}{2}$ and is not continuous everywhere else.

3. $f(x)$ is continuous everywhere except at $\dfrac{1}{n}$ for all natural numbers n.

4. $f(x)$ is not left continuous at $\dfrac{1}{2}$, not right continuous at 1, neither side continuous at $\dfrac{3}{2}$, and continuous everywhere else (including the right of $\dfrac{1}{2}$ and the left of 1).

Exercise 1.4.3. Prove that a function $f(x)$ is continuous at a if any only if there is l, such that for any $\epsilon > 0$, there is $\delta > 0$, such that

$$|x - a| < \delta \implies |f(x) - l| < \epsilon.$$

By (1.4.1), all you need to do here is to show $l = f(a)$.

Exercise 1.4.4. Prove that if a function is continuous on $(a, b]$ and $[b, c)$, then it is continuous on (a, c). What about other types of intervals?

Exercise 1.4.5. Suppose $f(x)$ and $g(x)$ are continuous. Prove $\max\{f(x), g(x)\}$ and $\min\{f(x), g(x)\}$ are also continuous.

Exercise 1.4.6. Suppose $f(x)$ is continuous on $[a, b]$ and $f(r) = 0$ for all rational numbers $r \in [a, b]$. Prove $f(x) = 0$ on the whole interval.

Exercise 1.4.7. Prove that a continuous function $f(x)$ on (a, b) is the restriction of a continuous function on $[a, b]$ if and only if the limits $\lim_{x \to a^+} f(x)$ and $\lim_{x \to b^-} f(x)$ exist.

Exercise 1.4.8. Suppose $f(x)$ is an increasing function on $[a, b]$. Prove that if any number in $[f(a), f(b)]$ can be the value of $f(x)$, then the function is continuous.

Exercise 1.4.9. Suppose $f(x)$ and $g(x)$ be continuous functions on (a, b). Find the places where the following function is continuous.

$$h(x) = \begin{cases} f(x) & \text{if } x \text{ is rational} \\ g(x) & \text{if } x \text{ is irrational} \end{cases}$$

Exercise 1.4.10. Suppose $f(x)$ is an increasing function on $[a, b]$. By Proposition 1.3.9, the limits $f(c^+) = \lim_{x \to c^+} f(x)$ and $f(c^-) = \lim_{x \to c^-} f(x)$ exist at any $c \in (a, b)$.

1. Prove that for any $\epsilon > 0$, there are finitely many c satisfying $f(c^+) - f(c^-) > \epsilon$.

2. Prove that $f(x)$ is not continuous only at countably many points.

1.4.2   Uniformly Continuous Function

Theorem 1.4.4. *Suppose $f(x)$ is a continuous function on a bounded closed interval $[a, b]$. Then for any $\epsilon > 0$, there is $\delta > 0$, such that for $x, y \in [a, b]$,*

$$|x - y| < \delta \implies |f(x) - f(y)| < \epsilon. \tag{1.4.4}$$

In the ϵ-δ formulation (1.4.1) of the continuity, only one variable x is allowed to change. This means that, in addition to being dependent on ϵ, the choice of δ may also be dependent on the location a of the continuity. The property (1.4.4) says that the choice of δ can be the *same* for all the points on the interval, so that it depends on ϵ only. Therefore the property described in the theorem is called the *uniform continuity*, and the theorem basically says a continuous function on a bounded closed interval is uniformly continuous.

Proof of Theorem 1.4.4. Suppose $f(x)$ is not uniformly continuous. Then there is $\epsilon > 0$, such that for any natural number n, there are $x_n, y_n \in [a, b]$, such that

$$|x_n - y_n| < \frac{1}{n}, \quad |f(x_n) - f(y_n)| > \epsilon. \tag{1.4.5}$$

Since the sequence $\{x_n\}$ is bounded by a and b, by Theorem 1.2.8, there is a convergent subsequence $\{x_{n_k}\}$. Then by $x_{n_k} - \frac{1}{n_k} < y_{n_k} < x_{n_k} + \frac{1}{n_k}$ and the sandwich rule, the subsequence $\{y_{n_k}\}$ also converges to the same limit. Denote $\lim_{k \to \infty} x_{n_k} = \lim_{k \to \infty} y_{n_k} = c$.

By $a \leq x_n, y_n \leq b$, we have $a \leq c \leq b$. Therefore by the assumption, $f(x)$ is continuous at c. By Proposition 1.4.3, we have

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

Then by the second inequality in (1.4.5), we have

$$\epsilon \le \left| \lim_{k \to \infty} f(x_{n_k}) - \lim_{k \to \infty} f(y_{n_k}) \right| = |f(c) - f(c)| = 0.$$

The contradiction shows that it was wrong to assume that the function is not uniformly continuous.

$\square$

Example 1.4.3. Consider the function x^2 on $[0, 2]$. For any $\epsilon > 0$, take $\delta = \dfrac{\epsilon}{4}$. Then for $0 \le x, y \le 2$, we have

$$|x - y| < \delta \implies |x^2 - y^2| = |x - y||x + y| \le 4|x - y| < \epsilon.$$

Thus x^2 is uniformly continuous on $[0, 2]$.

Now consider the same function on $[0, \infty)$. For any $\delta > 0$, take $x = \dfrac{1}{\delta}$, $y = \dfrac{1}{\delta} + \dfrac{\delta}{2}$. Then $|x - y| < \delta$, but $|x^2 - y^2| = x\delta + \dfrac{\delta^2}{4} > x\delta = 1$. Thus (1.4.4) fails for $\epsilon = 1$, and x^2 is not uniformly continuous on $[0, \infty)$.

Example 1.4.4. Consider the function $\sqrt{x}$ on $[1, \infty)$. For any $\epsilon > 0$, take $\delta = \epsilon$. Then for $x, y \ge 1$, we have

$$|x - y| < \delta \implies |\sqrt{x} - \sqrt{y}| = \frac{|x - y|}{|\sqrt{x} + \sqrt{y}|} \le \frac{|x - y|}{2} < \epsilon.$$

Thus $\sqrt{x}$ is uniformly continuous on $[1, \infty)$.

We also know $\sqrt{x}$ is continuous on $[0, 1]$, by Theorem 1.4.4. Then by using Exercise 1.4.12, $\sqrt{x}$ is uniformly continuous on $[0, \infty)$.

Exercise 1.4.11. Which are uniformly continuous?

1. $f(x) = x^2$ on $(0, 3)$.

2. $f(x) = \dfrac{1}{x}$ on $[1, 3]$.

3. $f(x) = \dfrac{1}{x}$ on $(0, 3]$.

4. $f(x) = \sqrt[3]{x}$ on $(-\infty, \infty)$.

5. $f(x) = x^{\frac{3}{2}}$ on $[1, \infty)$.

6. $f(x) = \sin x$ on $(-\infty, \infty)$.

7. $f(x) = \sin \dfrac{1}{x}$ on $(0, 1]$.

8. $f(x) = x \sin \dfrac{1}{x}$ on $(0, 1]$.

9. $f(x) = x^x$ on $(0, 1]$.

10. $f(x) = \left(1 + \dfrac{1}{x} \right)^x$ on $(0, \infty)$.

Exercise 1.4.12. Prove that if a function is uniformly continuous on $(a, b]$ and $[b, c)$, then it is uniformly continuous on (a, c). What about other types of intervals?

Exercise 1.4.13. A function $f(x)$ is called *Lipschitz*[11] if there is a constant L such that $|f(x) - f(y)| \le L|x - y|$ for any x and y. Prove that Lipschitz functions are uniformly continuous.

---

[11]Rudolf Otto Sigismund Lipschitz, born 1832 in Königsberg (Germany, now Kaliningrad, Russia), died 1903 in Bonn (Germany). He made important contributions in number theory, Fourier series, differential equations, and mechanics.

Exercise 1.4.14. A function $f(x)$ on the whole real line is called *periodic* if there is a constant p such that $f(x + p) = f(x)$ for any x (the number p is called the *period* of the function). Prove that continuous periodic functions are uniformly continuous.

Exercise 1.4.15. Is the sum of uniformly continuous functions uniformly continuous? What about the product, the maximum and the composition of uniformly continuous functions?

Exercise 1.4.16. Suppose $f(x)$ is a continuous function on $[a, b]$. Prove that $g(x) = \sup\{f(t) \colon a \le t \le x\}$ is continuous.

Exercise 1.4.17. Let $f(x)$ be a continuous function on (a, b).

1. Prove that if the limits $\lim_{x \to a^+} f(x)$, $\lim_{x \to b^-} f(x)$ exist, then $f(x)$ is uniformly continuous.

2. Prove that if (a, b) is bounded and $f(x)$ is uniformly continuous, then the limits $\lim_{x \to a^+} f(x)$, $\lim_{x \to b^-} f(x)$ exist.

3. Use $\sqrt{x}$ to show that the bounded condition in the second part is necessary.

4. Use the second part to show that $\sin \dfrac{1}{x}$ is not uniformly continuous on $(0, 1)$.

Exercise 1.4.18 (Dirichlet). For any continuous function $f(x)$ on a bounded and closed interval $[a, b]$ and $\epsilon > 0$, inductively define

$$c_0 = a, \quad c_n = \sup\{c \colon |f(x) - f(c_{n-1})| < \epsilon \text{ on } [c_{n-1}, c]\}.$$

Prove that the process must stop after finitely many steps, which means that there is n, such that $|f(x) - f(c_{n-1})| < \epsilon$ on $[c_n, b]$. Then use this to give another proof of Theorem 1.4.4.

1.4.3 Maximum and Minimum

Theorem 1.4.5. *A continuous function on a bounded closed interval must be bounded and reaches its maximum and minimum.*

The theorem says that there are $x_0, x_1 \in [a, b]$, such that $f(x_0) \le f(x) \le f(x_1)$ for any $x \in [a, b]$. Thus the values of the function are bounded by $f(x_0)$ and $f(x_1)$. Moreover, the function reaches its minimum at x_0 and reaches its maximum at x_1.

We also note that although the theorem tells us the *existence* of the maximum and the minimum, its does not tell us how to find the *extrema*. The method for finding the extrema will be developed in Section 2.2.1.

Proof of Theorem 1.4.5. Let $f(x)$ be a continuous function on a bounded closed interval $[a, b]$. We first prove the function is bounded. By Theorem 1.4.4, for $\epsilon = 1 > 0$, there is $\delta > 0$, such that $x, y \in [a, b]$ and $|x - y| < \delta$ implies $|f(x) - f(y)| < 1$. In other words, $f(x)$ is bounded by $f(y) + 1$ and $f(y) - 1$ on any open interval $(y - \delta, y + \delta)$ of length 2δ. Since the bounded interval $[a, b]$ can be covered by finitely many (any number bigger than $\dfrac{b - a}{2\delta}$ is enough) such intervals, the function is bounded on the whole interval $[a, b]$.
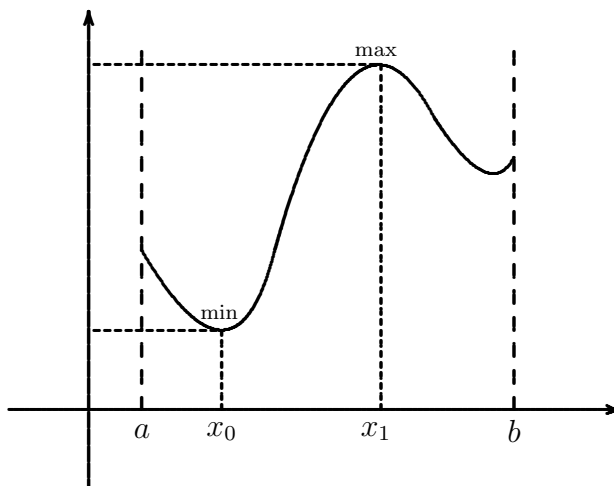
Figure 1.8: maximum and minimum

Let $\beta = \sup\{f(x)\colon x \in [a,b]\}$. Showing $f(x)$ reaches its maximum is the same as proving β is a value of $f(x)$. By the characterization of supremum, for any natural number n, there is $x_n \in [a,b]$, such that $\beta - \dfrac{1}{n} < f(x_n) \le \beta$. By the sandwich rule, we get

$$\lim_{n \to \infty} f(x_n) = \beta. \tag{1.4.6}$$

On the other hand, since the sequence $\{x_n\}$ is bounded by a and b, by Theorem 1.2.8, there is a convergent subsequence $\{x_{n_k}\}$ with $c = \lim_{k \to \infty} x_{n_k} \in [a,b]$. Then we have

$$\lim_{k \to \infty} f(x_{n_k}) = f(c). \tag{1.4.7}$$

Combining the limits (1.4.6), (1.4.7) and using Proposition 1.2.1, we get $f(c) = \beta$. The proof for the function to reach its minimum is similar.

$\square$

Exercise 1.4.19. Construct functions satisfying the requirements.

1. $f(x)$ is continuous and not bounded on $(0,1)$.

2. $f(x)$ is continuous and bounded on $(0,1)$ but does not reach its maximum.

3. $f(x)$ is continuous and bounded on $(0,1)$. Moreover, $f(x)$ also reaches its maximum and minimum.

4. $f(x)$ is not continuous and not bounded on $[0,1]$.

5. $f(x)$ is not continuous on $[0,1]$ but reaches its maximum and minimum.

6. $f(x)$ is continuous and bounded on $(-\infty, \infty)$ but does not reach its maximum.

7. $f(x)$ is continuous and bounded on $(-\infty, \infty)$. Moreover, $f(x)$ also reaches its maximum and minimum.

What do your examples say about Theorem 1.4.5?

Exercise 1.4.20. Suppose a continuous function $f(x)$ on (a, b) satisfies $\lim_{x \to a^+} f(x) = \lim_{x \to b^-} f(x) = -\infty$. Prove that the function reaches its maximum on the interval.

Exercise 1.4.21. Suppose $f(x)$ is continuous on a bounded closed interval $[a, b]$. Suppose for any $a \le x \le b$, there is $a \le y \le b$, such that $|f(y)| \le \frac{1}{2}|f(x)|$. Prove that $f(c) = 0$ for some $a \le c \le b$. Does the conclusion still hold if the closed interval is changed to an open one?

Exercise 1.4.22. Prove that a uniformly continuous function on a bounded interval must be bounded.

1.4.4 Intermediate Value Theorem

Theorem 1.4.6 (Intermediate Value Theorem). *Suppose $f(x)$ is a continuous function on a bounded closed interval $[a, b]$. If y is a number between $f(a)$ and $f(b)$, then there is $c \in [a, b]$, such that $f(c) = y$.*

By Theorem 1.4.5, for the function $f(x)$ in the theorem above, there are $x_0, x_1 \in [a, b]$, such that

$$f(x_0) = \min\{f(x) \colon a \le x \le b\}, \quad f(x_1) = \max\{f(x) \colon a \le x \le b\}.$$

Denote $\alpha = f(x_0)$ and $\beta = f(x_1)$. Applying the intermediate value theorem to $f(x)$ on $[x_0, x_1]$ (or $[x_1, x_0]$ if $x_0 > x_1$), we see that $f(x)$ reaches any value between α and β. This proves the following result.

Proposition 1.4.7. *Suppose $f(x)$ is a continuous function on a bounded closed interval $[a, b]$. Then the values of the function on the interval again form a bounded closed interval:*

$$f([a, b]) = \{f(x) \colon a \le x \le b\} = [\alpha, \beta].$$

Conversely, the proposition implies both Theorems 1.4.5 and 1.4.6.
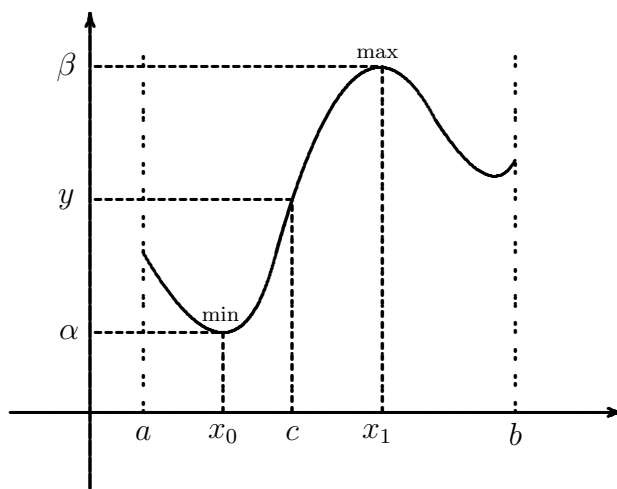
Figure 1.9: intermediate value theorem

Proof of Theorem 1.4.6. Without loss of generality, assume $f(a) \leq y \leq f(b)$. Let

$$X = \{x \in [a, b] \colon f(x) \leq y\}.$$

The set is not empty because $a \in X$. The set is also bounded by a and b. Therefore we have $c = \sup X \in [a, b]$. We expect $f(c) = y$.

If $f(c) > y$, then by the order property in Proposition 1.3.6, there is $\delta > 0$, such that $f(x) > y$ for any x satisfying $c - \delta < x \leq c$ (the right side of c may not be allowed in case $c = b$). On the other hand, by $c = \sup X$, there is $c - \delta < x' \leq c$, such that $f(x') \leq y$. Therefore we have a contradiction at x'.

If $f(c) < y$, then by $f(b) \geq y$, we have $c < b$. Again by the order property in Proposition 1.3.6, there is $\delta > 0$, such that $f(c) < y$ for any x satisfying $|x - c| < \delta$. In particular, any $x' \in (c, c + \delta)$ will satisfy $x' > c$ and $f(x') < y$. This contradicts with the assumption that c is an upper bound of X.

Thus we conclude that $f(c) = y$, and the proof is complete.

$\square$

Example 1.4.5. For the function $f(x) = x^5 + 2x^3 - 5x^2 + 1$, we have $f(0) = 1$, $f(1) = -1$, $f(2) = 29$. By the intermediate value theorem, there are $a \in [0, 1]$ and $b \in [1, 2]$ such that $f(a) = f(b) = 0$.

Example 1.4.6. A number a is a *root* of a function $f(x)$ if $f(a) = 0$. For a polynomial $f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$ with $a_n > 0$ and n odd, we have

$$f(x) = x^n g(x), \quad g(x) = a_n + \frac{a_{n-1}}{x} + \cdots + \frac{a_1}{x^{n-1}} + \frac{a_0}{x^n}.$$

Since $\lim_{x \to \infty} g(x) = a_n > 0$ and n is odd, we have $f(x) > 0$ for sufficiently big and positive x, and $f(x) < 0$ for sufficiently big and negative x. Then by the intermediate value theorem, $f(a) = 0$ for some a (between two sufficiently big numbers of opposite signs). Similar argument also works for the case $a_n < 0$. Therefore we conclude that a polynomial of odd degree must have a real root.

Example 1.4.7. The function $\tan x$ is continuous on $\left(-\frac{\pi}{2}, \frac{\pi}{2} \right)$. By

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

for any number y, we can find $-\frac{\pi}{2} < a < b < \frac{\pi}{2}$, such that $\tan a < y < \tan b$. Then by the intermediate value theorem, we have $y = f(c)$ for some $c \in [a, b]$. This shows that any real number is the tangent of some angle.

Exercise 1.4.23. Prove that for any polynomial of odd degree, any real number is the value of the polynomial. Moreover, estimate a solution of the equation $x^5 - 3x^4 + 5x = 4$ to the accuracy of the first decimal point (i.e., find n such that you are certain that there is a solution between $\frac{n}{10}$ and $\frac{n+1}{10}$).

Exercise 1.4.24. Show that $2^x = 3x$ has solution on $(0, 1)$. Show that $3^x = x^2$ has solution.

Exercise 1.4.25. Suppose a continuous function on an interval is never zero. Prove that it is always positive or always negative.

Exercise 1.4.26. Suppose a continuous function $f(x)$ on (a, b) satisfies $\lim_{x \to a^+} f(x) = \alpha$ and $\lim_{x \to b^-} f(x) = \beta$. Prove that any number in (α, β) is the value of the function. The statement holds even when some of a, b, α, β are infinities.

Exercise 1.4.27. Suppose $f(x)$ is a continuous function on (a, b). Prove that if $f(x)$ only takes rational numbers as values, then $f(x)$ is a constant.

Exercise 1.4.28. Suppose $f(x)$ and $g(x)$ are continuous functions on $[a, b]$. Prove that if $f(a) < g(a)$ and $f(b) > g(b)$, then $f(c) = g(c)$ for some $c \in (a, b)$.

Exercise 1.4.29. Suppose $f \colon [0, 1] \to [0, 1]$ is a continuous function. Prove that $f(c) = c$ for some $c \in [0, 1]$. c is called a *fixed point* of f.

Exercise 1.4.30. Suppose $f(x)$ is a two-to-one function on $[a, b]$. In other words, for any $x \in [a, b]$, there is exactly one other $y \in [a, b]$ such that $x \neq y$ and $f(x) = f(y)$. Prove that $f(x)$ is not continuous.

1.4.5 Invertible Continuous Function

Functions are maps. By writing a function in the form $f \colon [a, b] \to [\alpha, \beta]$, we mean the function is defined on the *domain* $[a, b]$ and its values lie in the *range* $[\alpha, \beta]$. The function is *onto* (or *surjective*) if any $\gamma \in [\alpha, \beta]$ can be written as $\gamma = f(c)$ for some $c \in [a, b]$. It is *one-to-one* (or *injective*) if $x_1 \neq x_2$ implies $f(x_1) \neq f(x_2)$. It is *invertible* (or *bijective*) if there is another function $g \colon [\alpha, \beta] \to [a, b]$ such that $g(f(x)) = x$ for any $x \in [a, b]$ and $f(g(y)) = y$ for any $y \in [\alpha, \beta]$. The function g is called the *inverse* of f and is denoted $g = f^{-1}$. It is a basic fact that a function is invertible if and only if it is onto and one-to-one. Moreover, the inverse function is unique.

In the discussion above, the domain and the range are expressed as closed intervals. The discussion is still valid if other kinds of intervals (or any sets of real numbers) are used.

Theorem 1.4.8. *A continuous function on an interval is invertible if and only if it is strictly monotone. Moreover, the inverse is also continuous and strictly monotone.*

Proof. We only prove the case $f(x)$ is a continuous function on a bounded and closed interval $[a, b]$.

Suppose $f(x)$ is strictly increasing. By Theorem 1.4.5, the function is an onto map $f \colon [a, b] \to [\alpha, \beta] = [f(a), f(b)]$. Moreover,

$$\begin{aligned} x_1 \neq x_2 &\iff x_1 > x_2 \text{ or } x_1 < x_2 \\ &\implies f(x_1) > f(x_2) \text{ or } f(x_1) < f(x_2) \\ &\iff f(x_1) \neq f(x_2). \end{aligned}$$

Thus the function is one-to-one. Since onto and one-to-one maps are invertible, we conclude that $f(x)$ invertible.

Conversely, suppose $f \colon [a, b] \to [\alpha, \beta]$ is invertible and $f(a) < f(b)$. We claim $f(x)$ is strictly increasing. If $f(x)$ is not strictly increasing, then there are $a \leq x_1 < x_2 \leq b$ satisfying $f(x_1) \geq f(x_2)$. Since the invertibility implies one-to-one, we must have $f(x_1) > f(x_2)$. Then we have two possibilities.

1. If $f(a) \geq f(x_1)$, then $f(b) > f(a) \geq f(x_1) > f(x_2)$, and we have $f(a) > f(x_2) < f(b)$.

2. If $f(a) < f(x_1)$, then $f(a) < f(x_1) > f(x_2)$.

By f being one-to-one, we have either

$$y_1 < y_2 < y_3, \quad f(y_1) > f(y_2) < f(y_3),$$

or

$$y_1 < y_2 < y_3, \quad f(y_1) < f(y_2) > f(y_3).$$

For the first case, we have either $f(y_1) \leq f(y_3)$ or $f(y_1) \geq f(y_3)$. If $f(y_1) \leq f(y_3)$, then by the intermediate value theorem, there is $y_1 < y_2 \leq y_4 \leq y_3$ satisfying $f(y_1) = f(y_4)$. Similarly, if $f(y_1) \geq f(y_3)$, then there is $y_1 \leq y_4 \leq y_2 < y_3$ satisfying $f(y_4) = f(y_3)$. Both contradict the one-to-one property. Thus the first case cannot happen. The proof for the impossibility of the second case is similar. This completes the proof that $f(x)$ must be strictly increasing.

It remains to prove that if $f(x)$ is continuous, strictly increasing and invertible, then its inverse $f^{-1}(y)$ is also continuous and strictly increasing.

Let $y_1 = f(x_1)$ and $y_2 = f(x_2)$. Then

$$x_1 \geq x_2 \implies y_1 \geq y_2.$$

The implication is the same as

$$y_1 < y_2 \implies x_1 < x_2,$$

which means exactly that $f^{-1}(x)$ is strictly increasing.

Finally, we prove the continuity of $f(x)$ at $\gamma \in [\alpha, \beta] = [f(a), f(b)]$. We just proved that f^{-1} is strictly increasing, so that by Proposition 1.3.9, the right side limit $\lim_{y \to \gamma^+} f^{-1}(y) = c$ exists. Applying the continuous function $f(x)$ and (1.4.3), we get

$$\gamma = \lim_{y \to \gamma^+} y = \lim_{y \to \gamma^+} f(f^{-1}(y)) = f\left(\lim_{y \to \gamma^+} f^{-1}(y)\right) = f(c).$$

Therefore $\lim_{y \to \gamma^+} f^{-1}(y) = c = f^{-1}(\gamma)$, which means f^{-1} is right continuous at γ. By the similar reason, the inverse function is left continuous. □

For a strictly increasing and continuous function $f(x)$ on an (not necessarily bounded) *open* interval (a, b), the same argument as before shows it is one-to-one. Moreover, by Proposition 1.3.9, the limits $\alpha = \lim_{x \to a^+} f(x)$ and $\beta = \lim_{x \to b^-} f(x)$ exist. Then by Exercise 1.4.26, the map $f \colon (a, b) \to (\alpha, \beta)$ is onto. Thus the function is invertible.

For an invertible and continuous function $f(x)$ on (a, b), the restriction on any closed interval $[a', b'] \subset (a, b)$ is still invertible and continuous. Then by Theorem 1.4.8 for the closed intervals, the restriction on $[a', b']$ is one-to-one. Since $[a', b']$ is arbitrary, we conclude that $f(x)$ is one-to-one on (a, b).

The proof of the continuity of the inverse function can also be applied to the strictly increasing and continuous functions on (a, b). Moreover, the proof also shows

$$\lim_{y \to \alpha^+} f^{-1}(y) = a, \quad \lim_{y \to \beta^-} f^{-1}(y) = b.$$

Exercise 1.4.31. Determine invertible functions.

1. $x^2 \colon [0, 1] \to [0, 1]$.

2. $x^2 \colon (0, 2) \to (0, 4)$.

3. $x^2 \colon (0, 2) \to (0, 5)$.

4. $x^2 \colon (-2, 2) \to (0, 4)$.

5. $x^2 \colon [-2, 0] \to [0, 4]$.

6. $x^2 \colon (0, \infty) \to (0, \infty)$.

7. $x^2 \colon (-\infty, 0] \to [0, \infty)$.

8. $x^2 \colon (-\infty, \infty) \to (-\infty, \infty)$.

Exercise 1.4.32. Suppose $f(x)$ is a strictly decreasing and continuous function on $[a, b)$. Let $\alpha = f(a)$ and $\beta = \lim_{x \to b^-} f(x)$. Prove that $f \colon [a, b) \to (\beta, \alpha]$ is invertible and $\lim_{y \to \beta^+} f^{-1}(y) = b$.

1.4.6   Inverse Trignometric and Logarithmic Functions

Theorem 1.4.8 may be used to introduce some important inverse functions. The functions

$$\sin \colon \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \to [-1, 1],$$
$$\cos \colon [0, \pi] \to [-1, 1],$$
$$\tan \colon \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \to (-\infty, \infty),$$

are onto, strictly monotone and continuous (see (1.3.20), (1.3.21), (1.3.22)). Therefore they are invertible, and the *inverses trigonometric functions*

$$\arcsin \colon [-1, 1] \to \left[-\frac{\pi}{2}, \frac{\pi}{2}\right],$$
$$\arccos \colon [-1, 1] \to [0, \pi],$$
$$\arctan \colon (-\infty, \infty) \to \left(-\frac{\pi}{2}, \frac{\pi}{2}\right),$$

are also strictly monotone and continuous. The other three inverse trigonometric functions may be defined similarly. The equality $\cos\left(\frac{\pi}{2} - x\right) = \sin x$ implies that

$$\arcsin x + \arccos x = \frac{\pi}{2}.$$

The similar equalities also imply the similar equations between other inverse trigonometric functions. Moreover, by the remark made after the proof of Theorem 1.4.8, we have

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

Similar limits hold for other inverse trigonometric functions.

Example 1.4.8. The continuity of inverse sine function says that $\lim_{x \to 0} \arcsin x = \arcsin 0 = 0$. Moreover, by the composition rule, the variable x in the limit (1.3.23) may be substituted by the continuous function $\arcsin x$ to get $\lim_{x \to 0} \dfrac{x}{\arcsin x} = 1$. Taking reciprocal, we get $\lim_{x \to 0} \dfrac{\arcsin x}{x} = 1$.

Exercise 1.4.33. Prove the equalities.

1. $\arcsin(-x) = -\arcsin x$.

2. $\arccos(-x) = \pi - \arccos x$.

3. $\arctan(-x) = -\arctan x$.

4. $\arcsin(\cos x) = \dfrac{\pi}{2} - x$ for $0 \le x \le \pi$.

5. $\arctan x + \arctan \dfrac{1}{x} = \dfrac{\pi}{2}$ for $x > 0$.

6. $\cos(\arcsin x) = \sqrt{1 - x^2}$.

7. $\tan(\arcsin x) = \dfrac{x}{\sqrt{1 - x^2}}$.

8. $\tan(\arccos x) = \dfrac{\sqrt{1 - x^2}}{x}$.

Exercise 1.4.34. Compute the limits.

1. $\lim_{x \to -1+} \dfrac{2\arcsin x + \pi}{x + 1}$.

2. $\lim_{x \to 1-} \dfrac{(\arccos x)^2}{x - 1}$.

3. $\lim_{x \to 0} \dfrac{\arctan x}{x}$.

4. $\lim_{x \to 0} \dfrac{\tan(\arcsin x)}{\sin(\arctan x)}$.

5. $\lim_{x \to 0+} (\sin x)^{\arcsin x}$.

6. $\lim_{x \to +\infty} x \left(\arctan x - \dfrac{\pi}{2} \right)$.

By (1.3.26), the exponential function α^x based on a constant $\alpha > 0$ is continuous. For $\alpha > 1$, the function is strictly increasing, and

$$\lim_{x \to -\infty} \alpha^x = 0, \quad \lim_{x \to +\infty} \alpha^x = +\infty.$$

For $0 < \alpha < 1$, the function is strictly decreasing, and

$$\lim_{x \to -\infty} \alpha^x = +\infty, \quad \lim_{x \to +\infty} \alpha^x = 0.$$

Thus the exponential function

$$\alpha^x \colon (-\infty, \infty) \to (0, \infty), \ \ 0 < \alpha \ne 1$$

is invertible. The inverse function

$$\log_\alpha x \colon (0, \infty) \to (-\infty, \infty), \ \ 0 < \alpha \ne 1$$

is the *logarithmic function*, which is also continuous, strictly increasing for $\alpha > 1$ and strictly decreasing for $0 < \alpha < 1$. Moreover,

$$\lim_{x \to 0+} \log_\alpha x = \begin{cases} -\infty & \text{if } \alpha > 1 \\ +\infty & \text{if } 0 < \alpha < 1 \end{cases}, \tag{1.4.8}$$

$$\lim_{x \to +\infty} \log_\alpha x = \begin{cases} +\infty & \text{if } \alpha > 1 \\ -\infty & \text{if } 0 < \alpha < 1 \end{cases}. \tag{1.4.9}$$

The following equalities for the exponential function

$$\alpha^0 = 1, \ \ \alpha^1 = \alpha, \ \ \alpha^x \alpha^y = \alpha^{x+y}, \ \ (\alpha^x)^y = \alpha^{xy}$$

imply the following equalities for the logarithmic function

$$\log_\alpha 1 = 0, \ \ \log_\alpha \alpha = 1, \ \ \log_\alpha(xy) = \log_\alpha x + \log_\alpha y,$$

$$\log_\alpha x^y = y \log_\alpha x, \ \ \log_\alpha x = \frac{\log x}{\log \alpha}.$$

The logarithmic function with the special base $\alpha = e$ is called the *natural logarithmic function* and is denoted by log or ln. In particular, we have $\log e = 1$.

Next we derive some important limits involving the exponential and the logarithmic functions.

Taking $\alpha = \dfrac{1}{e}$ in the limit (1.3.10), we have $\lim_{x \to +\infty} \dfrac{x^\beta}{e^x} = 0$. By $\lim_{x \to +\infty} \log x = +\infty$ and the composition rule, we may substitute x by $\log x$ and get

$$\lim_{x \to +\infty} \frac{(\log x)^\beta}{x} = 0. \tag{1.4.10}$$

This means that at $+\infty$, the logarithmic function goes to infinity much slower than x. Even raising the power of the logarithmic function does not help the speed of growth.

Substituting x by $\dfrac{1}{x}$ in the limit (1.3.30), we get $\lim_{x \to 0}(1+x)^{\frac{1}{x}} = e$. By the continuity of log, we have $\lim_{x \to 0} \log\left((1+x)^{\frac{1}{x}}\right) = \log \lim_{x \to 0}(1+x)^{\frac{1}{x}} = \log e = 1$. Then by the property of log, the limit is

$$\lim_{x \to 0} \frac{\log(1+x)}{x} = 1. \tag{1.4.11}$$

The continuity of e^x tells us $\lim_{x \to 0}(e^x - 1) = 0$. Substituting x in (1.4.11) by $e^x - 1$, we get $\lim_{x \to 0} \dfrac{x}{e^x - 1} = \lim_{x \to 0} \dfrac{\log(1 + e^x - 1)}{e^x - 1} = 1$. Taking reciprocal, we have

$$\lim_{x \to 0} \frac{e^x - 1}{x} = 1. \tag{1.4.12}$$

The continuity of log tells us $\lim_{x \to 0} \log(1 + x) = 0$. Substituting x in (1.4.12) by $\alpha \log(1+x)$ and using $e^{\alpha \log(1+x)} = (1+x)^\alpha$, we get $\lim_{x \to 0} \dfrac{(1 + x)^\alpha - 1}{\log(1 + x)} = \alpha$. Multiplying the limit with (1.4.11), we get

$$\lim_{x \to 0} \frac{(1 + x)^\alpha - 1}{x} = \alpha. \tag{1.4.13}$$

Example 1.4.9. For $\beta > 0$, taking $\dfrac{1}{\beta}$ power of the limit (1.4.10) gives us

$$\lim_{x \to +\infty} \frac{\log x}{x^\beta} = 0 \ \text{ for } \ \beta > 0. \tag{1.4.14}$$

Then substituting x by $\dfrac{1}{x}$ and using $\log \dfrac{1}{x} = -\log x$, we get

$$\lim_{x \to 0^+} x^\beta \log x = 0 \ \text{ for } \ \beta > 0. \tag{1.4.15}$$

The limits (1.4.10), (1.4.14), (1.4.15) can be extended to the logarithms based on any $\alpha > 0$ by using $\log_\alpha x = \dfrac{\log x}{\log \alpha}$.

Example 1.4.10. Substituting x by $x \log \alpha$ in (1.4.12) and using $e^{x \log \alpha} = \alpha^x$, we get a more general formula

$$\lim_{x \to 0} \frac{\alpha^x - 1}{x} = \log \alpha. \tag{1.4.16}$$

Example 1.4.11. We already know $\lim_{x \to 0^+} x^x = 1$. To further find out how close x^x is to 1, we let $y = x^x$. Then $x \log x = \log y$ and $\dfrac{x^x - 1}{x \log x} = \dfrac{y - 1}{\log y}$. As $x \to 0^+$, we have $y \to 1^+$, and $\lim_{y \to 1} \dfrac{y - 1}{\log y} = \lim_{z \to 0} \dfrac{z}{\log(1 + z)} = 1$ by (1.4.11). Thus we conclude that

$$x^x = 1 + x \log x + o(x \log x),$$

where $x \log x$ is an infinitesimal at 0^+.

Exercise 1.4.35. Compute the limits.

1. $\lim_{x \to 1} \dfrac{\log x}{\sqrt{x} - 1}$.

2. $\lim_{x \to 0^+} \dfrac{\log x}{x}$.

3. $\lim_{x \to +\infty} \dfrac{\log(\log x)}{x}$.

4. $\lim_{x \to 1^+} (x - 1) \log(\log x)$.

5. $\lim_{x \to 2} \dfrac{\log x - \log 2}{x - 2}$.

6. $\lim_{x \to +\infty} \dfrac{\log x}{\sqrt{x} + \log x}$.

7. $\lim_{x \to +\infty} \dfrac{\log(2 + 3x)}{\log(2 + 7x)}$.

8. $\lim_{x \to 0} \dfrac{\log(2 + 3x)}{\log(2 + 7x)}$.

9. $\lim_{x \to -\frac{2}{7}^+} \dfrac{\log(2 + 3x)}{\log(2 + 7x)}$.

10. $\lim_{x \to 0} \dfrac{\log(1 + \sqrt{x})}{\log(1 + x)}$.

11. $\lim_{x \to +\infty} \dfrac{\log(1 + \sqrt{x})}{\log(1 + x)}$.

12. $\lim_{x \to +\infty} \dfrac{\log_3(2 + x^3)}{\log_4(3 + x^4)}$.

13. $\lim_{x \to +\infty} \sqrt[x]{\sqrt{x} + \log x}$.

14. $\lim_{x \to 0} \dfrac{\log_\alpha(1 + x)}{\log_\alpha(1 + 2x - x^2)}$.

15. $\lim_{x \to 0} \dfrac{(1 + x + x^2)^{\frac{7}{9}} - 1}{x}$.

16. $\lim_{x \to 0} \dfrac{(1 + x + x^2)^{\frac{7}{9}} - (1 + x)^{\frac{7}{9}}}{x^2}$.

17. $\lim_{x \to 0^+} \dfrac{x^{x^2} - 1}{x \log x}$.

18. $\lim_{x \to +\infty} x(\sqrt[x]{x} - 1)$.

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

Exercise 1.4.36. Use Exercise 1.1.36 and the continuity of logarithmic function to prove that if $x_n > 0$ and $\lim_{n \to \infty} x_n = l$, then $\lim_{n \to \infty} \sqrt[n]{x_1 x_2 \cdots x_n} = l$. What about the case $\lim_{n \to \infty} x_n = +\infty$?

Exercise 1.4.37. Prove that if $x_n > 0$ and $\lim_{n \to \infty} \dfrac{x_n}{x_{n-1}} = l$, then $\lim_{n \to \infty} \sqrt[n]{x_n} = l$. Use the conclusion to compute the following limits

$$\lim_{n \to \infty} \frac{\sqrt[n]{n!}}{n}, \quad \lim_{n \to \infty} \frac{\sqrt[n]{(2n)!}}{n^2}, \quad \lim_{n \to \infty} \sqrt[n]{\frac{(2n)!}{(n!)^2}}.$$

Exercise 1.4.38. Use Exercise 1.2.19 to derive $\dfrac{1}{1+n} < \log\left(1 + \dfrac{1}{n}\right) < \dfrac{1}{n}$. Then use the inequality to show that $x_n = 1 + \dfrac{1}{2} + \cdots + \dfrac{1}{n} - \log n$ is strictly decreasing and $y_n = 1 + \dfrac{1}{2} + \cdots + \dfrac{1}{n} - \log(n+1)$ is strictly increasing. Moreover, prove that both x_n and y_n converge to the same limit, called the *Euler*[12]*-Mascheroni*[13] *constant*:

$$\lim_{n \to \infty} \left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} - \log n\right) = 0.577215669015328\cdots. \tag{1.4.17}$$

A vast extension of the exercise appears in Exercise 4.1.64.

1.4.7   Additional Exercise

Additive and Multiplicative Functions

Exercise 1.4.39. Suppose $f(x)$ is a continuous function on $\mathbb{R}$ satisfying $f(x + y) = f(x) + f(y)$.

1. Prove that $f(nx) = nf(x)$ for integers n.

2. Prove that $f(rx) = rf(x)$ for rational numbers r.

3. Prove that $f(x) = ax$ for some constant a.

Exercise 1.4.40. Suppose $f(x)$ is a continuous function on $\mathbb{R}$ satisfying $f(x + y) = f(x)f(y)$. Prove that either $f(x) = 0$ or $f(x) = a^x$ for some constant $a > 0$.

Cauchy Criterion for Continuity

Exercise 1.4.41. Prove that a function $f(x)$ is continuous at a if any only if for any $\epsilon > 0$, there is $\delta > 0$, such that

$$|x - a| < \delta, |y - a| < \delta \implies |f(x) - f(y)| < \epsilon.$$

Exercise 1.4.42. Prove that if a function $f(x)$ is not continuous at a, then one of the following will happen.

---

[12]Leonhard Paul Euler, born 1707 in Basel (Switzerland), died 1783 in St. Petersburg (Russia). Euler is one of the greatest mathematicians of all time. He made important discoveries in almost all areas of mathematics. Many theorems, quantities, and equations are named after Euler. He also introduced much of the modern mathematical terminology and notation, including $f(x)$, e, Σ (for summation), i (for $\sqrt{-1}$), and modern notations for trigonometric functions.

[13]Lorenzo Mascheroni, born 1750 in Lombardo-Veneto (now Italy), died 1800 in Paris (France). The Euler-Mascheroni constant first appeared in a paper by Euler in 1735. Euler calculated the constant to 6 decimal places in 1734, and to 16 decimal places in 1736. Mascheroni calculated the constant to 20 decimal places in 1790.

1. There is $\epsilon > 0$, such that for any $\delta > 0$, there are $a - \delta < x < a$ and $a < y < a + \delta$, such that $|f(x) - f(a)| \geq \epsilon$ and $|f(y) - f(a)| \geq \epsilon$.

2. There is $\epsilon > 0$, such that for any $\delta > 0$, there are $a - \delta < x < a$ and $a < y < a + \delta$, such that $|f(x) - f(y)| \geq \epsilon$.

Note that x and y are on different sides of a.

One Side Invertible Monotone Function

Let $f(x)$ be an increasing but not necessarily continuous function on $[a, b]$. By Exercise 1.4.10, the function has only countably many discontinuities. A function $g(x)$ on $[f(a), f(b)]$ is a *left inverse* of $f(x)$ if $g(f(x)) = x$, and is a *right inverse* if $f(g(y)) = y$.

Exercise 1.4.43. Prove that if $f(x)$ has a left inverse $g(y)$, then $f(x)$ is strictly increasing, and $g(f(x^-)) = g(f(x^+)) = x$ for any $x \in [a, b]$.

Exercise 1.4.44. Suppose $f(x)$ is strictly increasing and has only one discontinuity at $c \in (a, b)$. Prove that $f(x)$ has a unique left inverse $g(y)$. Moreover, $g(y)$ is strictly increasing on $[f(a), f(c^-)]$ and $[f(c^+), f(b)]$, and $g(y) = c$ on $[f(c^-), f(c^+)]$.

Exercise 1.4.45. Suppose $f(x)$ is strictly increasing. Prove that for any $y \in [f(a), f(b)]$, there is a unique $x \in [a, b]$ satisfying $f(x^-) \leq y \leq f(x^+)$. Moreover, for such unique x, the definition $g(y) = x$ gives the unique left inverse of $f(x)$.

Exercise 1.4.46. Prove that $f(x)$ has a right inverse if and only if $f(x)$ is continuous. Moreover, the right inverse must be strictly increasing, and the right inverse is unique if and only if $f(x)$ is strictly increasing.

Chapter 2

Differentiation

2.1   Approximation and Differentiation

The precalculus mathematics deals with *static* quantities which always have specific and fixed values. For example, the angle in an equilateral triangle is 60 degrees. The area of the rectangle with width 3 and height 2 is 6.

Differentiation is the mathematical tool for studying *changing* quantities. Such quantities often depend on other quantities. The dependence is often expressed as functions. For example, if two angles in a triangle are α and β degrees, then the third angle is $\gamma = 180 - \alpha - \beta$ degrees. Moreover, the area of a rectangle depends on the width and the height (and is given by the product function).

Of course, functions may be evaluated to get static values. For example, for width $= 3$ and height $= 2$, we get area $= 6$. Such substitutions simply fix the changing quantities and cannot answer the following questions:

1. Does the area increase when the width and the height are increasing?

2. What is the biggest area of a rectangle with circumference 10?

To answer such questions, we have to think of the whole function instead of specific evaluation. We have to study how the function changes. We have to consider the whole evolving process instead of specific moments of the process.

The change of general functions may be described in terms of approximations by simple functions. The simplest and the most useful function for the purpose of approximation is the linear function. The linear approximation is called the differential and the key coefficient in the linear approximation is called the derivative.

2.1.1   Approximation

Let us start with some common sense. Suppose $P(A)$ is a problem about an object A. To solve the problem, we may consider a simpler object B that closely approximates A. Because B is simpler, $P(B)$ is easier to solve. Moreover, because B is close to A, the (easily obtained) answer to $P(B)$ is also pretty much the answer to $P(A)$.

Example 2.1.1. To find the distance from Hong Kong to Shanghai, we take a map and use the ruler to measure the distance between the two cities and then multiply the scaling ratio of the map. In this process, P is the distance, A is the cities Hong Kong and Shanghai, and B is the two dots representing the two cities on a map.

The map is only an approximation of the actual world. Of course, the bigger the map, the more accurate the measurement is. An even better measurement can be obtained by a sphere shaped map (the globe found in stationary stores) because the sphere is a better approximation of the world than the flat plane (some ancient people may disagree, though).

On the other hand, how accurate we wish the answer to be depends on our purpose. If our purpose is only to estimate the time needed for flying from Hong Kong to Shanghai, a poster sized map is more than sufficient.

Example 2.1.2. Suppose we want to count how much money there is in a big bag of one dollar coins. We may weigh the whole bag, weigh one coin, and then divide the two numbers. In this process, A is the collection of real coins, with slight variation in size and weight, and B is the imaginary ideal collection in which each coin is identical to the one we choose to weigh. If the big bag is not too big (say two kilo), the error would be small enough and the answer would be quite reliable.

Exercise 2.1.1. Think again how you solve problems in everyday life. Do you actually solve the problem as is? Or what you solve is just an approximation of the problem?

Many real world problems can often be mathematically interpreted as quantities related by complicated functions. To solve the problem, we may try to approximate the complicated functions by simple ones and solve the same problem for the simple functions.

What are the simple functions? Although the answer could be rather subjective, most people would agree that the following functions are listed from simple to complicated.

type	one variable examples	two variable examples
zero	0	0
constant	$-1, -3$	$-1, -3$
linear	$3 + x, 4 - 5x$	$3 + x + 2y, 5x - 7y$
quadratic	$1 + x^2, -4 + 5x - 2x^2$	$xy, x^2 + xy - 3y^2 + y + 1$
cubic	$2x - 2x^3$	$2x^2 - 2y^3, x^2y + xy^2$
rational	$\dfrac{2 + x^2}{x(4 - x)}, \dfrac{1}{1 + x}$	$\dfrac{1}{x + y}, \dfrac{x + xy + 3y^2}{x(y - x)}$
algebraic	$\sqrt{x}, (1 + x^{1/2})^{1/3}$	$\sqrt{xy}, (x^3 - 3y^2)^{1/3}$
transcendental	$\sin x, e^x + \cos x$	$\sin xy, e^{x+y} + \cos(x + y)$

What do we mean by approximating a function with a simple (class of) functions? Consider measuring certain length (say the height of a person, for example) by a ruler with only centimeters. We expect to get an approximate reading of the length with the accuracy within millimeters, or significantly smaller than the base unit of 1cm for the ruler:

$$|\text{actual length } - \text{ reading from ruler}| \leq \epsilon(1\text{cm}).$$

Similarly, approximating a function $f(x)$ near x_0 by a function $p(x)$ of some class would mean the difference $|f(x) - p(x)|$ is significantly smaller than the "base unit" for the class. Specifically, for any $\epsilon > 0$, there is $\delta > 0$, such that

$$|x - x_0| < \delta \implies |f(x) - p(x)| \leq \epsilon(\text{unit}). \tag{2.1.1}$$

The "base unit" for the class of constant functions is 1. Thus the approximation of a function $f(x)$ by a constant function $p(x) = a$ near x_0 means that for any $\epsilon > 0$, there is $\delta > 0$, such that

$$|x - x_0| < \delta \implies |f(x) - a| \leq \epsilon. \tag{2.1.2}$$

The condition can be split into two parts. First, for $x = x_0$, we get $|f(x_0) - a| \leq \epsilon$ for any $\epsilon > 0$. This means exactly $a = f(x_0)$. Second, for $x \neq x_0$, we get

$$0 < |x - x_0| < \delta \implies |f(x) - a| \leq \epsilon.$$

This means exactly $\lim_{x \to x_0} f(x) = a$. Combined with $a = f(x_0)$, we conclude that a function $f(x)$ is approximated by a constant near x_0 if and only if it is continuous at x_0. Moreover, the approximating constant is $f(x_0)$.

The approximation by constant functions, which happens for continuous functions only, is good enough for the problem of the sign of the function.

Example 2.1.3. Consider the positivity of functions. The cubic function $f(x) = x^3 - 5x + 2$ is approximated by the constant $f(3) = 14$ near 3. Since the constant function is positive, the cubic function is also positive near 3.

Note that the farther away from 3, the less certain we are about the positivity for the cubic function. The reason is that the approximation becomes worse further away from 3. As a matter of fact, if we move away by distance 1 (say at $x = 2$), the positivity is lost.

Exercise 2.1.2. Rigorously prove that if $f(x)$ is continuous at x_0 and $f(x_0) > 0$, then there is $\delta > 0$, such that $f(x) > 0$ for any $x \in (x_0 - \delta, x_0 + \delta)$.

2.1.2   Differentiation

The approximation by constant functions is too crude for most problems. As the next simplest, the *linear functions $A + Bx$* often need to be used. For the convenience of discussion, we introduce

$$\Delta x = x - x_0$$

(Δ is the Greek equivalence of D, here used to stand for *D*ifference) and rewrite

$$A + Bx = A + B(x_0 + \Delta x) = (A + Bx_0) + B\Delta x = a + b\Delta x.$$

What is the base unit of $a + b\Delta x$ as x approaches x_0? The base unit of the constant term a is 1. The base unit of the difference term $b\Delta x$ is Δx, which is very small compared with the unit 1. Therefore the base unit for the linear function $a + b\Delta x$ is Δx. The discussion may be compared with the expression $am + bcm$ (a meters and b centimeters). Since 1cm is much smaller than 1m, the base unit for $am + bcm$ is 1cm.

The approximation of a function $f(x)$ by a linear function $a + b\Delta x$ at x_0 means the following.

Definition 2.1.1. A function $f(x)$ is *differentiable* at x_0 if there is a linear function $a + b\Delta x$, such that for any $\epsilon > 0$, there is $\delta > 0$, such that

$$|\Delta x| = |x - x_0| < \delta \implies |f(x) - a - b\Delta x| \leq \epsilon|\Delta x|. \qquad (2.1.3)$$

Example 2.1.4. To find the linear approximation of x^2 at $x_0 = 1$, we rewrite the function in terms of $\Delta x = x - 1$ near 1.

$$x^2 = (1 + \Delta x)^2 = 1 + 2\Delta x + \Delta x^2.$$

Therefore
$$|\Delta x| < \delta = \epsilon \implies |x^2 - 1 - 2\Delta x| = |\Delta x|^2 \leq \delta|\Delta x|.$$

This shows that $1 + 2\Delta x$ is the linear approximation of x^2 at 1.

Exercise 2.1.3. Show that x^2 is linearly approximated by $x_0^2 + 2x_0\Delta x$ at x_0. What about the linear approximation of x^3?

For $x = x_0$, the condition (2.1.3) tells us the constant coefficient $a = f(x_0)$. Let
$$\Delta f = f(x) - a = f(x) - f(x_0)$$

be the change of the function caused by the change Δx of the variable. Then the right side of (2.1.3) becomes $|\Delta f - b\Delta x| \leq \epsilon|\Delta x|$, so that the condition becomes
$$\Delta f = b\Delta x + o(\Delta x). \tag{2.1.4}$$

In other words, the scaling $b\Delta x$ of the change of variable is the linear approximation of the change of the function. The viewpoint is caught in the notation of the *differential*
$$df = bdx \tag{2.1.5}$$

of the function at x_0.

Note that the symbols df and dx have not yet been specified as numerical quantities. Thus bdx should be, at least for the moment, considered as an *integrated* notation instead of the product of two quantities. On the other hand, the notation is motivated from $b\Delta x$, which was indeed a product of two numbers. So it is allowed to add two differentials and to multiply numbers to differentials. In more advanced mathematics, the differential symbols will indeed be defined as quantities in some linear approximation spaces (called *tangent spaces*). However, one has to be careful in multiplying differentials together because this is not a valid operation within linear spaces. Finally, dividing differentials is not allowed.

2.1.3 Derivative

Taking $x = x_0$ in (2.1.3) tells us $a = f(x_0)$. Taking $x \neq x_0$, the condition (2.1.3) becomes
$$0 < |\Delta x| = |x - x_0| < \delta \implies \left|\frac{f(x) - f(x_0)}{x - x_0} - b\right| \leq \epsilon. \tag{2.1.6}$$

This tells us how to compute the coefficient b.

Definition 2.1.2. The *derivative* of a function $f(x)$ at x_0 is
$$f'(x_0) = \frac{df}{dx} = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}.$$

We already saw that the differentiability implies the existence of the derivative. Conversely, if the derivative exists, then we have the implication (2.1.6), which is the same as (2.1.3) with $a = f(x_0)$ and $0 < |x - x_0| < \delta$. Since (2.1.3) always holds with $a = f(x_0)$ and $x = x_0$, we conclude that the condition for differentiability holds with $a = f(x_0)$ and $|x - x_0| < \delta$.
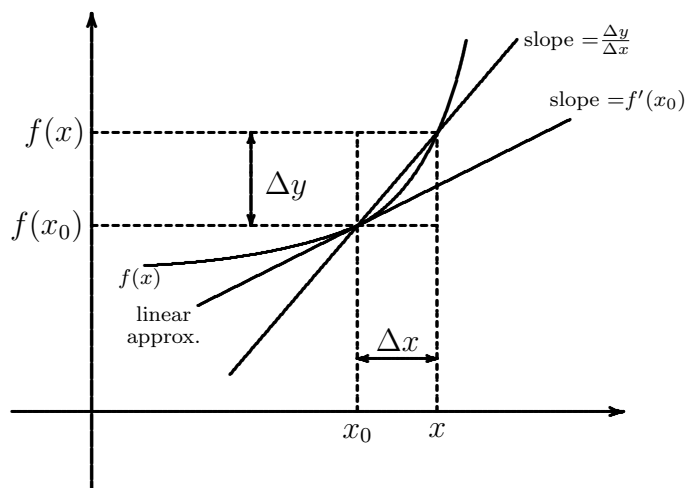
Figure 2.1: linear approximation and derivative of $f(x)$ at x_0

Proposition 2.1.3. *A function $f(x)$ is differentiable at x_0 if and only if it has the derivative at x_0. Moreover, the function is continuous at x_0 and its linear approximation is $f(x_0) + f'(x_0)\Delta x$.*

Proof. It remains to prove the continuity. Taking $\epsilon = 1$ in the definition of differentiability, we find $\delta > 0$, such that

$$|\Delta x| = |x - x_0| < \delta \implies |f(x) - f(x_0) - b\Delta x| \leq |\Delta x|$$
$$\implies |f(x) - f(x_0)| \leq (|b| + 1)|\Delta x|.$$

Then for any $\epsilon > 0$, we get

$$|x - x_0| < \max\left\{\delta, \frac{\epsilon}{|b| + 1}\right\} \implies |f(x) - f(x_0)| \leq (|b| + 1)|x - x_0| \leq \epsilon.$$

This shows $f(x)$ is continuous at x_0. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

We emphasis here that although the existence of linear approximation is equivalent to the existence of derivative, the two play different roles. Linear approximation is the motivation and the concept. Derivative is the computation of the linear approximation and is derived from it. Therefore linear approximation is much more important in understanding the essence of calculus. As a matter of fact, for multivariable functions, linear approximations may be similarly computed by *partial derivatives*. However, the existence of partial derivatives does not necessarily imply the existence of linear approximation.

Mathematical concepts are always derived from common sense. The formulas for computing the concepts are only obtained after mathematically analysing the common sense. Never equate the formula with the concept itself!

Example 2.1.5. The linear approximation to a linear function $f(x) = A + Bx$ must be the linear function itself. Moreover, since $\Delta f = f(x) - f(x_0) = B(x - x_0) = B\Delta x$, the derivative $f'(x_0) = B$ by comparing with (2.1.4).

Example 2.1.6. For a quadratic function $f(x) = A + Bx + Cx^2$, we have

$$\Delta f = f(x_0 + \Delta x) - f(x_0) = (B + 2Cx_0)\Delta x + C\Delta x^2 = (B + 2Cx_0)\Delta x + o(\Delta x).$$

By comparing with (2.1.4), the function is differentiable at x_0, with the differential $df = (B + 2Cx_0)dx$ at x_0. Usually we write $df = (B + 2Cx)dx$ for the differential at any x. We also have $f'(x) = B + 2Cx$.

Example 2.1.7. For $f(x) = \sin x$, the limit (1.3.23) tells us

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

Therefore the sine function is differentiable at 0, with $f(0) + f'(0)(x - 0) = x$ as the linear approximation.

Example 2.1.8. For $f(x) = |x|$, the limit $\lim_{x \to 0} \dfrac{f(x) - f(0)}{x} = \lim_{x \to 0} \dfrac{|x|}{x}$ diverges because the left and the right limits are different. Therefore the absolute value function is not differentiable at 0.

Exercise 2.1.4. Find the differential and the derivative for the cubic function $f(x) = x^3$ by computing $\Delta f = f(x_0 + \Delta x) - f(x_0)$.

Exercise 2.1.5. Determine the differentiability and find the linear approximations if possible $(\alpha, \beta > 0)$.

1. $\sqrt{|x|}$ at 0.

2. $\cos x$ at a.

3. $\sin^2 x$ at 0.

4. e^x at a.

5. $\log(1 + x)$ at 0.

6. $|\sin \pi x^3|$ at 0 and 1.

7. $\begin{cases} x & \text{if } x \neq 0 \\ 1 & \text{if } x = 0 \end{cases}$ at 0.

8. $\begin{cases} -x^\alpha & \text{if } x \geq 0 \\ (-x)^\beta & \text{if } x < 0 \end{cases}$ at 0.

9. $\begin{cases} x^\alpha & \text{if } x \geq 0 \\ (-x)^\beta & \text{if } x < 0 \end{cases}$ at 0.

Exercise 2.1.6. Study the differentiability of the function $|x^3(x - 1)(x - 2)^2|$.

Exercise 2.1.7. Study the differentiability of Thomae's function in Example 1.4.2.

Exercise 2.1.8. Suppose $f(0) = 0$ and $f(x) \geq |x|$. Show that $f(x)$ is not differentiable at 0. What if $f(x) \leq |x|$?

Exercise 2.1.9. Suppose $f(x)$ is differentiable on a bounded interval $(a - \epsilon, b + \epsilon)$. Suppose $f(x) = 0$ for infinitely many $x \in [a, b]$. Prove that there is $c \in [a, b]$, such that $f(c) = 0$ and $f'(c) = 0$.

Exercise 2.1.10. Suppose $f(x)$ is differentiable at x_0, with $f(x_0) \neq 0$. Find $\lim_{t \to 0} \left(\dfrac{f(x_0 + t)}{f(x_0)} \right)^{\frac{1}{t}}$.

Exercise 2.1.11. Prove that $f(x)$ is differentiable at x_0 if and only if $f(x) = f(x_0) + (x - x_0)g(x)$ for a function $g(x)$ continuous at x_0.

2.1.4   Tangent Line and Rate of Change

Geometrically, the differentiation may be understood through the graphs of functions. The graph of a function is a curve, and the graph of a linear function is a straight line. Therefore the linear approximation is the approximation of a curve by a straight line. Such straight line is the "best fit" among all the straight lines passing through $(x_0, f(x_0))$. To construct the best fit, we let L_x be the line passing through the two points $(x_0, f(x_0))$ and $(x, f(x))$ on the graph of the function. Then the limit of the line L_x as x approaches x_0 is the linear approximation. Since the slope of L_x is $\dfrac{\Delta f}{\Delta x}$, the slope of the linear approximation is the limit of the quotient, or the derivative.

   The differentiation can also be viewed as the rate of change. To understand how a quantity y depends on another quantity x, it suffices to understand how the *change* in x induces the change in y. To illustrate the idea, suppose x is the age of a person and $y = f(x) = 100 + x^2$ is the wage the person earns at age x. Then we know the wage changes by $Df(x) = f(x + 1) - f(x) = 2x + 1$ when the person becomes one year older. Conversely, suppose we know the starting salary $f(20) = 500$ at the age 20 and we also know that every year the wage changes by $Df(x) = 2x + 1$. Then we may recover the actual salary $f(x)$ by adding up the changes. For example, the salary at age 25 is

$$\begin{aligned} f(25) &= f(20) + Df(20) + Df(21) + Df(22) + Df(23) + Df(24) \\ &= 500 + 41 + 43 + 45 + 47 + 49 \\ &= 725. \end{aligned}$$

In general, it is not difficult to recover the formula $f(x) = 100 + x^2$. The process of measuring changes is differentiation, and the process of recovering the whole by adding up the changes is integration.

Exercise 2.1.12. Given the initial term x_1 of a sequence and the difference $x_{n+1} - x_n$ between consecutive terms, recover the sequence.

1. $x_1 = 1$ and $x_{n+1} - x_n = 1$.

2. $x_1 = 1$ and $x_{n+1} - x_n = n$.

3. $x_1 = 1$ and $x_{n+1} - x_n = (-1)^n$.

4. $x_1 = 1$ and $x_{n+1} - x_n = \dfrac{1}{n(n + 1)}$.

5. $x_1 = 1$ and $x_{n+1} - x_n = \dfrac{1}{\sqrt{n} + \sqrt{n + 1}}$.

   In the example above, the variable is taken to be integers only. In reality, many variables are real numbers and the changes are continuous. In this case, we have to consider the change $\Delta f = f(x + \Delta x) - f(x)$ of a function when x is changed by Δx. Since Δf usually becomes smaller as Δx gets smaller, it is more sensible to measure the rate of change $\dfrac{\Delta f}{\Delta x}$ for the period between

x and $x + \Delta x$. When Δx is approaching zero, the limit $\lim_{\Delta x \to 0} \dfrac{\Delta f}{\Delta x}$ then becomes the rate of change at the instant of x. The process of computing the rate of change is differentiation, and the process of recovering the whole from the rate of change is integration.

Example 2.1.9. The height of a free falling object on the face of the earth is

$$h(t) = h_0 - \frac{1}{2}gt^2,$$

where h_0 is the initial height, g is the gravitational constant, and t is the time. As the time is changed from t to $t + \Delta t$, the height is changed by

$$\Delta h = h(t + \Delta t) - h(t) = -gt\Delta t - \frac{1}{2}\Delta t^2.$$

The *average* speed of the falling object during the period is

$$\frac{\Delta h}{\Delta t} = -gt - \frac{1}{2}\Delta t.$$

As $\Delta t \to 0$, we get the *instantaneous* speed

$$h(t) = \lim_{\Delta t \to 0} \frac{\Delta h}{\Delta t} = -gt$$

at the time t. Note that the speed is getting bigger as time goes by, which is consistent with the observation.

As a matter of fact, the formula for the height $h(t)$ is obtained in the reverse way (i.e., by integration). From Newton[1]'s second law of mechanics, we know the speed of the falling object is $v(t) = -gt$. To recover the height of the object is the same as finding a function $h(t)$ satisfying $h'(t) = v(t) = -gt$. Indeed $h(t) = h_0 - \dfrac{1}{2}gt^2$ is the function with the required derivative.

Example 2.1.10. Let $f(t)$ be the amount of money in a savings account. The increase of the amount, due to the interest, can be measured by the rate of change, which at a particular time t is proportional to the amount of money $f(t)$ at the time. In other words, we have $f'(t) = cf(t)$, where the constant c is the interest rate. Now the question is what kind of function satisfies the (differential) equation.

To get some idea, let us consider the discrete case. Suppose the interest is paid out yearly instead of continuously. Then the amount at the $(t + 1)$-st year is $f(t + 1) = (1 + c)f(t)$. Thus for any natural number t we get

$$f(t) = (1 + c)f(t - 1) = (1 + c)^2 f(t - 2) = \cdots = (1 + c)^t f(0),$$

which is an exponential function of t. If the interest payment is made monthly at the yearly rate of c, then we have $f\left(t + \dfrac{1}{12}\right) = \left(1 + \dfrac{c}{12}\right)f(t)$, from which we can similarly get $f(t) = \left(1 + \dfrac{c}{12}\right)^{12t} f(0)$. If the interest payment is made n times

---

[1]Isaac Newton, born 1643 in Woolsthorpe (England), died 1727 in London (England). Newton is a giant of science and one of the most influential people in human history. Together with Gottfried Leibniz, he invented calculus. For Newton, differentiation is the fundamental concept, and integration only means to recover a function from its derivative.

a year, then we will get $f(t) = \left(1 + \dfrac{c}{n}\right)^{nt} f(0)$. As n goes to infinity, we get the formula

$$f(t) = \lim_{n \to \infty} \left(1 + \frac{c}{n}\right)^{nt} f(0) = f(0)e^{ct}$$

for the account for which the interest is paid continuously. We will rigorously show later that the solutions to $f'(t) = cf(t)$ are indeed scalar multiples of e^{ct}.

2.1.5   Rules of Computation

Proposition 2.1.4. *The sum, the product and the composition of differentiable functions are differentiable. Moreover,*

$$(f(x) + g(x))' = f'(x) + g'(x),$$
$$(f(x)g(x))' = f'(x)g(x) + f(x)g'(x),$$
$$(g(f(x)))' = g'(f(x))f'(x).$$

The rules can also be written as

$$\frac{d(f+g)}{dx} = \frac{df}{dx} + \frac{dg}{dx}, \quad \frac{d(fg)}{dx} = \frac{df}{dx}g + f\frac{dg}{dx}, \quad \frac{dg}{dx} = \frac{dg}{df}\frac{df}{dx},$$

or in the differential form,

$$d(f+g) = df + dg, \quad d(fg) = g\,df + f\,dg, \quad d(g \circ f) = (g' \circ f)df.$$

The formula for the product is called the *Leibniz*[2] *rule*. The formula for the composition is called the *chain rule*.

A special case of the Leibniz rule is that for any constant c, we have

$$(cf)' = c'f + cf' = 0f + cf' = cf'.$$

The formula for the derivative of the sum can be explained through the linear approximation. Suppose the functions $f(x)$ and $g(x)$ are approximated by the linear functions

$$L(x) = f(x_0) + f'(x_0)\Delta x, \quad K(x) = g(x_0) + g'(x_0)\Delta x$$

near x_0. Then we expect $f(x) + g(x)$ to be approximated by the sum

$$L(x) + K(x) = (f(x_0) + g(x_0)) + (f'(x_0) + g'(x_0))\Delta x$$

of the linear functions. In particular, this implies the derivative $(f+g)'(x_0)$ is the coefficient $f'(x_0) + g'(x_0)$ of Δx.

---

[2]Gottfried Wilhelm von Leibniz, born 1646 in Leipzig, Saxony (Germany), died 1716 in Hannover (Germany). Leibniz was a great scholar who contributed to almost all the subjects in the human knowledge of his time. He invented calculus independent of Newton. He also invented the binary system, the foundation of modern computer system. He was, along with René Descartes and Baruch Spinoza, one of the three greatest 17th-century rationalists. Leibniz was perhaps the first major European intellect who got seriously interested in Chinese civilization. His fascination of I Ching may be related to his invention of bindary system.

Similarly, we expect $f(x)g(x)$ to be approximated by the product

$$L(x)K(x) = f(x_0)g(x_0) + (f'(x_0)g(x_0) + f(x_0)g'(x_0))\Delta x + f'(x_0)g'(x_0)\Delta x^2$$

of linear functions. Although $L(x)K(x)$ is not linear, it is further approximated by the linear function

$$M(x) = f(x_0)g(x_0) + (f'(x_0)g(x_0) + f(x_0)g'(x_0))\Delta x.$$

Therefore $f(x)g(x)$ is also approximated by the linear function $M(x)$. In particular, $(fg)'(x_0)$ is the coefficient $f'(x_0)g(x_0) + f(x_0)g'(x_0)$ of Δx.

For the composition $g(f(x))$, the functions $f(x)$ and $g(y)$ are approximated by the linear functions

$$L(x) = f(x_0) + f'(x_0)\Delta x, \quad K(y) = g(y_0) + g'(y_0)\Delta y$$

near x_0 and $y_0 = f(x_0)$. Then we expect $g(f(x))$ to be approximated by the composition

$$K(L(x)) = g(y_0) + g'(y_0)f'(x_0)\Delta x$$

of linear functions. In particular, this implies the derivative $(g \circ f)'(x_0)$ is the coefficient $g'(y_0)f'(x_0) = g'(f(x_0))f'(x_0)$ of Δx.

Proof of Proposition 2.1.4. We prove the formulae by rigorously verifying the claims made above regarding the linear approximations.

Consider the sum first. For any $\epsilon_1 > 0$, there are $\delta_1, \delta_2 > 0$, such that

$$|\Delta x| < \delta_1 \implies |f(x) - L(x)| \le \epsilon_1 |\Delta x|, \qquad (2.1.7)$$
$$|\Delta x| < \delta_2 \implies |g(x) - K(x)| \le \epsilon_1 |\Delta x|. \qquad (2.1.8)$$

This implies

$$|\Delta x| < \min\{\delta_1, \delta_2\} \implies |(f(x) + g(x)) - (L(x) + K(x))| \le 2\epsilon_1 |\Delta x|. \quad (2.1.9)$$

Thus for any $\epsilon > 0$, we take $\epsilon_1 = \dfrac{\epsilon}{2}$ and find $\delta_1, \delta_2 > 0$, such that (2.1.7) and (2.1.8) hold. Then (2.1.9) holds and becomes

$$|\Delta x| < \min\{\delta_1, \delta_2\} \implies |(f(x) + g(x)) - (L(x) + K(x))| \le \epsilon |\Delta x|.$$

This proves that $f(x) + g(x)$ is approximated by the linear function $L(x) + K(x)$.

Now consider the product. For $|\Delta x| < \min\{\delta_1, \delta_2\}$, we have

$$\begin{aligned}
|f(x)g(x) - L(x)K(x)| &\le |f(x)g(x) - L(x)g(x)| + |L(x)g(x) - L(x)K(x)| \\
&\le \epsilon_1 |\Delta x||g(x)| + \epsilon_1 |L(x)||\Delta x| \\
&\le \epsilon_1(|g(x)| + |L(x)|)|\Delta x|.
\end{aligned}$$

Then

$$\begin{aligned}
|f(x)g(x) - M(x)| &\le |f(x)g(x) - L(x)K(x)| + |f'(x_0)g'(x_0)\Delta x^2| \\
&\le [\epsilon_1(|g(x)| + |L(x)|) + |f'(x_0)g'(x_0)\Delta x|] \ |\Delta x|.
\end{aligned}$$

Since $|g(x)|$ and $|L(x)|$ are continuous at x_0, they are bounded near x_0 by Proposition 1.3.6. Therefore for any $\epsilon > 0$, it is not difficult to find $\delta_3 > 0$ and $\epsilon_1 > 0$, such that

$$|\Delta x| < \delta_3 \implies \epsilon_1(|g(x)| + |L(x)|) + |f'(x_0)g'(x_0)\Delta x| \le \epsilon.$$

Next for this ϵ_1, we may find $\delta_1, \delta_2 > 0$, such that (2.1.7) and (2.1.8) hold. Then we have

$$|\Delta x| < \min\{\delta_1, \delta_2, \delta_3\} \implies |f(x)g(x) - M(x)| \le \epsilon|\Delta x|.$$

This proves that $f(x)g(x)$ is approximated by the linear function $M(x)$.

Finally consider the composition. For any $\epsilon_1, \epsilon_2 > 0$, there are $\delta_1, \delta_2 > 0$, such that

$$|\Delta x| = |x - x_0| < \delta_1 \implies |f(x) - L(x)| \le \epsilon_1|\Delta x|, \qquad (2.1.10)$$
$$|\Delta y| = |y - y_0| < \delta_2 \implies |g(y) - K(y)| \le \epsilon_2|\Delta y|. \qquad (2.1.11)$$

Then for $y = f(x)$, we have

$$|\Delta x| < \delta_1 \implies |\Delta y| = |f(x) - f(x_0)| \le |f(x) - L(x)| + |f'(x_0)\Delta x|$$
$$\le (\epsilon_1 + |f'(x_0)|)|\Delta x| < (\epsilon_1 + |f'(x_0)|)\delta_1. \qquad (2.1.12)$$

and

$$|\Delta x| < \delta_1, |\Delta y| < \delta_2 \implies |g(f(x)) - K(L(x))|$$
$$\le |g(f(x)) - K(f(x))| + |K(f(x)) - K(L(x))|$$
$$= |g(y) - K(y)| + |g'(x_0)||f(x) - L(x)|$$
$$\le \epsilon_2|\Delta y| + |g'(x_0)|\epsilon_1|\Delta x|$$
$$\le [\epsilon_2(\epsilon_1 + |f'(x_0)|) + \epsilon_1|g'(x_0)|]|\Delta x|. \qquad (2.1.13)$$

Suppose for any $\epsilon > 0$, we can find suitable $\delta_1, \delta_2, \epsilon_1, \epsilon_2$, such that

$$(\epsilon_1 + |f'(x_0)|)\delta_1 \le \delta_2, \qquad (2.1.14)$$
$$\epsilon_2(\epsilon_1 + |f'(x_0)|) + \epsilon_1|g'(x_0)| \le \epsilon, \qquad (2.1.15)$$

and (2.1.10), (2.1.11) hold. Then (2.1.12) tells us $|\Delta x| < \delta_1$ implies $|\Delta y| < \delta_2$, and (2.1.13) becomes

$$|\Delta x| < \delta_1 \implies |g(f(x)) - K(L(x))| \le [\epsilon_2(\epsilon_1 + |f'(x_0)|) + \epsilon_1|g'(x_0)|]|\Delta x| \le \epsilon|\Delta x|.$$

This proves that $f(g(x))$ is approximated by the linear function $K(L(x))$.

It remains to find $\delta_1, \delta_2, \epsilon_1, \epsilon_2$ such that (2.1.10), (2.1.11), (2.1.14) and (2.1.15) hold. For any $\epsilon > 0$, we first find $\epsilon_1, \epsilon_2 > 0$ satisfying (2.1.15). For this $\epsilon_2 > 0$, we find $\delta_2 > 0$, such that the implication (2.1.11) holds. Then for the $\epsilon_1 > 0$ we already found, it is easy to find $\delta_1 > 0$ such that the implication (2.1.10) holds and (2.1.14) is also satisfied.

$\square$

Exercise 2.1.13. A function is *odd* if $f(-x) = -f(x)$. It is *even* if $f(-x) = f(x)$. A function is *periodic* with period p if $f(x + p) = f(x)$. Prove that the derivatives of odd, even, periodic functions are respectively even, odd, periodic.

2.1.6   Basic Example

We establish the following derivatives.

$$(x^\alpha)' = \alpha x^{\alpha-1},$$
$$(\sin x)' = \cos x,$$
$$(\cos x)' = -\sin x,$$
$$(\tan x)' = \sec^2 x,$$
$$(e^x)' = e^x,$$
$$(\alpha^x)' = \alpha^x \log \alpha.$$

By the limit (1.4.13), for $x > 0$ we have

$$\frac{dx^\alpha}{dx} = \lim_{\Delta x \to 0} \frac{(x + \Delta x)^\alpha - x^\alpha}{\Delta x} = \lim_{u \to 0} \frac{(x + xu)^\alpha - x^\alpha}{xu}$$
$$= \lim_{u \to 0} x^{\alpha-1} \frac{(1 + u)^\alpha - 1}{u} = \alpha x^{\alpha-1}.$$

For $x < 0$ and integer α, let $y = -x$. By the chain rule,

$$\frac{dx^\alpha}{dx} = \frac{d((-y)^\alpha)}{dy} \frac{dy}{dx} = (-1)^\alpha \frac{d(y^\alpha)}{dy} \frac{d(-x)}{dx} = (-1)^\alpha \alpha y^{\alpha-1}(-1) = \alpha x^{\alpha-1}.$$

For natural number α, we also have

$$\frac{dx^\alpha}{dx}\bigg|_{x=0} = \lim_{\Delta x \to 0} \frac{(0 + \Delta x)^\alpha - 0^\alpha}{\Delta x} = \lim_{\Delta x \to 0} (\Delta x)^{\alpha-1} = \begin{cases} 0 & \text{if } \alpha > 1 \\ 1 & \text{if } \alpha = 1 \end{cases} = \alpha x^{\alpha-1}\big|_{x=0}.$$

Thus the formula for the derivative of the power function holds whenever the function is defined.

The derivative of power functions leads to the derivative for polynomials

$$(c_n x^n + c_{n-1} x^{n-1} + \cdots + c_1 x + c_0)' = n c_n x^{n-1} + (n-1) c_{n-1} x^{n-2} + \cdots + c_1.$$

On the other hand, by the derivative of x^{-1} and the chain rule, we get the derivative

$$\left(\frac{1}{f}\right)' = (-1)(f)^{-2} f' = -\frac{f'}{f^2}$$

of the reciprocal of any differentiable function. By further applying the Leibniz rule, we get the derivative

$$\left(\frac{f}{g}\right)' = f' \frac{1}{g} + f \left(\frac{1}{g}\right)' = \frac{f'g - fg'}{g^2} \qquad (2.1.16)$$

of the quotient of differentiable functions. In particular, we are able to compute the derivative of any rational function.

The limits (1.3.23) and (1.3.25) tell us

$$\frac{d \sin x}{dx}\bigg|_{x=0} = \lim_{x \to 0} \frac{\sin x - \sin 0}{x} = \lim_{x \to 0} \frac{\sin x}{x} = 1,$$

$$\frac{d \cos x}{dx}\bigg|_{x=0} = \lim_{x \to 0} \frac{\cos x - \cos 0}{x} = \lim_{x \to 0} x \frac{\cos x - 1}{x^2} = 0.$$

At any $x = a$, let $y = x - a$. Then by the chain rule,

$$\left.\frac{d\sin x}{dx}\right|_{x=a} = \left.\frac{d\sin(y+a)}{dy}\right|_{y=0}\left.\frac{dy}{dx}\right|_{x=a} = \left.\frac{d}{dy}(\sin y \cos a + \cos y \sin a)\right|_{y=0}$$

$$= \left.\frac{d\sin y}{dy}\right|_{y=0}\cos a + \left.\frac{d\cos y}{dy}\right|_{y=0}\sin a = 1\cos a + 0\sin a = \cos a.$$

Similar computation gives us the derivative of the cosine function. Further applying (2.1.16) gives us the derivative of the tangent function.

$$\frac{d\tan x}{dx} = \frac{d}{dx}\left(\frac{\sin x}{\cos x}\right) = \frac{\cos x \cos x - (-\sin x)\sin x}{\cos x^2} = \frac{1}{\cos x^2} = \sec^2 x.$$

The limit (1.4.12) tells us

$$\frac{de^x}{dx} = \lim_{\Delta x \to 0}\frac{e^{x+\Delta x} - e^x}{\Delta x} = \lim_{\Delta x \to 0}e^x\frac{e^{\Delta x} - 1}{\Delta x} = e^x.$$

The derivative of the other exponential functions can be obtained by the chain rule

$$\frac{d\alpha^x}{dx} = \frac{de^{x\log\alpha}}{dx} = \left.\frac{de^y}{dy}\right|_{y=x\log\alpha}\frac{d(x\log\alpha)}{dx} = e^{x\log\alpha}\log\alpha = \alpha^x\log\alpha.$$

Note that all the exponential functions are the solutions of the differential equations of the form $f' = cf$, where c is a constant (see Example 2.2.6 for the converse). Such functions are characterized by the important property that the rate of change is proportional to the value of the function itself (see the discussion on interest rate in Example 2.1.10). From this viewpoint, the case $c = 1$ is the purest and the most natural case. Only the exponential function based on e corresponds to $c = 1$. This is why e^x is called the *natural* exponential function.

Example 2.1.11. The function $f(x) = \sqrt{1 + \sin^2 x}$ is the composition of $f = \sqrt{u}$, $u = 1 + v^2$, $v = \sin x$. Therefore

$$\frac{df}{dx} = \frac{d\sqrt{u}}{du}\frac{d(1+v^2)}{dv}\frac{d(\sin x)}{dx} = \frac{1}{2\sqrt{u}}2v\cos x = \frac{\sin x \cos x}{\sqrt{1 + \sin^2 x}}.$$

Example 2.1.12. The equation $x^2 + \dfrac{y^2}{4} = 1$ is an ellipse. The ellipse has upper and lower parts given respectively by $y = 2\sqrt{1 - x^2}$ and $y = -2\sqrt{1 - x^2}$. By the chain rule, the derivative of the upper part is

$$\frac{dy}{dx} = \left.\frac{d(2u^{\frac{1}{2}})}{du}\right|_{u=1-x^2}\frac{d(1-x^2)}{dx} = 2\frac{1}{2}u^{-\frac{1}{2}}(-2x) = -\frac{2x}{\sqrt{1-x^2}}.$$

Alternatively, instead of finding the explicit formula for y in terms of x, we keep in mind that y is a function of x and take the equation to mean $x^2 + \dfrac{y(x)^2}{4} = 1$. Then taking $\dfrac{d}{dx}$ on both sides of the equation gives us

$$0 = \frac{d}{dx}\left(x^2 + \frac{y^2}{4}\right) = \frac{d(x^2)}{dx} + \frac{1}{4}\frac{d(y^2)}{dy}\frac{dy}{dx} = 2x + \frac{1}{4}2y\frac{dy}{dx}.$$

Solving the equation, we get
$$\frac{dy}{dx} = -\frac{4x}{y}.$$

Yet another way of finding $\dfrac{dy}{dx}$ is by the parametrization $x = \cos t$, $y = 2 \sin t$ of the ellipse. The chain rule $\dfrac{dy}{dt} = \dfrac{dy}{dx}\dfrac{dx}{dt}$ tells us

$$\frac{dy}{dx} = \frac{\dfrac{dy}{dt}}{\dfrac{dx}{dt}} = \frac{2 \cos t}{-\sin t} = -2 \cot t.$$

The reader is left to verify that the three ways give the same result.

Exercise 2.1.14. Compute the derivatives of the trigonometric functions $\cot x$, $\sec x$, $\csc x$.

Exercise 2.1.15. Compute the derivative of $\log x$ by making use of the limit (1.4.11).

Exercise 2.1.16. Compute the derivatives.

1. $\dfrac{1}{x^2 + 1}$.

2. $\dfrac{2x^2 + 1}{x^2 + 1}$.

3. $\left(\dfrac{2x^2 + 1}{x^2 + 1}\right)^4$.

4. $\sqrt{1 + x^2}$.

5. $\sqrt{1 + \sqrt{2 + \sqrt{3 + x}}}$.

6. $(\sin 2x + \cos 3x)^7$.

7. $\cos(\sin(\tan x))$.

8. $\sin^2 \dfrac{1}{x}$.

9. $2^{\sin 3x}$.

10. $e^{2x}(\cos x - 2 \sin x)$.

11. $\dfrac{2^{3x} - 3x^2}{3^{2x} + 2x^3}$.

Exercise 2.1.17. The *hyperbolic functions* are

$$\operatorname{sh} x = \frac{e^x - e^{-x}}{2}, \quad \operatorname{ch} x = \frac{e^x + e^{-x}}{2}, \quad \operatorname{th} x = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \quad \operatorname{coth} x = \frac{e^x + e^{-x}}{e^x - e^{-x}}.$$

Express the derivatives of hyperbolic functions in terms of hyperbolic functions.

Exercise 2.1.18. Suppose $f(x)$ and $g(x)$ are differentiable functions satisfying $f(2) = 3$, $g(2) = 2$, $f'(2) = 1$, $g'(2) = -1$. Compute the derivatives of the following functions at 2.

$$f(x)g(x), \quad e^{f(x)} \sin \pi g(x), \quad f(g(x)), \quad g(2f(x)^2 - 8g(x)).$$

Exercise 2.1.19. Suppose $u = u(x)$ and $v = v(x)$ are differentiable functions satisfying $u(0) = 1$, $v(0) = -1$ and the given equations. Compute the derivatives of u and v at 0.

1. $u^2 + uv + v^2 = 1$, $(1 + x^2)u + (1 - x^2)v = x$.

2. $xu + (x + 1)v = e^x$, $e^u + xe^{-v} = e^{x+1}$.

Exercise 2.1.20. Suppose y is a function of x satisfying given equation. Compute the derivative $\dfrac{dy}{dx}$ at the given point.

1. $y^3 + 2xy^2 - 2x^3 = 1$, at $x = 1$, $y = 1$.

2. $x \sin(x - y) = y \cos(x + y)$, at $x = \pi$, $y = 0$.

3. $e^y = xy$, at $x = -e^{-1}$, $y = -1$.

4. $(1 + y)e^x + (1 + x)e^y = xy + 2$, at $x = 0$, $y = 0$.

5. $y^2 \sin x = x^2 \cos y$, at $x = \dfrac{\pi}{4}$, $y = \dfrac{\pi}{4}$.

Exercise 2.1.21. For the parametrized curves, compute the derivative $\dfrac{dy}{dx}$ at the given point. Then find the equation for the tangent line of the curve.

1. Cycloid: $x = t - \sin t$, $y = 1 - \cos t$, at $t = \dfrac{\pi}{4}$.

2. Spiral: $x = t \cos t$, $y = t \sin t$, at $t = \pi$.

3. Involute of circle: $x = \cos t + t \sin t$, $y = \sin t - t \cos t$, at $t = \pi$.

4. Four-leaved rose: $x = \cos 2t \cos t$, $y = \cos 2t \sin t$, at $t = \dfrac{\pi}{4}$.

2.1.7   Derivative of Inverse Function

The inverse trigonometric functions and the logarithmic functions were defined as inverse functions. In order to compute their derivatives, we need to know how to compute the derivatives of inverse functions.

If a function $y = f(x)$ is approximated by a linear function $y = A + Bx$, then the inverse function $x = f^{-1}(y)$ should be approximated by the inverse linear function $x = -B^{-1}A + B^{-1}y$. This suggests that $(f^{-1})' = B^{-1} = \dfrac{1}{f'}$.

Proposition 2.1.5. *Suppose a continuous function $f(x)$ is invertible near x_0 and is differentiable at x_0. If $f'(x_0) \neq 0$, then the inverse function is also differentiable at $y_0 = f(x_0)$, with*

$$(f^{-1})'(y_0) = \frac{1}{f'(x_0)}. \tag{2.1.17}$$

The formula can also be written as

$$\frac{dx}{dy} = \left(\frac{dy}{dx} \right)^{-1} \quad \text{for } y = f(x), x = f^{-1}(y). \tag{2.1.18}$$

Proof. Let $f(x)$ be invertible and differentiable. By Proposition 2.1.3, it is continuous. By Theorem 1.4.8, the inverse function is also continuous. Therefore for $y = f(x)$, we have $x \to x_0$ if and only if $y \to y_0$. Moreover, because f is invertible, we have $x \neq x_0$ if and only if $y \neq y_0$. By the composition rule, we may substitute x, x_0 by $f^{-1}(y), f^{-1}(y_0)$ in the limit

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

and get

$$f'(x_0) = \lim_{y \to y_0} \frac{y - y_0}{f^{-1}(y) - f^{-1}(y_0)}.$$

If $f'(x_0) \neq 0$, then this implies

$$\frac{1}{f'(x_0)} = \lim_{y \to y_0} \frac{f^{-1}(y) - f^{-1}(y_0)}{y - y_0}.$$

Thus $f^{-1}(y)$ is differentiable at y_0 with $\dfrac{1}{f'(x_0)}$ as the derivative.

$\square$

For more detailed discussion on the relation between the invertibility of $f(x)$ near x_0 and $f'(x_0) \neq 0$, see Exercises 2.1.22, 2.1.24, and the remark after Proposition 2.2.4.

Exercise 2.1.22. Suppose $f(x)$ is invertible near x_0 and is differentiable at x_0. Prove that if the inverse function is differentiable at $y_0 = f(x_0)$, then $f'(x_0) \neq 0$. This is the "conditional" converse of Proposition 2.1.5.

Exercise 2.1.23. Suppose $f(x)$ is invertible near x_0 and is differentiable at x_0. Prove directly that if $L(x) = f(x_0) + b\Delta x$ is the linear approximation of $f(x)$ at x_0 with $b \neq 0$, then $L^{-1}(y) = x_0 + b^{-1}\Delta y$ is the linear approximation of $f^{-1}(x)$ at $f(x_0)$. This gives an alternative proof of Proposition 2.1.5.

Exercise 2.1.24. Consider the function

$$f(x) = \begin{cases} \dfrac{n}{n^2 + 1} & \text{if } x = \dfrac{1}{n}, n \in \mathbb{N} \\ x & \text{otherwise} \end{cases}.$$

Verify that $f'(0) = 1$ but $f(x)$ is not one-to-one. This shows that the invertibility condition is necessary in Proposition 2.1.5. In particular, $f'(x_0) \neq 0$ does not necessarily imply the function is invertible near x_0.

Now we are ready to derive the following derivatives.

$$(\arcsin x)' = \frac{1}{\sqrt{1 - x^2}},$$
$$(\arccos x)' = -\frac{1}{\sqrt{1 - x^2}},$$
$$(\arctan x)' = \frac{1}{1 + x^2},$$
$$(\operatorname{arcsec} x)' = \frac{1}{x\sqrt{x^2 - 1}},$$
$$(\log |x|)' = \frac{1}{x},$$
$$(\log_\alpha |x|)' = \frac{1}{x \log \alpha}.$$

Let $y = \arcsin x$. Then $x = \sin y$, $-\dfrac{\pi}{2} \leq y \leq \dfrac{\pi}{2}$, and

$$\frac{d \arcsin x}{dx} = \left(\frac{d \sin y}{dy} \right)^{-1} = \frac{1}{\cos y} = \frac{1}{\sqrt{1 - \sin^2 y}} = \frac{1}{\sqrt{1 - x^2}}.$$

The derivative of the inverse cosine can be derived similarly, or derived from $\arcsin x + \arccos x = \dfrac{\pi}{2}$. For the derivative of the inverse tangent, let $y = \arctan x$. Then $x = \tan y$, $-\dfrac{\pi}{2} \leq y \leq \dfrac{\pi}{2}$, and

$$\frac{d \arctan x}{dx} = \left(\frac{d \tan y}{dy}\right)^{-1} = \frac{1}{\sec^2 y} = \frac{1}{1 + \tan^2 y} = \frac{1}{1 + x^2}.$$

Using $\sec' x = \sec x \tan x$, we also get the derivative for inverse secant.

$$\frac{d \operatorname{arcsec} x}{dx} = \frac{1}{\sec y \tan y}\bigg|_{y=\operatorname{arcsec} x} = \frac{1}{\sec y \sqrt{\sec^2 y - 1}}\bigg|_{y=\operatorname{arcsec} x} = \frac{1}{x\sqrt{x^2 - 1}}.$$

The derivative of the logarithmic function can be obtained from the derivative of the exponential function. For $x > 0$,

$$\frac{d \log x}{dx} = \left(\frac{d e^y}{dy}\right)^{-1}\bigg|_{y=\log x} = \frac{1}{e^y}\bigg|_{y=\log x} = \frac{1}{x}.$$

For $x < 0$, $y = \log|x|$, we have $x = -|x| = -e^y$ and

$$\frac{d \log|x|}{dx} = \left(\frac{d(-e^y)}{dy}\right)^{-1}\bigg|_{y=\log|x|} = \frac{1}{-e^y}\bigg|_{y=\log|x|} = \frac{1}{-|x|} = \frac{1}{x}.$$

Thus the formula holds for any $x \neq 0$. The derivative for the other logarithmic functions can be then obtained from $\log_\alpha x = \dfrac{\log x}{\log \alpha}$.

Example 2.1.13. To compute the derivative of $f(x) = \dfrac{(x+2)^7 \sqrt{x^2+1}}{(x^2 - x + 1)^6}$, we take the log and get

$$\log f(x) = 7 \log(x+2) + \frac{1}{2} \log(x^2 + 1) - 6 \log(x^2 - x + 1).$$

Taking the derivative, we get

$$\frac{f'(x)}{f(x)} = 7\frac{1}{x+2} + \frac{2x}{2(x^2 + 1)} - 6\frac{2x - 1}{x^2 - x + 1}.$$

Thus

$$f'(x) = \frac{(x+2)^7 \sqrt{x^2+1}}{(x^2 - x + 1)^6}\left(\frac{7}{x+2} + \frac{x}{x^2 + 1} - \frac{12x - 6}{x^2 - x + 1}\right).$$

Exercise 2.1.25. Compute the derivatives.

1. $\arcsin \dfrac{1}{x}$.

2. $\dfrac{\arcsin x}{\arccos x}$.

3. $(\arctan(1 + x^2))^3$.

4. $x(\log x - 1)$.

5. $\log(\log x)$.

6. $\log(x + \sqrt{1 + x^2})$.

7. $\log|\sin x|$.

8. $\arctan(\log x)$.

9. $\arcsin(\arccos x)$.

Exercise 2.1.26. Suppose $f(x)$ and $g(x)$ are differentiable functions satisfying $f(2) = 3$, $g(2) = 2$, $f'(2) = 1$, $g'(2) = -1$. Compute the derivatives of the following functions at 2.

$$\log(f(x) + g(x)), \quad \arcsin \frac{f(x)}{3\sqrt{g(x)}}, \quad f(1 + \log_2 g(x)), \quad g\left(\frac{6\arctan\sqrt{f(x)}}{\pi}\right).$$

Exercise 2.1.27. Suppose y is a function of x satisfying $\arctan \dfrac{y}{x} = \log\sqrt{x^2 + y^2}$. Compute the derivative of y and express the result in terms of x and y.

Exercise 2.1.28. Compute the derivatives by making use of the logarithmic function.

1. $\dfrac{(x^2 + 1)^3 \sqrt{x - 2}}{(x + 3)^9 \sqrt[3]{x^3 - x + 2}}$.

2. x^x.

3. $\sqrt[x]{\dfrac{1 + x}{\sin x + \cos x}}$.

4. $f(x)^{g(x)}$.

5. $x^{(x^x)}$.

6. $(1 + \sin x)^{\arccos x}$.

Exercise 2.1.29. Find the inverse of hyperbolic functions in Exercise 2.1.17. Then compute the derivatives in two ways: by direct computation and by using Proposition 2.1.5.

2.1.8 Additional Exercise

One Side Derivative

Define *one side derivatives*

$$f'_+(x_0) = \lim_{x \to x_0^+} \frac{f(x) - f(x_0)}{x - x_0}, \quad f'_-(x_0) = \lim_{x \to x_0^-} \frac{f(x) - f(x_0)}{x - x_0}.$$

By Proposition 1.3.3, the usual (two side) derivative exists if and only if the left and right side derivatives exist and are equal.

Exercise 2.1.30. Define one side differentiability and prove the equivalence between the right differentiability and the existence of the right derivative.

Exercise 2.1.31. For $\alpha > 0$, compute the right derivative of the power function x^α at 0.

Exercise 2.1.32. Determine whether the function $\begin{cases} x^x & \text{if } x > 0 \\ 1 & \text{if } x = 0 \end{cases}$ is right differentiable at $x = 0$.

Exercise 2.1.33. Prove right differentiability implies right continuity.

Exercise 2.1.34. Prove the properties $(f + g)'_+(x) = f'_+(x) + g'_+(x)$ and $(fg)'_+(x) = f'_+(x)g(x) + f(x)g'_+(x)$ for the right derivative.

Exercise 2.1.35. Suppose $f'_+(x_0) > 0$. Prove that $f(x) > f(x_0)$ for $x > x_0$ and near x_0.

Exercise 2.1.36. Suppose $f(x)$ and $g(y)$ are right differentiable at x_0 and $y_0 = f(x_0)$. Prove that under one of the following conditions, the composition $g(f(x))$ is right differentiable at x_0, and we have the chain rule $(g \circ f)'_+(x_0) = g'_+(y_0)f'_+(x_0)$.

    1. $f(x) \geq f(x_0)$ for $x \geq x_0$.

   2. $g(y)$ is (two side) differentiable at y_0.

Note that by Exercise 2.1.35, the first condition is satisfied if $f'_+(x_0) > 0$. Can you find the other chain rules for one side derivatives?

Exercise 2.1.37. State and prove the one side derivative version of Proposition 2.1.5.

2.2   Application of Differentiation

Having defined and computed the linear approximations, we are ready to solve problems related to the change of functions.

2.2.1   Maximum and Minimum

A function $f(x)$ has a *local maximum* at x_0 if $f(x_0)$ is the biggest value among all the values near x_0. In other words, there is $\delta > 0$, such that

$$|x - x_0| < \delta \implies f(x_0) \geq f(x). \tag{2.2.1}$$

Similarly, $f(x)$ has a *local minimum* at x_0 if there is $\delta > 0$, such that

$$|x - x_0| < \delta \implies f(x_0) \leq f(x). \tag{2.2.2}$$

Local maxima and local minima are also called *local extrema.*

In the definition above, the function is assumed to be defined on both sides of x_0. Similar definition can be made when the function is defined on only one side of x. For example, suppose $f(x)$ is defined on a bounded closed interval $[a, b]$. Then a is a local maximum if there is $\delta > 0$, such that

$$0 \leq x - a < \delta \implies f(a) \geq f(x).$$
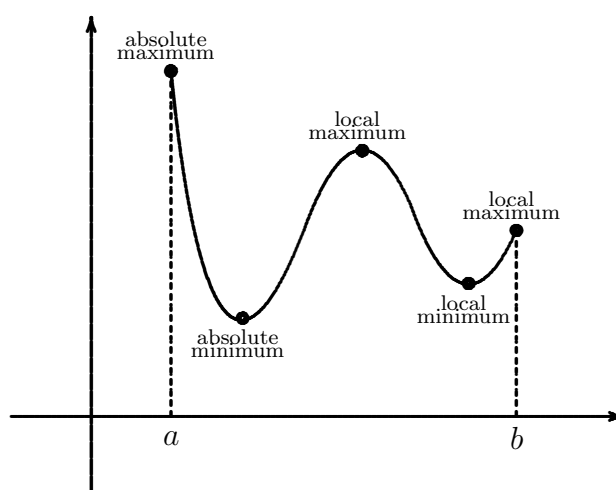
Figure 2.2: maximum and minimum

A function $f(x)$ has an *absolute maximum* at x_0 if $f(x_0)$ is the biggest value among all the values of $f(x)$. In other words, we have $f(x_0) \geq f(x)$

for any x in the domain of f. The concept of *absolute minimum* is similarly defined. Absolute extrema are clearly also local extrema.

The following result tells us how to find local extrema.

Proposition 2.2.1. *Suppose a function $f(x)$ is defined on both sides of x_0 and has a local extreme at x_0. If $f(x)$ is differentiable at x_0, then $f'(x_0) = 0$.*

Proof. Suppose $f(x)$ is differentiable at x_0, with $f'(x_0) > 0$. Take $\epsilon = \dfrac{f'(x_0)}{2}$. Then there is $\delta > 0$, such that

$$|x - x_0| < \delta \implies |f(x) - f(x_0) - f'(x_0)(x - x_0)| \leq \epsilon |x - x_0|.$$

For $\delta > x - x_0 > 0$, this tells us

$$f(x) - f(x_0) \geq f'(x_0)(x - x_0) - \epsilon |x - x_0| = (f'(x_0) - \epsilon)(x - x_0) > 0.$$

For $0 > x - x_0 > -\delta$, this tells us

$$f(x) - f(x_0) \leq f'(x_0)(x - x_0) + \epsilon |x - x_0| = (f'(x_0) - \epsilon)(x - x_0) < 0.$$

Thus x_0 is not a local extreme. Similarly, if $f'(x_0) < 0$, then x_0 is also not a local extreme.

$\square$

Note that the proof makes critical use of both sides of x_0. For a function $f(x)$ defined on a bounded closed interval $[a, b]$, this means that the proposition may be applied to the *interior* points of the interval where the function is differentiable. Therefore the proposition tells us that a local extreme point x_0 must be one of the following three cases.

1. $a < x_0 < b$, $f'(x_0)$ does not exist.

2. $a < x_0 < b$, $f'(x_0)$ exists and is 0.

3. $x_0 = a$ or b.

Typically, the three possibilities would provide finitely many candidates for the potential local extrema. If we take the maximum and minimum of the values at these points, we will get the absolute maximum and minimum.

Example 2.2.1. The function $f(x) = x^2 e^x$ is differentiable everywhere, with the derivative $f'(x) = x(x + 2)e^x$. From $f'(x) = 0$, we get two possible local extrema $x = -2$ and $x = 0$. By $\lim_{x \to -\infty} f(x) = 0$, $\lim_{x \to \infty} f(x) = +\infty$, $f(x) \geq 0$ for any x, $f(-2) = 4e^{-2}$, $f(0) = 0$, we conclude that the function has the absolute minimum at 0, and the function has no absolute maximum. As to whether $x = -2$ is a local extreme, see Example 2.2.7.

Example 2.2.2. For the function $f(x) = (x + 1)x^{\frac{2}{3}}$ on $[-1, 1]$, we have $f'(x) = \dfrac{1}{3}(5x + 2)x^{-\frac{1}{3}}$. The derivative vanishes at $x = -\dfrac{2}{5}$ and does not exist at $x = 0$. Thus the possible local extrema are the two points and the end points -1, 1. By $f(-1) = 0$, $f\left(-\dfrac{2}{5}\right) = \dfrac{3\sqrt[3]{20}}{25}$, $f(0) = 0$, $f(1) = 2$, we conclude that the function has absolute minimum at -1 and 0 and has absolute maximum at 1.

Example 2.2.3. What is the biggest rectangle with circumference 1?

Let one side of the rectangle be x. From practical consideration, we have $0 \le x \le \dfrac{1}{2}$, and the area is $A(x) = x\left(\dfrac{1}{2} - x\right)$. Thus the problem becomes finding the maximum of $A(x)$ on the interval $\left[0, \dfrac{1}{2}\right]$. Since $A(x)$ is differentiable, the potential candidates are either the end points of the interval or the interior points where $A'(x) = \dfrac{1}{2} - 2x = 0$. So the possible local extrema are 0, $\dfrac{1}{4}$, $\dfrac{1}{2}$. The values of A at the three points are 0, $\dfrac{1}{16}$, 0, respectively. The biggest value $\dfrac{1}{16}$ appears at $x = \dfrac{1}{4}$, in which case the rectangle is a square.

Example 2.2.4. The function $f(x) = x^3$ satisfies $f'(0) = 0$. However, we have $f(x) < f(0)$ for $x < 0$ and $f(x) > f(0)$ for $x > 0$. Therefore 0 is not a local extreme for $f(x)$. The example shows that $f'(x_0) = 0$ is a necessary instead of a sufficient condition for local extrema (of differentiable functions).

Exercise 2.2.1. Find the maximum and the minimum on the given range.

1. $x(x + 1)(x + 2)$ on $[-3, 3]$.

2. $x(x + 1)(x + 2)$ on $[-3, 0]$.

3. $|x|e^x$ on $[-2, 1]$.

4. $\sin x + \cos x$ on $[0, 2\pi]$.

5. $\sin x + \cos x$ on whole $\mathbb{R}$.

6. $(x + 1)\cos x + (x - 1)\sin x$ on $[-\pi, \pi]$.

Exercise 2.2.2. How much of the corners of a square with side length 1 should be cut so that the tray made from it has the biggest volume?

Exercise 2.2.3. Given the surface area of a cylinder, when is the volume biggest?

Exercise 2.2.4. Find the distance from the point $(1, 1)$ to the parabola $y^2 = 2x$.

Exercise 2.2.5. An arrow is shot at the angle α. The location of the arrow at time t is
$$x = tv\cos\alpha, \;\; y = tv\sin\alpha - \frac{1}{2}gt^2,$$
where v is the initial speed, x is the distance traveled and y is the height. For what angle α does the arrow travel the furthest?

Exercise 2.2.6. Suppose $f(x)$ is left differentiable at x_0 (see Exercise 2.1.30). Prove that if x_0 is a local maximum, then $f'(x_0^-) \ge 0$. What about the right differentiable case?

2.2.2   Mean Value Theorem

Theorem 2.2.2 (Main Value Theorem). *Suppose $f(x)$ is continuous on $[a, b]$ and differentiable on (a, b). Then there is $a < c < b$, such that*

$$f'(c) = \frac{f(b) - f(a)}{b - a}. \tag{2.2.3}$$

The quotient on the right is the slope of the line segment that connects the end points of the graph of $f(x)$ on $[a, b]$. The theorem says the line segment is parallel to some tangent line.

The conclusion of the theorem can also be written as

$$f(b) - f(a) = f'(c)(b - a) \text{ for some } a < c < b, \qquad (2.2.4)$$

or

$$f(x + \Delta x) - f(x) = f'(x + \theta \Delta x)\Delta x \text{ for some } 0 < \theta < 1. \qquad (2.2.5)$$

Also note that a and b may be exchanged in the theorem, so that we do not have to insist $a < b$ (or $\Delta x > 0$) for the theorem to hold.
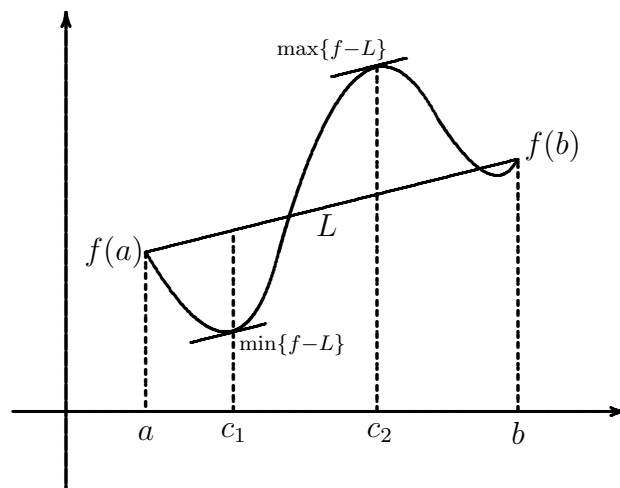
Figure 2.3: mean value theorem

Proof. The line connecting $(a, f(a))$ and $(b, f(b))$ is

$$L(x) = f(a) + \frac{f(b) - f(a)}{b - a}(x - a).$$

As suggested by Figure 2.3, the tangent lines parallel to $L(x)$ can be found where the difference

$$h(x) = f(x) - L(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b - a}(x - a)$$

reaches maximum or minimum.

Since $h(x)$ is continuous, by Theorem 1.4.5, it reaches maximum and minimum at $c_1, c_2 \in [a, b]$. If both c_1 and c_2 are end points a and b, then the maximum and the minimum of $h(x)$ are $h(a) = h(b) = 0$. This implies $h(x)$ is constantly zero on the interval, so that $h'(c) = 0$ for any $a < c < b$. If one of c_1 and c_2, denoted c, is not an end point, then by Proposition 2.2.1, we have $h'(c) = 0$. In any case, we have $a < c < b$ satisfying

$$h'(c) = f'(c) - \frac{f(b) - f(a)}{b - a} = 0.$$

$\square$

Example 2.2.5. We prove $\dfrac{x}{1+x} \leq \log(1+x) \leq x$ for $x > -1$. Taking $f(x)$, a, b in the mean value theorem to be $\log(1+x)$, 0, x, we have

$$\log(1+x) = \log(1+x) - \log(1+0) = \left.\frac{d\log(1+u)}{du}\right|_{u=\theta x}(x-0) = \frac{x}{1+\theta x},$$

where $0 < \theta < 1$. For $x > 0$, we have $1 < 1 + \theta x < 1 + x$. For $-1 < x < 0$, we have $1 > 1 + \theta x > 1 + x$. In both cases, we conclude

$$\frac{x}{1+x} < \frac{x}{1+\theta x} < x.$$

Exercise 2.2.7. Find c in the mean value theorem.

1. x^3 on $[-1, 1]$.   2. $\dfrac{1}{x}$ on $[1, 2]$.   3. 2^x on $[0, 1]$.

Exercise 2.2.8. Let $f(x) = |x - 1|$. Is there $0 < c < 3$ such that $f(3) - f(0) = f'(c)(3 - 0)$? Does your conclusion contradict the means value theorem?

Exercise 2.2.9. Suppose $f(x)$ is continuous on $[a, b]$ and differentiable on (a, b). Prove that $f(x)$ is Lipschitz (see Exercise 1.4.13) if and only if $f'(x)$ is bounded on (a, b).

Exercise 2.2.10. Use the main value theorem to prove $|\sin x - \sin y| < |x - y|$ and $|\arctan x - \arctan y| < |x - y|$.

Exercise 2.2.11. Prove that $\dfrac{x}{1+x^2} < \arctan x < x$ for $x > 0$.

Exercise 2.2.12 (Rolle[3]'s Theorem). Suppose $f(x)$ is continuous on $[a, b]$ and differentiable on (a, b). Prove that if $f(a) = f(b)$, then $f'(c) = 0$ for some $a < c < b$.

Exercise 2.2.13. Suppose $f(x)$ has continuous derivative on a bounded and closed interval $[a, b]$. Prove that for any $\epsilon > 0$, there is $\delta > 0$, such that $|\Delta x| < \delta$ implies $|f(x + \Delta x) - f(x) - f'(x)\Delta x| \leq \epsilon|\Delta x|$. In other words, $f(x)$ is *uniformly differentiable*.

Exercise 2.2.14. Suppose $f(x)$ is continuous on $[a, b]$ and differentiable on (a, b). Suppose $f(a) = 0$ and $|f'(x)| \leq A|f(x)|$ for some constant A. First prove that $f(x) = 0$ on $\left[a, a + \dfrac{1}{2A}\right]$. Then further prove that $f(x) = 0$ on the whole interval $[a, b]$.

Exercise 2.2.15. Suppose $f(x)$ is continuous on $[a, b]$ and is left and right differentiable on (a, b) (see Exercise 2.1.30). Prove that there is $a < c < b$, such that $\dfrac{f(b) - f(a)}{b - a}$ lies between $f'_-(c)$ and $f'_+(c)$. A further extension of the mean value theorem appears in Exercise 2.2.37.

The mean value theorem has the following important consequence, which basically says that a non-changing quantity must be a constant.

Proposition 2.2.3. *Suppose $f'(x) = 0$ for all x on an interval. Then $f(x)$ is a constant on the interval.*

---

[3]Michel Rolle, born 1652 in Ambert (France), died 1719 in Paris (France). Rolle invented the notion $\sqrt[n]{x}$ for the n-th root of x. His theorem appeared in an obscure book in 1691.

For any two points x_1, x_2 in the interval, we applying the mean value theorem to get

$$f(x_1) - f(x_2) = f'(c)(x_1 - x_2) = 0(x_1 - x_2) = 0 \text{ for some } x_1 < c < x_2.$$

This proves the proposition.

Example 2.2.6. Suppose $f(x)$ satisfies $f'(x) = af(x)$, where a is a constant. Then

$$(e^{-ax}f(x))' = -ae^{-ax}f(x) + e^{-ax}f'(x) = -ae^{-ax}f(x) + ae^{-ax}f(x) = 0.$$

Therefore $e^{-ax}f(x) = c$ is a constant, and $f(x) = ce^{ax}$ is a multiple of the exponential function.

Exercise 2.2.16. Suppose $f(x)$ satisfies $f'(x) = xf(x)$. Prove that $f(x) = ce^{\frac{x^2}{2}}$ for some constant c.

Exercise 2.2.17. Suppose $f(x)$ satisfies $|f(x) - f(y)| \le |x - y|^\alpha$ for some constant $\alpha > 1$. Prove that $f(x)$ is constant.

Exercise 2.2.18. Prove that if $f'(x) = g'(x)$ for all x, then $f(x) = g(x) + c$ for some constant c. Then prove the following equalities.

1. $2\arctan x + \arcsin \dfrac{2x}{1 + x^2} = \pi$ for $x \ge 1$.

2. $3\arccos x - \arccos(3x - 4x^3) = \pi$ for $|x| \le \dfrac{1}{2}$.

2.2.3 Monotone Function

To find out whether a function $f(x)$ is increasing near x_0, we consider the linear approximation $f(x_0) + f'(x_0)\Delta x$. The linear approximation is increasing if and only if the coefficient $f'(x_0) \ge 0$. Because of the approximation, the condition $f'(x_0) \ge 0$ very likely will imply that $f(x)$ is also increasing near x_0. This leads to the following criterion for a function to be increasing.

Proposition 2.2.4. *Suppose $f(x)$ is continuous on an interval and is differentiable on the interior of the interval. Then $f(x)$ is increasing if and only if $f'(x) \ge 0$. Moreover, if $f'(x) > 0$, then $f(x)$ is strictly increasing.*

The similar statement for decreasing functions is also true. Moreover, by combining with Theorem 1.4.8, we find that $f'(x) \ne 0$ near x_0 implies $f(x)$ is invertible near x_0 (compare Proposition 2.1.5).

We note that in the proposition, it is not sufficient for the derivative to be non-negative at a single point. The derivative needs to be non-negative everywhere on an interval.

Proof. Suppose $f(x)$ is increasing. Then either $f(x) = f(y)$ or $f(x) - f(y)$ has the same sign as $x - y$. Therefore $\dfrac{f(y) - f(x)}{y - x} \ge 0$ for any $x \ne y$. This implies $f'(x) \ge 0$ for any x.

Conversely, for $x > y$, the mean value theorem tells us $f(x) - f(y) = f'(c)(x - y)$ for some $x > c > y$. Then the condition $f'(c) \ge 0$ implies $f(x) \ge f(y)$, and the condition $f'(c) > 0$ implies $f(x) > f(y)$.

$\square$
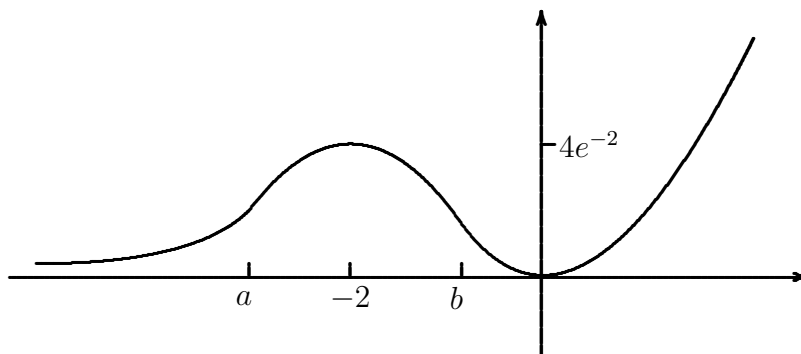
Figure 2.4: graph of $x^2 e^x$

Example 2.2.7. The derivative of $f(x) = x^2 e^x$ is $f'(x) = x(x+2)e^x$. The possible local extrema found in Example 2.2.1 divide the whole real line into three sections $(-\infty, -2]$, $[-2, 0]$ and $[0, \infty)$, on the interiors of which we have respectively $f'(x) > 0$, $f'(x) < 0$ and $f'(x) > 0$. Therefore the function is strictly increasing, strictly decreasing and strictly increasing again on the three sections.

Since the function changes from increasing to decreasing at $x = -2$, $f(-2) = 4e^{-2}$ is a local maximum. Since the function changes from decreasing to increasing at $x = 0$, $f(0) = 0$ is a local minimum. Combined with $\lim_{x \to -\infty} f(x) = 0$, $\lim_{x \to +\infty} f(x) = +\infty$, we get a rough sketch for the graph of the function in Figure 2.4. The numbers a and b in the graph will be explained in Example 2.3.21.

Example 2.2.8. By taking the derivative of the function $f(x) = (x+1)x^{\frac{2}{3}}$, we found $x = -\frac{2}{5}$ and $x = 0$ are the possible local extrema in Example 2.2.2. To determine whether they are indeed local extrema, we note that $f'(x) > 0$, $f'(x) < 0$ and $f'(x) > 0$ in the interiors of the intervals $\left(-\infty, -\frac{2}{5}\right]$, $\left[-\frac{2}{5}, 0\right]$ and $[0, \infty)$. Thus the function changes from increasing to decreasing at $x = -\frac{2}{5}$ and changes from decreasing to increasing at $x = 0$. This implies that $f\left(-\frac{2}{5}\right) = \frac{3\sqrt[3]{20}}{25}$ is a local maximum and $f(0) = 0$ is a local minimum.

Example 2.2.9. We prove $e^x > 1 + x$ for $x \neq 0$. The problem is the same as showing $f(x) = e^x - x > 1 = f(0)$. For $x > 0$, we have $f'(x) = e^x - 1 > 0$ for $x > 0$. Thus $f(x)$ is strictly increasing and $f(x) > f(0)$ for $x > 0$. The argument for the case $x < 0$ is similar.

Exercise 2.2.19. Study the monotone property of functions and find local maxima and minima.

1. $x(x+1)(x+2)$.

2. $\dfrac{x}{1 + x^2}$.

3. $x - \dfrac{1}{x}$.

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

5. $\sin x + \cos x$.

6. $|x|e^x$.

7. $(x+1)\cos x + (x-1)\sin x$.

8. $x - 2\sin x$.

9. $x^2 - \log x$.

10. $e^x \cos x$.

12. x^x.

11. $x - 2 \arctan x$.

Exercise 2.2.20. Prove the inequalities.

1. $\tan x > x$ for $0 < x < \dfrac{\pi}{2}$. (Hint: Consider $\sin x - x \cos x$)

2. $\sin x > \dfrac{2}{\pi} x$ for $0 < x < \dfrac{\pi}{2}$. (Hint: Consider $\dfrac{\sin x}{x}$)

3. $\sin x + \cos x > 1 + x - \dfrac{x^2}{\sqrt{2}}$ for $x > 0$.

4. $\cos x > 1 - \dfrac{x^2}{2}$ for $0 < x < \dfrac{\pi}{2}$.

5. $x - \dfrac{x^2}{2(1+x)} > \log(x+1) > x - \dfrac{x^2}{2}$ for $x > 0$.

Exercise 2.2.21. Prove that $\left(1 + \dfrac{1}{x}\right)^x$ is strictly increasing and $\left(1 + \dfrac{1}{x}\right)^{x+1}$ is strictly decreasing on $(0, +\infty)$. Moreover, prove that $\left(1 + \dfrac{1}{x}\right)^x < e < \left(1 + \dfrac{1}{x}\right)^{x+1}$ and $e - \left(1 + \dfrac{1}{x}\right)^x < \dfrac{e}{x}$ for $x > 0$.

Exercise 2.2.22. The following steps show that the smallest α such that $\left(1 + \dfrac{1}{x}\right)^{x+\alpha} > e$ for all $x > 0$ is $\alpha = \dfrac{1}{2}$.

1. Convert the problem to $\alpha > f\left(\dfrac{1}{x}\right)$, with $f(x) = \dfrac{1}{\log(1+x)} - \dfrac{1}{x}$. Thus the problem is the same as finding the supremum of $f(x)$ on $(0, +\infty)$.

2. Prove that $u - \dfrac{1}{u} > 2 \log u$ for $u > 1$ and $u - \dfrac{1}{u} < 2 \log u$ for $u < 1$. Then use the inequality to show that f is a decreasing function.

3. Show that the supremum of $f(x)$ is $\lim_{x \to 0^+} f(x) = \dfrac{1}{2}$.

4. Find the smallest α and biggest β such that $\left(1 + \dfrac{1}{n}\right)^{n+\alpha} > e > \left(1 + \dfrac{1}{n}\right)^{n+\beta}$ for all natural numbers n.

Exercise 2.2.23. Suppose $f(x)$ is continuous for $x \geq 0$ and differentiable for $x > 0$. Prove that if $f'(x)$ is strictly increasing and $f(0) = 0$, then $\dfrac{f(x)}{x}$ is also strictly increasing.

Exercise 2.2.24. Suppose $f(x)$ is left and right differentiable on an interval. Prove that if $f'_+(x) \geq 0$ and $f'_-(x) \geq 0$, then $f(x)$ is increasing. Moreover, if the inequalities are strict, then $f(x)$ is strictly increasing.

2.2.4   L'Hôspital's Rule

The derivative is a useful tool for the computation of function limits of the types $\dfrac{0}{0}$ or $\dfrac{\infty}{\infty}$.

Proposition 2.2.5 (L'Hôspital[4]'s Rule). *Suppose $f(x)$ and $g(x)$ are differentiable on $(a - \delta, a) \cup (a, a + \delta)$ for some $\delta > 0$. Assume*

1. *Either $\lim_{x \to a} f(x) = \lim_{x \to a} g(x) = 0$ or $\lim_{x \to a} f(x) = \lim_{x \to a} g(x) = \infty$.*

2. *$\lim_{x \to a} \dfrac{f'(x)}{g'(x)}$ exists.*

Then $\lim_{x \to a} \dfrac{f(x)}{g(x)}$ also exists and $\lim_{x \to a} \dfrac{f(x)}{g(x)} = \lim_{x \to a} \dfrac{f'(x)}{g'(x)}$.

The following counterexample

$$\lim_{x \to 0} \frac{1 + x}{2 + x} = \frac{1}{2}, \quad \lim_{x \to 0} \frac{(1 + x)'}{(2 + x)'} = \lim_{x \to 0} \frac{1}{1} = 1.$$

shows the necessity of the first condition. The second condition means that the existence of the limit for the derivative quotient implies the existence of the limit for the original quotient. The converse of this implication is not necessarily true.

The L'Hôspital's rule as stated above is only about the finite limit at a finite a. The subsequent proof shows that the technique also applies to one side limits. By converting x to $\dfrac{1}{x}$, it is easy to show that the technique also holds for x approaching various kinds of infinities. Moreover, the rule can also be applied if the limit of the quotient is some kind of infinity. This can be shown by reverting the quotient $\dfrac{f(x)}{g(x)}$ to $\dfrac{g(x)}{f(x)}$.

To prove L'Hôspital's rule, we need the following extension of the mean value theorem.

Theorem 2.2.6 (Cauchy's Mean Value Theorem). *Suppose $f(x)$ and $g(x)$ are continuous on $[a, b]$ and differentiable on (a, b). If $g'(x)$ is never zero, then there is $a < c < b$, such that*

$$\frac{f'(c)}{g'(c)} = \frac{f(b) - f(a)}{g(b) - g(a)}. \tag{2.2.6}$$

Geometrically, consider $(g(x), f(x))$ as a parametrized curve in $\mathbb{R}^2$. The vector from one point at $x = a$ to another point at $x = b$ is $(g(b) - g(a), f(b) -$

---

[4]Guillaume Francois Antoine Marquis de L'Hôspital, born 1661 in Paris (France), died 1704 in Paris (France). His famous rule was found in his 1696 book "Analyse des infiniment petits pour l'intelligence des lignes courbes", which was the first textbook in differential calculus.
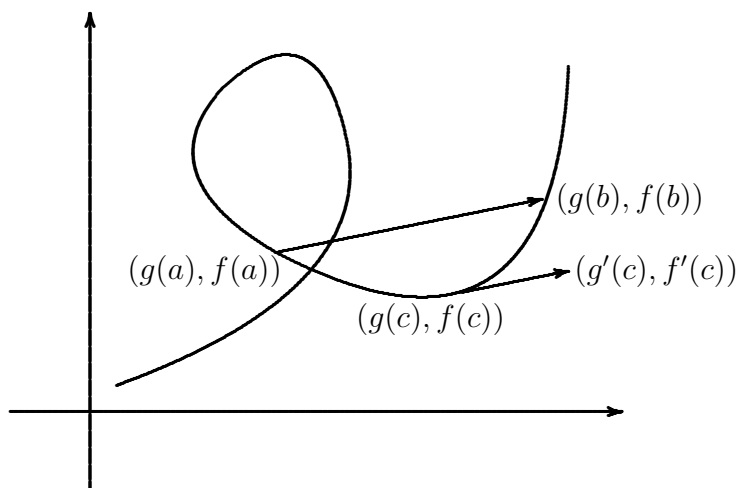
Figure 2.5: Cauchy's mean value theorem

$f(a))$. Cauchy's mean value theorem says that it should be parallel to a tangent vector $(g'(c), f'(c))$ at another point $x = c$ on the curve. This suggests that the theorem may be proved by imitating the proof of the mean value theorem, by considering

$$h(x) = f(x) - f(a) - \frac{f(b) - f(a)}{g(b) - g(a)}(g(x) - g(a)).$$

The details are left to the reader.

Proof of L'Hôspital's Rule. We will prove for the limit of the type $\lim_{x \to a^+}$ only, with a a finite number. Thus $f(x)$ and $g(x)$ are assumed to be differentiable on $(a, a + \delta)$.

First assume $\lim_{x \to a^+} f(x) = \lim_{x \to a^+} g(x) = 0$. Then $f(x)$ and $g(x)$ can be extended to continuous functions on $[a, a+\delta)$ by defining $f(a) = g(a) = 0$. Cauchy's mean value theorem then tells us that for any $a < x < a + \delta$, we have

$$\frac{f(x)}{g(x)} = \frac{f(x) - f(a)}{g(x) - g(a)} = \frac{f'(c)}{g'(c)} \tag{2.2.7}$$

for some c satisfying $a < c < x$ (and c depends on x). As $x \to a^+$, we have $c \to a^+$. Therefore if the limit of the right of (2.2.7) exists, so is the limit on the left, and the two limits are the same.

Now consider the case $\lim_{x \to a^+} f(x) = \lim_{x \to a^+} g(x) = \infty$. The technical difficulty here is that the functions cannot be extended to $x = a$ as before. Still, we try to establish something similar to (2.2.7) by replacing $\dfrac{f(x) - f(a)}{g(x) - g(a)}$ with $\dfrac{f(x) - f(b)}{g(x) - g(b)}$, where $b > a$ is very close to a. The second equality in (2.2.7) still holds. Although the first equality no longer holds, it is sufficient to show that $\dfrac{f(x)}{g(x)}$ and $\dfrac{f(x) - f(b)}{g(x) - g(b)}$ are very close. Of course all these should be put together in logical order.

Denote $\lim_{x \to a^+} \dfrac{f'(x)}{g'(x)} = l$. For any $\epsilon > 0$, there is $\delta_1 > 0$, such that

$$a < x \leq b = a + \delta_1 \implies \left| \frac{f'(x)}{g'(x)} - l \right| < \epsilon.$$

Then by Cauchy's mean value theorem,

$$a < x < b \implies \left| \frac{f(x) - f(b)}{g(x) - g(b)} - l \right| = \left| \frac{f'(c)}{g'(c)} - l \right| < \epsilon, \qquad (2.2.8)$$

where $a < x < c < b$. In particular, the quotient $\dfrac{f(x) - f(b)}{g(x) - g(b)}$ is bounded. Moreover, by the assumption $\lim_{x \to a^+} f(x) = \lim_{x \to a^+} g(x) = \infty$, we have

$$\lim_{x \to a^+} \frac{1 - \dfrac{g(b)}{g(x)}}{1 - \dfrac{f(b)}{f(x)}} = 1.$$

Therefore, there is $\delta_1 \geq \delta > 0$, such that

$$a < x < a + \delta \implies \left| \frac{f(x)}{g(x)} - \frac{f(x) - f(b)}{g(x) - g(b)} \right|$$

$$= \left| \frac{f(x) - f(b)}{g(x) - g(b)} \right| \left| \frac{1 - \dfrac{g(b)}{g(x)}}{1 - \dfrac{f(b)}{f(x)}} - 1 \right| < \epsilon.$$

Since $a < x < a + \delta$ implies $a < x < b$, the conclusion of (2.2.8) also holds. Thus

$$a < x < a + \delta \implies \left| \frac{f(x)}{g(x)} - l \right|$$

$$\leq \left| \frac{f(x)}{g(x)} - \frac{f(x) - f(b)}{g(x) - g(b)} \right| + \left| \frac{f(x) - f(b)}{g(x) - g(b)} - l \right| < 2\epsilon.$$

$\square$

Example 2.2.10. To find $\lim_{x \to 0} \dfrac{\sin x}{x}$, we note that both $\sin x$ and x are continuous and vanish at $x = 0$. Moreover,

$$\lim_{x \to 0} \frac{(\sin x)'}{x'} = \lim_{x \to 0} \frac{\cos x}{1} = \cos 0 = 1.$$

Therefore by L'Hôspital's rule, $\lim_{x \to 0} \dfrac{\sin x}{x} = \lim_{x \to 0} \dfrac{(\sin x)'}{x'} = 1$.

Unfortunately, the argument above is circular, because the limit $\lim_{x \to 0} \dfrac{\sin x}{x} = 1$ was used to compute the derivative of $\sin x$. Similarly, the computations of $\lim_{x \to 0} \dfrac{e^x - 1}{x}$ and $\lim_{x \to 0} \dfrac{\log(x + 1)}{x}$ by L'Hôspital's rule are also circular.

Example 2.2.11. The limit $\lim_{x \to 1} \left(\dfrac{1}{x-1} - \dfrac{1}{\log x} \right)$ may be found by making use of L'Hôspital's rule twice.

$$\lim_{x \to 1} \left(\frac{1}{x-1} - \frac{1}{\log x} \right) = \lim_{x \to 1} \frac{\log x - x + 1}{(x-1)\log x}$$

$$=^{(1)} \lim_{x \to 1} \frac{\dfrac{1}{x} - 1}{\log x - \dfrac{x-1}{x}} = \lim_{x \to 1} \frac{1-x}{x \log x + x - 1}$$

$$=^{(2)} \lim_{x \to 1} \frac{-1}{\log x + 2} = -\frac{1}{2}.$$

Note that one should not be blindly apply L'Hôspital's rule without checking the conditions. For the computation above, since $1 - x = x \log x + x - 1 = 0$ at $x = 0$ and $\lim_{x \to 1} \dfrac{-1}{\log x + 2}$ exists, the equality $=^{(2)}$ holds. Then since $\log x - x + 1 = (x-1)\log x = 0$ at $x = 0$ and $\lim_{x \to 1} \dfrac{\dfrac{1}{x} - 1}{\log x - \dfrac{x-1}{x}}$ exists, the equality $=^{(1)}$ holds.

Example 2.2.12. The limit $\lim_{x \to +\infty} (x + \sqrt{1+x^2})^{\frac{1}{\log x}}$ may be found by first applying L'Hôspital's rule to the log of the function.

$$\lim_{x \to +\infty} \frac{\log(x + \sqrt{1+x^2})}{\log x} = \lim_{x \to +\infty} \frac{\dfrac{1 + \dfrac{x}{\sqrt{1+x^2}}}{x + \sqrt{1+x^2}}}{\dfrac{1}{x}} = \lim_{x \to +\infty} \frac{x}{\sqrt{1+x^2}} = 1.$$

Note that L'Hôspital's rule may be applied because $\lim_{x \to +\infty} \log(x + \sqrt{1+x^2}) = \lim_{x \to +\infty} \log x = \infty$ and the limit in the middle exists. By taking the exponential, we get

$$\lim_{x \to +\infty} (x + \sqrt{1+x^2})^{\frac{1}{\log x}} = e^1 = e.$$

Exercise 2.2.25. Compute the limits.

1. $\lim_{x \to 0} \dfrac{x - \sin x}{x^3}$.

2. $\lim_{x \to 0} \dfrac{e^x - 1 - x}{x^2}$.

3. $\lim_{x \to 0} \left(\dfrac{\cos x - 1}{x^4} + \dfrac{1}{2x^2} \right)$.

4. $\lim_{x \to 0} \sin x \log x$.

5. $\lim_{x \to 0} \left(\dfrac{1}{x} - \cot x \right)$.

6. $\lim_{x \to 0} (\cos x)^{\frac{1}{x^2}}$.

7. $\lim_{x \to 0} \dfrac{\sqrt[x]{1+x} - e}{x}$.

8. $\lim_{x \to +\infty} (\pi - 2 \arctan x) \log x$.

9. $\lim_{x \to \infty} \left(x \sin \dfrac{1}{x} \right)^{x^2}$.

Exercise 2.2.26. Discuss whether L'Hôspital's rule can be applied to the limits.

1. $\lim_{x \to \infty} \dfrac{x + \sin x}{x - \cos x}$.

2. $\lim_{x \to +\infty} \dfrac{x}{x + \sin x}$.

3. $\lim_{x \to 0} \dfrac{x^2 \sin \dfrac{1}{x}}{\sin x}$.

Exercise 2.2.27. The mean value theorem tells us

$$\log(1+x) - \log 1 = x\frac{1}{1+\theta x}, \quad e^x - 1 = xe^{\theta x},$$

for some $0 < \theta < 1$. Prove that in both cases, $\lim_{x \to 0} \theta = \frac{1}{2}$.

Exercise 2.2.28. Prove L'Hospital's rule for the case $a = +\infty$. Moreover, discuss L'Hospital's rule for the case $l = \infty$.

2.2.5   Additional Exercise

Ratio Rule

By specifying the ratio rule in Exercise 1.1.32 to $y_n = a^n$, we get the limits in Exercises 1.1.33 and 1.1.34. By making other choices of y_0 and using the linear approximations to estimate the quotient $\frac{y_{n+1}}{y_n}$, we get other concrete forms of the ratio rule.

Exercise 2.2.29. Prove that if $a > b > c > 0$, then $1 - ax < (1-x)^b < 1 - cx$ and $1 + ax > (1+x)^b > 1 + cx$ for small and positive x. Then prove the following

1. If $\left|\frac{x_{n+1}}{x_n}\right| \le 1 - \frac{a}{n}$ for some $a > 0$ and big n, then $\lim_{n \to \infty} x_n = 0$.

2. If $\left|\frac{x_{n+1}}{x_n}\right| \ge 1 + \frac{a}{n}$ for some $a > 0$ and big n, then $\lim_{n \to \infty} x_n = \infty$.

Exercise 2.2.30. Study the limits.

1. $\lim_{n \to \infty} \frac{(n!)^2 a^n}{(2n)!}$.   2. $\lim_{n \to \infty} \frac{(n+a)^{n+b}}{c^n n!}$.

Exercise 2.2.31. Rephrase the rules in Exercise 2.2.29 in terms of the quotient $\left|\frac{x_n}{x_{n+1}}\right|$. Then prove that if $\underline{\lim}_{n \to \infty} n\left(\left|\frac{x_n}{x_{n+1}}\right| - 1\right) > 0$, then $\lim_{n \to \infty} x_n = 0$. Find the similar condition for the conclusion $\lim_{n \to \infty} x_n = \infty$.

Exercise 2.2.32. Prove that if $a > b > c > 0$, then $1 - \frac{a}{x \log x} < \frac{(\log(x-1))^b}{(\log x)^b} < 1 - \frac{c}{x \log x}$ and $1 + \frac{a}{x \log x} > \frac{(\log(x+1))^b}{(\log x)^b} > 1 + \frac{c}{x \log x}$ for big and positive x. Then prove the following.

1. If $\left|\frac{x_{n+1}}{x_n}\right| \le 1 - \frac{a}{n \log n}$ for some $a > 0$ and big n, then $\lim_{n \to \infty} x_n = 0$.

2. If $\left|\frac{x_{n+1}}{x_n}\right| \ge 1 + \frac{a}{n \log n}$ for some $a > 0$ and big n, then $\lim_{n \to \infty} x_n = \infty$.

Darboux[5]'s Intermediate Value Theorem

---

[5]Jean Gaston Darboux, born 1842 in Nimes (France), died 1917 in Paris (France). Darboux made important contributions to differential geometry and analysis.

Exercise 2.2.33. Suppose $f(x)$ is differentiable on $[a, b]$. By considering the extrema of $f(x) - \gamma x$, prove that for any γ between $f'(a)$ and $f'(b)$, there is $c \in [a, b]$, such that $f'(c) = \gamma$.

Exercise 2.2.34. Find a function $f(x)$ differentiable everywhere on $[0, 1]$, such that $f'(x)$ is not continuous. The examples shows that Darboux's intermediate value theorem is not a consequence of the usual intermediate value theorem.

Extension of Mean Value Theorem[6]

In Exercise 2.2.15, the mean value theorem is extended to continuous functions that are both left and right differentiable. In what follows, we consider continuous functions that are left *or* right differentiable.

Exercise 2.2.35. Prove that for any function $f(x)$ on $[a, b]$ and $l < \dfrac{f(b) - f(a)}{b - a}$, there is a linear function $L(x)$ satisfying $L'(x) = l$ and $L(a) > f(a)$, $L(b) < f(b)$.

Exercise 2.2.36. Suppose $f(x)$ is a continuous function $f(x)$ on $[a, b]$ and L is a linear function with the properties in Exercise 2.2.35. Prove that

$$c = \sup\{x \in (a, b) \colon f \leq L \text{ on } [a, x]\}$$

satisfies $a < c < b$ and $f(c) = L(c)$. Moreover, if $f(x)$ has any one side derivative at c, then the one side derivative is no less than $L'(c) = l$.

Exercise 2.2.37. Suppose $f(x)$ is a continuous function on $[a, b]$ and is left *or* right differentiable at any point on (a, b). Let $f'_*(x)$ be one of the one side derivatives at x. Prove that

$$\inf_{(a,b)} f'_* \leq \frac{f(b) - f(a)}{b - a} \leq \sup_{(a,b)} f'_*.$$

Exercise 2.2.38. Use Exercise 2.2.37 to further extend the criterion for the monotonicity in Exercise 2.2.24.

Basic Inequalities

The monotone property can be used to prove some important basic inequalities. Let p and q be real numbers satisfying

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Exercise 2.2.39. For $x > 0$, prove that $x^{\frac{1}{p}} \leq \dfrac{1}{p} x + \dfrac{1}{q}$ in case $p > 1$ and $x^{\frac{1}{p}} \geq \dfrac{1}{p} x + \dfrac{1}{q}$ in case $p < 1$.

Exercise 2.2.40 (Young[7] Inequality). For $a, b > 0$, prove

$$ab \leq \frac{1}{p} a^p + \frac{1}{q} b^q \text{ for } p > 1$$

and

$$ab \geq \frac{1}{p} a^p + \frac{1}{q} b^q \text{ for } p < 1.$$

When does the equality hold?

---

[6]See "Some Remarks on Functions with One-Sided Derivatives" by Miller and Výborný, American Math Monthly **93** (1986) 471-475.

[7]William Henry Young, born 1863 in London (England), died 1942 in Lausanne (Switzerland). Young discovered a form of Lebesgue integration independently. He wrote an influential book "The fundamental theorems of the differential calculus" in 1910.

Exercise 2.2.41 (Hölder[8] Inequality). Suppose $p, q > 0$. For positive numbers a_1, $a_2, \ldots, a_n, b_1, b_2, \ldots, b_n$, by taking $a = \dfrac{a_i}{\left(\sum a_i^p\right)^{\frac{1}{p}}}$ and $b = \dfrac{b_i}{\left(\sum b_i^q\right)^{\frac{1}{q}}}$ in the Young inequality, prove

$$\sum a_i b_i \le \left(\sum a_i^p\right)^{\frac{1}{p}} \left(\sum b_i^q\right)^{\frac{1}{q}}. \qquad (2.2.9)$$

When does the equality hold?

Exercise 2.2.42 (Minkowski[9] Inequality). Suppose $p > 1$. By applying the Hölder inequality to $a_1, a_2, \ldots, a_n, (a_1 + b_1)^{p-1}, (a_2 + b_2)^{p-1}, \ldots, (a_n + b_n)^{p-1}$ and then to $b_1, b_2, \ldots, b_n, (a_1 + b_1)^{p-1}, (a_2 + b_2)^{p-1}, \ldots, (a_n + b_n)^{p-1}$, prove

$$\left(\sum (a_i + b_i)^p\right)^{\frac{1}{p}} \le \left(\sum a_i^p\right)^{\frac{1}{p}} + \left(\sum b_i^p\right)^{\frac{1}{p}}. \qquad (2.2.10)$$

When does the equality hold?

2.3 High Order Approximation

The linear approximation has many nice properties and can be used to solve many problems. Still, there are problems that require approximations more refined than the linear one. In this case, polynomials of higher and higher degrees can be used. This leads to high order derivatives and Taylor series.

2.3.1 Quadratic Approximation

A function is approximated by a *quadratic function* $a + b\Delta x + c\Delta x^2$ at x_0, if for any $\epsilon > 0$, there is $\delta > 0$, such that

$$|\Delta x| < \delta \implies |f(x) - a - b\Delta x - c\Delta x^2| \le \epsilon |\Delta x|^2. \qquad (2.3.1)$$

Similar to the linear approximation, the condition (2.3.1) for quadratic approximation means exactly $a = f(x_0)$, the derivative $b = f'(x_0)$ exists, and the limit

$$\begin{aligned}
c &= \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0) - f'(x_0)\Delta x}{\Delta x^2} \\
&= \lim_{x \to x_0} \frac{f(x) - f(x_0) - f'(x_0)(x - x_0)}{(x - x_0)^2}
\end{aligned} \qquad (2.3.2)$$

exists. It is rather tempting to define c to be the second order derivative. But the following result suggests that it is better to call $2c$ the second order derivative.

---

[8]Otto Ludwig Hölder, born 1859 in Stuttgart (Germany), died 1937 in Leipzig (Germany). He discovered the inequality in 1884. Hölder also made fundamental contributions to the group theory.

[9]Hermann Minkowski, born 1864 in Alexotas (Russia, now Kaunas of Lithuania), died 1909 in Göttingen (Germany). Minkowski's fundamental contribution to geometry provided the mathematical foundation of Einstein's theory of relativity.

Proposition 2.3.1. *Suppose $f(x)$ is differentiable near x_0, and the derivative function $f'(x)$ has further derivative $f''(x_0)$ at x_0. Then $f(x)$ has a quadratic approximation at x_0, given by $f(x_0) + f'(x_0)(x - x_0) + \dfrac{f''(x_0)}{2}(x - x_0)^2$.*

Proof. Consider the difference (called *remainder*)

$$R_2(x) = f(x) - f(x_0) - f'(x_0)(x - x_0) - \frac{f''(x_0)}{2}(x - x_0)^2$$

between the function and the expected quadratic approximation. The difference satisfies

$$R_2(x_0) = R_2'(x_0) = R_2''(x_0) = 0.$$

By Cauchy's mean value theorem, we have

$$\frac{R_2(x)}{(x - x_0)^2} = \frac{R_2(x) - R_2(x_0)}{(x - x_0)^2 - (x_0 - x_0)^2} = \frac{R_2'(c)}{2(c - x_0)} = \frac{R_2'(c) - R_2'(x_0)}{2(c - x_0)}$$

for some c between x_0 and x. As $x \to x_0$, we have $c \to x_0$, so that

$$\lim_{x \to x_0} \frac{R_2(x)}{(x - x_0)^2} = \lim_{c \to x_0} \frac{R_2'(c) - R_2'(x_0)}{2(c - x_0)} = \frac{R_2''(x_0)}{2} = 0.$$

$\square$

Because of the proposition, we define the *second order differential* to be

$$d^2 f = 2c\,dx^2,$$

where dx^2 is indeed considered as the square of dx, at least symbolically. We have $d^2 f = f''(x_0)dx^2$ if the second order derivative exists at x_0.

Example 2.3.1. Suppose $f(x)$ has second order derivative at x_0. Then we have

$$f(x_0 + \Delta x) = f(x_0) + f'(x_0)\Delta x + \frac{f''(x_0)}{2}\Delta x^2 + o(\Delta x^2)$$

$$f(x_0 + 2\Delta x) = f(x_0) + 2f'(x_0)\Delta x + 2f''(x_0)\Delta x^2 + o(\Delta x^2).$$

This implies

$$f(x_0 + 2\Delta x) - 2f(x_0 + \Delta x) + f(x_0) = f''(x_0)\Delta x^2 + o(\Delta x^2),$$

and we get another way of expressing the second order derivative as a limit.

$$f''(x_0) = \lim_{\Delta x \to 0} \frac{f(x_0 + 2\Delta x) - 2f(x_0 + \Delta x) + f(x_0)}{\Delta x^2}.$$

Exercise 2.3.1. Find suitable conditions among constants a, b, λ, μ so that $\lambda f(x_0 + a\Delta x) + \mu f(x_0 + b\Delta x) + f(x_0) = f''(x_0)\Delta x^2 + o(\Delta x^2)$ holds for twice differentiable functions. Then derive some formula for the second order derivative similar to the one in Example 2.3.1.

Exercise 2.3.2. Suppose $f''(0)$ exists and $f''(0) \neq 0$. Prove that in the mean value theorem $f(x) - f(0) = xf'(\theta x)$, we have $\lim_{x \to 0} \theta = \dfrac{1}{2}$. This generalizes the observation in Exercise 2.2.27.

Exercise 2.3.3. Suppose $f(x)$ has second order derivative at x_0. Let h and k be small, distinct and nonzero numbers. Find the quadratic function $q(x) = a + b\Delta x + c\Delta x^2$ satisfying

$$q(x_0) = f(x_0), \quad q(x_0 + h) = f(x_0 + h), \quad q(x_0 + k) = f(x_0 + k).$$

Then prove that $\lim_{h,k \to 0} b = f'(x_0)$ and $\lim_{h,k \to 0} c = \dfrac{f''(x_0)}{2}$ as long as $\dfrac{h}{h-k}$ is kept bounded. This provides the geometrical interpretation of the quadratic approximation.

Exercise 2.3.4. Proposition 2.3.1 basically says that the existence of the second order derivative implies the second order differentiability. Show that the converse is not true (in contrast what Proposition 2.1.3 says about the first order case) by considering the following functions at $x_0 = 0$.

1. $f(x) = \begin{cases} x^3 & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational} \end{cases}$.   2. $f(x) = \begin{cases} x^3 \sin \dfrac{1}{x^2} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$.

Exercise 2.3.5. Determine the existence of the quadratic approximation and the existence of the second order derivative of functions in Exercise 2.1.5.

Exercise 2.3.6. Study the existence of the quadratic approximation and the existence of the second order derivative of function $|x^3(x-1)(x-2)^2|$.

Exercise 2.3.7. Suppose $P(x)$ and $Q(x)$ are quadratic approximations of $f(x)$ and $g(x)$ at x_0.

1. Prove that $P(x) + Q(x)$ is the quadratic approximation of $f(x) + g(x)$ at x_0.

2. Prove that although $P(x)Q(x)$ has degree 4, the second order truncation of the product is the quadratic approximation of $f(x)g(x)$.

3. Suppose $f(x)$ and $g(x)$ have second order derivatives. What do the two conclusions tell you about the second order derivatives of $f(x) + g(x)$ and $f(x)g(x)$?

Exercise 2.3.8. Suppose $P(x)$ is the quadratic approximations of $f(x)$ at x_0. Suppose $Q(y)$ is the quadratic approximations of $g(y)$ at $y_0 = f(x_0)$.

1. Prove that the second order truncation of $Q(P(x))$ is the quadratic approximation of $g(f(x))$.

2. Suppose $f(x)$ and $g(y)$ have second order derivatives at x_0 and y_0. What can you say about the second order derivative of $g(f(x))$?

2.3.2 High Order Derivative

The quadratic approximation is computed by taking the derivative twice. Approximations by higher degree polynomials are expected to be computed by taking the derivative many times.

Given a differentiable function $f(x)$, its derivative $f'(x)$ is also a function. If the function $f'(x)$ is also differentiable, then we have the second order derivative $f''(x)$. If the function $f''(x)$ is again differentiable, then we get the

third order derivative $f'''(x)$. In general, the *n-th order derivative* of $f(x)$ is denoted $f^{(n)}(x)$ or $\dfrac{d^n f}{dx^n}$.

It is easy to show (by induction, for example), that the high order derivatives of the power, the exponential, and the logarithmic functions are

$$(x^\alpha)^{(n)} = \alpha(\alpha - 1)\cdots(\alpha - n + 1)x^{\alpha-n},$$
$$(e^x)^{(n)} = e^x,$$
$$(\alpha^x)^{(n)} = \alpha^x(\log \alpha)^n,$$
$$(\log |x|)^{(n)} = (-1)^{n-1}(n-1)!x^{-n}.$$

Note that if α is a natural number, then $(x^\alpha)^{(n)} = 0$ for $n > \alpha$. The high order derivatives of the sine and the cosine functions have periodic pattern.

$$\sin' x = \cos x, \quad \sin'' x = -\sin x, \quad \sin''' x = -\cos x, \quad \sin'''' x = \sin x, \quad \ldots$$
$$\cos' x = -\sin x, \quad \cos'' x = -\cos x, \quad \cos''' x = \sin x, \quad \cos'''' x = \cos x, \quad \ldots$$

However, there is no clear pattern for the derivatives of the tangent function.

$$\tan' x = \sec^2 x,$$
$$\tan'' x = 2\sec^2 x \tan x,$$
$$\tan''' x = 4\sec^2 x \tan^2 x + 2\sec^4 x = 6\sec^4 x - 4\sec^2 x,$$
$$\tan'''' x = (24\sec^4 x - 8\sec^2 x)\tan x.$$

The formulae for the derivatives of the sum and the scalar multiplication can be directly extended to the high order derivatives.

$$(f + g)^{(n)} = f^{(n)} + g^{(n)}, \quad (cf)^{(n)} = cf^{(n)}.$$

The Leibniz rule can also be extended.

$$(fg)' = f'g + fg',$$
$$(fg)'' = f''g + 2f'g' + fg'',$$
$$(fg)''' = f'''g + 3f''g' + 3f'g'' + fg''',$$
$$(fg)'''' = f''''g + 4f'''g' + 6f''g'' + 4f'g''' + fg''''.$$

By induction, it is not hard to show that the coefficients in the extended Leibniz rule are the same as the ones in the binomial expansion.

$$(fg)^{(n)} = f^{(n)} + nf^{(n-1)}g' + \frac{n(n-1)}{2}f^{(n-2)}g'' + \cdots + g^{(n)}.$$

There is no clean extension of the chain rule. The following is the chain rule

for the second order derivative of $z = g(y) = g(f(x))$.

$$\frac{d^2z}{dx^2} = \frac{d}{dx}\left(\frac{dz}{dx}\right) \qquad\qquad\qquad\qquad \text{(definition of } f'')$$

$$= \frac{d}{dx}\left(\frac{dz}{dy}\frac{dy}{dx}\right) \qquad\qquad\qquad\qquad \text{(chain rule)}$$

$$= \frac{d}{dx}\left(\frac{dz}{dy}\right)\frac{dy}{dx} + \frac{dz}{dy}\frac{d}{dx}\left(\frac{dy}{dx}\right) \qquad \text{(Leibniz rule)}$$

$$= \frac{d}{dy}\left(\frac{dz}{dy}\right)\frac{dy}{dx}\frac{dy}{dx} + \frac{dz}{dy}\frac{d^2y}{dx^2} \qquad \text{(chain rule and definition of } f'')$$

$$= \frac{d^2z}{dy^2}\left(\frac{dy}{dx}\right)^2 + \frac{dz}{dy}\frac{d^2y}{dx^2}. \qquad\qquad \text{(definition of } f'')$$

If the n-th order derivative exists, then the *n-th order differential* is

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

where dx^n is symbolically the n-th power of dx. The high order differential can be generally computed by $d^n f = d(d^{n-1}f)$, $d(dx^n) = 0$ for the variable x, and the Leibniz rule $d(\alpha\beta) = (d\alpha)\beta + \alpha d\beta$. For example,

$$d^2 f = d(df) = d(f'dx) = (df')dx + f'd(dx) = (f''dx)dx = f''dx^2,$$

and

$$d^3 f = d(d^2 f) = d(f''dx^2) = (df'')dx^2 + f''d(dx^2) = (f'''dx)dx^2 = f'''dx^3.$$

Example 2.3.2. Let

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

Then

$$f'(x) = \begin{cases} 2x & \text{if } x \geq 0 \\ -2x & \text{if } x < 0 \end{cases} = 2|x|.$$

By Example 2.1.8, the function $f'(x)$ is not differentiable at 0. Therefore $f(x)$ has derivative at 0 only up to the first order, and has all the high order derivatives at any $x \neq 0$.

Example 2.3.3. We have $\left(\sqrt{1+\sin^2 x}\right)' = \dfrac{\sin x \cos x}{\sqrt{1+\sin^2 x}}$ from Example 2.1.11. Then

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

$$= \frac{(\cos^2 x - \sin^2 x)\sqrt{1+\sin^2 x} - \sin x \cos x \dfrac{\sin x \cos x}{\sqrt{1+\sin^2 x}}}{1+\sin^2 x}$$

$$= \frac{(1 - 2\sin^2 x)(1+\sin^2 x) - \sin^2 x(1 - \sin^2 x)}{\left(1+\sin^2 x\right)^{\frac{3}{2}}}$$

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

Example 2.3.4. By $(x^3)^{(n)} = 0$ for $n > 3$ and the Leibniz rule,

$$(x^3 e^x)^{(n)} = x^3 (e^x)^{(n)} + n(x^3)'(e^x)^{(n-1)} + \frac{n(n-1)}{2}(x^3)''(e^x)^{(n-2)}$$

$$+ \frac{n(n-1)(n-2)}{6}(x^3)'''(e^x)^{(n-3)}$$

$$= (x^3 + 3nx^2 + 3n(n-1)x + n(n-1)(n-2))e^x.$$

Example 2.3.5. From the formula for the high order derivatives of x^α, it is easy to deduce

$$((ax+b)^\alpha)^{(n)} = \alpha(\alpha-1)\cdots(\alpha-n+1)a^n(ax+b)^{\alpha-n}.$$

Then by $\dfrac{1+x}{\sqrt{1-x}} = 2(1-x)^{-\frac{1}{2}} - (1-x)^{\frac{1}{2}}$, we get

$$\left(\frac{1+x}{\sqrt{1-x}}\right)^{(n)} = 2\frac{1}{2}\frac{3}{2}\cdots\frac{2n-1}{2}(1-x)^{-\frac{2n+1}{2}} - \left(-\frac{1}{2}\right)\frac{1}{2}\cdots\frac{2n-3}{2}(1-x)^{-\frac{2n-1}{2}}$$

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

Example 2.3.6. In Example 2.1.12, the derivative $\dfrac{dy}{dx}$ for the ellipse $x^2 + \dfrac{y^2}{4} = 1$ was computed in three ways. Continuing the three ways, the second order derivative can also be computed.

In the first way, the upper part $y = 2\sqrt{1-x^2}$ has the first order derivative

$$\frac{dy}{dx} = -\frac{2x}{\sqrt{1-x^2}},$$

and the second order derivative

$$\frac{d^2y}{dx^2} = -\frac{(2x)'\sqrt{1-x^2} - 2x(\sqrt{1-x^2})'}{(\sqrt{1-x^2})^2} = -\frac{2\sqrt{1-x^2} + 2x\dfrac{x}{\sqrt{1-x^2}}}{1-x^2}$$

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

In the second way, y is implicitly considered as a function of x. Taking $\dfrac{d}{dx}$ on both sides of the equation $x^2 + \dfrac{y^2}{4} = 1$ and using the chain rule, we get

$$2x + \frac{1}{2}y\frac{dy}{dx} = 0.$$

Taking $\dfrac{d}{dx}$ again, we get

$$2 + \frac{1}{2}\left(\frac{dy}{dx}\right)^2 + \frac{1}{2}y\frac{d^2y}{dx^2} = 0.$$

In Example 2.1.12, the first equation was solved to give $\dfrac{dy}{dx} = -\dfrac{4x}{y}$. Substituting this into the second equation and solve for the second order derivative, we get

$$\frac{d^2y}{dx^2} = -\frac{16\left(x^2 + \dfrac{y^2}{4}\right)}{y^3} = -\frac{16}{y^3}.$$

In the third way, the ellipse is parametrized by $x = \cos t$, $y = 2 \sin t$. The first order derivative $\dfrac{dy}{dx} = -2 \cot t$ was computed by the chain rule in Example 2.1.12. Continuing with the same idea, we get

$$\frac{d^2 y}{dx^2} = \frac{d(-2 \cot t)}{dx} = \frac{\dfrac{d(-2 \cot t)}{dt}}{\dfrac{dx}{dt}} = \frac{2 \csc^2 t}{-\sin t} = -\frac{2}{\sin^3 t}.$$

Example 2.3.7. Let $p(x)$ be a polynomial and

$$f(x) = \begin{cases} p\left(\dfrac{1}{x}\right) e^{-\frac{1}{x^2}} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}.$$

At $x \neq 0$, we have

$$f'(x) = \left[-p'\left(\frac{1}{x}\right) \frac{1}{x^2} + p\left(\frac{1}{x}\right) \frac{2}{x^3} \right] e^{-\frac{1}{x^2}} = q\left(\frac{1}{x}\right) e^{-\frac{1}{x^2}},$$

where $q(x) = -p'(x)x^2 - 2p(x)x^3$ is also a polynomial. Moreover, by

$$\lim_{x \to 0} x^k e^{-\frac{1}{x^2}} = \lim_{x \to \infty} x^{-k} e^{-x^2} = \lim_{x \to +\infty} \frac{x^{-\frac{k}{2}}}{e^x} = 0$$

for any k, we have

$$f'(0) = \lim_{x \to 0} \frac{f(x)}{x} = \lim_{x \to 0} \frac{1}{x} p\left(\frac{1}{x}\right) e^{-\frac{1}{x^2}} = 0.$$

Therefore $f'(x)$ is of the same type as $f(x)$, with another polynomial $q(x)$ in place of $p(x)$. In particular, we conclude that $f(x)$ has derivatives of all orders, and $f^{(n)}(0) = 0$ for any n.

Exercise 2.3.9. Compute the derivatives up to the third order.

1. $\dfrac{1}{x^2 + 1}$.

2. $\dfrac{2x^2 + 1}{x^2 + 1}$.

3. $\cot x$.

4. $\sec x$.

5. $\csc x$.

6. $\arcsin x$.

7. $\arctan x$.

8. $\operatorname{arcsec} x$.

9. $(\sin 2x + \cos 3x)^7$.

10. $2^{\sin 3x}$.

11. $e^{2x}(\cos x - 2 \sin x)$.

12. $\arcsin \dfrac{1}{x}$.

13. $x(\log x - 1)$.

14. $\log(\log x)$.

15. $\log(x + \sqrt{1 + x^2})$.

16. $\log |\sin x|$.

Exercise 2.3.10. Compute all the high order derivatives.

1. $\log(2 - 3x)$.

2. $\dfrac{x^2 + x + 1}{4x^2 - 1}$.

3. $\dfrac{x}{\sqrt{2x + 1}}$.

4. $x^4 \log x$.

Exercise 2.3.11. Prove $\dfrac{d^n (f(ax + b))}{dx^n} = a^n f^{(n)}(ax + b)$.

Exercise 2.3.12. Suppose $u = u(x)$ and $v = v(x)$ are differentiable functions satisfying $u(0) = 1$, $v(0) = -1$ and the given equations. Compute the second order derivatives of u and v at 0.

1. $u^2 + uv + v^2 = 1$, $(1 + x^2)u + (1 - x^2)v = x$.

2. $xu + (x + 1)v = e^x$, $e^u + xe^{-v} = e^{x+1}$.

Exercise 2.3.13. Suppose y is a function of x satisfying the given equation. Compute the second order derivative $\dfrac{d^2 y}{dx^2}$ at the given point.

1. $y^3 + 2xy^2 - 2x^3 = 1$, at $x = 1$, $y = 1$.

2. $x \sin(x - y) = y \cos(x + y)$, at $x = \pi$, $y = 0$.

3. $e^y = xy$, at $x = -e^{-1}$, $y = -1$.

4. $(1 + y)e^x + (1 + x)e^y = xy + 2$, at $x = 0$, $y = 0$.

Exercise 2.3.14. For the parametrized curves, compute the second order derivative $\dfrac{d^2 y}{dx^2}$ at the given point.

1. Cycloid: $x = t - \sin t$, $y = 1 - \cos t$, at $t = \dfrac{\pi}{4}$.

2. Spiral: $x = t \cos t$, $y = t \sin t$, at $t = \pi$.

3. Involute of circle: $x = \cos t + t \sin t$, $y = \sin t - t \cos t$, at $t = \pi$.

4. Four-leaved rose: $x = \cos 2t \cos t$, $y = \cos 2t \sin t$, at $t = \dfrac{\pi}{4}$.

Exercise 2.3.15. Find the formulae for the derivatives of quotient, composition and inverse functions up to the third order.

Exercise 2.3.16. Prove the functions $y = \arctan x$ and $y = (\arcsin x)^2$ satisfy the equations $(1 + x^2)y'' + 2xy' = 0$ and $(1 - x^2)y'' - xy' = 2$. Then use the equations to compute all the high order derivatives of $\arctan x$ and $(\arcsin x)^2$ at 0.

Exercise 2.3.17. Suppose $x = x(t)$ and $y = y(t)$ is a parametrized curve. Prove that

$$\frac{d^2 y}{dx^2} = \frac{y'' x' - y' x''}{x'^3},$$

where x', x'', y', y'' are the derivatives with respect to t. Also find the formula for the third order derivative.

2.3.3 Taylor Expansion

Using high order derivatives, the discussion on the quadratic approximations can be extended to high order approximations.

Theorem 2.3.2. *Suppose $f(x)$ has the n-th order derivative $f^{(n)}(x_0)$ at x_0. Then for the n-th degree polynomial*

$$T_n(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n,$$
$$(2.3.3)$$

we have

$$\lim_{x \to x_0} \frac{f(x) - T_n(x)}{(x - x_0)^n} = 0.$$

Note that the existence of $f^{(n)}(x_0)$ implicitly assumes the existence of $f^{(k)}(x)$ for all $k < n$ and all x near x_0. The function $T_n(x)$ is a polynomial of degree n characterized by the property

$$f(x) = T_n(x) + o(\Delta x^n), \quad \Delta x = x - x_0. \tag{2.3.4}$$

Because of the property, we say $f(x)$ is *n-th order differentiable*. The polynomial $T_n(x)$ is called the *n-th order Taylor expansion*.

Proof of Theorem 2.3.2. The theorem can be proved similar to Proposition 2.3.1. The remainder $R_n(x) = f(x) - T_n(x)$ satisfies

$$R_n(x_0) = R_n'(x_0) = R_n''(x_0) = \cdots = R_n^{(n)}(x_0) = 0.$$

Therefore by Cauchy's mean value theorem,

$$\begin{aligned}
\frac{R_n(x)}{(x - x_0)^n} &= \frac{R_n(x) - R_n(x_0)}{(x - x_0)^n - (x_0 - x_0)^n} \\
&= \frac{R_n'(c_1)}{n(c_1 - x_0)^{n-1}} = \frac{R_n'(c_1) - R_n'(x_0)}{n((c_1 - x_0)^{n-1} - (x_0 - x_0)^{n-1})} = \cdots \\
&= \frac{R_n^{(n-1)}(c_{n-1})}{n(n-1)\cdots 2(c_{n-1} - x_0)}
\end{aligned} \tag{2.3.5}$$

for some c_1 between x_0 and x, c_2 between x_0 and c_1, $\ldots$, and c_{n-1} between x_0 and c_{n-2}. Then we have

$$\lim_{x \to x_0} \frac{R_n(x)}{(x - x_0)^n} = \lim_{c_{n-1} \to x_0} \frac{R_n^{(n-1)}(c_{n-1}) - R_n^{(n-1)}(x_0)}{n!(c_1 - x_0)} = \frac{R_n^{(n)}(x_0)}{n!} = 0.$$

$\square$

The computation of the high order derivatives in Section 2.3.2 immediately gives the following Taylor expansions at 0.

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

$$+ \cdots + \frac{\alpha(\alpha - 1)\cdots(\alpha - n + 1)}{n!}x^n + o(x^n),$$

$$\frac{1}{1 - x} = 1 + x + x^2 + x^3 + \cdots + x^n + o(x^n),$$

$$e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots + \frac{1}{n!}x^n + o(x^n),$$

$$\log(1 + x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 + \cdots + \frac{(-1)^{n-1}}{n}x^n + o(x^n),$$

$$\sin x = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 + \cdots + \frac{(-1)^{n+1}}{(2n-1)!}x^{2n-1} + o(x^{2n}),$$

$$\cos x = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 + \cdots + \frac{(-1)^n}{(2n)!}x^{2n} + o(x^{2n+1}).$$

Example 2.3.8. By rewriting the function x^4 as a polynomial in $(x-1)$

$$x^4 = (1 + (x-1))^4 = 1 + 4(x-1) + 6(x-1)^2 + 4(x-1)^3 + (x-1)^4,$$

and using the characterization (2.3.4), we get the Taylor expansion of various orders of x^4 at 1.

$$T_1(x) = 1 + 4(x-1),$$
$$T_2(x) = 1 + 4(x-1) + 6(x-1)^2,$$
$$T_3(x) = 1 + 4(x-1) + 6(x-1)^2 + 4(x-1)^3,$$
$$T_n(x) = 1 + 4(x-1) + 6(x-1)^2 + 4(x-1)^3 + (x-1)^4, \text{ for } n \geq 4.$$

Example 2.3.9. To find the Taylor expansion of $\log x$ at 2, we use the Taylor expansion of $\log(1+x)$ at 0 to get

$$\log x = \log 2 + \log\left(1 + \frac{x-2}{2}\right)$$

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

$$+ \cdots + \frac{(-1)^{n-1}}{n}\frac{(x-2)^n}{2^n} + o((x-2)^n).$$

Thus the Taylor expansion of $\log x$ at 2 is

$$T_n(x) = \log 2 + \frac{1}{2}(x-2) - \frac{1}{8}(x-2)^2 + \frac{1}{24}(x-2)^3 + \cdots + \frac{(-1)^{n-1}}{n2^n}(x-2)^n.$$

Example 2.3.10. The Taylor expansion of $e^x \sin x$ at 0 can be obtained by multiplying the Taylor expansions of e^x and $\sin x$ together. In particular, to get the 5-th order Taylor expansion, we have

$$e^x \sin x = \left(1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + o(x^4)\right)\left(x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 + o(x^6)\right)$$

$$= \left(1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \frac{1}{4!}x^4\right)\left(x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5\right) + o(x^5)$$

$$= x + x^2 + \frac{1}{2!}x^3 + \frac{1}{3!}x^4 + \frac{1}{4!}x^5 - \frac{1}{3!}x^3 - \frac{1}{3!}x^4 - \frac{1}{2! \cdot 3!}x^5 + \frac{1}{5!}x^5 + o(x^5).$$

The second equality uses $x^m o(x^n) = o(x^{m+n})$. The third equality uses $x^m = o(x^n)$ for $m > n$. Thus the 5-th order Taylor expansion of $e^x \sin x$ at 0 is

$$T_5(x) = x + x^2 + \frac{1}{3}x^3 - \frac{1}{30}x^5.$$

Example 2.3.11. To find the 5-th order Taylor expansion of $\sec x$ at 0, we use the Taylor expansions of $\cos x$ and $(1-x)^{-1}$ at 0 to get

$$\sec x = \frac{1}{\cos x} = \left(1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 + o(x^5)\right)^{-1}$$

$$= 1 + \left(\frac{1}{2!}x^2 - \frac{1}{4!}x^4 + o(x^5)\right) + \left(\frac{1}{2!}x^2 - \frac{1}{4!}x^4 + o(x^5)\right)^2$$

$$+ \left(\frac{1}{2!}x^2 - \frac{1}{4!}x^4 + o(x^5)\right)^3 + o\left(\left(\frac{1}{2!}x^2 - \frac{1}{4!}x^4 + o(x^5)\right)^3\right)$$

$$= 1 + \frac{1}{2!}x^2 - \frac{1}{4!}x^4 + \left(\frac{1}{2!}x^2\right)^2 + o(x^5).$$

Thus the Taylor expansion of $\sec x$ at 0 is

$$T_5(x) = 1 + \frac{1}{2}x^2 + \frac{5}{24}x^4.$$

As a consequence of the Taylor expansion, we get a limit

$$\lim_{x \to 0} \frac{\cos x + \sec x - 2}{(e^{x^2} - 1)^2} = \lim_{x \to 0} \frac{1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 + 1 + \frac{1}{2}x^2 + \frac{5}{24}x^4 - 2 + o(x^5)}{(1 + x^2 + o(x^3) - 1)^2}$$

$$= \lim_{x \to 0} \frac{\frac{1}{4}x^4 + o(x^5)}{x^4 + o(x^5)} = \frac{1}{4}.$$

Example 2.3.12. Here is one more example of using Taylor expansions to compute the limit.

$$\lim_{x \to 0} \left(\frac{1}{x} - \frac{1}{\sin x} \right) = \lim_{x \to 0} \left(\frac{1}{x} - \frac{1}{x - \frac{1}{3!}x^3 + o(x^4)} \right)$$

$$= \lim_{x \to 0} \frac{1}{x} \left(1 - \frac{1}{1 - \frac{1}{3!}x^2 + o(x^3)} \right)$$

$$= \lim_{x \to 0} \frac{1}{x} \left(1 - 1 - \frac{1}{3!}x^2 + o(x^2) \right)$$

$$= \lim_{x \to 0} \left(-\frac{1}{6}x + o(x) \right) = 0.$$

Example 2.3.13. The derivative of $f(x) = \arcsin x$ is

$$g(x) = f'(x) = (1 - x)^{-\frac{1}{2}} = 1 + \frac{1}{2}x + \cdots + \frac{1}{n!}\frac{1 \cdot 3 \cdots (2n - 1)}{2^n}x^n + o(x^n)$$

$$= 1 + \frac{1}{2}x + \cdots + \frac{(2n)!}{2^{2n}(n!)^2}x^n + o(x^n),$$

This tells us

$$\frac{f^{(2n)}(0)}{(2n-1)!} = \frac{g^{(2n-1)}(0)}{(2n-1)!} = 0, \quad \frac{f^{(2n+1)}(0)}{(2n)!} = \frac{g^{(2n)}(0)}{(2n)!} = \frac{(2n)!}{2^{2n}(n!)^2}.$$

Thus we get

$$f^{(2n)}(0) = 0, \quad f^{(2n+1)}(0) = (2n)!\frac{(2n)!}{2^{2n}(n!)^2} = \left(\frac{(2n)!}{2^n n!} \right)^2.$$

Example 2.3.14 (Cauchy). From Example 2.3.7, we know the derivatives of the function

$$f(x) = \begin{cases} e^{-\frac{1}{x^2}} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

are 0 at any order. Thus the Taylor expansion of the function is 0 at any order.

Exercise 2.3.18. Find the Taylor expansions.

1. $x^3 + 5x - 1$ at -1.   3. a^x at 1.   5. $\sin 2x$ at $\dfrac{\pi}{4}$.

2. $x^3 + 5x - 1$ at 0.   4. x^a at 1.   6. $\log \dfrac{3 + x^2}{2 + x}$ at 0.

Exercise 2.3.19. Find the 5-th order Taylor expansions at 0.

1. $\sqrt{x + 1}e^{x^2} \tan x$.   3. $\arcsin x$.   5. $\log \cos x$.

2. $(1 + x)^x$.   4. $\log \dfrac{\sin x}{x}$.   6. $\dfrac{x}{e^x - 1}$.

Exercise 2.3.20. Use Taylor expansions to compute limits.

1. $\lim_{x \to 0} \dfrac{x - \tan x}{x - \sin x}$.

7. $\lim_{x \to 0} \dfrac{xe^x - \log(1 + x)}{x^2}$.

2. $\lim_{x \to 0} \left(\dfrac{1}{x^2} - \dfrac{1}{\tan^2 x} \right)$.

8. $\lim_{x \to \infty} x^2 \log \left(x \sin \dfrac{1}{x} \right)$.

3. $\lim_{x \to 0} \left(\dfrac{\sin x}{x} \right)^{\frac{1}{\log(1 - x^2)}}$.

9. $\lim_{x \to 1} \dfrac{(x - 1) \log x}{1 + \cos \pi x}$.

4. $\lim_{x \to 0} (\cos x + \sin x)^{\frac{1}{x(x+1)}}$.

10. $\lim_{x \to 1} \left(\dfrac{1}{x - 1} - \dfrac{1}{\log x} \right)$.

5. $\lim_{x \to 0} (\cos x)^{\frac{1}{x^2}}$.

11. $\lim_{x \to 0} \left(\dfrac{1}{x} - \cot x \right)$.

6. $\lim_{x \to \infty} x \left(e - \left(1 + \dfrac{1}{x} \right)^x \right)$.

12. $\lim_{x \to 0} \dfrac{(1 + 2x + x^2)^{\frac{1}{x}} - (1 + 2x - x^2)^{\frac{1}{x}}}{x}$.

Exercise 2.3.21. Find the high order derivatives of $\arctan x$ at 0.

Exercise 2.3.22. Suppose $f(x)$ has second order derivative at 0 and satisfies

$$\lim_{x \to 0} \left(1 + x + \frac{f(x)}{x} \right)^{\frac{1}{x}} = e^\lambda.$$

1. Find $f(0)$, $f'(0)$ and $f''(0)$.

2. Find $\lim_{x \to 0} \left(1 + \dfrac{f(x)}{x} \right)^{\frac{1}{x}}$.

Exercise 2.3.23. Suppose $f(x)$ has derivatives of any order. How are the high order derivatives of $f(x)$ and $f(x^2)$ at 0 related?

Exercise 2.3.24. Prove that the Taylor expansion at 0 of an odd function contains only terms of odd power. What about an even function?

Exercise 2.3.25. Suppose $f(x)$ and $g(x)$ are approximated by linear functions $a_0 + a_1 \Delta x$ and $b_0 + b_1 \Delta x$ at x_0. Suppose $a_1 > 0$. Without computing the derivatives, find the linear approximation of $f(x)^{g(x)}$ at x_0. Moreover, extend the result to quadratic approximations.

As explained after the statement of Theorem 2.3.2, like the quadratic situation, the existence of n-th order derivative at x_0 implies the n-th order differentiability at x_0. On the other hand, Exercise 2.3.4 shows that the

n-th order differentiability does not necessarily imply the existence of n-th order derivative. This is rather different from the first order case, when the differentiability is equivalent to the existence of the derivative.

Example 2.3.15. It is easy to see that $|x|^\alpha$ is n-th order differentiable at 0 when one of the following happens:

1. $\alpha > n$: The n-th order approximation is 0.

2. α is an even natural number: We have $|x|^\alpha = x^\alpha$. The n-th order approximation is x^α if $\alpha \le n$ and is 0 if $\alpha > n$.

We claim the converse is also true.

Let m be the unique natural number satisfying $m + 1 \ge \alpha > m$. Then for any natural number $n \ge \alpha$, we have $n \ge m + 1$. Therefore n-th order differentiability implies $(m + 1)$-st order differentiability. Moreover, by the first statement above, the m-th order approximation is 0, so that the $(m + 1)$-st order approximation is bx^{m+1}. In other words, for any $\epsilon > 0$, there is $\delta > 0$, such that

$$|x| < \delta \implies \left||x|^\alpha - bx^{m+1}\right| < \epsilon|x|^{m+1}.$$

This is equivalent to the existence of the limit $b = \lim_{x\to0} \dfrac{|x|^\alpha}{x^{m+1}}$. Since $m + 1 \ge \alpha$, this happens exactly when $\alpha = m + 1$ is an even number. Therefore the converse is proved.

As for the existence of the n-th order derivative, because this implies the n-th order differentiability, the conditions above are necessary. Conversely, if $\alpha > n \ge k > 0$, then we have $(|x|^\alpha)^{(k)} = \alpha(\alpha - 1)\cdots(\alpha - k + 1)x^{\alpha-k}$ for $x > 0$ or $x = 0^+$ and $(|x|^\alpha)^{(k)} = (-1)^k\alpha(\alpha - 1)\cdots(\alpha - k + 1)(-x)^{\alpha-k}$ for $x < 0$ or $x = 0^-$. Therefore $|x|^\alpha$ has n-th order derivative 0 at 0. Moreover, if α is an even natural number, then $|x|^\alpha = x^\alpha$ has derivative of any order. Therefore the condition for the existence of the n-order derivative is the same as the condition for the n-th order differentiability.

Exercise 2.3.26. Study the n-th order differentiability and the existence of the n-th order derivative of the functions at 0 ($\alpha, \beta > 0$, $a, b \ne 0$).

1. $\begin{cases} ax^\alpha & \text{if } x \ge 0 \\ b(-x)^\beta & \text{if } x < 0 \end{cases}$.

2. $\begin{cases} |x|^\alpha \sin \dfrac{1}{|x|^\beta} & \text{if } x \ne 0 \\ 0 & \text{if } x = 0 \end{cases}$.

2.3.4 Remainder

Let $T_n(x)$ be the n-th order Taylor expansion of $f(x)$ at x_0. Under the condition of Theorem 2.3.2, all we know about the remainder is $R_n(x) = f(x) - T_n(x) = o(\Delta x^n)$. Under slightly stronger assumption, however, more can be said about the remainder.

Proposition 2.3.3 (Lagrange[10]). *Suppose $f(t)$ has $(n+1)$-st order derivative*

---

[10]Joseph-Louis Lagrange, born 1736 in Turin (Italy), died 1813 in Paris (France). In analysis, Lagrange invented calculus of variations, Lagrange multipliers, and Lagrange interpolation. He invented the method of solving differential equations by variation of parameters. In number theory, he proved that every natural number is a sum of four squares. He also transformed Newtonian mechanics into a branch of analysis, now called Lagrangian mechanics.

between x and x_0. Then there is c between x and x_0, such that

$$R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x - x_0)^{n+1}. \tag{2.3.6}$$

Note that when $n = 0$, the conclusion of the proposition is exactly the mean value theorem.

The expression for $R_n(x)$ in Proposition 2.3.3 is called the *Lagrange form* of the remainder. In Exercise 2.3.47 and Example 3.3.1, two other formulae for the remainder will be given.

Proof. Under the assumption that f has $(n+1)$-st order derivative between x_0 and x, we have the following computation similar to (2.3.5),

$$\frac{R_n(x)}{(x-x_0)^{n+1}} = \frac{R_n(x) - R_n(x_0)}{(x-x_0)^{n+1} - (x_0 - x_0)^{n+1}}$$

$$= \frac{R_n'(c_1)}{(n+1)(c_1 - x_0)^n} = \frac{R_n'(c_1) - R_n'(x_0)}{(n+1)((c_1 - x_0)^n - (x_0 - x_0)^n)} = \cdots$$

$$= \frac{R_n^{(n)}(c_n)}{(n+1)n(n-1)\cdots 2(c_n - x_0)} = \frac{R_n^{(n)}(c_n) - R_n^{(n)}(x_0)}{(n+1)!(c_n - x_0)} = \frac{R_n^{(n+1)}(c)}{(n+1)!},$$

where c is between x_0 and x. Since T_n is a polynomial of degree n, its $(n+1)$-st order derivative is zero. Therefore $R_n^{(n+1)}(c) = f^{(n+1)}(c)$. The formula for the remainder then follows. □

Example 2.3.16. The remainder of the Taylor expansion for e^x at 0 is

$$R_n(x) = \frac{e^c}{(n+1)!}x^{n+1}.$$

Since $|c| < |x|$, for each fixed x, we have $|R_n(x)| < \dfrac{e^{|x|}|x|^{n+1}}{(n+1)!}$, which converges to 0 as $n \to \infty$. Therefore

$$e^x = \lim_{n\to\infty}\left(1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots + \frac{1}{n!}x^n\right)$$

$$= 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots + \frac{1}{n!}x^n + \cdots.$$

Moreover, since $|R_9(1)| < \dfrac{e}{10!} < \dfrac{3}{10!} < 10^{-6}$, we find

$$e \approx 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{9!} \approx 2.718285$$

is accurate up to the 6th digit.

Example 2.3.17. Suppose $f(x)$ has second order derivative on $[0,1]$. Suppose $|f(0)| \le 1$, $|f(1)| \le 1$ and $|f''(x)| \le 1$. We would like to estimate the size of $f'(x)$.

Fix any $0 < x < 1$. By the second order Taylor expansion at x and the remainder formula, we have

$$f(0) = f(x) + f'(x)(x - 0) + \frac{f''(c_1)}{2}(x - 0)^2, \qquad 0 < c_1 < x$$

$$f(1) = f(x) + f'(x)(x - 1) + \frac{f''(c_2)}{2}(x - 1)^2, \qquad x < c_2 < 1$$

Subtracting the two, we get

$$f'(x) = f(1) - f(0) + \frac{f''(c_1)}{2}x^2 - \frac{f''(c_2)}{2}(x - 1)^2.$$

Therefore by the assumption on the sizes of $f(1)$, $f(0)$ and $f''(x)$, we get

$$|f'(x)| \le 2 + \frac{1}{2}(x^2 + (x - 1)^2) \le \frac{5}{2}.$$

Exercise 2.3.27. Estimate the errors of approximations for $|x| \le 0.2$.

1. $\cos x \approx 1 - \dfrac{1}{2}x^2 + \dfrac{1}{24}x^4$.

3. $\log(1 + x) \approx x - \dfrac{1}{2}x^2$.

2. $\dfrac{1}{\sqrt{1 + x}} \approx 1 - \dfrac{1}{2}x + \dfrac{3}{8}x^2$.

4. $2^x \approx 1 + x \log 2 + \dfrac{(\log 2)^2}{2}x^2$.

Exercise 2.3.28. Compute the values up to the 4-th digit.

1. $\sqrt{e}$.

2. $\log 0.9$.

3. $\sqrt[5]{30}$.

Exercise 2.3.29. Show that the Taylor expansions of $\sin x$ and $\cos x$ converge to the respective functions for any x as $n \to \infty$.

Exercise 2.3.30. Prove that if there is M, such that $|f^{(n)}(x)| \le M$ for all n and $x \in [a, b]$, then the Taylor expansion of $f(x)$ converges to $f(x)$ for $x \in [a, b]$.

Exercise 2.3.31. Prove that for $x > 0$, we have

$$x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots + \frac{x^{2k}}{2k} - \frac{x^{2k+1}}{(2k+1)(1+x)^{2k+1}} > \log(1+x) > x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots + \frac{x^{2k}}{2k} - \frac{x^{2k+1}}{2k+1},$$

and

$$x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots - \frac{x^{2k-1}}{2k-1} + \frac{x^{2k}}{2k(1+x)^{2k}} < \log(1+x) < x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots - \frac{x^{2k-1}}{2k-1} + \frac{x^{2k}}{2k}.$$

This extends the inequality in Exercise 2.2.20. Also derive the similar inequalities for $0 > x > -1$. Then use the inequalities to discuss the convergence of the Taylor series of $\log(1 + x)$.

Exercise 2.3.32. Suppose $f(x)$ has the third order derivative on $[-1, 1]$, such that $f(-1) = 0$, $f(0) = 0$, $f(1) = 1$, $f'(0) = 0$. Prove that there are $-1 < x < 0$ and $0 < y < 1$, such that $f'''(x) + f'''(y) = 6$.

2.3.5   Maximum and Minimum

Suppose $f(x)$ is differentiable at x_0. By Proposition 2.2.1, a necessary condition for x_0 to be a local extreme is $f'(x_0) = 0$. To further determine whether x_0 is indeed a local extreme, high order approximations can be used.

Proposition 2.3.4. *Suppose $f(x)$ has n-th order approximation $a + b(x - x_0)^n$ at x_0, with $b \neq 0$.*

1. *If n is odd and $b \neq 0$, then x_0 is not a local extreme.*

2. *If n is even and $b > 0$, then x_0 is a local minimum. If n is even and $b < 0$, then x_0 is a local maximum.*

Proof. The n-th order approximation means that for any $\epsilon > 0$, there is $\delta > 0$, such that

$$|x - x_0| < \delta \implies |f(x) - a - b(x - x_0)^n| \leq \epsilon |(x - x_0)^n|.$$

This implies $a = f(x_0)$ by taking $x = x_0$. In what follows, we will fix ϵ to be any number satisfying $0 < \epsilon < |b|$, so that $b - \epsilon$ and b have the same sign.

Suppose n is odd and $b > 0$. Then for $\delta > x - x_0 > 0$, we have

$$f(x) - f(x_0) > b(x - x_0)^n - \epsilon |(x - x_0)^n| = (b - \epsilon)(x - x_0)^n > 0,$$

and for $0 > x - x_0 > -\delta$, we have

$$f(x) - f(x_0) < b(x - x_0)^n + \epsilon |(x - x_0)^n| = (b - \epsilon)(x - x_0)^n < 0,$$

Thus x_0 is not a local extreme. Similarly, if $b < 0$, then x_0 is also not a local extreme.

Suppose n is even and $b > 0$. Then $|x - x_0| < \delta$ implies

$$f(x) - f(x_0) \geq b(x - x_0)^n - \epsilon |(x - x_0)^n| = (b - \epsilon)(x - x_0)^n \geq 0.$$

Thus x_0 is a local minimum. Similarly, if $b < 0$, then x_0 is a local maximum. $\square$

Suppose $f(x)$ has derivatives of sufficiently high order. Then the function has high order approximations by the Taylor expansion. To apply the proposition, we assume

$$f'(x_0) = f''(x_0) = \cdots = f^{(n-1)}(x_0) = 0, \quad f^{(n)}(x_0) \neq 0.$$

Then we conclude the following.

1. If n is odd, then x_0 is not a local extreme.

2. If n is even, then x_0 is a local minimum when $f^{(n)}(x_0) > 0$, and is a local maximum when $f^{(n)}(x_0) < 0$.

On the other hand, since the n-th order differentiability is weaker than the existence of the n-th order derivative, Proposition 2.3.4 is a technically stronger statement.

Example 2.3.18. By Example 2.2.1, we know the possible local extrema of the function $f(x) = x^2 e^x$ are 0 and -2. Then by $f''(x) = (x^2 + 4x + 2)e^x$, $f''(0) = 2 > 0$, $f''(-2) = -2e^{-2} < 0$, we conclude that 0 is a local minimum and -2 is a local maximum. This confirms the same conclusion in Example 2.2.7, which was obtained by studying the monotone properties of the function.

Example 2.3.19. For the function $f(x) = (x+1)x^{\frac{8}{3}}$, we have $f'(x) = \dfrac{1}{3}(11x+8)x^{\frac{5}{3}}$, $f''(x) = \dfrac{8}{9}(11x+5)x^{\frac{2}{3}}$. From $f'(x)$ we find two candidates $-\dfrac{11}{8}$ and 0 for the local extrema. Since $f''\left(-\dfrac{11}{8}\right) = -\dfrac{9}{4}11^{\frac{2}{3}} < 0$, we know $-\dfrac{11}{8}$ is a local maximum. Since $f''(0) = 0$, no immediate conclusion can be drawn from the other candidate 0 by Proposition 2.3.4. Moreover, since $f'''(0)$ does not exist, the proposition cannot be applied to determine whether $x = 0$ is a local extreme. On the other hand, for x close to 0, we do have $f(x) \ge 0 = f(0)$, so that 0 is indeed a local minimum by the definition.

Exercise 2.3.33. Find local extrema.

1. $x + \dfrac{1}{x}$.

2. $6x^{10} - 10x^6$.

3. $\sin x + \cos x$.

4. $\dfrac{\sin x}{2 + \cos x}$.

5. $\cos x + \dfrac{1}{2}\cos 2x$.

6. $(x^2 + 1)e^x$.

7. $x \log x$.

8. $\dfrac{(\log x)^2}{x}$.

9. $\left(1 + x + \dfrac{1}{2}x^2\right)e^{-x}$.

10. $\left(1 + x + \dfrac{1}{2}x^2 + \dfrac{1}{6}x^3\right)e^{-x}$.

11. $\arctan x - \dfrac{1}{2}\log(1 + x^2)$.

Exercise 2.3.34. Does the second part of Proposition 2.3.4 hold if $>$ and $<$ are replaced by $\ge$ and $\le$?

Exercise 2.3.35. Study the local extrema of the function

$$f(x) = \begin{cases} \dfrac{1}{x^4}e^{-\frac{1}{x^2}} & \text{if } x \ne 0 \\ 0 & \text{if } x = 0 \end{cases}.$$

2.3.6   Convex and Concave

A function is *convex* if the line segment $L_{x,y}$ connecting any two points $(x, f(x))$ and $(y, f(y))$ on the graph of f lies above the graph of f. In other words, for any $x < z < y$, the point $(z, f(z))$ lies below $L_{x,y}$. See Figure 2.6. From the picture, it is easy to see that the condition is equivalent to any one of the following.

1. slope of $L_{z,y} \ge$ slope of $L_{x,y}$.

2. slope of $L_{x,y} \ge$ slope of $L_{x,z}$.

3. slope of $L_{z,y} \ge$ slope of $L_{x,z}$.

Algebraically, the slope of $L_{x,y}$ is $\dfrac{f(y) - f(x)}{y - x}$, and it is not difficult to verify by direct computation that the three conditions are equivalent. Moreover, the line segment $L_{x,y}$ is given by

$$L_{x,y}(z) = f(y) + \frac{f(y) - f(x)}{y - x}(z - y).$$

Thus the convexity of $f(x)$ means $L_{x,y}(z) \geq f(z)$, which is easily seen to be equivalent to the first condition.
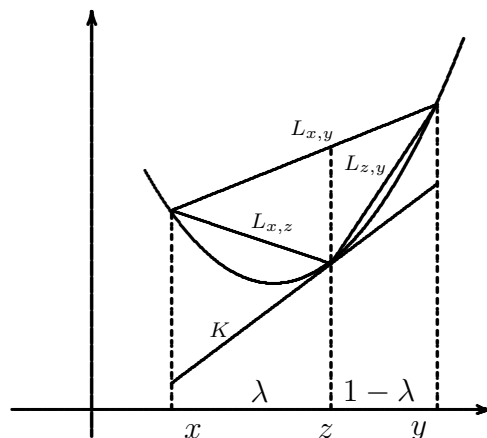
Figure 2.6: convex function

Convex functions can also be characterized by straight lines below the graph.

Proposition 2.3.5. *A function $f(x)$ on an open interval is convex if and only if for any z, there is a linear function $K(x)$ such that $K(z) = f(z)$ and $K(x) \leq f(x)$ for all x.*

Proof. For a convex function f and fixed z, by the third convexity condition, we have

$$\sup_{x<z}(\text{slope of } L_{x,z}) \leq \inf_{y>z}(\text{slope of } L_{z,y}).$$

Let $B = B(z)$ be a number between the supremum and the infimum. Then

$$\text{slope of } L_{x,z} \leq B \leq \text{slope of } L_{z,y}$$

for any $x < z < y$. Let $K(x) = f(z) + B(x - z)$ be the linear function with slope B and satisfies $K(z) = f(z)$. Then the relation above between B and the slopes tells us that $f(x) \geq K(x)$ for any x.

Conversely, suppose the linear function K exists with the claimed property. Then geometrically it is clear that the properties of K implies slope of $L_{x,z} \leq B$ and slope of $L_{z,y} \geq B$ for any $x < z < y$. This implies the third convexity condition.

$\square$

The convexity condition can also be rephrased as follows. Write $z = (1 - \lambda)x + \lambda y$, where $x < z < y$ is equivalent to $0 < \lambda < 1$ (and $z < x$, $z = x$, $x < z < y$, $z = y$, $z > y$ are respectively equivalent to $\lambda < 0$, $\lambda = 0$, $0 < \lambda < 1$, $\lambda = 1$, $\lambda > 1$). Either geometrical consideration or algebraic computation tells us

$$L_{x,y}(z) = (1 - \lambda)f(x) + \lambda f(y).$$

Thus the convexity is the same as

$$0 < \lambda < 1 \implies (1 - \lambda)f(x) + \lambda f(y) \geq f((1 - \lambda)x + \lambda y). \qquad (2.3.7)$$

A function is *concave* if the line segment connecting two points on the graph of f lies below the graph of f. By exchanging the directions of the inequalities, the characterizations of convex functions become the characterization of concave functions.
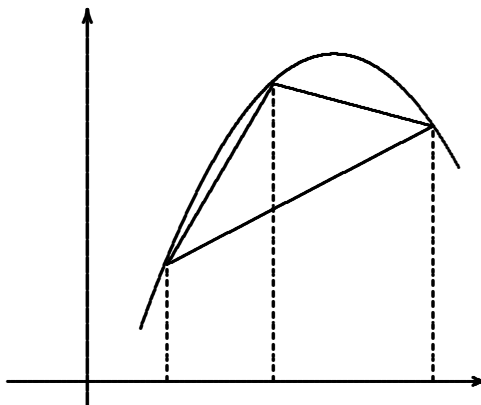
Figure 2.7: concave function

A *point of inflection* is where the function is changed from concave to convex, or from convex to concave.

Proposition 2.3.6. *Suppose $f(x)$ is differentiable on an interval. Then $f(x)$ is convex if and only if $f'(x)$ is increasing.*

Combining Propositions 2.2.4 and 2.3.6, a function $f(x)$ with second order derivative is convex if and only if $f''(x) \geq 0$. The similar statement for concave functions is also true. The points of inflection are then the places where $f''(x)$ changes the sign.

Proof. Suppose $f(x)$ is convex. Then for fixed $x < y$ and changing z between x and y, we have

$$f'(x) = \lim_{z \to x^+} (\text{slope of } L_{x,z}) \leq \text{slope of } L_{x,y} \leq \lim_{z \to y^-} (\text{slope of } L_{z,y}) = f'(y).$$

Conversely, suppose f' is increasing and $x < z < y$. By the mean value theorem, we have

$$\text{slope of } L_{x,z} = f'(c), \ \ \text{slope of } L_{z,y} = f'(d),$$

for some $x < c < z$ and $z < d < y$. Since $c < d$, we have $f'(c) \leq f'(d)$, so that the third condition for the convexity holds.

$\square$

Example 2.3.20. Since $(-\log x)'' = \dfrac{1}{x^2} > 0$, the derivative $(-\log x)'$ is an increasing function and $-\log x$ is convex. Therefore if $p, q > 0$ satisfy $\dfrac{1}{p} + \dfrac{1}{q} = 1$, then we have

$$\log xy = \frac{1}{p}\log x^p + \frac{1}{q}\log y^q \leq \log\left(\frac{1}{p}x^p + \frac{1}{q}y^q\right).$$

Taking the exponential, we get the Young equality in Exercise 2.2.40

$$xy \leq \frac{1}{p}x^p + \frac{1}{q}y^q.$$

Example 2.3.21. The function $f(x) = x^2 e^x$ in Examples 2.2.7 and 2.3.18 has $f''(x) = (x^2 + 4x + 2)e^x = (x-a)(x-b)e^x$, where $a = -2 - \sqrt{2}$, $b = -2 + \sqrt{2}$. From the signs of the second order derivative, we know that the function is convex on $(-\infty, a]$ and $[b, \infty)$, and is concave on $[a, b]$. The points a and b, where concavity and convexity are exchanged, are the points of inflection. This adds additional information to the graph of the function in Figure 2.4.

Example 2.3.22. Let $f(x) = \dfrac{x}{1+x^2}$. From $f'(x) = \dfrac{1-x^2}{(1+x^2)^2}$, we find $f(x)$ is decreasing on $(-\infty, -1]$, increasing on $[-1, 1]$, and decreasing again on $[1, \infty)$. Therefore $f(-1) = -\dfrac{1}{2}$ is a local minimum, and $f(1) = \dfrac{1}{2}$ is a local maximum.

From $f''(x) = \dfrac{-2x(3-x^2)}{(1+x^2)^3}$, we find $f(x)$ is concave on $(-\infty, -\sqrt{3}]$, convex on $[-\sqrt{3}, 0]$, concave again on $[0, \sqrt{3}]$, and then convex again on $[\sqrt{3}, \infty)$. Therefore $f(-\sqrt{3}) = -\dfrac{\sqrt{3}}{4}$, $f(0) = 0$ and $f(\sqrt{3}) = \dfrac{\sqrt{3}}{4}$ are points of inflection.

Combining the information with $\lim_{x\to\infty} f(x) = 0$, we get a rough sketch of the function in Figure 2.8.
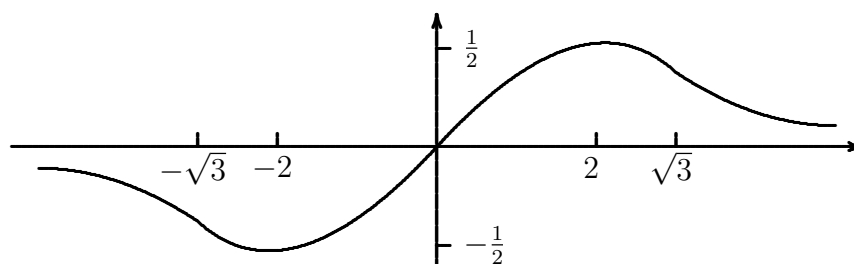
Figure 2.8: graph of $\dfrac{x}{1+x^2}$

Example 2.3.23. Consider the function $f(x) = (x+1)x^{\frac{2}{3}}$ in Example 2.2.8. From $f'(x) = \dfrac{1}{3}(5x+2)x^{-\frac{1}{3}}$, we find $f(x)$ is increasing on $\left(-\infty, -\dfrac{2}{5}\right]$, decreasing on $\left[-\dfrac{2}{5}, 0\right]$, and increasing again on $[0, \infty)$. Therefore $f\left(-\dfrac{2}{5}\right) = \dfrac{3\sqrt[3]{20}}{25}$ is a local maximum, and $f(0) = 0$ is a local minimum.

From $f''(x) = \dfrac{2}{9}(5x - 1)x^{-\frac{4}{3}}$, we find $f(x)$ is concave on $\left(-\infty, \dfrac{1}{5}\right]$ and convex

on $\left[\dfrac{1}{5}, \infty\right)$. Therefore $f\left(\dfrac{1}{5}\right) = \dfrac{6\sqrt[3]{5}}{25}$ is a point of inflection.

Combined with $\lim_{x \to \infty} \dfrac{f(x)}{x^{\frac{5}{3}}} = 1$, we get a rough sketch of the function in Figure 2.9.
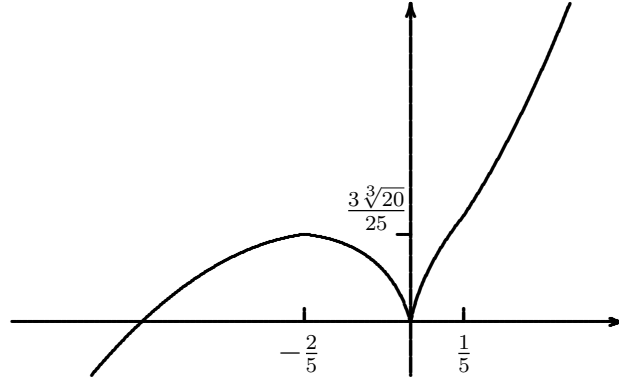
Figure 2.9: graph of $(x + 1)x^{\frac{2}{3}}$

Exercise 2.3.36. Sketch the graph of functions.

1. $x^3 + 6x^2 - 15x - 20$.

2. $x - \dfrac{1}{x}$.

3. $\dfrac{1}{1 + x^2}$.

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

5. $\log(1 + x^2)$.

6. xe^{-x}.

7. $x - \arctan x$.

8. $\dfrac{\log x}{x}$.

9. $e^{-x}\cos x$.

Exercise 2.3.37. Are the sum, product, composition, maximum, minimum of two convex functions still convex?

Exercise 2.3.38. Verify the convexity of $x \log x$ and then use the property to prove the inequality $\left(\dfrac{x + y}{2}\right)^{x+y} \le x^x y^y$.

Exercise 2.3.39. Suppose $p \ge 1$. Show that x^p is convex. Then for non-negative $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n$, take

$$x = \frac{a_i}{\left(\sum a_i^p\right)^{\frac{1}{p}}}, \quad y = \frac{b_i}{\left(\sum b_i^p\right)^{\frac{1}{p}}}, \quad \lambda = \frac{\left(\sum a_i^p\right)^{\frac{1}{p}}}{\left(\sum a_i^p\right)^{\frac{1}{p}} + \left(\sum b_i^p\right)^{\frac{1}{p}}},$$

in the inequality (2.3.7) and derive the Minkowski inequality in Exercise 2.2.42.

Exercise 2.3.40. Suppose $f(x)$ is a convex function. For any $\lambda_1, \lambda_2, \ldots, \lambda_n$ satisfying $\lambda_1 + \lambda_2 + \cdots + \lambda_n = 1$ and $0 < \lambda_i < 1$, prove *Jensen*[11] *inequality*

$$f(\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_n x_n) \le \lambda_1 f(x_1) + \lambda_2 f(x_2) + \cdots + \lambda_n f(x_n).$$

---

[11]Johan Jensen, born 1859 in Nakskov (Denmark), died 1925 in Copenhagen (Denmark). He proved the inequality in 1906.

Then use this to prove that for $x_i > 0$, we have

$$\sqrt[n]{x_1 x_2 \cdots x_n} \le \frac{x_1 + x_2 + \cdots + x_n}{n} \le \sqrt[p]{\frac{x_1^p + x_2^p + \cdots + x_n^p}{n}} \text{ for } p \ge 1,$$

and

$$(x_1 x_2 \cdots x_n)^{\frac{x_1 + x_2 + \cdots + x_n}{n}} \le x_1^{x_1} x_2^{x_2} \cdots x_n^{x_n}.$$

Exercise 2.3.41. Prove that if $f''(x_0) = 0$ and $f'''(x_0) \ne 0$, then x_0 is a point of inflection.

Exercise 2.3.42. Prove that a continuous function on an interval is convex if and only if $\dfrac{f(x) + f(y)}{2} \ge f\left(\dfrac{x+y}{2}\right)$ for any x and y on the interval.

Exercise 2.3.43. Prove that a function $f(x)$ on an open interval (a, b) is convex if and only if for any $a < x < y < b$, we have $f(z) \ge L_{x,y}(z)$ for any $z \in (a, x)$ and $z \in (y, b)$. Then prove that a convex function on an open interval must be continuous.

2.3.7   Additional Exercise

Estimation of $\sin x$ **and** $\cos x$

We know the inequality $\sin x > \dfrac{2}{\pi}x$ for $0 < x < \dfrac{\pi}{2}$ from (1.3.18) and Exercise 2.2.20. The subsequent exercises extend the estimation to higher order and to $\cos x$.

Exercise 2.3.44. Let

$$f_k(x) = x - \frac{x^3}{3!} + \cdots + (-1)^{k-1}\frac{x^{2k-1}}{(2k-1)!} - \sin x,$$

$$g_k(x) = 1 - \frac{x^2}{2!} + \cdots + (-1)^k \frac{x^{2k}}{(2k)!} - \cos x.$$

Verify that $g_k' = -f_k$ and $f_{k+1}' = g_k$. Then use the equalities to prove that

$$x - \frac{x^3}{3!} + \cdots - \frac{x^{4k-1}}{(4k-1)!} < \sin x < x - \frac{x^3}{3!} + \cdots - \frac{x^{4k-1}}{(4k-1)!} + \frac{x^{4k+1}}{(4k+1)!}$$

for $x > 0$. Also derive the similar inequalities for $\cos x$.

Exercise 2.3.45. Prove that

$$x - \frac{x^3}{3!} + \cdots - \frac{x^{4k-1}}{(4k-1)!} + \frac{2}{\pi}\frac{x^{4k+1}}{(4k+1)!} < \sin x < x - \frac{x^3}{3!} + \cdots + \frac{x^{4k+1}}{(4k+1)!} - \frac{2}{\pi}\frac{x^{4k+3}}{(4k+3)!},$$

and

$$1 - \frac{x^2}{2!} + \cdots - \frac{x^{4k-2}}{(4k-2)!} + \frac{2}{\pi}\frac{x^{4k}}{(4k)!} < \cos x < 1 - \frac{x^2}{2!} + \cdots + \frac{x^{4k}}{(4k)!} - \frac{2}{\pi}\frac{x^{4k+2}}{(4k+2)!}$$

for $0 < x < \dfrac{\pi}{2}$.

Exercise 2.3.46. Let

$$f_n(x) = 1 + x - \frac{x^2}{2!} - \frac{x^3}{3!} + \cdots + s_1(n)\frac{x^n}{n!}, \quad s_1(n) = \begin{cases} 1 & \text{if } n = 4k, 4k+1 \\ -1 & \text{if } n = 4k+2, 4k+3 \end{cases},$$

$$g_n(x) = 1 - x - \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + s_2(n)\frac{x^n}{n!}, \quad s_2(n) = \begin{cases} 1 & \text{if } n = 4k-1, 4k \\ -1 & \text{if } n = 4k+1, 4k+2 \end{cases}.$$

Prove that

$$f_{4k+1}(x) - \sqrt{2}\frac{x^{4k+2}}{(4k+2)!} < \cos x + \sin x < f_{4k-1}(x) + \sqrt{2}\frac{x^{4k}}{(4k)!},$$

and

$$g_{4k}(x) - \sqrt{2}\frac{x^{4k+1}}{(4k+1)!} < \cos x - \sin x < g_{4k+2}(x) + \sqrt{2}\frac{x^{4k+3}}{(4k+3)!}.$$

Moreover, derive similar inequalities for $a\cos x + b\sin x$.

Cauchy Form of the Remainder

The Lagrange form (2.3.6) is the simplest form of the remainder. However, for certain functions, it is more suitable to use the *Cauchy form*

$$R_n(x) = \frac{f^{(n+1)}(c)}{n!}(x - c)^n(x - x_0) \tag{2.3.8}$$

of the remainder. The proof makes use of the function

$$F(t) = f(x) - f(t) - f'(t)(x - t) - \frac{f''(t)}{2}(x - t)^2 - \cdots - \frac{f^{(n)}(t)}{n!}(x - t)^n$$

defined for any fixed x and x_0.

Exercise 2.3.47. By applying the mean value theorem to $F(t)$ for t between x_0 and x, prove the Cauchy form of the remainder.

Exercise 2.3.48. By applying Cauchy's mean value theorem to $F(t)$ and $G(t) = (x - t)^{n+1}$, derive the Lagrange form (2.3.6) for the remainder.

Exercise 2.3.49. Prove the remainder of the Taylor series of $(1 + x)^\alpha$ satisfies

$$|R_n| \le \rho_n = A\left|\frac{\alpha(\alpha - 1)\cdots(\alpha - n)}{n!}x^{n+1}\right| \quad \text{for } |x| < 1,$$

where $A = (1 + |x|)^{\alpha-1}$ for $\alpha \ge 1$ and $A = (1 - |x|)^{\alpha-1}$ for $\alpha < 1$. Then use Exercise 1.1.33 to show that $\lim_{n\to\infty} \rho_n = 0$. This shows that the Taylor series of $(1 + x)^\alpha$ converges for $|x| < 1$.

Exercise 2.3.50. Study the convergence of the Taylor series of $\log(1 + x)$.

Relation between the Bounds of a Function and its Derivatives

In Example 2.3.17, we saw bounds on a function and its second order derivative will impose a bound on the first order derivative. The subsequent exercises provide more examples.

Exercise 2.3.51. Suppose $f(x)$ is a function on $[0,1]$ with second order derivative and satisfying $f(0) = f'(0) = 0$, $f(1) = 1$. Prove that if $f''(x) \leq 2$ for any $0 < x < 1$, then $f(x) = x^2$. In other words, unless $f(x) = x^2$, we will have $f''(x) > 2$ somewhere on $(0,1)$.

Exercise 2.3.52. Consider functions $f(x)$ on $[0,1]$ with second order derivative and satisfying $f(0) = f(1) = 0$ and $\min_{[0,1]} f(x) = -1$. What would be the "lowest bound" a for $f''(x)$? In other words, find biggest a, such that any such function $f(x)$ will have $f''(x) \geq a$ somewhere on $(0,1)$.

Exercise 2.3.53. Study the constraint on the second order derivative for functions on $[a,b]$ satisfying $f(a) = A$, $f(b) = B$ and $\min_{[a,b]} f(x) = m$.

Exercise 2.3.54. Suppose $f(x)$ has the second order derivative on (a,b). Suppose M_0, M_1, M_2 are the suprema of $|f(x)|$, $|f'(x)|$, $|f''(x)|$ on the interval. By rewriting the remainder formula as an expression of f' in terms of f and f'', prove that

$$|f'(x)| \leq \frac{h}{2}M_2 + \frac{2}{h}M_0$$

for any $a < x < b$ and $0 < h < \max\{x - a, b - x\}$. Then prove $M_1 \leq 2\sqrt{M_2 M_0}$ in case $b = +\infty$. Moreover, verify that the equality happens for

$$f(x) = \begin{cases} 2x^2 - 1 & \text{if } -1 < x < 0 \\ \dfrac{x^2 - 1}{x^2 + 1} & \text{if } x \geq 0 \end{cases}.$$

Exercise 2.3.55. Suppose $f(x)$ has the second order derivative on $(a, +\infty)$. Prove that if $f''(x)$ is bounded and $\lim_{x \to +\infty} f(x) = 0$, then $\lim_{x \to +\infty} f'(x) = 0$.

Convexity Criterion by the One Side Derivative

Convex functions are continuous (see Exercise 2.3.43) but not necessarily always differentiable ($|x|$ is convex, for example). So a convexity criterion more general than Proposition 2.3.6 is needed.

Exercise 2.3.56. Prove that a convex function $f(x)$ on an open interval is left and right differentiable, and the one side derivatives satisfy

$$f'_+(x) \leq \frac{f(y) - f(x)}{y - x} \leq f'_-(y).$$

Exercise 2.3.57. Prove the following are equivalent for a function $f(x)$ on an open interval.

1. $f(x)$ is convex.

2. $f(x)$ is left and right differentiable, with $x < y$ implying $f'_-(x) \leq f'_+(x) \leq f'_-(y) \leq f'_+(y)$.

3. $f(x)$ is left continuous and right differentiable, with increasing $f'_+(x)$.

Chapter 3

Integration

3.1   Riemann Integration

Discovered independently by Newton and Leibniz, the integration was originally the method of using the antiderivative to find the areas under curves. Thus the method started as an application of the differentiation. Then Riemann[1] studied the limiting process leading to the area and established the integration as an independent subject. The new viewpoint further led to other integration theories, among which the most significant is the Lebesgue integration.

3.1.1   Riemann Sum

Let $f(x)$ be a function on a bounded interval $[a, b]$. To compute the area of the region between the graph of the function and the x-axis, we choose a *partition*

$$P \colon a = x_0 < x_1 < x_2 < \cdots < x_n = b \qquad (3.1.1)$$

of the interval and approximate the region by a sequence of rectangles with base $[x_{i-1}, x_i]$ and height $f(x_i^*)$, where $x_i^* \in [x_{i-1}, x_i]$. The total area of the rectangles is the *Riemann sum*

$$S(P, f) = \sum_{i=1}^{n} f(x_i^*)(x_i - x_{i-1}) = \sum_{i=1}^{n} f(x_i^*)\Delta x_i. \qquad (3.1.2)$$

Note that $S(P, f)$ also depends on the choices of x_i^*, although the choice does not explicitly appear in the notation.
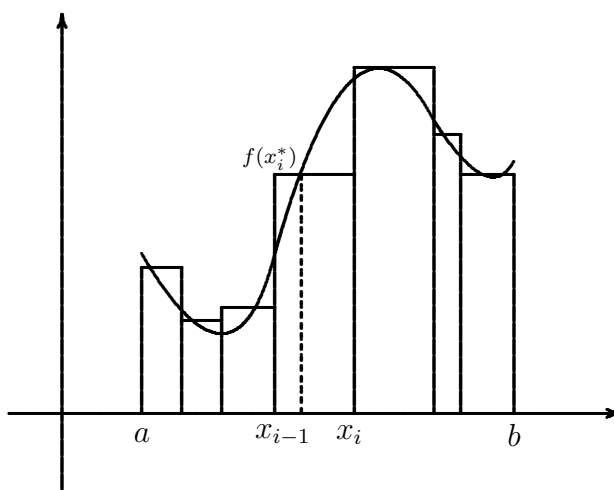
Figure 3.1: Riemann sum

Example 3.1.1. To find the area under the function $f(x) = x$ over the interval $[0, 1]$, we choose a partition P_n given by $x_i = \dfrac{i}{n}$ and choose $x_i^* = \dfrac{i}{n}$. Then

$$S(P_n, x) = \sum_{i=1}^{n} \frac{i}{n}\frac{1}{n} = \frac{1}{n^2}\frac{n(n+1)}{2} = \frac{n+1}{2n}.$$

---

[1]Georg Friedrich Bernhard Riemann, born 1826 in Breselenz (Germany), died 1866 in Selasca (Italy).

If we instead choose the middle point $x_i^* = \dfrac{1}{2}\left(\dfrac{i-1}{n} + \dfrac{i}{n}\right) = \dfrac{2i-1}{2n}$ of the interval $[x_{i-1}, x_i]$, then the Riemann sum

$$S(P_n, x) = \sum_{i=1}^{n} \frac{2i-1}{2n}\frac{1}{n} = \frac{n(n+1)-n}{2n^2} = \frac{1}{2}.$$

Exercise 3.1.1. Compute the Riemann sums.

1. $f(x) = x$, $x_i = \dfrac{i}{n}$, $x_i^* = \dfrac{i-1}{n}$.

2. $f(x) = x^2$, $x_i = \dfrac{i}{n}$, $x_i^* = \dfrac{i}{n}$.

3. $f(x) = x^2$, $x_i = \dfrac{i}{n}$, $x_i^* = \dfrac{2i-1}{2n}$.

4. $f(x) = \alpha^x$, $x_i = \dfrac{i}{n}$, $x_i^* = \dfrac{i-1}{n}$.

The Riemann sum is only an approximation of the area of the region. We expect the approximation to get more accurate when the size of the partition

$$\|P\| = \max_{1 \le i \le n} \Delta x_i$$

gets smaller. This leads to the definition of the *definite integral*.

Definition 3.1.1. A function $f(x)$ on a bounded interval $[a, b]$ is *Riemann integrable*, with *integral I*, if for any $\epsilon > 0$, there is $\delta > 0$, such that

$$\|P\| < \delta \implies |S(P, f) - I| < \epsilon. \tag{3.1.3}$$

Because of the similarity to the definition of limits, we write

$$I = \int_a^b f(x)dx = \lim_{\|P\| \to 0} S(P, f).$$

The numbers a and b are called the *lower limit* and the *upper limit* of the integral.

As pointed out at the beginning of the chapter, there are several integration theories. Therefore there are different meanings of integrability. In this course, unless otherwise indicated, the integrability will always mean Riemann integrability.

Example 3.1.2. For the constant function $f(x) = c$ on $[a, b]$ and any partition P, we have

$$S(P, c) = \sum_{i=1}^{n} c\Delta x_i = c\sum_{i=1}^{n} \Delta x_i = c(b - a).$$

Therefore the constant function is integrable, with

$$\int_a^b cdx = c(b - a).$$

Example 3.1.3. Consider the function

$$d_c(x) = \begin{cases} 0 & \text{if } x \ne c \\ 1 & \text{if } x = c \end{cases}$$

that is constantly zero except at $x = c$. For any partition P of a bounded closed interval $[a, b]$ containing c, we have

$$S(P, d_c) = \begin{cases} 0 & \text{if no } x_i^* = c \\ \Delta x_k & \text{if } x_{k-1} < x_k^* = c < x_k \\ \Delta x_k & \text{if } x_1^* = a = c, k = 1 \ \text{ or } \ x_n^* = b = c, k = n \\ \Delta x_k + \Delta x_{k+1} & \text{if } x_k^* = x_{k+1}^* = x_k = c, k \neq 0, k \neq n \end{cases}.$$

This implies that $d_c(x)$ is integrable, and $\int_a^b d_c(x)dx = 0$.

Example 3.1.4. For the Dirichlet function $D(x)$ in Example 1.3.24 and any partition P of $[a, b]$, we have $S(P, D) = b - a$ if all x_i^* are rational numbers and $S(P, D) = 0$ if all x_i^* are irrational numbers. Thus the Dirichlet function is not integrable.

Example 3.1.5. Consider Thomae's function $R(x)$ in Example 1.4.2. For any natural number N, let A_N be the set of rational numbers in $[0, 1]$ with denominators $\leq N$. Then A_N is finite, containing say ν_N numbers. For any partition P of $[0, 1]$ and choices of x_i^*, the Riemann sum $S(P, R)$ can be divided into two parts. The first part consists of those intervals with $x_i^* \in A_N$, and the second part has $x_i^* \notin A_N$. The number of terms in the first part is $\leq 2\nu_N$ (the factor 2 takes into account of the possibility that $x_i^* = x_{i+1}^* = x_i \in A_N$), so that the total length of the intervals in the first part is no more than $2\nu_N \|P\|$. Moreover, the total length of the intervals in the second part is no more than the total length 1 of the whole interval $[0, 1]$. Since we have $0 < R(x_i^*) \leq 1$ the first part and we also have $0 \leq R(x_i^*) \leq \dfrac{1}{N}$ in the second part, we conclude

$$0 \leq S(P, R) \leq 2\nu_N \|P\| + \frac{1}{N} 1 = 2\nu_N \|P\| + \frac{1}{N}.$$

By taking $\|P\| < \delta = \dfrac{1}{2N\nu_N}$, for example, we get $0 \leq S(P, R) < \dfrac{2}{N}$. Thus we conclude that the function is integrable on $[0, 1]$, with $\int_0^1 R(x)dx = 0$.

Exercise 3.1.2. Study the integrability. For the integrable ones, find the integrals.

1. $\begin{cases} 0 & \text{if } 0 \leq x < 1 \\ 1 & \text{if } 1 \leq x \leq 2 \end{cases}$ on $[0, 2]$.

3. $\begin{cases} 1 & \text{if } x = \frac{1}{n}, n \in \mathbb{Z} \\ 0 & \text{if otherwise} \end{cases}$ on $[-1, 1]$.

2. $\begin{cases} x & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational} \end{cases}$ on $[0, 1]$.

4. $\begin{cases} \frac{1}{n} & \text{if } x = \frac{1}{2^n}, n \in \mathbb{N} \\ 0 & \text{if otherwise} \end{cases}$ on $[0, 1]$.

Exercise 3.1.3. Prove that $\int_0^1 xdx = \dfrac{1}{2}$ in the following steps.

1. For any partition P, if $x_i^* = \dfrac{x_i + x_{i-1}}{2}$ is the middle point of the intervals, then the Riemann sum $S_{\mathrm{mid}}(P, x) = \dfrac{1}{2}$.

2. For any partition P and choice of x_i^*, we have $|S(P, x) - S_{\mathrm{mid}}(P, x)| \leq \dfrac{1}{2}\|P\|$.

Finally, we remark that since the Riemann sum $S(P, f)$ takes into account of the sign of the function $f(x)$, the integration is actually the *signed* area, which counts the part of the area corresponding to $f(x) < 0$ as negative.
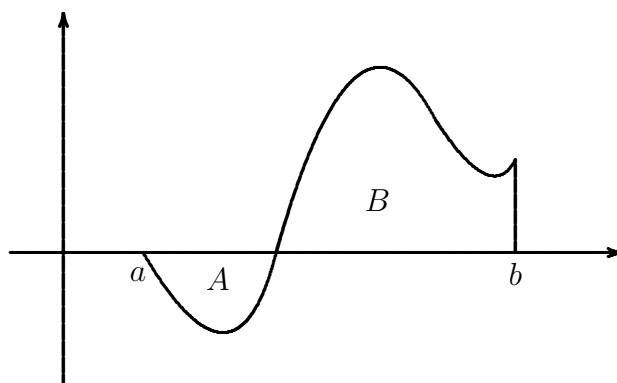
Figure 3.2: $\displaystyle\int_a^b f(x)dx = -\text{area}(A) + \text{area}(B)$

3.1.2 Integrability Criterion

In Examples 3.1.2 through 3.1.5, we saw a function may or may not be integrable. The following is a simple condition for integrability.

Proposition 3.1.2. *Riemann integrable functions are bounded.*

Example 3.1.4 shows the converse is not true.

Proof. Let $f(x)$ be integrable on a bounded interval $[a, b]$ and let I be the integral. Then for $\epsilon = 1 > 0$, there is a partition P, such that

$$\left| \sum_{i=1}^n f(x_i^*)\Delta x_i - I \right| = |S(P, f) - I| < 1$$

for any choice of x_i^*. Now we fix $x_2^*, x_3^*, \ldots, x_n^*$, so that $\sum_{i=2}^n f(x_i^*)\Delta x_i$ is a fixed bounded number. Then

$$|f(x_1^*)\Delta x_1| \leq \left| \sum_{i=2}^n f(x_i^*)\Delta x_i - I \right| + 1$$

for any $x_1^* \in [x_0, x_1]$. In particular, this shows that $f(x)$ is bounded by $\dfrac{1}{\Delta x_1}\left(|\sum_{i=2}^n f(x_i^*)\Delta x_i - I| + 1\right)$ on the first interval $[x_0, x_1]$ of the partition. Similar argument shows that the function is bounded on any interval of the partition. Since the partition contains finitely many intervals, the function is bounded on $[a, b]$.

$\square$

Similar to the convergence of sequences, a more refined criterion for integrability can be obtained by considering the Cauchy criterion. By a proof similar to the convergence of sequences and functions, the Riemann sum converges if and only if for any $\epsilon > 0$, there is $\delta > 0$, such that

$$\|P\|, \|P'\| < \delta \implies |S(P, f) - S(P', f)| < \epsilon. \tag{3.1.4}$$

Note that hidden in the notation is the choices of x_i^* and $x_i'^*$ for P and P'. For the special case $P = P'$, we have

$$S(P, f) - S(P', f) = \sum_{i=1}^{n} (f(x_i^*) - f(x_i'^*))\Delta x_i,$$

and the supremum of the difference for all possible choices of x_i^* and $x_i'^*$ is

$$\sup_{\text{all } x_i^*} S(P, f) - \inf_{\text{all } x_i^*} S(P, f) = \sum_{i=1}^{n} \left(\sup_{[x_{i-1}, x_i]} f(x) - \inf_{[x_{i-1}, x_i]} f(x) \right) \Delta x_i.$$

Define the *oscillation* of a bounded function $f(x)$ on an interval $[a, b]$ to be

$$\omega_{[a,b]}(f) = \sup_{a \le x \le y \le b} |f(x) - f(y)| = \sup_{[a,b]} f(x) - \inf_{[a,b]} f(x).$$

Then

$$\sup_{\text{all } x_i^*} S(P, f) - \inf_{\text{all } x_i^*} S(P, f) = \sum_{i=1}^{n} \omega_{[x_{i-1}, x_i]}(f)\Delta x_i,$$

is the Riemann sum of the oscillations. Therefore the specialization of the Cauchy criterion (3.1.4) to the case $P = P'$ becomes

$$\|P\| < \delta \implies \sum_{i=1}^{n} \omega_{[x_{i-1}, x_i]}(f)\Delta x_i < \epsilon. \tag{3.1.5}$$
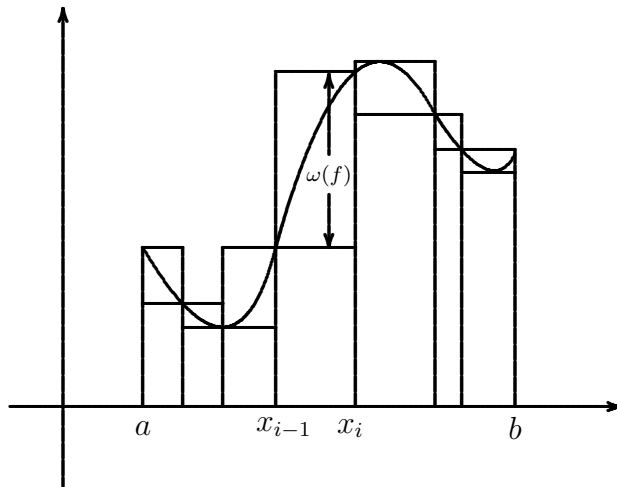
Figure 3.3: Riemann sum of oscillations

It turns out this specialized Cauchy criterion also implies the general Cauchy criterion (3.1.4). As a preparation for the proof, note that for any $a \le c \le b$, we have

$$|f(c)(b-a) - S(P,f)| = \left| \sum_{i=1}^{n} (f(c) - f(x_i^*))\Delta x_i \right| \le \sum_{i=1}^{n} |f(c) - f(x_i^*)|\Delta x_i$$

$$\le \sum_{i=1}^{n} \omega_{[a,b]}(f)\Delta x_i \le \omega_{[a,b]}(f)(b-a). \qquad (3.1.6)$$

Theorem 3.1.3 (Riemann Criterion). *A bounded function $f(x)$ on a bounded interval $[a,b]$ is Riemann integrable if and only if for any $\epsilon > 0$, there is $\delta > 0$, such that $\|P\| < \delta$ implies $\sum_{i=1}^{n} \omega_{[x_{i-1},x_i]}(f)\Delta x_i < \epsilon$.*

Proof. Assume the implication (3.1.5) holds. Let P and P' be partitions satisfying $\|P\|, \|P'\| < \delta$. Let Q be the partition obtained by combining the partition points in P and P' together. Make an arbitrary choice of x_i^* for Q and form the Riemann sum $S(Q,f)$.

Note that Q is obtained by adding more points into P (Q is called a *refinement* of P). For any interval $[x_{i-1}, x_i]$ in the partition P, denote by $Q_{[x_{i-1},x_i]}$ the part of the partition Q lying inside the interval. Then

$$S(Q,f) = \sum_{i=1}^{n} S(Q_{[x_{i-1},x_i]}, f).$$

By the inequality (3.1.6), we have

$$|S(P,f) - S(Q,f)| = \left| \sum_{i=1}^{n} f(x_i^*)\Delta x_i - \sum_{i=1}^{n} S(Q_{[x_{i-1},x_i]}, f) \right|$$

$$\le \sum_{i=1}^{n} \left| f(x_i^*)(x_i - x_{i-1}) - S(Q_{[x_{i-1},x_i]}, f) \right|$$

$$\le \sum_{i=1}^{n} \omega_{[x_{i-1},x_i]}(f)\Delta x_i.$$

Since $\|P\| < \delta$, the implication (3.1.5) tells us that the right side is less than ϵ. By the same reason, we get $|S(P',f) - S(Q,f)| < \epsilon$. Therefore

$$|S(P,f) - S(P',f)| \le |S(P,f) - S(Q,f)| + |S(P',f) - S(Q,f)| < 2\epsilon.$$

$\square$

Example 3.1.6. For the function $f(x) = x$ in Example 3.1.1, we have $\omega_{[x_{i-1},x_i]}(f) = x_i - x_{i-1} \le \|P\|$ and $\sum \omega_{[x_{i-1},x_i]}(f)\Delta x_i \le \sum \|P\|\Delta x_i = (b-a)\|P\|$. Then it is easy to see that the Riemann criterion is satisfied, so that $f(x) = x$ is integrable. Moreover, the computation in Example 3.1.1 tells us $\int_0^1 x\,dx = \frac{1}{2}$.

Example 3.1.7. For the Dirichlet function in Exercise 1.3.24, we have $\omega_{[x_{i-1},x_i]}(f) = 1$ and $\sum \omega_{[x_{i-1},x_i]}(f)\Delta x_i = b-a$. By the Riemann criterion, the function is not integrable.

Exercise 3.1.4. Prove that the function $f(x) = x^2$ is integrable on $[0, 1]$.

Exercise 3.1.5. Study the integrability of the functions in Exercise 3.1.2 again by using Theorem 3.1.3.

Exercise 3.1.6. Prove the inequality $\omega(|f|) \leq \omega(f)$. Then prove that the integrability of $f(x)$ implies the integrability of $|f(x)|$.

Exercise 3.1.7. Prove that the integrability of $f(x)$ implies the integrability of $f(x)^2$.

Exercise 3.1.8. Suppose a bounded function $f(x)$ on $[a, b]$ is integrable on $[c, b]$ for any $a < c < b$. Prove that $f(x)$ is integrable on $[a, b]$. In fact, we also have $\int_a^b f(x)dx = \lim_{c \to a^+} \int_c^b f(x)dx$ by Theorem 3.2.1.

3.1.3   Integrability of Continuous and Monotone Functions

By using the integrability criterion in Theorem 3.1.3, we may identify some important classes of integrable functions.

Proposition 3.1.4. *Continuous functions on bounded closed intervals are Riemann integrable.*

Proof. By Theorem 1.4.4, for any $\epsilon > 0$, there is $\delta > 0$, such that $|x - y| < \delta$ implies $|f(x) - f(y)| < \epsilon$. Suppose $\|P\| < \delta$. Then for any $x, y \in [x_{i-1}, x_i]$, we have $|x - y| \leq |x_i - x_{i-1}| < \delta$, which implies $|f(x) - f(y)| < \epsilon$. Therefore the oscillation $\omega_{[x_{i-1}, x_i]}(f) \leq \epsilon$ and $\sum \omega_{[x_{i-1}, x_i]}(f)\Delta x_i \leq \epsilon(b - a)$. By Theorem 3.1.3, the function is integrable. $\square$

Proposition 3.1.5. *Monotone functions on bounded closed intervals are Riemann integrable.*

Proof. Let $f(x)$ be an increasing function on a bounded closed interval $[a, b]$. Then $\omega_{[x_{i-1}, x_i]}(f) = f(x_i) - f(x_{i-1})$, and

$$\sum \omega_{[x_{i-1}, x_i]}(f)\Delta x_i = \sum (f(x_i) - f(x_{i-1}))\Delta x_i$$
$$\leq \|P\| \sum (f(x_i) - f(x_{i-1})) = \|P\|(f(b) - f(a)).$$

By Theorem 3.1.3, this implies that the function is integrable. $\square$

Example 3.1.8. The function x is integrable on $[a, b]$ by either Proposition 3.1.4 or Proposition 3.1.5. The computation in Example 3.1.1 tells us $\int_0^1 xdx = \lim_{n \to \infty} S(P_n, x) = \frac{1}{2}$. An extension of the computation shows

$$\int_a^b xdx = \frac{1}{2}(b^2 - a^2).$$

The function x^2 is integrable on $[a, b]$ by Proposition 3.1.4. To find the integral, we take a partition P_n to consist of $x_i = a + \dfrac{i}{n}(b - a)$ and choose $x_i^* = x_i$. Then

$$
\begin{aligned}
S(P_n, x^2) &= \sum_{i=1}^{n} \left(a + \frac{i}{n}(b - a) \right)^2 \frac{1}{n}(b - a) \\
&= \sum_{i=1}^{n} \left(\frac{1}{n}a^2 + 2\frac{i}{n^2}a(b - a) + \frac{i^2}{n^3}(b - a)^2 \right)(b - a) \\
&= \frac{n}{n}a^2(b - a) + \frac{n(n + 1)}{n^2}a(b - a)^2 + \frac{n(n + 1)(2n + 1)}{6n^3}(b - a)^3.
\end{aligned}
$$

Taking the limit as $n \to \infty$, we get

$$
\int_a^b x^2 dx = \frac{1}{3}(b^3 - a^3).
$$

Exercise 3.1.9. Suppose $f(x)$ is a convex function on $[a, b]$. Use the inequalities in Exercises 2.3.56 and 2.3.57 to prove that $f(b) - f(a) = \displaystyle\int_a^b f'(x^-)dx = \displaystyle\int_a^b f'(x^+)dx$.

Proposition 3.1.6. *Suppose $f(x)$ is an integrable function on a bounded closed interval $[a, b]$. Suppose the values of $f(x)$ lie in a finite union U of closed intervals, and $\phi(y)$ is a continuous function on U. Then the composition $\phi(f(x))$ is integrable.*

Proof. By Theorems 1.4.4 and 1.4.5, the continuous function ϕ is uniformly continuous and bounded on each interval inside U. Since U contains finitely many closed intervals, ϕ is also uniformly continuous and bounded on U. In other words, for any $\epsilon > 0$, there is $\delta > 0$, such that $|y - y'| < \delta$ implies $|\phi(y) - \phi(y')| < \epsilon$. By taking $y = f(x)$ and $y' = f(x')$, for any interval $[c, d] \subset [a, b]$, we have

$$
\begin{aligned}
\omega_{[c,d]}(f) < \delta &\implies |f(x) - f(x')| < \delta \text{ for any } x, x' \in [c, d] \\
&\implies |\phi(f(x)) - \phi(f(x'))| < \epsilon \text{ for any } x, x' \in [c, d] \\
&\implies \omega_{[c,d]}(\phi \circ f) \leq \epsilon.
\end{aligned}
$$

By Theorem 3.1.3, since $f(x)$ is integrable, there is $\delta' > 0$ satisfying $\delta' < \delta$, such that

$$
\|P\| < \delta' \implies \sum \omega_{[x_{i-1}, x_i]}(f)\Delta x_i < \delta\epsilon.
$$

Then the sum

$$
\sum \omega_{[x_{i-1}, x_i]}(\phi \circ f)\Delta x_i = \sum_{<\delta} + \sum_{\geq \delta},
$$

where $\sum_{<\delta}$ and $\sum_{\geq \delta}$ consist respectively of the terms with $\omega_{[x_{i-1}, x_i]}(f) < \delta$ and the terms with $\omega_{[x_{i-1}, x_i]}(f) \geq \delta$. Since $\omega_{[x_{i-1}, x_i]}(f) < \delta$ implies $\omega_{[x_{i-1}, x_i]}(\phi \circ f) < \epsilon$, we have

$$
\sum_{<\delta} \omega_{[x_{i-1}, x_i]}(\phi \circ f)\Delta x_i \leq \sum_{<\delta} \epsilon \Delta x_i \leq \sum \epsilon \Delta x_i = (b - a)\epsilon.
$$

For those intervals with $\omega_{[x_{i-1}, x_i]}(f) \geq \delta$, the total length $\sum_{\geq \delta} \Delta x_i$ of such intervals can be estimated as follows

$$\delta \sum_{\geq \delta} \Delta x_i \leq \sum_{\geq \delta} \omega_{[x_{i-1}, x_i]}(f) \Delta x_i \leq \sum \omega_{[x_{i-1}, x_i]}(f) \Delta x_i \leq \delta \epsilon.$$

If ϕ is bounded by B on U, then $\omega_{[x_{i-1}, x_i]}(\phi \circ f) \leq 2B$ and

$$\sum_{\geq \delta} \omega_{[x_{i-1}, x_i]}(\phi \circ f) \Delta x_i \leq 2B \sum_{\geq \delta} \Delta x_i \leq 2B \epsilon.$$

Thus we conclude

$$\sum \omega_{[x_{i-1}, x_i]}(\phi \circ f) \Delta x_i = \sum_{<\delta} + \sum_{\geq \delta} \leq (b-a)\epsilon + 2B\epsilon.$$

Since a, b, B are all fixed constants, by Theorem 3.1.3, this implies that $\phi \circ f$ is integrable.

$\square$

Example 3.1.9. If $f(x)$ is integrable, then $f(x)^2$ and $|f(x)|$ are also integrable. If we further have $f(x) \geq 0$, then $\sqrt{f(x)}$ is integrable. If we further have $|f(x)| > c > 0$ for a constant c, then by taking $U = [-B, -c] \cup [c, B]$, where B is the bound for f, we find $\dfrac{1}{f}$ to be integrable.

Exercise 3.1.10. Does the integrability of $|f(x)|$ imply the integrability of $f(x)$? What about $f(x)^2$? What about $f(x)^3$?

Exercise 3.1.11. Suppose $\phi(x)$ satisfies $A(x' - x) < \phi(x') - \phi(x) < B(x' - x)$ for some constants $A, B > 0$ and all $a \leq x < x' \leq b$.

1. Prove that $\omega_{[x, x']}(f \circ \phi) = \omega_{[\phi(x), \phi(x')]}(f)$.

2. Prove that if $f(y)$ is integrable on $[\phi(a), \phi(b)]$, then $f(\phi(x))$ is integrable on $[a, b]$.

Moreover, prove that if $\phi(x)$ is continuous on $[a, b]$ and differentiable on (a, b), satisfying $A < \phi'(x) < B$ for all $x \in (a, b)$, then $A(x'-x) < \phi(x')-\phi(x) < B(x'-x)$ for some constants $A, B > 0$ and all $a \leq x < x' \leq b$.

3.1.4   Properties of Integration

Being defined as a certain type of limit, the integration has similar properties as the limit. The properties $\lim_{n \to \infty}(x_n + y_n) = \lim_{n \to \infty} x_n + \lim_{n \to \infty} y_n$ and $\lim_{n \to \infty} c x_n = c \lim_{n \to \infty} x_n$ have the following analogues.

Proposition 3.1.7. *Suppose $f(x)$ and $g(x)$ are integrable on $[a, b]$. Then $f(x) + g(x)$ and $cf(x)$ are also integrable on $[a, b]$, and*

$$\int_a^b (f(x) + g(x))dx = \int_a^b f(x)dx + \int_a^b g(x)dx, \qquad (3.1.7)$$

$$\int_a^b cf(x)dx = c\int_a^b f(x)dx. \qquad (3.1.8)$$

Proof. Denote $I = \displaystyle\int_a^b f(x)dx$ and $J = \displaystyle\int_a^b g(x)dx$. For any $\epsilon > 0$, there is $\delta > 0$, such that

$$\|P\| < \delta \implies |S(P, f) - I| < \epsilon, \quad |S(P, g) - J| < \epsilon.$$

On the other hand, by choosing the same partition P and the same x_i^* for both $f(x)$ and $g(x)$, we have

$$S(P, f + g) = \sum (f(x_i^*) + g(x_i^*))\Delta x_i$$
$$= \sum f(x_i^*)\Delta x_i + \sum g(x_i^*)\Delta x_i = S(P, f) + S(P, g).$$

Thus we conclude

$$\|P\| < \delta \implies |S(P, f + g) - I - J| \le |S(P, f) - I| + |S(P, g) - J| < 2\epsilon.$$

This shows that $f + g$ is also integrable, and $\displaystyle\int_a^b (f(x) + g(x))dx = I + J$.

The proof of $\displaystyle\int_a^b cf(x)dx = cI$ is similar.

$\square$

Example 3.1.10. Let $L(x) = A + Bx$ be the linear function satisfying $L(a) = h$ and $L(b) = k$. Then $k + l = 2A + B(a + b)$. By Examples 3.1.2 and 3.1.8, we have

$$\int_a^b L(x)dx = A\int_a^b dx + B\int_a^b xdx = A(b - a) + \frac{B}{2}(b^2 - a^2) = \frac{1}{2}(b - a)(k + l).$$

This is indeed the area of the trapezoid under the line.

Example 3.1.11. Suppose $f(x)$ and $g(x)$ are integrable. By Proposition 3.1.7, $f(x) + g(x)$ is integrable. By Proposition 3.1.6, $f(x)^2$, $g(x)^2$ and $(f(x) + g(x))^2$ are also integrable. Then by Proposition 3.1.7 again, the product

$$f(x)g(x) = \frac{1}{2}\left[(f(x) + g(x))^2 - f(x)^2 - g(x)^2\right]$$

is also integrable. However, there is no formula expressing the integral of $f(x)g(x)$ in terms of the integrals of $f(x)$ and $g(x)$.

Moreover, if $|g(x)| > c > 0$ for a constant c, then by the discussion in Example 3.1.9, the quotient $\dfrac{f(x)}{g(x)} = f(x)\dfrac{1}{g(x)}$ is integrable.

Example 3.1.12. Suppose $f(x)$ is integrable on $[a, b]$. Suppose $g(x) = f(x)$ for all x except at $c \in [a, b]$. Then $g(x) = f(x) + \lambda d_c(x)$, where $d_c(x)$ is the function in Example 3.1.3 and $\lambda = g(c) - f(c)$. In particular, by the computation in Example 3.1.3, $g(x)$ is also integrable and

$$\int_a^b g(x)dx = \int_a^b f(x)dx + \lambda \int_a^b d_c(x)dx = \int_a^b f(x)dx.$$

The example shows that changing an integrable function at finitely many places does not change the integrability and the integral. In particular, it makes sense to talk about the integrability of a function $f(x)$ on a bounded open interval (a, b) because any numbers may be assigned as $f(a)$ and $f(b)$ without affecting the integrability of (the extended) $f(x)$ on $[a, b]$.

Example 3.1.13. In Example 3.1.5, we showed that the Thomae's function $R(x)$ in Example 1.4.2 is integrable. By Example 3.1.12, we also know that the function

$$f(x) = \begin{cases} 1 & \text{if } 0 < x \le 1 \\ 0 & \text{if } x = 0 \end{cases}$$

is integrable. However, the composition $f(R(x))$ is the Dirichlet function, which is not integrable by Example 3.1.4.

Exercise 3.1.12. Prove that if $f(x)$ and $g(x)$ are integrable, then $\max\{f(x), g(x)\}$ and $\min\{f(x), g(x)\}$ are also integrable.

Exercise 3.1.13. Find a function $f(x)$ on $[0, 1]$ that is never zero and is integrable, such that $\dfrac{1}{f(x)}$ is not integrable.

Proposition 3.1.8. *Suppose $f(x)$ and $g(x)$ are integrable on $[a, b]$. If $f(x) \le g(x)$, then*

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

Proof. Denote $I = \int_a^b f(x)dx$ and $J = \int_a^b g(x)dx$. For any $\epsilon > 0$, there is $\delta > 0$, such that

$$\|P\| < \delta \implies |S(P, f) - I| < \epsilon, \ \ |S(P, g) - J| < \epsilon.$$

By choosing the same x_i^* for both functions, we have

$$I - \epsilon < S(P, f) = \sum f(x_i^*)\Delta x_i \le \sum g(x_i^*)\Delta x_i = S(P, g) < J + \epsilon.$$

Therefore we have the inequality $I - \epsilon < J + \epsilon$ for any $\epsilon > 0$. This implies $I \le J$.

$\square$

Example 3.1.14. If $f(x)$ is integrable, then $|f(x)|$ is integrable, and $-|f(x)| \le f(x) \le |f(x)|$. By Proposition 3.1.8, we have $-\int_a^b |f(x)|dx \le \int_a^b f(x)dx \le \int_a^b |f(x)|dx$. This is the same as

$$\left| \int_a^b f(x)dx \right| \le \int_a^b |f(x)|dx. \tag{3.1.9}$$

Exercise 3.1.14. Suppose $f(x)$ is integrable on $[a, b]$. Prove that

$$(b - a) \inf_{[a,b]} f \le \int_a^b f(x)dx \le (b - a) \sup_{[a,b]} f. \tag{3.1.10}$$

Exercise 3.1.15. Suppose $f(x)$ is continuous on $[a, b]$. Prove that there is $a < c < b$, such that

$$\int_a^b f(x)dx = f(c)(b - a). \tag{3.1.11}$$

More generally, prove the *first integral mean value theorem*: For any non-negative integrable function $g(x)$ on $[a, b]$, there is $a < c < b$, such that

$$\int_a^b g(x)f(x)dx = f(c) \int_a^b g(x)dx. \tag{3.1.12}$$

Exercise 3.1.16. Suppose $f(x)$ is integrable on $[a, b]$. Prove that

$$\left| f(c)(b-a) - \int_a^b f(x)dx \right| \le \omega_{[a,b]}(f)(b-a) \tag{3.1.13}$$

for any $a \le c \le b$. Moreover, prove that

$$\left| S(P, f) - \int_a^b f(x)dx \right| \le \sum \omega_{[x_{i-1}, x_i]}(f)\Delta x_i. \tag{3.1.14}$$

Exercise 3.1.17. Suppose $f(x)$ is a continuous function on an open interval containing $[a, b]$. Prove that $\lim_{t \to 0} \int_a^b |f(x+t) - f(x)|dx = 0$. We will see in Exercise 3.1.46 that the continuity assumption is not needed.

If a region is divided into non-overlapping parts, then the whole area is the sum of the areas of the parts. The following result reflects the intuition.

Proposition 3.1.9. *Suppose $f(x)$ is a function on $[a, c]$ and $a < b < c$. Then $f(x)$ is integrable on $[a, c]$ if and only if its restrictions on $[a, b]$ and $[b, c]$ are integrable. Moreover,*

$$\int_a^c f(x)dx = \int_a^b f(x)dx + \int_b^c f(x)dx. \tag{3.1.15}$$

Proof. The proof is based on the study of the relation of the Riemann sums of the function on $[a, b]$, $[b, c]$ and $[a, c]$. Let P be a partition of $[a, c]$.

If P contains b as a partition point, then P is obtained by combining a partition P' of $[a, b]$ and a partition P'' of $[b, c]$ together. For any choice of x_i^* for P and the same choice for P' and P'', we have $S(P, f) = S(P', f) + S(P'', f)$.

If P does not contain b as a partition point, then $x_{k-1} < b < x_k$ for some k, and the new partition $\tilde{P} = P \cup \{b\}$ is still obtained by combining a partition P' of $[a, b]$ and a partition P'' of $[b, c]$ together. For any choice of x_i^* for P, we keep all x_i^* with $i \ne k$ and introduce $x_{k-1} < x'^*_k < b$, $b < x''^*_k < x_k$ for $\tilde{P}$. Then $S(\tilde{P}, f) = S(P', f) + S(P'', f)$ as before, and

$$|S(P, f) - S(P', f) - S(P'', f)| = |S(P, f) - S(\tilde{P}, f)|$$
$$= |f(x_k^*)(x_k - x_{k-1}) - f(x'^*_k)(b - x_{k-1}) - f(x''^*_k)(x_k - b)|$$
$$\le 2 \sup_{[x_{k-1}, x_k]} |f| \, \|P\|.$$

Suppose f is integrable on $[a, b]$ and $[b, c]$. Then by Proposition 3.1.2, $f(x)$ is bounded on the two intervals. Thus $|f(x)| < B$ for some constant B and all $x \in [a, c]$. Denote $I = \int_a^b f(x)dx$ and $J = \int_b^c f(x)dx$. For any $\epsilon > 0$, there is

$\delta > 0$, such that for partitions P' of $[a, b]$ and P'' of $[b, c]$ satisfying $\|P'\| < \delta$ and $\|P''\| < \delta$, we have $|S(P', f) - I| < \epsilon$ and $|S(P'', f) - J| < \epsilon$. Then for any partition P of $[a, c]$ satisfying $\|P\| < \delta$, we always have (regardless we are in the first or the second case above)

$$
\begin{aligned}
&|S(P, f) - I - J| \\
\leq &|S(P, f) - S(P', f) - S(P'', f)| + |S(P', f) - I| + |S(P'', f) - I| \\
< &2\delta B + 2\epsilon.
\end{aligned}
$$

This implies that $f(x)$ is integrable on $[a, c]$, with $\displaystyle\int_a^c f(x)dx = I + J$.

It remains to show that the integrability of $f(x)$ on $[a, c]$ implies the integrability of f on $[a, b]$ and $[b, c]$. By Cauchy criterion, for any $\epsilon > 0$, there is $\delta > 0$, such that for partitions P and Q of $[a, c]$ satisfying $\|P\| < \delta$ and $\|Q\| < \delta$, we have $|S(P, f) - S(Q, f)| < \epsilon$. Now suppose P' and Q' are partitions of $[a, b]$ satisfying $\|P'\| < \delta$ and $\|Q'\| < \delta$. Let R be any partition of $[b, c]$ satisfying $\|R\| < \delta$. By adding R to P' and Q', we get partitions P and Q of $[a, b]$ satisfying $\|P\| < \delta$ and $\|Q\| < \delta$. Moreover, the choices of x_i^* (which may be different for P' and Q') may be extended by adding the same x_i^* for P''. Then we get

$$
\begin{aligned}
|S(P', f) - S(Q', f)| &= |(S(P', f) + S(R, f)) - (S(Q', f) + S(R, f))| \\
&= |S(P, f) - S(Q, f)| < \epsilon.
\end{aligned}
$$

This proves the integrability for f on $[a, b]$. The proof of the integrability on $[b, c]$ is similar.

$\square$

Example 3.1.15. A function $f(x)$ on $[a, b]$ is a *step function* if there is a partition P and constants c_i, such that $f(x) = c_i$ for $x_{i-1} < x < x_i$ (it does not matter what $f(x_i)$ are). Then

$$
\begin{aligned}
\int_a^b f(x)dx &= \sum \int_{x_{i-1}}^{x_i} f(x)dx && \text{(Theorem 3.1.9)} \\
&= \sum \int_{x_{i-1}}^{x_i} c_i dx && \text{(Example 3.1.12)} \\
&= \sum c_i(x_i - x_{i-1}). && \text{(Example 3.1.2)}
\end{aligned}
$$

Example 3.1.16. Combining Example 3.1.10 with Proposition 3.1.9, we know the integral of a piecewise linear function is equal to the geometrical area of the region under the graph.

Exercise 3.1.18. Suppose $f(x)$ is a continuous function on $[a, b]$. Prove that the following are equivalent.

1. $f(x) = 0$ for all x.

2. $\displaystyle\int_a^b |f(x)|dx = 0$.

3. $\int_c^d f(x)dx = 0$ for any $[c,d] \subset [a,b]$.

4. $\int_a^b f(x)g(x)dx = 0$ for any continuous function $g(x)$.

Exercise 3.1.19. Suppose $f(x)$ is a continuous function on $[a,b]$. Prove that $\left| \int_a^b f(x)dx \right| = \int_a^b |f(x)|dx$ if and only if $f(x)$ does not change sign.

Exercise 3.1.20. Suppose $f(x)$ is a continuous function on $[a,b]$. Prove that if $\int_a^b f(x)dx = \int_a^b xf(x)dx = 0$, then there are at least two distinct points $a < c_1, c_2 < b$, such that $f(c_1) = f(c_2) = 0$. Extend the result to more points.

Exercise 3.1.21. Suppose $f(x) \geq 0$ is a concave function on $[a,b]$. Then for any $y \in [a,b]$, $f(x)$ is bigger than the function obtained by connecting straight lines from $(a,0)$ to $(y, f(y))$ and then to $(b,0)$. Use this to prove that $f(y) \leq \dfrac{2}{b-a} \int_a^b f(x)dx$. Moreover, determine when the equality holds.

Exercise 3.1.22. Suppose $f(x)$ is continuous on $[a,b]$ and differentiable on (a,b). Suppose $m \leq f' \leq M$ on (a,b) and denote $\mu = \dfrac{f(b) - f(a)}{b-a}$. By comparing $f(x)$ with suitable piecewise linear functions, prove that

$$\left| \int_a^b f(x)dx - \frac{f(a) + f(b)}{2}(b-a) \right| \leq \frac{(M-\mu)(\mu - m)}{2(M-m)}(b-a)^2. \qquad (3.1.16)$$

Exercise 3.1.23. Suppose $f(x)$ is a non-negative and strictly increasing function on $[a,b]$.

1. Prove that if $f(b) \leq 1$, then $\lim_{n \to \infty} \int_a^b f(x)^n dx = 0$.

2. Prove that if $f(b) > 1$ and $f(x)$ is continuous at b, then $\lim_{n \to \infty} \int_a^b f(x)^n dx = +\infty$.

Extend the result to $\lim_{n \to \infty} \int_a^b f(x)^n g(x)dx$, where $g(x)$ is non-negative and integrable on $[a,b]$.

Exercise 3.1.24. Suppose $f(x)$ is continuous on $[a,b]$ and $f(x) > 0$ on (a,b). Suppose $g(x)$ is integrable on $[a,b]$. Prove that $\lim_{n \to \infty} \int_a^b g(x) \sqrt[n]{f(x)}dx = \int_a^b g(x)dx$.

Exercise 3.1.25. Suppose $f(x) \geq 0$ is continuous on $[a,b]$. Prove that

$$\lim_{p \to +\infty} \left(\int_a^b f(x)^p dx \right)^{\frac{1}{p}} = \max_{[a,b]} f(x).$$

Exercise 3.1.26. Suppose $f(x)$ satisfies the Lipschitz condition $|f(x) - f(x')| < L|x - x'|$ on $[a,b]$. Prove that for any partition P and any choice of x_i^*, we have

$$\left| S(P, f) - \int_a^b f(x)dx \right| \leq \frac{L}{2} \sum \Delta x_i^2.$$

This gives an estimate of how close the Riemann sum is to the actual integral.

Exercise 3.1.27. Suppose $g(x)$ is integrable on $[a, b]$. Prove that for any partition P of $[a, b]$ and choice of x_i^*, we have

$$\left| \sum f(x_i^*) \int_{x_{i-1}}^{x_i} g(x)dx - \int_a^b f(x)g(x)dx \right| \le \sup_{[a,b]} |g| \sum \omega_{[x_{i-1}, x_i]}(f)\Delta x_i.$$

In particular, if $f(x)$ is integrable, then

$$\lim_{\|P\| \to 0} \sum f(x_i^*) \int_{x_{i-1}}^{x_i} g(x)dx = \int_a^b f(x)g(x)dx.$$

The definition of the Riemann integral $\int_a^b f(x)dx$ implicitly assumes $a < b$. If $a > b$, then we also define

$$\int_a^b f(x)dx = - \int_b^a f(x)dx. \tag{3.1.17}$$

Moreover, we define $\int_a^a f(x)dx = 0$ (which can be considered as a special case of the original definition of the Riemann integral). Then the equality

$$\int_a^c f(x)dx = \int_a^b f(x)dx + \int_b^c f(x)dx$$

still holds for any order between a, b, c. Proposition 3.1.7 still holds for $a \ge b$, and the direction of the inequality in Proposition 3.1.8 needs to be reversed for $a \ge b$.

3.1.5 Additional Exercise

Modified Riemann Sum and Riemann Product

Exercise 3.1.28. Let $\phi(t)$ be a function defined near 0. For any partition P of $[a, b]$ and choice of x_i^*, define the "modified Riemann sum"

$$S_\phi(P, f) = \sum_{i=1}^n \phi(f(x_i^*)\Delta x_i).$$

Prove that if ϕ is differentiable at 0, such that $\phi(0) = 0$ and $\phi'(0) = 1$, and $f(x)$ is integrable on $[a, b]$, then $\lim_{\|P\| \to 0} S_\phi(P, f) = \int_a^b f(x)dx$.

Exercise 3.1.29. For any partition P of $[a, b]$ and choice of x_i^*, define the "Riemann product"

$$\Pi(P, f) = (1 + f(x_1^*)\Delta x_1)(1 + f(x_2^*)\Delta x_2) \cdots (1 + f(x_n^*)\Delta x_n).$$

Prove that if $f(x)$ is integrable on $[a, b]$, then $\lim_{\|P\| \to 0} \Pi(P, f) = e^{\int_a^b f(x)dx}$.

Integrability and Continuity

Riemann integrable functions are not necessarily continuous. How much discontinuity can Riemann integrable functions have?

Exercise 3.1.30. Suppose $f(x)$ is integrable on $[a, b]$. Prove that for any $\epsilon > 0$, there is $\delta > 0$, such that for any partition P satisfying $\|P\| < \delta$, we have $\omega_{[x_{i-1}, x_i]}(f) < \epsilon$ for some interval $[x_{i-1}, x_i]$ in the partition.

Exercise 3.1.31. Suppose there is a sequence of intervals $[a, b] \supset [a_1, b_1] \supset [a_2, b_2] \supset \cdots$, such that $a_n < c < b_n$ for all n and $\lim_{n \to \infty} \omega_{[a_n, b_n]}(f) = 0$. Prove that $f(x)$ is continuous at c.

Exercise 3.1.32. Prove that an integrable function must be continuous somewhere. In fact, prove that for any $(c, d) \subset [a, b]$, an integrable function on $[a, b]$ is continuous somewhere in (c, d). In other words, the continuous points of the integrable function must be dense.

Exercise 3.1.33. Define the *oscillation*

$$\omega(x) = \lim_{\delta \to 0^+} \omega_{[x - \delta, x + \delta]}(f) = \overline{\lim_{y \to x}} f(x) - \underline{\lim_{y \to x}} f(x)$$

of a function at a point (see the definition before Exercise 1.3.40 for the upper and lower limit of functions). Prove that $f(x)$ is continuous at x_0 if and only if $\omega(x_0) = 0$.

Exercise 3.1.34. Prove that if $f(x)$ is integrable, then for any $\epsilon > 0$ and $\delta > 0$, there is a union U of finitely many intervals, such that the sum of the lengths of the intervals in U is $< \epsilon$, and $\omega(x) \geq \delta$ implies $x \in U$.

Exercise 3.1.35 (Hankel[2]). Prove that if $f(x)$ is integrable, then for any $\epsilon > 0$, there is a union U of countably many intervals, such that the sum of the lengths of the intervals in U is $< \epsilon$, and all discontinuous points of $f(x)$ lie inside U. This basically says that the set of discontinuous points of a Riemann integrable function has Lebesgue measure 0. The converse is also true.

Strict Inequality in Integration

The existence of the continuous points for integrable functions (see Exercise 3.1.32) enables us to change the inequalities in Proposition 3.1.8 to become strict.

Exercise 3.1.36. Prove that if $f(x) > 0$ is integrable on $[a, b]$, then $\int_a^b f(x) dx > 0$. In particular, this shows

$$f(x) < g(x) \implies \int_a^b f(x) dx < \int_a^b g(x) dx.$$

Exercise 3.1.37. Suppose $f(x)$ is an integrable function on $[a, b]$. Prove that the following are equivalent.

1. $\int_c^d f(x) dx = 0$ for any $[c, d] \subset [a, b]$.

2. $\int_a^b |f(x)| dx = 0$.

---

[2]Hermann Hankel, born 1839 in Halle (Germany), died 1873 in Schramberg (Germany). Hankel was Riemann's student, and his study of Riemann's integral prepared for the discovery of Lebesgue integral.

3. $\int_a^b f(x)g(x)dx = 0$ for any continuous function $g(x)$.

4. $\int_a^b f(x)g(x)dx = 0$ for any integrable function $g(x)$.

5. $f(x) = 0$ at continuous points.

Refinement of Partition and Integrability Criterion

The integrability criterion in Theorem 3.1.3 requires the Riemann sum of oscillations to be small for *all* partitions P satisfying $\|P\| < \delta$. The following exercises show that it is sufficient for this to happen for just *one* partition P.

Exercise 3.1.38. Suppose a partition $Q\colon a = y_0 < y_1 < y_2 < \cdots < y_m = b$ is obtained by adding k points to a partition $P\colon a = x_0 < x_1 < x_2 < \cdots < x_n = b$. Prove that

$$\sum \omega_{[y_{j-1}, y_j]}(f)\Delta y_j \leq \sum \omega_{[x_{i-1}, x_i]}(f)\Delta x_i$$

and

$$\sum \omega_{[x_{i-1}, x_i]}(f)\Delta x_i \leq \sum \omega_{[y_{j-1}, y_j]}(f)\Delta y_j + 2k\|P\|\omega_{[a,b]}(f).$$

Exercise 3.1.39. Prove that a function $f(x)$ is integrable if and only if for any $\epsilon > 0$, there is a partition P, such that $\sum \omega_{[x_{i-1}, x_i]}(f)\Delta x_i < \epsilon$.

Darboux Sum and Darboux Integral

For a function $f(x)$ on $[a, b]$ and a partition P of $[a, b]$, the upper and lower *Darboux sums* are

$$U(P, f) = \sup_{\text{all } x_i^*} S(P, f) = \sum_{i=1}^n \sup_{[x_{i-1}, x_i]} f(x)\Delta x_i, \qquad (3.1.18)$$

$$L(P, f) = \inf_{\text{all } x_i^*} S(P, f) = \sum_{i=1}^n \inf_{[x_{i-1}, x_i]} f(x)\Delta x_i. \qquad (3.1.19)$$

For bounded $f(x)$, the upper and lower *Darboux integrals* are

$$\overline{\int_a^b} f(x)dx = \inf_{\text{all } P} U(P, f), \quad \underline{\int_a^b} f(x)dx = \sup_{\text{all } P} L(P, f). \qquad (3.1.20)$$

Exercise 3.1.40. Prove that if Q is a refinement of P, then $U(P, f) \geq U(Q, f) \geq L(Q, f) \geq L(P, f)$.

Exercise 3.1.41. Prove that

$$\overline{\int_a^b} f(x)dx = \lim_{\|P\| \to 0} U(P, f), \quad \underline{\int_a^b} f(x)dx = \lim_{\|P\| \to 0} L(P, f).$$

Exercise 3.1.42. Prove that

$$\overline{\int_a^b} f(x)dx \geq \underline{\int_a^b} f(x)dx,$$

and the equality holds if and only if $f(x)$ is Riemann integrable on $[a, b]$. Moreover, the Riemann integral $\int_a^b f(x)dx$ is the common value.

Exercise 3.1.43. Prove that if $f(x)$ is integrable, then

$$\int_a^b f(x)dx = \lim_{\|P\| \to 0} \sum \phi_i \Delta x_i,$$

where ϕ_i is any number satisfying $\inf_{[x_{i-1}, x_i]} f(x) \le \phi_i \le \sup_{[x_{i-1}, x_i]} f(x)$.

Exercise 3.1.44. Study the properties in Section 3.1.4 for the Darboux integral.

Integral Continuity

The equality $\lim_{t \to 0} |f(x+t) - f(x)| = 0$ means the continuity of f at t. Therefore the equality $\lim_{t \to 0} \int_a^b |f(x+t) - f(x)|dx = 0$ means the "integral continuity" of f on $[a, b]$. Exercise 3.1.17 says that continuity implies integral continuity. The following exercises show that the continuity of f is unnecessary for the integral continuity.

Exercise 3.1.45. Suppose f is integrable on an open interval containing $[a, b]$. Suppose P be a partition of $[a, b]$ by intervals of equal length δ. Prove that if $|t| < \delta$, then

$$\int_{x_{i-1}}^{x_i} |f(x+t) - f(x)|dx \le \delta(\omega_{[x_{i-2}, x_{i-1}]}(f) + \omega_{[x_{i-1}, x_i]}(f) + \omega_{[x_i, x_{i+1}]}(f)).$$

Exercise 3.1.46. Suppose f is integrable on an open interval containing $[a, b]$. Prove that $\lim_{t \to 0} \int_a^b |f(x+t) - f(x)|dx = 0$ and $\lim_{t \to 1} \int_a^b |f(tx) - f(x)|dx = 0$.

Integral Inequalities for Convex Functions

Exercise 3.1.47. Suppose $f(x)$ is a convex function on $[a, b]$. By comparing $f(x)$ with linear functions in Figure 2.6, prove that

$$f\left(\frac{a+b}{2}\right)(b-a) \le \int_a^b f(x)dx \le \frac{f(a) + f(b)}{2}(b-a).$$

Exercise 3.1.48. A *weight* on $[a, b]$ is a function $\lambda(x)$ satisfying

$$\lambda(x) \ge 0, \quad \frac{1}{b-a} \int_a^b \lambda(x) = 1. \tag{3.1.21}$$

We have $\frac{1}{b-a} \int_a^b \lambda(x)x\,dx = (1-\mu)a + \mu b$ for some $0 < \mu < 1$. For a convex function on $[a, b]$, prove that

$$f((1-\mu)a + \mu b) \le \frac{1}{b-a} \int_a^b \lambda(x)f(x)dx \le (1-\mu)f(a) + \mu f(b).$$

The left inequality is the *integral version* of the Jensen inequality in Exercise 2.3.40. What do you get by applying the integral Jensen inequality to x^2, e^x and $\log x$?

Exercise 3.1.49. Suppose $f(x)$ is a convex function on $[a, b]$ and $\phi(t)$ is an integrable function on $[\alpha, \beta]$ satisfying $a \le \phi(t) \le b$. Suppose $\lambda(x)$ is a weight function on $[\alpha, \beta]$ as defined in Exercise 3.1.48. Prove that

$$f\left(\frac{1}{\beta - \alpha} \int_\alpha^\beta \lambda(t)\phi(t)dt\right) \le \frac{1}{\beta - \alpha} \int_\alpha^\beta \lambda(t)f(\phi(t))dt.$$

This further extends the integral Jensen inequality.

Hölder and Minkowski Inequalities

Exercise 3.1.50. Suppose $p, q > 0$ satisfy $\dfrac{1}{p} + \dfrac{1}{q} = 1$. Prove the integral versions of the Hölder and Minkowski inequalities in Exercises 2.2.41 and 2.2.42.

$$\int_a^b |f(x)g(x)|dx \leq \left(\int_a^b |f(x)|^p dx\right)^{\frac{1}{p}} \left(\int_a^b |g(x)|^q dx\right)^{\frac{1}{q}}, \qquad (3.1.22)$$

$$\left(\int_a^b |f(x) + g(x)|^p dx\right)^{\frac{1}{p}} \leq \left(\int_a^b |f(x)|^p dx\right)^{\frac{1}{p}} + \left(\int_a^b |g(x)|^p dx\right)^{\frac{1}{p}}. \qquad (3.1.23)$$

Estimation of Integral

Exercise 3.1.51. Suppose $f(x)$ is continuous on $[a, b]$ and differentiable on (a, b). By comparing $f(x)$ with the straight line $L(x) = f(a) + m(x - a)$ for $m = \sup_{(a,b)} f'$ or $m = \inf_{(a,b)} f'$, prove that

$$\frac{\inf_{(a,b)} f'}{2}(b - a)^2 \leq \int_a^b f(x)dx - f(a)(b - a) \leq \frac{\sup_{(a,b)} f'}{2}(b - a)^2. \qquad (3.1.24)$$

Then use Darboux's intermediate value theorem in Exercise 2.2.33 to show

$$\int_a^b f(x)dx = f(a)(b - a) + \frac{f'(c)}{2}(b - a)^2 \qquad (3.1.25)$$

for some $a < c < b$.

Exercise 3.1.52. Suppose $f(x)$ is continuous on $[a, b]$ and differentiable on (a, b). Suppose $g(x)$ is non-negative and integrable on $[a, b]$. Prove that

$$\int_a^b f(x)g(x)dx = f(a)\int_a^b g(x)dx + f'(c)\int_a^b (x - a)g(x)dx \qquad (3.1.26)$$

for some $a < c < b$.

Exercise 3.1.53. Suppose $f(x)$ is continuous on $[a, b]$ and differentiable on (a, b). Use Exercise 3.1.52 to prove that

$$\left|\int_a^b f(x)dx - \frac{f(a) + f(b)}{2}(b - a)\right| \leq \frac{\omega_{(a,b)}(f')}{8}(b - a)^2. \qquad (3.1.27)$$

In fact, this estimation can also be derived from Exercise 3.1.22.

Exercise 3.1.54. Suppose $f(x)$ is continuous on $[a, b]$ and has second order derivative on (a, b). Use the Taylor expansion at $\dfrac{a + b}{2}$ to prove that

$$\int_a^b f(x)dx = f\left(\frac{a + b}{2}\right)(b - a) + \frac{f''(c)}{24}(b - a)^3 \qquad (3.1.28)$$

for some $a < c < b$.

3.2   Antiderivative

Riemann's definition of integration is independent of the differentiation. Before Riemann, however, Newton and Leibniz considered the integration as the inverse of the differentiation. The relation enables us to compute the integrals.

3.2.1 Fundamental Theorem of Calculus

Let us consider the integral of $f(x)$ as a function by changing the upper limit. The result is a function with $f(x)$ as the derivative.

Theorem 3.2.1. *Suppose $f(x)$ is integrable. Then $F(x) = \int_a^x f(t)dt$ is a continuous function. Moreover, if $f(x)$ is continuous at x_0, then $F(x)$ is differentiable at x_0, with $F'(x_0) = f(x_0)$.*

Proof. By Proposition 3.1.2, we have $|f(x)| < B$ for a constant B and all x. Then by Propositions 3.1.8 (see also (3.1.9)) and Proposition 3.1.9, we have

$$|F(x) - F(x_0)| = \left| \int_{x_0}^x f(t)dt \right| \le B|x - x_0|.$$

This implies $\lim_{x \to x_0} F(x) = F(x_0)$.

Now further assume that $f(x)$ is continuous at x_0. For any $\epsilon < 0$, there is $\delta > 0$, such that $|x - x_0| < \delta$ implies $|f(x) - f(x_0)| < \epsilon$. Then

$$|F(x) - F(x_0) - f(x_0)(x - x_0)| = \left| \int_{x_0}^x f(t)dt - f(x_0)(x - x_0) \right|$$
$$= \left| \int_{x_0}^x (f(t) - f(x_0))dt \right| \le \epsilon|x - x_0|.$$

This means that $F(x_0) + f(x_0)(x - x_0)$ is the linear approximation of $F(x)$ at x_0, with $F'(x_0) = f(x_0)$.

$\square$

The theorem suggests that, to compute the integral of a continuous function $f(x)$, we may consider a function $\phi(x)$ satisfying $\phi'(x) = f(x)$. Although the functions $\phi(x)$ and $F(x)$ may not be equal, the property $\phi'(x) = f(x) = F'(x)$ implies, by Proposition 2.2.3, that $\phi(x) = F(x) + C$ for a constant C. Then

$$\int_a^b f(x)dx = F(b) = F(b) - F(a) = \phi(b) - \phi(a).$$

For the obvious reason, the function $\phi(x)$ is called an *antiderivative* of $f(x)$. It is unique up to adding a constant.

Example 3.2.1. For the sign function $\text{sign}(x)$ in Example 1.4.1, we have

$$\int_0^x \text{sign}(t)dt = \begin{cases} \int_0^x 1dx = x & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ \int_0^x -1dx = -x & \text{if } x < 0 \end{cases}$$

by considering the antiderivative of 1 for $x > 0$ and the antiderivative of -1 for $x < 0$. Thus $\int_0^x \text{sign}(t)dt = |x|$ is continuous and is differentiable except at $x = 0$, which is the place where $\text{sign}(t)$ is not continuous.

Example 3.2.2. The functions c, x, x^2 are continuous, with cx, $\dfrac{1}{2}x^2$, $\dfrac{1}{3}x^3$ as antiderivatives. Therefore

$$\int_a^b c\,dx = cb - ca, \quad \int_a^b x\,dx = \frac{1}{2}b^2 - \frac{1}{2}a^2, \quad \int_a^b x^2\,dx = \frac{1}{3}b^3 - \frac{1}{3}a^3.$$

Example 3.2.3. The functions $\cos^2 x = \dfrac{1}{2}(1 + \cos 2x)$, $\tan^2 x = \sec^2 x - 1$ are continuous, with $\dfrac{1}{4}(2x + \sin 2x)$, $\tan x - x$ as antiderivatives. Therefore

$$\int_0^{\frac{\pi}{2}} \cos^2 x\,dx = \frac{1}{4}\left(2\frac{\pi}{2} + \sin\pi\right) - \frac{1}{4}\left(2\cdot 0 + \sin 0\right) = \frac{\pi}{4},$$
$$\int_0^{\frac{\pi}{4}} \tan^2 x\,dx = \left(\tan\frac{\pi}{4} - \frac{\pi}{4}\right) - (\tan 0 - 0) = 1 - \frac{\pi}{4}.$$

Example 3.2.4. The function $\dfrac{1}{x}$ is integrable on $[1,2]$. Consider the partition $P: 1 < \dfrac{n+1}{n} < \dfrac{n+2}{n} < \cdots < \dfrac{2n}{n} = 2$ and take the right ends $\dfrac{n+1}{n}, \dfrac{n+2}{n}, \ldots, \dfrac{2n}{n}$ as x_i^*. Then the Riemann sum

$$S\left(P, \frac{1}{x}\right) = \frac{1}{\frac{n+1}{n}}\frac{1}{n} + \frac{1}{\frac{n+2}{n}}\frac{1}{n} + \cdots + \frac{1}{\frac{2n}{n}}\frac{1}{n} = \frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{2n}.$$

Thus

$$\lim_{n\to\infty}\left(\frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{2n}\right) = \lim_{\|P\|\to 0} S\left(P, \frac{1}{x}\right) = \int_1^2 \frac{1}{x}\,dx = \log 2,$$

where the last equality is obtained from the fact that $\log x$ is an antiderivative of $\dfrac{1}{x}$.

Example 3.2.5. The function $g(x) = \displaystyle\int_0^{x^2} \frac{t\,dt}{1+t^3} = h(x^2)$, with $h(y) = \displaystyle\int_0^y \frac{t\,dt}{1+t^3}$. By Theorem 3.2.1 and the chain rule, we have

$$g'(x) = h'(x^2)2x = 2x\frac{x^2}{1+(x^2)^3} = \frac{2x^3}{1+x^6}.$$

By the similar idea, we have

$$\frac{d}{dx}\left(\int_{\log x}^{e^x} \frac{t\,dt}{1+t^3}\right) = \frac{d}{dx}\left(\int_0^{e^x} \frac{t\,dt}{1+t^3} - \int_0^{\log x} \frac{t\,dt}{1+t^3}\right)$$
$$= e^x\frac{e^x}{1+(e^x)^3} - \frac{1}{x}\frac{\log x}{1+(\log x)^3}$$
$$= \frac{e^{2x}}{1+e^{3x}} - \frac{\log x}{x(1+(\log x)^3)}.$$

Exercise 3.2.1. Compute the integrals.

1. $\displaystyle\int_0^b x^n dx$, $n \in \mathbb{N}$.

2. $\displaystyle\int_1^b x^\alpha dx$, $b > 0$.

3. $\displaystyle\int_0^b \sin x dx$.

4. $\displaystyle\int_0^b \cos 2x dx$.

5. $\displaystyle\int_0^b e^x dx$.

6. $\displaystyle\int_0^b 2^x dx$.

Exercise 3.2.2. Compute the derivatives.

1. $\displaystyle\int_0^x \sin t^2 dt$.

2. $\displaystyle\int_0^{x^2} \sin t dt$.

3. $\displaystyle\int_0^{\sin x} t^2 dt$.

4. $\displaystyle\int_{\sin x}^x \sin t^2 dt$.

5. $\displaystyle\int_0^{|x|} \sin t^2 dt$.

6. $\displaystyle\int_0^{\sin x} |t| dt$.

Exercise 3.2.3. For non-negative integers m and n, prove that

$$\int_0^{2\pi} \cos mx \sin nx dx = 0,$$

$$\int_0^{2\pi} \cos mx \cos nx dx = \begin{cases} 0 & \text{if } m \neq n \\ \pi & \text{if } m = n \neq 0 \\ 2\pi & \text{if } m = n = 0 \end{cases}, \tag{3.2.1}$$

$$\int_0^{2\pi} \sin mx \sin nx dx = \begin{cases} 0 & \text{if } m \neq n \text{ or } m = n = 0 \\ \pi & \text{if } m = n \neq 0 \end{cases}.$$

Exercise 3.2.4. Compute the limits by relating to the Riemann sums of suitable integrals.

1. $\displaystyle\lim_{n \to \infty} \frac{1^2 \cdot 1 + 2^2 \cdot 3 + \cdots + n^2 \cdot (2n - 1)}{n^4}$.

2. $\displaystyle\lim_{n \to \infty} \frac{1}{n} \left(\cos \frac{\pi}{n} + \cos \frac{2\pi}{n} + \cdots + \cos \frac{(n-1)\pi}{n} \right)$.

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

4. $\displaystyle\lim_{n \to \infty} \frac{1^\alpha + 3^\alpha + \cdots + (2n + 1)^\alpha}{n^{\alpha+1}}$, $\alpha > 0$.

5. $\displaystyle\lim_{n \to \infty} \frac{1^\alpha + 3^\alpha + \cdots + (2n - 1)^\alpha}{2^\alpha + 4^\alpha + \cdots + (2n)^\alpha}$, $\alpha > 0$.

Exercise 3.2.5. If $f(x)$ is not continuous at x_0, is it true that $F(x) = \displaystyle\int_a^x f(t)dt$ is not differentiable at x_0?

Exercise 3.2.6. Suppose $f(x)$ is integrable and $F(x) = \displaystyle\int_a^x f(t)dt$. Prove that if $f(x)$ has left limit at x_0, then $F'(x_0^-) = f(x_0^-)$. In particular, this shows that if f has different left and right limits at x_0, then $F(x)$ is not differentiable at x_0.

Exercise 3.2.7. Suppose $f(x)$ is integrable on $[a, b]$. Prove that there is $a < c < b$ such that $\displaystyle\int_a^c f(x)dx = \frac{1}{2} \int_a^b f(x)dx$.

Exercise 3.2.8. Find continuous functions on given intervals satisfying the equalities.

1. $\displaystyle\int_0^x f(t)dt = \int_x^1 f(t)dt$ on $[0,1]$.

2. $\displaystyle A\int_0^x tf(t)dt = x\int_0^x f(t)dt$ on $(0,+\infty)$.

3. $\displaystyle (f(x))^2 = 2\int_0^x f(t)dt$ on $(-\infty,+\infty)$.

Exercise 3.2.9. Find continuous functions $f(x)$ on $(0,+\infty)$, such that for all $b > 0$, the integral $\displaystyle\int_a^{ab} f(x)dx$ is independent of $a > 0$.

Exercise 3.2.10. Suppose $f(x)$ is continuous on $[a,b]$, differentiable on (a,b), and satisfies $f(a) = 0$, $1 \ge f'(x) \ge 0$. Prove that $\displaystyle\left(\int_a^b f(x)dx\right)^2 \ge \int_a^b f(x)^3 dx$.

As a matter of fact, the continuity assumption can be weakened in order for the integral to be the antiderivative.

Theorem 3.2.2. *Suppose $f(x)$ is integrable on $[a,b]$. Suppose $F(x)$ is continuous on $[a,b]$ and is differentiable on (a,b). If $F'(x) = f(x)$, then $\displaystyle\int_a^b f(x)dx = F(b) - F(a)$.*

The theorem is almost the converse of Theorem 3.2.1. Put together, they are called the *Fundamental Theorem of Calculus.*

Proof. For a partition P of $[a,b]$, we have

$$F(b) - F(a) = \sum(F(x_i) - F(x_{i-1})) = \sum f(x_i^*)(x_i - x_{i-1}),$$

where the second equality comes from the mean value theorem and $F'(x) = f(x)$. Therefore $F(b) - F(a)$ is a Riemann sum with suitable choice of x_i^*. When $\|P\| \to 0$, by the integrability of $f(x)$, we get $F(b) - F(a) = \displaystyle\int_a^b f(x)dx$.
$\square$

Example 3.2.6. By Exercise 3.1.8 and Proposition 3.1.4, the function

$$f(x) = \begin{cases} 2x\sin\dfrac{1}{x} - \cos\dfrac{1}{x} & \text{if } x \ne 0 \\ 0 & \text{if } x = 0 \end{cases}$$

is not continuous but still integrable on any bounded interval. Moreover, the function

$$F(x) = \begin{cases} x^2\sin\dfrac{1}{x} & \text{if } x \ne 0 \\ 0 & \text{if } x = 0 \end{cases}$$

is an antiderivative of $f(x)$. Therefore Theorem 3.2.2 can be applied to give us $\displaystyle\int_0^1 f(x)dx = F(1) - F(0) = \sin 1$. Note that the discussion after Theorem 3.2.1 does not apply to the example because of the discontinuity.

Example 3.2.7. Consider $F(x) = x^2 \sin \dfrac{1}{x^2}$ for $x \neq 0$ and $F(0) = 0$. The function is differentiable with the derivative

$$F'(x) = \begin{cases} 2x \sin \dfrac{1}{x^2} - \dfrac{2}{x} \cos \dfrac{1}{x^2} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}.$$

However, $F'(x)$ is not integrable on $[0,1]$ because it is not bounded. So the existence of antiderivative does not necessarily imply the integrability.

On the other hand, we note that the limit $\lim_{a \to 0^+} \int_a^1 F'(x)dx = \lim_{a \to 0^+} (F(1) - F(a)) = \sin 1$ exists. What we have here is that $\int_0^1 F'(x)dx$ is an *improper integral* with $\sin 1$ as the value.

Exercise 3.2.11. Prove that Theorem 3.2.2 still holds if $F(x)$ is differentiable at all but finitely many points (so $F(x)$ is *piecewise differentiable*).

Exercise 3.2.12. Suppose $f(x)$ is differentiable. Prove that $f'(x)$ is integrable if and only if there is an integrable function $g(x)$, such that $f(x) = f(a) + \int_a^x g(t)dt$.

Exercise 3.2.13. Suppose $f(x)$ is differentiable and $f'(x)$ is integrable. Compute $\int_a^b [x]f'(x)dx$, where $[x]$ is defined in Exercise 1.3.5.

Exercise 3.2.14. Suppose $f(x)$ has integrable derivative on $[a,b]$.

1. Prove that if $f(x)$ vanishes somewhere on $[a,b]$, then $|f(x)| \leq \int_a^b |f'(x)|dx$.

2. Prove that $\int_a^b |f(x)|dx \leq \max \left\{ \left| \int_a^b f(x)dx \right|, (b-a) \int_a^b |f'(x)|dx \right\}$.

Exercise 3.2.15. Suppose $f(x)$ is continuous on $[a,b]$ and differentiable on (a,b), such that $f'(x)$ is integrable on $[a,b]$. Use Hölder inequality (3.1.22) in Exercise 3.1.50 to prove that

$$(f(b) - f(a))^2 \leq (b-a) \int_a^b f'(x)^2 dx. \tag{3.2.2}$$

Then prove the following.

1. If $f(a) = 0$, then $\int_a^b f(x)^2 dx \leq \dfrac{(b-a)^2}{2} \int_a^b f'(x)^2 dx$.

2. If $f(a) = f(b) = 0$, then $\int_a^b f(x)^2 dx \leq \dfrac{(b-a)^2}{4} \int_a^b f'(x)^2 dx$.

3. If $f\left(\dfrac{a+b}{2}\right) = 0$, then $\int_a^b f(x)^2 dx \leq \dfrac{(b-a)^2}{4} \int_a^b f'(x)^2 dx$.

3.2.2   Antiderivative

Because of the fundamental theorem of calculus, we use $\int f(x)dx$ to denote all the antiderivatives of $f(x)$. In other words, if $F(x)$ is a differentiable function satisfying $F'(x) = f(x)$, then

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

where C is an arbitrary constant. Here are some basic examples.

$$\int x^\alpha dx = \begin{cases} \dfrac{x^{\alpha+1}}{\alpha + 1} + C & \text{if } \alpha \neq -1 \\ \log|x| + C & \text{if } \alpha = -1 \end{cases},$$

$$\int \sin x dx = -\cos x + C,$$

$$\int \cos x dx = \sin x + C,$$

$$\int \tan x dx = -\log|\cos x| + C,$$

$$\int \sec^2 x dx = \tan x + C,$$

$$\int \sec x \tan x dx = \sec x + C,$$

$$\int e^x dx = e^x + C,$$

$$\int a^x dx = \frac{a^x}{\log a} + C,$$

$$\int \log|x| dx = x \log|x| - x + C,$$

$$\int \frac{dx}{\sqrt{1 - x^2}} = \arcsin x + C,$$

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

$$\int \frac{dx}{x\sqrt{x^2 - 1}} = \operatorname{arcsec} x + C.$$

Example 3.2.8. By $\dfrac{1}{1 - x^2} = \dfrac{1}{2}\left(\dfrac{1}{1 - x} + \dfrac{1}{1 + x}\right)$, we get

$$\int \frac{dx}{1 - x^2} = \frac{1}{2}\left(\int \frac{dx}{1 - x} + \int \frac{dx}{1 + x}\right)$$

$$= \frac{1}{2}(-\log|1 - x| + \log|1 + x|) + C = \frac{1}{2}\log\left|\frac{1 + x}{1 - x}\right| + C.$$

Example 3.2.9. By $\sin x \sin 2x = \dfrac{1}{2}(\cos 3x - \cos x)$, we get

$$\int \sin x \sin 2x dx = \frac{1}{6}\sin 3x - \frac{1}{2}\sin x + C.$$

By $\sin^2 x + \cos^2 x = 1$, we get

$$\int \frac{dx}{\sin^2 x \cos^2 x} = \int \frac{\sin^2 x + \cos^2 x}{\sin^2 x \cos^2 x} dx$$

$$= \int (\sec^2 x + \csc^2 x) dx = \tan x - \cot x + C.$$

Exercise 3.2.16. Verify the antiderivatives.

1. $\int \dfrac{dx}{\sqrt{x^2 + a}} = \log |x + \sqrt{x^2 + a}| + C.$

2. $\int \sqrt{1 - x^2} dx = \dfrac{1}{2}(\arcsin x + x\sqrt{1 - x^2}) + C.$

3. $\int e^{ax} \cos bx\, dx = e^{ax} \dfrac{a \cos bx + b \sin bx}{a^2 + b^2} + C.$

4. $\int \log |x| dx = x \log |x| - x + C.$

Exercise 3.2.17. Compute the antiderivatives.

1. $\int \dfrac{dx}{x(1 + x)}.$

2. $\int x(1 + x)^9 dx.$

3. $\int \dfrac{x^2}{(1 + x)^9} dx.$

4. $\int x^2(1 + x)^n dx.$

5. $\int \dfrac{x^2 - x + 1}{(1 + x)^n} dx.$

6. $\int (2^x + 2^{-x})^2 dx.$

7. $\int |x| dx.$

8. $\int \sin 2x \cos 3x\, dx.$

9. $\int \cos^2 x\, dx.$

10. $\int \tan^2 x\, dx.$

11. $\int \sin^3 x\, dx.$

12. $\int \dfrac{\cos 2x\, dx}{\sin^2 x \cos^2 x}.$

The formulae for derivatives can be directly translated into formulae for antiderivatives. For example, the formula $(f(x) + g(x))' = f'(x) + g'(x)$ corresponds to the formula

$$\int (f(x) + g(x)) dx = \int f(x) dx + \int g(x) dx. \qquad (3.2.3)$$

To translate the Leibniz rule, we start with

$$\int f(x) dx = F(x) + C, \quad \int g(x) dx = G(x) + C.$$

In other words, $F'(x) = f(x)$ and $G'(x) = g(x)$. Then

$$(F(x)G(x))' = F'(x)g(x) + F(x)G'(x) = f(x)G(x) + F(x)g(x).$$

In terms of the antiderivative, this means

$$\int F(x)g(x) dx = F(x)G(x) - \int f(x)G(x) dx. \qquad (3.2.4)$$

This is the formula for the *integration by parts.*

Using the differential notation, we have $dF(x) = f(x)dx$ and $dG(x) = g(x)dx$. Thus the formula can also be written as

$$\int F(x)dG(x) = F(x)G(x) - \int G(x)dF(x). \qquad (3.2.5)$$

In other words, the integration by parts is a way of exchanging the functions "inside" and "outside" the differential notation. The viewpoint suggests the usefulness of including dx in the integration (or antiderivative) notation $\int f(x)dx$. Indeed, in the more advanced mathematics, integrations are carried out for the differential forms $f(x)dx$ instead of the functions only.

Example 3.2.10. The antiderivative of the logarithmic function is

$$\int \log|x|dx = x\log|x| - \int xd\log|x| = x\log|x| - \int x\frac{1}{x}dx = x\log|x| - x + C.$$

Example 3.2.11. To find the antiderivative of $e^x \sin x$, we apply the integration by parts twice.

$$\int e^x \sin x dx = -\int e^x d\cos x$$

$$= -e^x \cos x + \int \cos x de^x$$

$$= -e^x \cos x + \int e^x \cos x dx$$

$$= -e^x \cos x + \int e^x d\sin x$$

$$= -e^x \cos x + e^x \sin x - \int \sin x de^x$$

$$= -e^x \cos x + e^x \sin x - \int e^x \sin x dx.$$

Solving for $\int e^x \sin x dx$, we get

$$\int e^x \sin x dx = \frac{1}{2}e^x(\sin x - \cos x) + C.$$

Exercise 3.2.18. Compute the antiderivatives.

1. $\int e^{ax} \sin bx dx.$ 3. $\int x^2 2^x dx.$ 5. $\int x^2 e^{-x} \cos 2x dx.$

2. $\int a^x \cos bx dx.$ 4. $\int xe^x \sin x dx.$ 6. $\int x^n \arctan x dx.$

Exercise 3.2.19. Prove the recursive relations.

1. $\int \sin^n x dx = -\frac{1}{n}\sin^{n-1} x \cos x + \frac{n-1}{n}\int \sin^{n-2} x dx.$

2. $\int \cos^n x dx = \frac{1}{n}\cos^{n-1} x \sin x + \frac{n-1}{n}\int \cos^{n-2} x dx.$

3. $\int \tan^n x dx = \dfrac{\tan^{n-1} x}{n-1} - \int \tan^{n-2} x dx.$

4. $\int \sec^n x dx = \dfrac{\sec^{n-2} x \tan x}{n-1} - \dfrac{n-2}{n-1} \int \sec^{n-2} x dx.$

5. $\int x^n e^x dx = x^n e^x - n \int x^{n-1} e^x dx.$

6. $\int x^\alpha (\log|x|)^n dx = \dfrac{1}{\alpha+1} x^{\alpha+1} (\log|x|)^n - \dfrac{n}{\alpha+1} \int x^\alpha (\log|x|)^{n-1} dx.$

7. $\int e^x \sin^n x dx = \dfrac{1}{n^2-1} e^x \sin^{n-1} x(n \cos x - \sin x) + \dfrac{n}{n+1} \int e^x \sin^{n-2} x dx.$

8. $\int x^n \sin x dx = -x^n \cos x + nx^{n-1} \sin x - n(n-1) \int x^{n-2} \sin x dx.$

9. $\int x^n \cos x dx = x^n \sin x + nx^{n-1} \cos x - n(n-1) \int x^{n-2} \cos x dx.$

10. $\int (1+ax^2)^n dx = \dfrac{x(1+ax^2)^n}{2n+1} + \dfrac{2n}{2n+1} \int (1+ax^2)^{n-1} dx.$

Then compute the antiderivatives.

1. $\int \sin^6 x dx.$

2. $\int \sin^5 x \cos^4 x dx.$

3. $\int \tan^6 x dx.$

4. $\int \tan^{-6} x dx$

5. $\int x^3(x+1)e^x dx.$

6. $\int \sqrt{x}(\log x)^2 dx.$

7. $\int e^x \sin^4 x dx.$

8. $\int x^4 \sin x dx.$

9. $\int \dfrac{dx}{(1+x^2)^2}.$

Exercise 3.2.20. Let $I(m,n) = \int \cos^m x \sin^n x dx.$ Use

$$d\sin^n x = n \sin^{n-1} x \cos x dx, \quad d\cos^n x = -n \cos^{n-1} x \sin x dx$$

to derive the recursive relations.

$$\begin{aligned} I(m,n) &= -\frac{\cos^{m+1} x \sin^{n-1} x}{m+1} + \frac{n-1}{m+1} I(m+2, n-2) && \text{if } m \neq -1 \\ &= -\frac{\cos^{m+1} x \sin^{n-1} x}{m+n} + \frac{n-1}{m+n} I(m, n-2) && \text{if } m+n \neq 0. \end{aligned}$$

Find the similar relations by exchanging the sine and cosine functions. Then compute the antiderivatives $\int \cos^4 x \sin^5 x dx, \int \dfrac{\sin^4 x}{\cos^3 x} dx.$

Exercise 3.2.21. Let $I(m,n) = \int (x-a)^m (x-b)^n dx.$ Find a recursive relation between $I(m,n)$ and $I(m-1, n+1)$. Then use the relation to compute the integral $\int_{-1}^1 (x-1)^3(x+1)^{10} dx.$

Now we translate the chain rule. Suppose $F'(y) = f(y)$, or

$$\int f(y)dy = F(y) + C.$$

Then for any differentiable function $\phi(x)$, we have

$$F(\phi(x))' = F'(\phi(x))\phi'(x) = f(\phi(x))\phi'(x).$$

Therefore $F(\phi(x))$ is the antiderivative of $f(\phi(x))\phi'(x)$, and we have

$$\int f(y)dy\bigg|_{y=\phi(x)} = F(\phi(x)) + C = \int f(\phi(x))\phi'(x)dx. \qquad (3.2.6)$$

This is the formula for the *change of variable.*

Example 3.2.12. To compute $\int \dfrac{dx}{1 + \sqrt{x}}$, we introduce $x = t^2$.

$$\int \frac{dx}{1 + \sqrt{x}} = \int \frac{d(t^2)}{1 + t} = \int \frac{2tdt}{1 + t} = \int 2\left(1 - \frac{1}{1 + t}\right)dt$$
$$= 2(t - \log|1 + t|) + C = 2(\sqrt{x} - \log|1 + \sqrt{x}|) + C.$$

Example 3.2.13. To compute the antiderivative of $\dfrac{x}{\sqrt{3 + 4x - x^2}}$, we note $5 + 4x - x^2 = 9 - (x - 2)^2$ and introduce $x - 2 = 3\sin t$.

$$\int \frac{xdx}{\sqrt{3 + 4x - x^2}} = \int \frac{(2 + 3\sin t)3d\sin t}{\sqrt{9 - 9\sin^2 t}} = \int 3(2 + 3\sin t)dt$$
$$= 3(2t - 3\cos t) + C = 6\arcsin \frac{x - 2}{3} - 3\sqrt{3 + 4x - x^2} + C.$$

In general, to compute an antiderivative of the form $\int f(x, \sqrt{ax^2 + bx + c})dx$, we may complete the square for the quadratic function $ax^2 + bx + c$. Then a suitable trigonometric function can be used to change the variable.

Example 3.2.14. We compute the antiderivative of the secant function.

$$\int \sec xdx = \int \frac{\cos xdx}{\cos^2 x} = \int \frac{d\sin x}{1 - \sin^2 x} = \frac{1}{2}\log\left|\frac{1 + \sin x}{1 - \sin x}\right| + C.$$

In the last step, the computation of Example 3.2.8 is used. Note that by

$$\frac{1 + \sin x}{1 - \sin x} = \frac{(1 + \sin x)^2}{(1 - \sin x)(1 + \sin x)} = \frac{(1 + \sin x)^2}{\cos^2 x} = (\sec x + \tan x)^2,$$

the antiderivative can also be written as

$$\int \sec xdx = \log|\sec x + \tan x| + C.$$

Example 3.2.15. The antiderivative of $\dfrac{1}{\sqrt{x^2 + 1}}$ may be computed by introducing $x = \tan t$.

$$\int \frac{dx}{\sqrt{x^2 + 1}} = \int \frac{\sec^2 tdt}{\sec t} = \int \sec tdt = \log|\sec x + \tan x| + C = \log(x + \sqrt{x^2 + 1}) + C.$$

The integration of $\sec t$ is computed in Example 3.2.14. Similarly, the antiderivative of $\dfrac{1}{\sqrt{x^2 - 1}}$ may be computed by introducing $x = \sec t$.

$$\int \frac{dx}{\sqrt{x^2 - 1}} = \int \frac{\sec t \tan t dt}{\tan t} = \int \sec t dt = \log|x + \sqrt{x^2 - 1}| + C.$$

Example 3.2.16. To compute the antiderivative of the inverse sine function, we introduce $x = \sin t$.

$$\int \arcsin x dx = \int t d \sin t = t \sin t - \int \sin t dt$$

$$= t \sin t + \cos t + C = x \arcsin x + \sqrt{1 - x^2} + C.$$

Note that the integration by parts is also used.

Exercise 3.2.22. Compute the antiderivatives.

1. $\displaystyle\int \frac{f'(x)}{f(x)^\alpha} dx.$
2. $\displaystyle\int \frac{f'(x)}{1 + f(x)^2} dx.$
3. $\displaystyle\int 2^{f(x)} f'(x) dx.$

Exercise 3.2.23. Compute the antiderivatives.

1. $\displaystyle\int \frac{dx}{\sqrt{x} + \sqrt[3]{x}}.$

2. $\displaystyle\int \frac{\sqrt{x} dx}{1 - \sqrt[3]{x}}.$

3. $\displaystyle\int (1 + \sqrt[3]{x})^{10} dx.$

4. $\displaystyle\int \frac{dx}{x^2 + 2x + 5}.$

5. $\displaystyle\int \frac{x dx}{x^2 + 2x + 5}.$

6. $\displaystyle\int \frac{x^3 dx}{x^2 + 2x + 5}.$

7. $\displaystyle\int \frac{dx}{(x^2 + 2x + 5)^2}.$

8. $\displaystyle\int \frac{x^3 dx}{(x^2 + 2x + 5)^2}.$

9. $\displaystyle\int \sqrt{x^2 + 2x + 5} dx.$

10. $\displaystyle\int \frac{dx}{\sqrt{2x - x^2}}.$

11. $\displaystyle\int x\sqrt{5 + 4x - x^2} dx.$

12. $\displaystyle\int \frac{x dx}{(x^2 + 2x + 2)^{\frac{3}{2}}}.$

13. $\displaystyle\int \frac{(2x + 1) dx}{\sqrt{x(x + 1)}}.$

14. $\displaystyle\int \frac{dx}{x \log x}.$

15. $\displaystyle\int \frac{\log x}{x} dx.$

16. $\displaystyle\int \frac{dx}{e^x + e^{-x}}.$

17. $\displaystyle\int \cot x dx.$

18. $\displaystyle\int \csc x dx.$

19. $\displaystyle\int \tan^3 x dx.$

20. $\displaystyle\int \sec^3 x dx.$

21. $\displaystyle\int \frac{\sin^4 x}{\cos^3 x} dx.$

22. $\displaystyle\int \frac{\sin^5 x}{\cos^3 x} dx.$

23. $\displaystyle\int (\arcsin x)^2 dx.$

24. $\displaystyle\int x(\arcsin x)^2 dx.$

Exercise 3.2.24. Compute the antiderivatives ($a > 0$).

1. $\displaystyle\int (ax + b)^\alpha dx.$

2. $\displaystyle\int \frac{dx}{a^2 + x^2}.$

3. $\displaystyle\int \frac{dx}{(a^2 + x^2)^{\frac{3}{2}}}.$

4. $\displaystyle\int \frac{dx}{\sqrt{a^2 - x^2}}.$

5. $\displaystyle\int \frac{dx}{\sqrt{a^2 + x^2}}.$

6. $\displaystyle\int \frac{dx}{\sqrt{x^2 - a^2}}.$

7. $\displaystyle\int \sqrt{a^2 - x^2} dx.$

8. $\displaystyle\int \sqrt{a^2 + x^2} dx.$

9. $\displaystyle\int \sqrt{x^2 - a^2} dx.$

10. $\displaystyle\int \frac{xdx}{\sqrt{x^2-a^2}}.$ 12. $\displaystyle\int \frac{\sqrt{x^2-a^2}}{x}dx.$ 14. $\displaystyle\int x\sqrt{a^2+x^2}dx.$

11. $\displaystyle\int \frac{dx}{x\sqrt{x^2-a^2}}.$ 13. $\displaystyle\int x\sqrt{a^2-x^2}dx.$ 15. $\displaystyle\int x\sqrt{x^2-a^2}dx.$

3.2.3   Integration by Parts

By the fundamental theorem of calculus, the integration by parts formula (3.2.4) for the antiderivatives gives the integration by parts formula for the Riemann integral.

Theorem 3.2.3 (Integration by Parts). *Suppose $F(x)$ and $G(x)$ are continuous on $[a,b]$ and differentiable on (a,b). Suppose $f(x) = F'(x)$ and $g(x) = G'(x)$ are integrable on $[a.b]$. Then*

$$\int_a^b F(x)g(x)dx = F(b)G(b) - F(a)G(a) - \int_a^b f(x)G(x)dx. \qquad (3.2.7)$$

Note that although $f(x)$ and $g(x)$ are not defined at a and b, arbitrary numbers may be assigned as the values at the two points. By Example 3.1.12, this does not affect the integrability and the integral.

The conditions in the theorem are included to make sure the antiderivative can be used to compute the integrals. The condition is weakened in Exercise 3.2.57. The reason behind the weakened condition will be revealed in Theorem 3.3.6.

Finally, similar to (3.2.5), the integration by parts formula can also be written as

$$\int_a^b F(x)dG(x) = F(b)G(b) - F(a)G(a) - \int_a^b G(x)dF(x). \qquad (3.2.8)$$

Example 3.2.17. Suppose $f(x)$ has integrable $(n+1)$-st order derivative. Then

$$f(x) = f(x_0) + \int_{x_0}^x f'(t)dt = f(x_0) + \int_{x_0}^x f'(t)d(t-x)$$

$$= f(x_0) + f'(x)(x-x) - f'(x_0)(x_0-x) - \int_{x_0}^x (t-x)f''(t)dt$$

$$= f(x_0) + f'(x_0)(x-x_0) - \frac{1}{2}\int_{x_0}^x f''(t)d(t-x)^2$$

$$= f(x_0) + f'(x_0)(x-x_0) + \frac{1}{2}f''(x_0)(x-x_0)^2 - \frac{1}{2}\int_{x_0}^x (t-x)^2 f'''(t)dt$$

$$= \cdots$$

$$= f(x_0) + f'(x_0)(x-x_0) + \frac{1}{2!}f''(x_0)(x-x_0)^2 + \cdots + \frac{1}{n!}f^{(n)}(x_0)(x-x_0)^n$$

$$+ (-1)^n \frac{1}{n!}\int_{x_0}^x (t-x)^n f^{(n+1)}(t)dt.$$

This gives the *integral form*

$$R_n(x) = \frac{1}{n!}\int_{x_0}^x (x-t)^n f^{(n+1)}(t)dt. \qquad (3.2.9)$$

of the remainder of the Taylor expansion.

Exercise 3.2.25 (Jean Bernoulli[3]). Suppose $f(t)$ has continuous n-th order derivative on $[0, x]$. Prove that

$$\int_0^x f(t)dt = xf(x) - \frac{x^2}{2!}f'(x) + \cdots + (-1)^{n-1}\frac{x^n}{n!}f^{(n-1)}(x) + (-1)^n\frac{1}{n!}\int_0^x t^n f^{(n)}(t)dt.$$

Exercise 3.2.26. Suppose $u(x)$ and $v(x)$ have continuous n-th order derivative on $[a, b]$. Prove that

$$\int_a^b uv^{(n)}dx = \left[uv^{(n-1)} - u'v^{(n-2)} + \cdots + (-1)^{n-1}u^{(n-1)}v\right]_{x=a}^{x=b} + (-1)^n\int_a^b u^{(n)}vdx.$$

Then apply the formula to $\int_{x_0}^x (x-t)^n f^{(n+1)}(t)dt$ to prove the integral form (3.2.9) of the remainder.

Exercise 3.2.27. Suppose $f'(x)$ is integrable and $\lim_{x\to+\infty} f'(x)dx = 0$. Prove that $\lim_{b\to+\infty}\frac{1}{b}\int_a^b f(x)\sin x dx = 0$. Moreover, extend the result to high order derivatives.

3.2.4 Change of Variable

By the fundamental theorem of calculus, the change of variable formula (3.2.6) for the antiderivatives gives the change of variable formula for the Riemann integral.

Theorem 3.2.4 (Change of Variable). *Suppose $\phi(x)$ is differentiable, with $\phi'(x)$ integrable on $[a, b]$. Suppose $f(y)$ is continuous on $\phi([a, b])$. Then*

$$\int_{\phi(a)}^{\phi(b)} f(y)dy = \int_a^b f(\phi(x))\phi'(x)dx. \tag{3.2.10}$$

Similar to the integration by parts, the differential notation can be used in the change of variable formula to get

$$\int_{\phi(a)}^{\phi(b)} f(y)dy = \int_a^b f(\phi(x))d\phi(x). \tag{3.2.11}$$

The formulation makes the meaning of the change of variable rather clear. In particular, the change should also be made for the variable y inside the differential dy. This suggests again the usefulness of including the differential notation in the integration.

Proof. By Theorem 3.2.1, the continuity of $f(y)$ implies that $F(t) = \int_{\phi(a)}^t f(y)dy$ satisfies $F'(t) = f(t)$. Since ϕ is differentiable, by the chain rule, we have

$$F(\phi(x))' = F'(\phi(x))\phi'(x) = f(\phi(x))\phi'(x).$$

---

[3]Jean Bernoulli, born 1667 and died 1748 in Basel (Switzerland).

Since f and ϕ are continuous, the composition $f(\phi(x))$ is continuous and is therefore integrable. Moreover, $\phi'(x)$ is assumed to be integrable. Therefore the product $f(\phi(x))\phi'(x)$ is integrable, and we may apply Theorem 3.2.2 to get

$$F(\phi(b)) - F(\phi(a)) = \int_a^b f(\phi(x))\phi'(x)dx.$$

By the definition of $F(t)$, this is the formula (3.2.10).   □

The following result shows that the change of variable formula also holds under more strict condition on $\phi(x)$ and less strict condition on $f(y)$. However, the result cannot be proved by applying the fundamental theorem of calculus. Riemann sum has to be used.

Theorem 3.2.5 (Change of Variable). *Suppose $\phi(x)$ is increasing and differentiable, with $\phi'(x)$ integrable on $[a, b]$. Suppose $f(y)$ is integrable on $[\phi(a), \phi(b)]$. Then*

$$\int_{\phi(a)}^{\phi(b)} f(y)dy = \int_a^b f(\phi(x))\phi'(x)dx.$$

Proof. Since ϕ' and f are integrable, they are both bounded. Assume $|\phi'(x)| < A$ and $|f(y)| < B$ for all $x \in [a, b]$ and $y \in [\phi(a), \phi(b)]$.

Let

$$P\colon a = x_0 < x_1 < x_2 < \cdots < x_n = b$$

be a partition of $[a, b]$. Denote by

$$\phi(P)\colon \phi(a) = \phi(x_0) \le \phi(x_1) \le \phi(x_2) \le \cdots \le \phi(x_n) = \phi(b)$$

the corresponding partition of $[\phi(a), \phi(b)]$. Note that although it might happen that $\phi(x_{i-1}) = \phi(x_i)$ for some i, it is not hard to see that the definition of Riemann sum is not changed if some equality in the partition is allowed. The assumption $|\phi'(x)| < A$ implies that $\|\phi(P)\| \le A\|P\|$.

For a Riemann sum

$$S(P, f(\phi(x))\phi'(x)) = \sum f(\phi(x_i^*))\phi'(x_i^*)(x_i - x_{i-1})$$

of $f(\phi(x))\phi'(x)$ with respect to the partition P, we consider a corresponding Riemann sum

$$S(\phi(P), f(y)) = \sum f(\phi(x_i^*))(\phi(x_i) - \phi(x_{i-1}))$$
$$= \sum f(\phi(x_i^*))\phi'(x_i^{**})(x_i - x_{i-1})$$

of $f(x)$ with respect to the partition $\phi(P)$, where the second equality is from the mean value theorem. Then

$$|S(P, f(\phi(x))\phi'(x)) - S(\phi(P), f(y))|$$
$$\le \sum |f(\phi(x_i^*))||\phi'(x_i^*) - \phi'(x_i^{**})|(x_i - x_{i-1}) \le B\sum \omega_{[x_{i-1}, x_i]}(\phi')\Delta x_i.$$

Since ϕ' and f are integrable, for any $\epsilon > 0$, there are $\delta_1, \delta_2 > 0$, such that $\|P\| < \delta_1$ and $\|\phi(P)\| \leq \delta_2$ imply

$$\sum \omega_{[x_{i-1}, x_i]}(\phi') \Delta x_i < \epsilon, \quad \left| S(\phi(P), f(y)) - \int_{\phi(a)}^{\phi(b)} f(y) dy \right| < \epsilon.$$

Combined with $\|\phi(P)\| \leq A\|P\|$, we conclude that $\|P\| < \min\left\{\delta_1, \dfrac{\delta_2}{A}\right\}$ implies

$$\left| S(P, f(\phi(x))\phi'(x)) - \int_{\phi(a)}^{\phi(b)} f(y) dy \right| < B\epsilon + \epsilon.$$

This proves that $f(\phi(x))\phi'(x)$ is integrable and the change of variable formula holds.

$\square$

Example 3.2.18. For the special case $\phi(x) = x + c$, we get

$$\int_{a+c}^{b+c} f(x) dx = \int_a^b f(x+c) dx.$$

For the special case $\phi(x) = cx$, $c \neq 0$, we get

$$\int_{ca}^{cb} f(x) dx = c \int_a^b f(cx) dx.$$

Example 3.2.19. To compute $I = \displaystyle\int_0^\pi \frac{x \sin x dx}{1 + \cos^2 x}$, we introduce $x = \pi - t$.

$$I = -\int_\pi^0 \frac{(\pi - t) \sin(\pi - t) dt}{1 + \cos^2(\pi - t)} = \int_0^\pi \frac{(\pi - t) \sin t dt}{1 + \cos^2 t} = \pi \int_0^\pi \frac{\sin t dt}{1 + \cos^2 t} - I.$$

Thus

$$I = \frac{\pi}{2} \int_0^\pi \frac{\sin t dt}{1 + \cos^2 t} = -\frac{\pi}{2} \int_0^\pi \frac{d\cos t}{1 + \cos^2 t} = -\frac{\pi}{2} \int_1^{-1} \frac{dx}{1 + x^2}$$
$$= \frac{\pi}{2}(\arctan 1 - \arctan(-1)) = \frac{\pi^2}{4}.$$

Exercise 3.2.28. Compute the integrals.

1. $\displaystyle\int_0^1 xe^{x^2} dx.$ 2. $\displaystyle\int_0^2 \frac{xdx}{1 + x^2}.$ 3. $\displaystyle\int_1^3 \frac{dx}{x\sqrt{x+1}}.$

Exercise 3.2.29. Suppose $f(x)$ is integrable on $[-a, a]$.

1. If $f(x)$ is an even function, prove that $\displaystyle\int_{-a}^a f(x) dx = 2 \int_0^a f(x) dx.$

2. If $f(x)$ is an odd function, prove that $\displaystyle\int_{-a}^a f(x) dx = 0.$

Exercise 3.2.30. Suppose $f(x)$ is a continuous function on $[-a, a]$. Prove the following are equivalent.

1. $f(x)$ is an odd function.

2. $\int_{-b}^{b} f(x)dx = 0$ for any $0 < b < a$.

3. $\int_{-a}^{a} f(x)g(x)dx = 0$ for any even continuous function $g(x)$.

4. $\int_{-a}^{a} f(x)g(x)dx = 0$ for any even integrable function $g(x)$.

Exercise 3.2.31. Use the formula $\log x = \int_{1}^{x} \dfrac{dt}{t}$ for $x > 0$ to prove the property $\log x + \log y = \log(xy)$ of the logarithmic function.

Exercise 3.2.32. Suppose $f(x)$ is integrable on an open interval containing $[a,b]$ and is continuous at a and b. Prove that

$$\lim_{h \to 0} \int_{a}^{b} \frac{f(x+h) - f(x)}{h} dx = f(b) - f(a).$$

The result should be compared with the equality

$$\int_{a}^{b} \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} dx = \int_{a}^{b} f'(x)dx = f(b) - f(a),$$

which by Theorem 3.2.2 holds when $f(x)$ is differentiable and $f'(x)$ is integrable.

3.2.5   Additional Exercise

Estimation of Integral

The estimations on the integrals in Exercises 3.1.51, 3.1.53 and 3.1.54 may be extended by using high order derivatives.

Exercise 3.2.33. Suppose $f(x)$ is continuous on $[a,b]$ and has n-th order derivative on (a,b). By either integrating the Taylor expansion at a or considering the Taylor expansion of the function $F(x) = \int_{a}^{x} f(t)dt$, prove that

$$\int_{a}^{b} f(x)dx = f(a)(b-a) + \frac{f'(a)}{2!}(b-a)^2 + \cdots + \frac{f^{(n-1)}(a)}{n!}(b-a)^n + \frac{f^{(n)}(c)}{(n+1)!}(b-a)^{n+1},$$
$$(3.2.12)$$

where $a < c < b$.

Exercise 3.2.34. Suppose $f(x)$ is continuous on $[a,b]$ and has n-th order derivative on (a,b). Prove that

$$\left| \int_{a}^{b} f(x)dx - \sum_{k=0}^{n-1} \frac{f^{(k)}(a) + (-1)^k f^{(k)}(b)}{(k+1)!2^k}(b-a)^{k+1} \right| \leq \frac{\omega_{(a,b)}(f^{(n)})}{(n+1)!2^{n+1}}(b-a)^{n+1}$$
$$(3.2.13)$$

for odd n and

$$\int_{a}^{b} f(x)dx = \sum_{k=0}^{n-1} \frac{f^{(k)}(a) + (-1)^k f^{(k)}(b)}{(k+1)!2^k}(b-a)^{k+1} + \frac{f^{(n)}(c)}{(n+1)!2^n}(b-a)^{n+1} \quad (3.2.14)$$

for even n and some $a < c < b$.

Exercise 3.2.35. Suppose $f(x)$ is continuous on $[a, b]$ and has $2n$-th order derivative on (a, b). Prove

$$\int_a^b f(x)dx = \sum_{k=0}^{n-1} \frac{1}{(2k+1)!2^{2k}} f^{(2k)}\left(\frac{a+b}{2}\right)(b-a)^{2k+1} + \frac{f^{(2n)}(c)}{(2n+1)!2^{2n}}(b-a)^{2n+1}.$$

(3.2.15)

for some $a < c < b$.

Estimation of Special Riemann Sums

Consider the partition of $[a, b]$ by evenly distributed partition points $x_i = a + \frac{i}{n}(b-a)$. We can form the Riemann sums by choosing the left, right and middle points of the partition intervals.

$$S_{\text{left},n}(f) = \sum f(x_{i-1})\Delta x_i = \frac{b-a}{n}\sum f(x_{i-1}),$$

$$S_{\text{right},n}(f) = \sum f(x_i)\Delta x_i = \frac{b-a}{n}\sum f(x_i),$$

$$S_{\text{middle},n}(f) = \sum f\left(\frac{x_{i-1}+x_i}{2}\right)\Delta x_i = \frac{b-a}{n}\sum f\left(\frac{x_{i-1}+x_i}{2}\right).$$

The question is how close these are to the actual integral.

Exercise 3.2.36. Suppose $f(x)$ is continuous on $[a, b]$ and differentiable on (a, b), such that $f'(x)$ is integrable. Use the estimation in Exercise 3.1.51 to prove that

$$\lim_{n\to\infty} n\left(\int_a^b f(x)dx - S_{\text{left},n}(f)\right) = \frac{1}{2}(f(b) - f(a))(b-a),$$

$$\lim_{n\to\infty} n\left(\int_a^b f(x)dx - S_{\text{right},n}(f)\right) = -\frac{1}{2}(f(b) - f(a))(b-a).$$

Exercise 3.2.37. Suppose $f(x)$ is continuous on $[a, b]$ and has second order derivative on (a, b), such that $f''(x)$ is integrable on $[a, b]$. Use the estimation in Exercise 3.1.54 to prove that

$$\lim_{n\to\infty} n^2\left(\int_a^b f(x)dx - S_{\text{middle},n}(f)\right) = \frac{1}{24}(f'(b) - f'(a))(b-a)^2.$$

Exercise 3.2.38. Use Exercises 3.2.33, 3.2.34 and 3.2.35 to derive higher order approximation formulae for the integral $\int_a^b f(x)dx$.

Average of Functions

The average of an integrable function on $[a, b]$ is

$$\text{Av}_{[a,b]}(f) = \frac{1}{b-a}\int_a^b f(x)dx.$$

(3.2.16)

Exercise 3.2.39. Prove the properties of the average.

1. $\text{Av}_{[a+c,b+c]}(f(x+c)) = \text{Av}_{[a,b]}(f(x))$, $\text{Av}_{[\lambda a,\lambda b]}(f(\lambda x)) = \text{Av}_{[a,b]}(f(x))$.

2. If $c = \lambda a + (1 - \lambda)b$, then $\mathrm{Av}_{[a,b]}(f) = \lambda \mathrm{Av}_{[a,c]}(f) + (1 - \lambda)\mathrm{Av}_{[c,b]}(f)$. In particular, $\mathrm{Av}_{[a,b]}(f)$ lies between $\mathrm{Av}_{[a,c]}(f)$ and $\mathrm{Av}_{[c,b]}(f)$.

3. $f \geq g$ implies $\mathrm{Av}_{[a,b]}(f) \geq \mathrm{Av}_{[a,b]}(g)$.

4. If $f(x)$ is continuous, then $\mathrm{Av}_{[a,b]}(f) = f(c)$ for some $a < c < b$.

Exercise 3.2.40. Suppose $f(x)$ is integrable on $[0, a]$ for any $a > 0$. Consider the average function $g(x) = \mathrm{Av}_{[0,x]}(f) = \dfrac{1}{x} \displaystyle\int_0^x f(t)dt$.

1. Prove that if $\lim_{x \to +\infty} f(x) = l$, then $\lim_{x \to +\infty} g(x) = l$ (compare Exercise 1.1.36).

2. Prove that if $f(x)$ is increasing, then $g(x)$ is also increasing.

3. Prove that if $f(x)$ is convex, then $g(x)$ is also convex.

Exercise 3.2.41. For a weight function $\lambda(x)$ defined in (3.1.21) in Exercise 3.1.47, the *weighted average* of an integrable function $f(x)$ is

$$\mathrm{Av}_{[a,b]}^\lambda(f(x)) = \frac{1}{b - a} \int_a^b \lambda(x)f(x)dx \qquad (3.2.17)$$

Can you extend the properties of average in Exercise 3.2.39 to weighted average?

Integration of Periodic Function and Riemann-Lebesgue Lemma

Suppose $f(x)$ is a periodic integrable function with period T.

Exercise 3.2.42. Prove that $\displaystyle\int_a^{a+T} f(x)dx = \int_0^T f(x)dx$.

Exercise 3.2.43. Prove that $\lim_{b \to +\infty} \dfrac{1}{b} \displaystyle\int_a^b f(x)dx = \dfrac{1}{T} \displaystyle\int_0^T f(x)dx$. This says that the limit of the average on bigger and bigger intervals is the average on an interval of the period length.

Exercise 3.2.44 (Riemann-Lebesgue Lemma). Suppose $f(x)$ is a periodic integrable function with period T and $g(x)$ is integrable on $[a, b]$. Prove that

$$\lim_{t \to \infty} \int_a^b f(tx)g(x)dx = \frac{1}{T} \int_0^T f(x)dx \int_a^b g(x)dx.$$

Trigonometric Integration

Exercise 3.2.45. Let a, b, n be given.

1. Prove that there are A, B, C, such that

$$\int \frac{dx}{(a \sin x + b \cos x)^n} = \frac{A \sin x + B \cos x}{(a \sin x + b \cos x)^{n-1}} + C \int \frac{dx}{(a \sin x + b \cos x)^{n-2}}.$$

2. Prove that if $|a| \neq |b|$, then there are A, B, C, such that

$$\int \frac{dx}{(a + b \cos x)^n} = \frac{A \sin x}{(a + b \cos x)^{n-1}} + B \int \frac{dx}{(a + b \cos x)^{n-1}} + C \int \frac{dx}{(a + b \cos x)^{n-2}}.$$

Exercise 3.2.46. Compute the antiderivative of $\displaystyle\int \frac{dx}{\cos(x + a)\cos(x + b)}$ by using

$$\tan(x + a) - \tan(x + b) = \frac{\sin(a - b)}{\cos(x + a)\cos(x + b)}.$$

Use the similar idea to compute the following antiderivatives.

1. $\displaystyle\int \frac{dx}{\sin(x+a)\cos(x+b)}.$ 3. $\displaystyle\int \frac{dx}{\sin x - \sin a}.$

2. $\displaystyle\int \tan(x+a)\tan(x+b)dx.$ 4. $\displaystyle\int \frac{dx}{\cos x + \cos a}.$

Wallis[4] Formula

For $p, q \geq 0$, define $w(p,q) = \displaystyle\int_0^1 (1 - x^{\frac{1}{p}})^q dx.$

Exercise 3.2.47. Prove $w(p,q) = \dfrac{q}{p+q} w(p, q-1).$

Exercise 3.2.48. Prove $w(p,q) = w(q,p)$ by changing the variable $y = (1 - x^{\frac{1}{p}})^q.$

Exercise 3.2.49. Prove $w(m,n) = \dfrac{m!n!}{(m+n)!}$ for natural numbers m and n.

Exercise 3.2.50. Prove w is strictly decreasing in p and in q.

Exercise 3.2.51. Show that $w\left(\dfrac{1}{2}, \dfrac{1}{2}\right) = \dfrac{\pi}{4}.$

Exercise 3.2.52. Use $w(n+1, n+1) < w\left(n + \dfrac{1}{2}, n + \dfrac{1}{2}\right) < w(n, n)$ to prove

$$\frac{2^2 \cdot 4^2 \cdot 6^2 \cdots (2n)^2}{1 \cdot 3^2 \cdot 5^2 \cdots (2n-1)^2 \cdot (2n+1)} < \frac{\pi}{2} < \frac{2^2 \cdot 4^2 \cdot 6^2 \cdots (2n)^2 \cdot (2n+2)}{1 \cdot 3^2 \cdot 5^2 \cdots (2n-1)^2 \cdot (2n+1)^2}.$$

Exercise 3.2.53. Prove that in the inequality in Exercise 3.2.52, the left is increasing, the right is decreasing, and both have $\dfrac{\pi}{2}$ as the limit. This leads to Wallis infinite product formula

$$\frac{\pi}{2} = \frac{2}{1} \cdot \frac{2}{3} \cdot \frac{3}{4} \cdot \frac{4}{5} \cdot \frac{5}{6} \cdot \frac{6}{7} \cdots.$$

Integration by Parts under Weaker Condition

Suppose $f(x)$ and $g(x)$ are integrable on $[a, b]$, and

$$F(x) = \int_a^x f(t)dt, \ \ G(x) = \int_a^x g(t)dt.$$

The subsequent exercises will prove the integration by parts formula (3.2.7) under the only assumption that f and g are integrable. Note that up to adding constants to F and G, the assumption is strictly weaker than the one in Theorem 3.2.3.

For a partition $P: a = x_0 < x_1 < \cdots < x_n = b$ and a choice of $x_i^* = x_i$, define the "partial Riemann sums" $S_i(P, f) = \sum_{k=1}^i f(x_k)\Delta x_k$ for $1 \leq i \leq n$.

Exercise 3.2.54. The partial Riemann sum is the Riemann sum for the integral $F(x_i) = \displaystyle\int_a^{x_i} f(t)dt$ and should approximate the integral. Then by replacing $F(x_i)$ with the approximation $S_i(P, f)$, the sum $\sum_{i=1}^n S_i(P, f)g(x_i)\Delta x_i$ should approximate the Riemann sum $S(P, Fg)$. By using the estimation (3.1.14) in Exercise 3.1.16, prove the estimation

$$\left| S(P, Fg) - \sum_{i=1}^n S_i(P, f)g(x_i)\Delta x_i \right| \leq \left(\sum_{i=1}^n \omega_{[x_{i-1}, x_i]}(f)\Delta x_i \right) S(P, |g|).$$

[4] John Wallis, born 1616 in Kent (England), died 1703 in Oxford (England).

Exercise 3.2.55. Exercise 3.2.54 provides approximations for the Riemann sums of of Fg and fG. Prove the sum of the two approximations is

$$\sum_{i=1}^{n} S_i(P, f)g(x_i)\Delta x_i + \sum_{i=1}^{n} f(x_i)\Delta x_i S_i(g, P) = S(P, f)S(P, g) + \sum_{i=1}^{n} f(x_i)g(x_i)\Delta x_i^2.$$

The geometric meaning of the eqality is given in Figure 4.1.

Exercise 3.2.56. Prove the integration by parts formula

$$\int_a^b (F(x)g(x) + f(x)G(x))dx = F(b)G(b)$$

for the special case $F(a) = G(a) = 0$.

Exercise 3.2.57. Suppose $f(x)$ and $g(x)$ are integrable on $[a, b]$ and

$$F(x) = \int_a^x f(t)dt + c, \quad G(x) = \int_a^x g(t)dt + d,$$

where c and d are some constants. Prove the integration by parts formula (3.2.7).

Second Integral Mean Value Theorem

The first integral mean value theorem appeared in Exercise 3.1.15.

Suppose $f(x)$ is differentiable with integrable $f'(x)$ on $[a, b]$. Suppose $g(x)$ is continuous on $[a, b]$.

Exercise 3.2.58. Suppose $f(x) \geq 0$ and is decreasing. Suppose $G(x) = \int_a^x g(x)dx$ satisfies $m \leq G(x) \leq M$ for $x \in [a, b]$. Prove that

$$f(a)m \leq \int_a^b f(x)g(x)dx \leq f(a)M.$$

Then use this to prove that

$$\int_a^b f(x)g(x)dx = f(a)\int_a^c g(x)dx$$

for some $a < c < b$. What if $f(x)$ is increasing?

Exercise 3.2.59. Suppose $f(x)$ is monotone. Prove that

$$\int_a^b f(x)g(x)dx = f(a)\int_a^c g(x)dx + f(b)\int_c^b g(x)dx$$

for some $a < c < b$.

Young Inequality

Exercise 3.2.60. Suppose $f(x)$ is differentiable for $x \geq 0$ and satisfies $f'(x) > 0$, $f(0) = 0$. Prove that for any $a, b > 0$, we have the *Young inequality*

$$\int_0^a f(x)dx + \int_0^b f^{-1}(y)dy \geq ab. \tag{3.2.18}$$

Then apply to the function x^{p-1} for $p > 1$ and derive the Young inequality in Exercise 2.2.40.

The inequality actually holds under weaker condition. See Exercise 3.3.30.

3.3 Topics on Integration

3.3.1 Integration of Rational Functions

It is a basic algebraic fact that any rational function is a sum of functions of the forms Ax^n, $\dfrac{A}{(x-a)^n}$ and $\dfrac{Bx+C}{(x^2+ax+b)^n}$, where x^2+ax+b has no real roots. Thus the antiderivative of the rational function is the sum of the antiderivatives of these three types of functions. The antiderivative of the first two types of functions are very simple. To compute the antiderivative of the last type, we complete the square and get $x^2+ax+b=(x+\alpha)^2+\beta^2$. Then

$$\int \frac{(Bx+C)dx}{(x^2+ax+b)^n} = b \int \frac{(x+\alpha)dx}{((x+\alpha)^2+\beta^2)^n} + c \int \frac{dx}{((x+\alpha)^2+\beta^2)^n},$$

where $b=B$, $c=C-B\alpha$. The first antiderivative is

$$\int \frac{(x+\alpha)dx}{((x+\alpha)^2+\beta^2)^n} = \begin{cases} \dfrac{1}{2}\log(x^2+ax+b) + C & \text{if } n=1 \\[2ex] -\dfrac{1}{2(n-1)(x^2+ax+b)^{n-1}} + C & \text{if } n>1 \end{cases}.$$

The change of variable $\beta t = x+\alpha$ reduces the computation of the second antiderivative into $\displaystyle\int \frac{dx}{(t^2+1)^n}$. This may be computed by the last recursive relation in Exercise 3.2.19.

Example 3.3.1. To compute the antiderivative of the rational function $f(x) = \dfrac{x^4+x^3+x^2-2x+1}{x^5+x^4-2x^3-2x^2+x+1}$, we find the factorization $x^5+x^4-2x^3-2x^2+x+1 = (x-1)^2(x+1)^3$ of the denominator and write

$$f(x) = \frac{A_1}{x-1} + \frac{A_2}{(x-1)^2} + \frac{B_1}{x+1} + \frac{B_2}{(x+1)^2} + \frac{B_3}{(x+1)^3}.$$

This is the same as

$$\begin{aligned} x^4+x^3+x^2-2x+1 = {} & A_1(x-1)(x+1)^3 + A_2(x+1)^3 \\ & + B_1(x-1)^2(x+1)^2 + B_2(x-1)^2(x+1) + B_3(x-1)^2. \end{aligned}$$

Then we get the following equations.

$$\begin{aligned} 2 &= 8A_2, && (\text{at } x=1) \\ 4 &= 4B_3, && (\text{at } x=-1) \\ 7 &= 8A_1 + 12A_2, && (\tfrac{d}{dx} \text{ at } x=1) \\ -5 &= 4B_2 - 4B_3, && (\tfrac{d}{dx} \text{ at } x=-1) \\ 1 &= A_1 + A_2 + B_1 + B_2 + B_3. && (\text{coefficient of } x^4) \end{aligned}$$

It is easy to solve the equations and get

$$\int f(x)dx = \int \left(\frac{1}{2(x-1)} + \frac{1}{4(x-1)^2} + \frac{1}{2(x+1)} - \frac{1}{4(x+1)^2} + \frac{1}{(x+1)^3} \right) dx$$

$$= \frac{1}{2}\log|x-1| - \frac{1}{4(x-1)} + \frac{1}{2}\log|x+1| + \frac{1}{4(x+1)} - \frac{1}{2(x+1)^2} + C$$

$$= \frac{1}{2}\log|x^2-1| - \frac{x}{(x-1)(x+1)^2} + C.$$

Example 3.3.2. To compute the antiderivative of the rational function $f(x) = \dfrac{4x^3 - x^2 + 2x}{8x^6 + 4x^5 - 4x^4 - 8x^3 - 2x^2 + x + 1}$, we find the factorization $(x-1)(2x-1)(2x^2 + 2x + 1)^2$ of the denominator and write

$$f(x) = \frac{A_1}{x-1} + \frac{A_2}{2x-1} + \frac{B_1 x + C_1}{2x^2 + 2x + 1} + \frac{B_2 x + C_2}{(2x^2 + 2x + 1)^2}.$$

This is the same as

$$4x^3 - x^2 + 2x = A_1(2x-1)(2x^2 + 2x + 1)^2 + A_2(x-1)(2x^2 + 2x + 1)^2$$
$$+ (B_1 x + C_1)(x-1)(2x-1)(2x^2 + 2x + 1)$$
$$+ (B_2 x + C_2)(x-1)(2x-1).$$

By various methods (comparing coefficients of x^k, evaluating at specific values of x, taking derivative and evaluating at specific values of x, etc.), we get a system of equations. Solving the system, we get

$$f(x) = \frac{1}{5(x-1)} - \frac{2}{5(2x-1)} - \frac{1}{5(2x^2 + 2x + 1)} + \frac{x}{(2x^2 + 2x + 1)^2}.$$

We have

$$\int \frac{dx}{x-1} = \log|x-1| + C,$$

$$\int \frac{dx}{2x-1} = \frac{1}{2}\int \frac{d(2x-1)}{2x-1} = \frac{1}{2}\log|2x-1| + C,$$

$$\int \frac{dx}{2x^2 + 2x + 1} = \int \frac{d(2x+1)}{(2x+1)^2 + 1} = \arctan(2x+1) + C,$$

$$\int \frac{xdx}{(2x^2 + 2x + 1)^2} = \int \frac{4xdx}{((2x+1)^2 + 1)^2} = \int \frac{(\tan t - 1)d\tan t}{(\tan^2 t + 1)^2}$$

$$= \int (\sin t \cos t - \cos^2 t)dt = \frac{1}{2}(\sin^2 t - \sin t \cos t - t)$$

$$= \frac{1}{2}(\cos^2 t(\tan^2 t - \tan t) - t) + C$$

$$= \frac{(2x+1)^2 - (2x+1)}{2(2x+1)^2 + 1} - \frac{1}{2}\arctan(2x+1) + C$$

$$= \frac{x(2x+1)}{2x^2 + 2x + 1} - \frac{1}{2}\arctan(2x+1) + C.$$

Thus we conclude

$$\int f(x)dx = \frac{1}{5}\log\left|\frac{x-1}{2x-1}\right| - \frac{7}{10}\arctan(2x+1) + \frac{x(2x+1)}{2x^2 + 2x + 1} + C.$$

Exercise 3.3.1. Compute the antiderivatives of rational functions.

1. $\int \dfrac{dx}{x^2(1+x)}.$

8. $\int \dfrac{dx}{(1-x^4)^2}.$

2. $\int \dfrac{dx}{x(1+x)(2+x)}.$

9. $\int \dfrac{xdx}{(x-1)^2(x^2+2x+2)}.$

3. $\int \dfrac{x^4 dx}{1-x^2}.$

10. $\int \dfrac{(x^4+4x^3+4x^2+4x+4)dx}{x(x+2)(x^2+2x+2)^2}.$

4. $\int \dfrac{xdx}{x^2+x-2}.$

11. $\int \dfrac{(2x^2+3)dx}{x^3+x^2-2}.$

5. $\int \dfrac{x^5 dx}{x^2+x-2}.$

12. $\int \dfrac{dx}{x^4+4}.$

6. $\int \dfrac{dx}{1-x^4}.$

13. $\int \dfrac{xdx}{x^3+1}.$

7. $\int \dfrac{x^2 dx}{(1-x^4)^2}.$

14. $\int \dfrac{dx}{(x+1)(x^2+1)(x^3+1)}.$

By suitable change of variables, many antiderivatives may be converted to the antiderivatives of rational functions.

Example 3.3.3. By the change $t = \tan\dfrac{x}{2}$, we have

$$\int R(\sin x, \cos x)dx = \int R\left(\frac{2t}{1+t^2}, \frac{1-t^2}{1+t^2}\right)\frac{2dt}{1+t^2}.$$

If R is a rational function, then the computation is reduced to the antiderivative of some rational function.

As a concrete example, for $a \neq 0$, we have

$$\int \frac{dx}{a+\sin x} = \int \frac{2dt}{\left(a+\dfrac{2t}{1+t^2}\right)(1+t^2)} = \int \frac{2dt}{a\left(\left(t+\dfrac{1}{a}\right)^2 + 1 - \dfrac{1}{a^2}\right)}.$$

If $|a| > 1$, then with $t + \dfrac{1}{a} = \sqrt{1 - \dfrac{1}{a^2}}\,u$, we further have

$$\int \frac{dx}{a+\sin x} = \int \frac{2\sqrt{1-\dfrac{1}{a^2}}\,du}{a\left(1-\dfrac{1}{a^2}\right)(u^2+1)} = \frac{2}{\text{sign}(a)\sqrt{a^2-1}}\arctan\frac{a\tan\dfrac{x}{2}+1}{\sqrt{a^2-1}} + C.$$

If $|a| = 1$, then

$$\int \frac{dx}{a+\sin x} = \int \frac{2dt}{a(t+a)^2} = \frac{-2}{a(t+a)} + C = \frac{-2}{a\tan\dfrac{x}{2}+1} + C.$$

If $|a| < 1$, then

$$\left(t+\frac{1}{a}\right)^2 + 1 - \frac{1}{a^2} = (t-\alpha)(t-\beta), \quad \alpha = -\frac{1}{a} + \sqrt{\frac{1}{a^2}-1}, \quad \beta = -\frac{1}{a} - \sqrt{\frac{1}{a^2}-1},$$

and we further have

$$\int \frac{dx}{a + \sin x} = \int \frac{2dt}{a(t - \alpha)(t - \beta)} = \frac{2}{a(\alpha - \beta)} \log \left| \frac{t - \alpha}{t - \beta} \right| + C$$

$$= \frac{1}{\sqrt{1 - a^2}} \log \left| \frac{a \tan \frac{x}{2} + 1 - \sqrt{1 - a^2}}{a \tan \frac{x}{2} + 1 + \sqrt{1 - a^2}} \right| + C.$$

Example 3.3.4. Although the change $t = \tan \frac{x}{2}$ can always be used for rational functions of trigomonetric functions, it may not be the most efficient one. For example, by $t = \cos x$, we have

$$\int \frac{dx}{(a + \cos x) \sin x} = \int \frac{-d \cos x}{(a + \cos x) \sin^2 x} = \int \frac{-dt}{(a + t)(1 - t^2)}.$$

If $a \neq \pm 1$, then

$$\frac{-1}{(a + t)(1 - t^2)} = \frac{1}{(a^2 - 1)(t + a)} - \frac{1}{2(a - 1)(t + 1)} + \frac{1}{2(a + 1)(t - 1)},$$

and

$$\int \frac{dx}{(a + \cos x) \sin x} = \frac{\log |t + a|}{a^2 - 1} - \frac{\log |t + 1|}{2(a - 1)} + \frac{\log |t - 1|}{2(a + 1)} + C$$

$$= \frac{\log |\cos x + a|}{a^2 - 1} - \frac{\log |\cos x + 1|}{2(a - 1)} + \frac{\log |\cos x - 1|}{2(a + 1)} + C.$$

Example 3.3.5. By the change $t = \sqrt[n]{\frac{ax + b}{cx + d}}$, we have

$$\int R \left(x, \sqrt[n]{\frac{ax + b}{cx + d}} \right) dx = \int R \left(\frac{\alpha t^n + \beta}{\gamma t^n + \delta}, t \right) \frac{(\alpha \delta - \beta \gamma) n t^{n-1}}{(\gamma t^n + \delta)^2} dt$$

Applying the idea to the antiderivative $\int \sqrt{\frac{x}{(x + 1)^3}} dx$, we introduce $t = \sqrt{\frac{x}{x + 1}}$. Then $x = \frac{t^2}{1 - t^2}$, $dx = \frac{2t dt}{(1 - t^2)^2}$, and

$$\int \sqrt{\frac{x}{(x + 1)^3}} dx = \int |1 - t^2| t \frac{2t dt}{(1 - t^2)^2} = \pm \int \left(\frac{1}{t + 1} - \frac{1}{t - 1} - 2 \right) dt$$

$$= \pm \left(\log \left| \frac{t + 1}{t - 1} \right| - 2t \right) + C$$

$$= \pm \left(2 \log(\sqrt{x} + \sqrt{x + 1}) - 2 \sqrt{\frac{x}{x + 1}} \right) + C.$$

The sign is positive if $x > 0$, and is negative if $x < -1$.

Example 3.3.6. To compute the antiderivative $\int \sqrt{\frac{e^x - 1}{e^x + 1}} dx$, we introduce $t = e^x$. Then $dt = e^x dx = t dx$, and the antiderivative becomes $\int \sqrt{\frac{t - 1}{t + 1}} \frac{dt}{t}$. By further

introducing $u = \sqrt{\dfrac{t-1}{t+1}} = \sqrt{\dfrac{e^x-1}{e^x+1}}$, we get

$$\int \sqrt{\frac{e^x-1}{e^x+1}}dx = \int \frac{4u^2 du}{(1+u^2)(1-u^2)} = \int \left(\frac{1}{u+1} - \frac{1}{u-1} - \frac{2}{u^2+1} \right) du$$

$$= \log\left| \frac{u+1}{u-1} \right| - 2\arctan u + C$$

$$= 2\log(\sqrt{e^x-1} + \sqrt{e^x+1}) - 2\arctan\sqrt{\frac{e^x-1}{e^x+1}} + C.$$

Exercise 3.3.2. Compute the antiderivatives.

1. $\displaystyle\int \frac{(1+\sin x)dx}{\sin x(1+\cos x)}.$

2. $\displaystyle\int \sqrt{\tan x}\,dx.$

3. $\displaystyle\int \frac{(\sin x + \cos x)dx}{\sin x(\sin x - \cos x)}.$

4. $\displaystyle\int \frac{dx}{a+\cos x}.$

5. $\displaystyle\int \frac{dx}{(a+\cos^2 x)\sin x}.$

6. $\displaystyle\int \frac{dx}{a+\tan x}.$

7. $\displaystyle\int \frac{dx}{a\sin x + b\cos x + c}.$

8. $\displaystyle\int \frac{dx}{a^2\sin^2 x + b^2\cos^2 x}.$

9. $\displaystyle\int \sqrt{\frac{1-x}{1+x}}dx.$

10. $\displaystyle\int \frac{1}{x^2}\sqrt{\frac{1-x}{1+x}}dx.$

11. $\displaystyle\int \frac{dx}{\sqrt{1+e^x} + \sqrt{1-e^x}}.$

12. $\displaystyle\int \frac{dx}{\sqrt{a^x+b}}.$

Exercise 3.3.3. Suppose R is a rational function. Suppose r, s are rational numbers such that $r+s$ is an integer. Find a suitable change of variable, such that $\displaystyle\int R(x, (ax+b)^r, (cx+d)^s)dx$ is changed into the antiderivative of a rational function.

Exercise 3.3.4. Suppose r, s, t are rational numbers. For each of the following cases, find a suitable change of variable, such that $\displaystyle\int x^r(a+bx^s)^t dx$ is changed into the antiderivative of a rational function.

1. t is an integer.

2. $\dfrac{r+1}{s}$ is an integer.

3. $\dfrac{r+1}{s} + t$ is an integer.

A theorem by Chebyshev[5] says that these are the only cases that the antiderivative can be changed to the antiderivative of a rational function.

---

[5]Pafnuty Lvovich Chebyshev, born 1821 in Okatovo (Russia), died 1894 in St Petersburg (Russia). Chebyshev's work touches many fields of mathematics, including analysis, probability, number theory and mechanics. Chebyshev introduced his famous polynomials in 1854 and later generalized to the concept of orthogonal polynomials.

3.3.2   Improper Integration

So far the integrability was defined only for bounded functions on bounded intervals. When the functions or the intervals are unbounded, we may integrate on the bounded part and then take limit.

Suppose $f(x)$ is integrable on $[a,c]$ for any $a < c < b$. If $\lim_{c \to b^-} \int_a^c f(x)dx$ converges, then we say the *improper integral* $\int_a^b f(x)dx$ converges and write

$$\int_a^b f(x)dx = \lim_{c \to b^-} \int_a^c f(x)dx.$$

The definition also applies to the case $b = +\infty$, and similar definition can be made when $f(x)$ is integrable on $[c,b]$ for any $a < c < b$.

By Exercise 3.1.8, if $f(x)$ is a bounded function on bounded interval $[a,b]$ and is integrable on $[a,c]$ for any $a < c < b$, then it is integrable on $[a,b]$. Therefore an integral on a bounded interval becomes improper only if it is not bounded.

Example 3.3.7. For any $\alpha \neq -1$, the limit $\lim_{a \to 0^+} \int_a^1 x^\alpha dx = \lim_{a \to 0^+} \dfrac{1 - a^{\alpha+1}}{\alpha + 1}$ converges if and only if $\alpha + 1 > 0$. Therefore the improper integral $\int_0^1 x^\alpha dx$ (which is in fact improper only when $\alpha < 0$) converges if and only if $\alpha > -1$, and $\int_0^1 x^\alpha dx = \dfrac{1}{\alpha + 1}$.

On the other hand, the limit $\lim_{a \to +\infty} \int_1^a x^\alpha dx = \lim_{a \to +\infty} \dfrac{a^{\alpha+1} - 1}{\alpha + 1}$ converges if and only if $\alpha + 1 < 0$. Therefore the improper integral $\int_1^{+\infty} x^\alpha dx$ converges if and only if $\alpha < -1$, and $\int_1^{+\infty} x^\alpha dx = \dfrac{-1}{\alpha + 1}$.

Example 3.3.8. The integral $\int_{-\infty}^{+\infty} \dfrac{dx}{1 + x^2}$ is improper at $-\infty$ and $+\infty$. Since

$$\lim_{a \to -\infty, b \to +\infty} \int_a^b \frac{dx}{1 + x^2} = \lim_{a \to -\infty, b \to +\infty} (\arctan b - \arctan a) = \frac{\pi}{2} - \left(-\frac{\pi}{2}\right) = \pi,$$

the improper integral converges, and $\int_{-\infty}^{+\infty} \dfrac{dx}{1 + x^2} = \pi$.

Exercise 3.3.5. Determine the convergence of the improper integrals and evaluate the convergent ones.

1. $\displaystyle\int_2^\infty \frac{dx}{x(\log x)^\alpha}$.

2. $\displaystyle\int_0^1 \frac{dx}{x(-\log x)^\alpha}$.

3. $\displaystyle\int_0^{+\infty} x^\alpha dx$.

4. $\displaystyle\int_0^1 \log x\, dx$.

5. $\displaystyle\int_0^{+\infty} a^x dx$.

6. $\displaystyle\int_{-\infty}^0 a^x dx$.

7. $\int_{-1}^{1} \dfrac{dx}{1-x^2}$.

9. $\int_{-1}^{1} \dfrac{dx}{\sqrt{1-x^2}}$.

11. $\int_{0}^{+\infty} e^{-x} \sin x \, dx$.

8. $\int_{2}^{+\infty} \dfrac{dx}{1-x^2}$.

10. $\int_{0}^{\frac{\pi}{2}} \tan x \, dx$.

12. $\int_{0}^{+\infty} e^{-x} |\sin x| \, dx$.

Exercise 3.3.6. Suppose $f(x)$ is continuous for $x \geq 0$, and $\lim_{x \to +\infty} f(x) = l$. Prove that for any $a, b > 0$, $\int_{0}^{+\infty} \dfrac{f(ax) - f(bx)}{x} dx = (f(0) - l) \log \dfrac{b}{a}$.

The convergent improper integrals have properties similar to the normal integrals. For example, the integration by parts and the change of variable still hold.

The convergence of improper integrals can be determined by the Cauchy criterion. The following is the criterion stated for the case that the integral becomes improper at $+\infty$.

Proposition 3.3.1 (Cauchy Criterion). *Suppose $f(x)$ is integrable on $[a, b]$ for any $b \in [a, +\infty)$. Then the improper integral $\int_{a}^{+\infty} f(x) dx$ converges if and only if for any $\epsilon > 0$, there is N, such that*

$$b, c > N \implies \left| \int_{b}^{c} f(x) dx \right| < \epsilon.$$

A consequence of the criterion is the following useful method for determining the convergence, again stated for integrals that are improper at $+\infty$.

Proposition 3.3.2 (Comparison Test). *Suppose $f(x)$ and $g(x)$ are integrable on $[a, b]$ for any $b > a$. If $|f(x)| \leq g(x)$ and $\int_{a}^{+\infty} g(x) dx$ converges, then $\int_{a}^{+\infty} f(x) dx$ also converges.*

Proof. For any $\epsilon > 0$, applying the Cauchy criterion to the convergence of $\int_{a}^{+\infty} g(x) dx$ tells us that there is N, such that $b, c > N$ implies $\int_{b}^{c} g(x) dx < \epsilon$. Then $b, c > N$ implies

$$\left| \int_{b}^{c} f(x) dx \right| \leq \int_{b}^{c} |f(x)| dx \leq \int_{b}^{c} g(x) dx < \epsilon.$$

Thus the Cauchy criterion for the convergence of $\int_{a}^{+\infty} f(x) dx$ is verified.

$\square$

A special case of the comparison test is that if both $f(x)$ and $g(x)$ are positive, and $\lim_{x \to +\infty} \dfrac{f(x)}{g(x)} = l$ exists, then the convergence of $\int_{a}^{+\infty} g(x) dx$ implies the convergence of $\int_{a}^{+\infty} f(x) dx$. If $l \neq 0$, then the two convergences are equivalent. More generally, if $c_1 g(x) \leq f(x) \leq c_2 g(x)$ for some $c_1, c_2 > 0$, then the two convergences are equivalent.

Example 3.3.9. Let $r(x) = \dfrac{p(x)}{q(x)}$ be a rational function. Let m and n be the

degrees of the polynomials $p(x)$ and $q(x)$. Then $\lim_{x\to\infty} \dfrac{r(x)}{x^{m-n}}$ converges to a

nonzero number (the quotient of the leading coefficients). By changing the sign of the leading coefficients if necessary, we may assume $p(x)$ and $q(x)$ are positive.

Since $\displaystyle\int_a^{+\infty} x^{m-n}dx$ converges if and only if $m - n \le -2$, by the comparison test,

we conclude that $\displaystyle\int_a^{+\infty} r(x)dx$ converges if and only if $m - n \le -2$.

Example 3.3.10. The integral $\displaystyle\int_0^2 \dfrac{\sin x\,dx}{\sqrt{|x(x-1)|}}$ is improper at 1 (in fact at both 1^+

and 1^-). Since $\lim_{x\to 1} \dfrac{\sin x}{\sqrt{|x|}} = \sin 1 \ne 0$, the convergence of the improper integral

is equivalent to the convergence of $\displaystyle\int_0^2 \dfrac{dx}{\sqrt{|x-1|}}$. The later converges (on both

sides of 1) because $\displaystyle\int_0^1 \dfrac{dx}{\sqrt{x}}$ converges. Therefore $\displaystyle\int_0^2 \dfrac{\sin x\,dx}{\sqrt{|x(x-1)|}}$ converges.

Example 3.3.11. Since $\lim_{x\to 0^+} \sqrt{x}\log x = 0$ and $\displaystyle\int_0^1 \dfrac{dx}{\sqrt{x}}$ converges, by the compar-

ison test, the improper integral $\displaystyle\int_0^1 \log x\,dx$ converges. Moreover, by $0 < \log\sin x <$

$\log x$ for $0 < x < 1$ and the comparison test, the improper integral $\displaystyle\int_0^{\frac{\pi}{2}} \log\sin x\,dx$

also converges.

   To compute the improper integral, we change the variable.

$$\int_0^{\frac{\pi}{2}} \log\sin x\,dx = \int_0^{\frac{\pi}{2}} \log\left(2\sin\frac{x}{2}\cos\frac{x}{2}\right) dx = 2\int_0^{\frac{\pi}{4}} \log(2\sin x\cos x)dx$$

$$= \frac{\pi}{2}\log 2 + \int_0^{\frac{\pi}{4}} \log\sin x\,dx + \int_0^{\frac{\pi}{4}} \log\cos x\,dx$$

$$= \frac{\pi}{2}\log 2 + 2\int_0^{\frac{\pi}{4}} \log\sin x\,dx - 2\int_{\frac{\pi}{2}}^{\frac{\pi}{4}} \log\sin x\,dx$$

$$= \frac{\pi}{2}\log 2 + 2\int_0^{\frac{\pi}{2}} \log\sin x\,dx.$$

Therefore $\displaystyle\int_0^{\frac{\pi}{2}} \log\sin x\,dx = -\frac{\pi}{4}\log 2.$

   The comparison test always concludes that the integral $\displaystyle\int_a^b |f(x)|dx$ con-

verges. In this case, we way that the improper integral $\displaystyle\int_a^b f(x)dx$ *absolutely*

converges. By taking $g(x) = |f(x)|$ in the comparison test, we see that the

absolute convergence of $\displaystyle\int_a^b f(x)dx$ implies the convergence of $\displaystyle\int_a^b f(x)dx$.

However, the converse is not necessarily true, as shown by the next example.

Example 3.3.12. Consider the improper integral $\int_1^{+\infty} \dfrac{\sin x}{x} dx$. The integral of the corresponding absolute value function satisfies

$$\int_{(n-1)\pi}^{n\pi} \left|\frac{\sin x}{x}\right| dx \geq \frac{1}{n\pi} \int_{(n-1)\pi}^{n\pi} |\sin x| dx = \frac{2}{n\pi}.$$

The divergence of $\sum \dfrac{1}{n}$ then implies that $\int_1^a \left|\dfrac{\sin x}{x}\right| dx$ is not bounded. Therefore $\int_1^{+\infty} \left|\dfrac{\sin x}{x}\right| dx$ diverges, and the comparison test cannot be used.

On the other hand, for $a, b > 1$, we use integration by parts to get

$$\int_a^b \frac{\sin x}{x} dx = -\int_a^b \frac{1}{x} d\cos x = -\frac{\sin b}{b} + \frac{\sin a}{a} + \int_a^b \frac{\cos x}{x^2} dx.$$

Therefore if $b > a > \dfrac{1}{\epsilon}$, then

$$\left|\int_a^b \frac{\sin x}{x} dx\right| \leq \left|\frac{\sin b}{b}\right| + \left|\frac{\sin a}{a}\right| + \left|\int_a^b \frac{\cos x}{x^2} dx\right| \leq \frac{1}{b} + \frac{1}{a} + \int_a^b \frac{1}{x^2} dx < 3\epsilon.$$

By the Cauchy criterion, the improper integral $\int_1^{+\infty} \dfrac{\sin x}{x} dx$ converges.

If $\int_a^b f(x) dx$ converges but $\int_a^b |f(x)| dx$ diverges, then we say the im-proper integral $\int_a^b f(x) dx$ *conditionally converges*. The idea used in deriving the (conditional) convergence of the improper integral $\int_1^{+\infty} \dfrac{\sin x}{x} dx$ can be elaborated to become the following tests.

Proposition 3.3.3 (Dirichlet Test). *Suppose $f(x)$ is monotone and satisfies $\lim_{x\to+\infty} f(x) = 0$. Suppose there is M, such that $\left|\int_a^b g(x) dx\right| < M$ for all $b \in [a, +\infty)$. Then $\int_a^{+\infty} f(x)g(x) dx$ converges.*

Proposition 3.3.4 (Abel[6] Test). *Suppose $f(x)$ is monotone and bounded. Suppose $\int_a^{+\infty} g(x) dx$ converges. Then $\int_a^{+\infty} f(x)g(x) dx$ converges.*

Proof. We prove the tests by making the additional assumption that $f(x)$ is differentiable and $f'(x)$ is integrable on $[a, b]$ for any $b \in [a, +\infty)$.

Let $G(x) = \int_a^x g(t) dt$. Using integration by parts, for $b, c > a$, we have

$$\int_b^c f(x)g(x) dx = \int_b^c f(x) dG(x) = f(b)G(b) - f(c)G(c) - \int_b^c G(x)f'(x) dx.$$

---

[6]Niels Henrik Abel, born 1802 in Frindoe (Norway), died 1829 in Froland (Norway). In 1824, Abel proved the impossibility of solving the general equation of fifth degree in radicals. Abel also made contributions to elliptic functions. Abel's name is enshrined in the term *abelian*, which describes the commutative property.

Under the assumption of either test, we have $|G(x)| \leq M$ for a constant M and all $x > a$. Moreover, $f'(x)$ does not change sign. Therefore

$$\left| \int_b^c G(x)f'(x)dx \right| \leq M \left| \int_b^c |f'(x)|dx \right| = M \left| \int_b^c f'(x)dx \right| = M|f(b) - f(c)|.$$

Assuming the condition of the Dirichlet test, for any $\epsilon > 0$, there is N, such that $x > N$ implies $|f(x)| < \epsilon$. Then $b, c > N$ implies

$$\left| \int_b^c f(x)g(x)dx \right| \leq \epsilon|G(b)| + \epsilon|G(c)| + 2M\epsilon \leq 4M\epsilon.$$

This verifies the Cauchy criterion for the convergence of $\int_a^{+\infty} f(x)g(x)dx$.

Assuming the condition of the Abel test, we know both $\lim_{x\to+\infty} f(x)$ and $\lim_{x\to+\infty} f(x)G(x)$ converge. The Cauchy criterion for both limits tells us that for any $\epsilon > 0$, there is N, such that $b, c > N$ implies $|f(b) - f(c)| < \epsilon$ and $|f(b)G(b) - f(c)G(c)| < \epsilon$. Then $b, c > N$ implies

$$\left| \int_b^c f(x)g(x)dx \right| \leq \epsilon + M\epsilon = (M + 1)\epsilon.$$

This also verifies the Cauchy criterion for the convergence of $\int_a^{+\infty} f(x)g(x)dx$.

$\square$

The assumption on the differentiability of $f(x)$ can be removed by using Theorem 3.3.6, which is the more general version of the integration by parts in terms of the Riemann-Stieltjes integral. The relevant inequality used for estimation can be found in Exercises 3.3.24 and 3.3.25. The detailed proof is left as Exercise 3.3.28.

Exercise 3.3.7. Compute improper integrals.

1. $\displaystyle\int_0^\pi \frac{x \sin x dx}{1 - \cos x}.$

2. $\displaystyle\int_0^\pi x \log \sin x dx.$

3. $\displaystyle\int_0^{+\infty} x^n e^{-x}dx.$

4. $\displaystyle\int_0^1 (\log x)^n dx.$

5. $\displaystyle\int_0^1 \frac{x^n dx}{\sqrt{1 - x}}.$

6. $\displaystyle\int_{-\infty}^{+\infty} \frac{dx}{(1 + x^2)^n}.$

Exercise 3.3.8. Determine the convergence of improper integrals. If possible, also find out whether the convergence is absolute or conditional.

1. $\displaystyle\int_2^\infty x^\alpha |\log x|^\beta dx.$

2. $\displaystyle\int_0^1 x^\alpha |\log x|^\beta dx.$

3. $\displaystyle\int_0^{+\infty} x^\alpha a^x dx.$

4. $\displaystyle\int_0^\pi \frac{1}{\sin x}dx.$

5. $\displaystyle\int_0^{+\infty} x^\alpha |\sin x|dx.$

6. $\displaystyle\int_0^1 x^\alpha (1 - x)^\beta dx.$

7. $\displaystyle\int_1^{+\infty} \frac{\sin x}{x^\alpha}dx.$

8. $\displaystyle\int_{-\infty}^{+\infty} \frac{x \sin x}{x^2 + 1}dx.$

9. $\displaystyle\int_0^{+\infty} \sin x^2 dx.$

Exercise 3.3.9. Prove that

$$\int_0^{+\infty} f\left(ax + \frac{b}{x}\right) dx = \frac{1}{a} \int_0^{+\infty} f(\sqrt{x^2 + 4ab})dx,$$

provided $a, b > 0$ and both sides converge.

Exercise 3.3.10. Suppose $f(x) < g(x) < h(x)$. Prove that if $\int_a^b f(x)dx$ and $\int_a^b h(x)dx$ converge, then $\int_a^b g(x)dx$ converges.

Exercise 3.3.11. Prove that if $f(x) \geq 0$ and $\int_a^{+\infty} f(x)dx$ converges, then there is an increasing sequence $\{x_n\}$ diverging to $+\infty$, such that $\lim_{n \to \infty} f(x_n) = 0$. Moreover, prove that in the special case $f(x)$ is monotone, we have $\lim_{x \to +\infty} f(x) = 0$.

Exercise 3.3.12. Suppose $f(x) \geq 0$.

1. If $[a, b]$ is a bounded interval and $\int_a^b f(x)^2 dx$ converges, prove that $\int_a^b f(x)dx$ also converges.

2. If $\lim_{x \to +\infty} f(x) = 0$ and $\int_a^{+\infty} f(x)dx$ converges, prove that $\int_a^{+\infty} f(x)^2 dx$ also converges.

Exercise 3.3.13. Prove that $\lim_{a \to +\infty} a^\beta \int_a^{+\infty} \frac{\sin x dx}{x^\alpha} = 0$ when $\alpha > \beta > 0$. Then prove that $\lim_{a \to 0} \frac{1}{a} \int_0^a \sin \frac{1}{x}dx = 0$.

Exercise 3.3.14. Derive the Abel test from the Dirichlet test.

Exercise 3.3.15. Can you extend the Dirichlet and Abel tests to other kinds of improper integrals, such as unbounded function on bounded interval?

3.3.3 Riemann-Stieltjes Integration

Let α be a function on a bounded interval $[a, b]$. The Riemann sum may be extended to the *Riemann-Stieltjes*[7] *sum*

$$S(P, f, \alpha) = \sum_{i=1}^n f(x_i^*)(\alpha(x_i) - \alpha(x_{i-1})) = \sum_{i=1}^n f(x_i^*)\Delta\alpha_i. \qquad (3.3.1)$$

We say f has *Riemann-Stieltjes integral* (or simply *Stieltjes integral*) I and denote $\int_a^b fd\alpha = I$, if for any $\epsilon > 0$, there is $\delta > 0$, such that

$$\|P\| < \delta \implies |S(P, f, \alpha) - I| < \epsilon. \qquad (3.3.2)$$

When $\alpha(x) = x$, we get the Riemann integral.

---

[7]Thomas Jan Stieltjes, born 1856 in Zwolle (Netherland), died 1894 in Toulouse (France). He is often called the father of the analytic theory of continued fractions and is best remembered for his integral.

Example 3.3.13. Suppose $f(x) = c$ is a constant. Then $S(P, f, \alpha) = c(\alpha(b) - \alpha(a))$. Therefore $\int_a^b c \, d\alpha = c(\alpha(b) - \alpha(a))$.

Example 3.3.14. Suppose $\alpha(x) = \alpha_0$ is a constant. Then $\Delta\alpha_i = 0$. Therefore $\int_a^b f \, d\alpha_0 = 0$. Since f can be arbitrary, we see that Proposition 3.1.2 cannot be extended without additional conditions.

Example 3.3.15. Suppose $a < c < b$ and

$$\alpha(x) = \begin{cases} \alpha_- & \text{if } x < c \\ \alpha_0 & \text{if } x = c \\ \alpha_+ & \text{if } x > c \end{cases}$$

is a step function, in which the three values are not the same. Then

$$S(P, f, \alpha) = \begin{cases} f(x_i^*)(\alpha_+ - \alpha_-) & \text{if } x_{i-1} < c < x_i \\ f(x_{i+1}^*)(\alpha_+ - \alpha_0) + f(x_i^*)(\alpha_0 - \alpha_-) & \text{if } c = x_i \end{cases}.$$

If $f(x)$ is continuous at c, then it is easy to see that $\|P\|$ small implies $|S(P, f, \alpha) - f(c)(\alpha_+ - \alpha_-)|$ small. Therefore $\int_a^b f(x) \, d\alpha = f(c)(\alpha_+ - \alpha_-)$.

If $f(x)$ is not continuous at c and $\alpha_+ \neq \alpha_-$, then we choose P with $x_{i-1} < c < x_i$. The discontinuity at c tells us that there is $\epsilon > 0$, such that no matter how small $\|P\|$ is, there are x, y satisfying $x_{i-1} < x, y < x_i$ and $|f(x) - f(y)| \geq \epsilon$. Taking x and y as x_i^* respectively and keeping all the other x_j^* the same, we get two Riemann-Stieltjes sums. The difference of the two sums is $|(f(x) - f(y))(\alpha_+ - \alpha_-)| \geq \epsilon|\alpha_+ - \alpha_-|$. Therefore $f(x)$ is not Riemann-Stieltjes integrable with respect to α.

If $f(x)$ is not continuous at c and $\alpha_+ = \alpha_-$, then $\alpha_0 \neq \alpha_+ = \alpha_-$ and we choose P with $x_i = c$. Assume $f(x)$ is not left continuous at c. Then there is $\epsilon > 0$, such that no matter how small $\|P\|$ is, there are x, y satisfying $x_{i-1} < x, y \leq x_i = c$ and $|f(x) - f(y)| \geq \epsilon$. Taking x and y as x_i^* respectively and keeping all the other x_j^* the same, we get two Riemann-Stieltjes sums. The difference of the two sums is $|(f(x) - f(y))(\alpha_0 - \alpha_-)| \geq \epsilon|\alpha_0 - \alpha_-|$. Therefore $f(x)$ is not Riemann-Stieltjes integrable with respect to α. The case $f(x)$ is not right continuous at c is similar.

We conclude that f is Riemann-Stieltjes integrable with respect to α if and only if f is continuous at c.

Exercise 3.3.16. Prove that if α and f have a common point of discontinuity in $[a, b]$, then f is not Riemann-Stieltjes integrable with respect to α. You may need the characterization of discontinuity in Exercise 1.4.42.

Exercise 3.3.17. Find suitable α on $[0, 2]$, such that $\int_0^2 f \, d\alpha = f(0) + f(1) + f(2)$ for any continuous f on $[0, 2]$.

Exercise 3.3.18. Prove that if the Dirichlet function is Riemann-Stieltjes integrable with respect α, then α is constant. In particular, this shows that $\int_a^b f \, d\alpha = 0$ for any function α with respect to which f is Riemann-Stieltjes integrable does not necessarily imply that f is zero.

Exercise 3.3.19. Prove that the only function Riemann-Stieltjes integrable with respect to the Dirichlet function is the constant function. In particular, this shows that $\int_a^b f d\alpha = 0$ for any function f that is Riemann-Stieltjes integrable with respect to α does not necessarily imply that α is a constant.

The Riemann-Stieltjes integrability can be verified by Cauchy criterion. The results from Section 3.1.4 remain mostly true. Proposition 3.1.7 can be adopted without much change (see Exercise 3.3.21). Proposition 3.1.8 still holds if α is an increasing function (see Exercise 3.3.24). Proposition 3.1.9 holds in one direction (see Exercises 3.3.22 and 3.3.23).

Exercise 3.3.20. Suppose f is Riemann-Stieltjes integrable with respect to α and β. Suppose c is a constant. Prove that f is Riemann-Stieltjes integrable with respect to $\alpha + \beta$ and $c\alpha$, and

$$\int_a^b f d(\alpha + \beta) = \int_a^b f d\alpha + \int_a^b f d\beta, \quad \int_a^b f d(c\alpha) = c \int_a^b f d\alpha. \qquad (3.3.3)$$

Exercise 3.3.21. Suppose f and g are Riemann-Stieltjes integrable with respect to α. Prove that $f + g$ and cf are Riemann-Stieltjes integrable with respect to α and

$$\int_a^b (f + g) d\alpha = \int_a^b f d\alpha + \int_a^b g d\alpha, \quad \int_a^b c f d\alpha = c \int_a^b f d\alpha. \qquad (3.3.4)$$

Exercise 3.3.22. Suppose $a < b < c$ and f is Riemann-Stieltjes integrable with respect to α on $[a, c]$. Prove that f is Riemann-Stieltjes integrable with respect to α on $[a, b]$ and $[b, c]$, and

$$\int_a^c f d\alpha = \int_a^b f d\alpha + \int_b^c f d\alpha. \qquad (3.3.5)$$

Exercise 3.3.23. Suppose $a < b < c$ and f is Riemann-Stieltjes integrable with respect to α on $[a, b]$ and $[b, c]$. Assume f and α are bounded.

1. Prove that if f is continuous at b, then f is Riemann-Stieltjes integrable on $[a, c]$.

2. Prove that if α is continuous at b, then f is Riemann-Stieltjes integrable on $[a, c]$.

3. Construct functions f and α on $[a, c]$, such that f is Riemann-Stieltjes integrable with respect to α on $[a, b]$ and $[b, c]$, but both f and α are not continuous at b. By Exercise 3.3.16, f is not Riemann-Stieltjes integrable with respect to α on $[a, c]$.

Exercise 3.3.24. Suppose f and g are Riemann-Stieltjes integrable with respect to an increasing α on $[a, b]$. Prove that

$$f \le g \implies \int_a^b f d\alpha \le \int_a^b g d\alpha.$$

Moreover, if α is strictly increasing and f and g are continuous, then the equality holds if and only if $f = g$.

Exercise 3.3.25. Suppose f is Riemann-Stieltjes integrable with respect to an increasing α on $[a, b]$. Prove that

$$\left| \int_a^b f d\alpha \right| \le \int_a^b |f| d\alpha,$$

$$(\alpha(b) - \alpha(a)) \inf_{[a,b]} f \le \int_a^b f d\alpha \le (\alpha(b) - \alpha(a)) \sup_{[a,b]} f,$$

$$\left| f(c)(\alpha(b) - \alpha(a)) - \int_a^b f d\alpha \right| \le \omega_{[a,b]}(f)(\alpha(b) - \alpha(a)),$$

$$\left| S(P, f, \alpha) - \int_a^b f d\alpha \right| \le \sum \omega_{[x_{i-1}, x_i]}(f) \Delta \alpha_i.$$

Moreover, extend the first integral mean theorem in Exercise 3.1.15 to Riemann-Stieltjes integral.

Exercise 3.3.26. Suppose f is Riemann-Stieltjes integrable with respect to α, and $F(x) = \int_a^x f d\alpha$. Prove that if α is strictly monotone and $f(x)$ is continuous at x_0, then $\lim_{h \to 0} \dfrac{F(x_0 + h) - F(x_0)}{\alpha(x_0 + h) - \alpha(x_0)} = f(x_0)$. This extends the fundamental theorem of calculus.

The Riemann-Stieltjes integral can often be computed by the ordinary Riemann integral.

Theorem 3.3.5. *Suppose f is bounded, β is Riemann integrable and $\alpha(x) = \int_a^x \beta(t)dt$. Then f is Riemann-Stieltjes integrable with respect to α if and only if $f\beta$ is Riemann integrable. Moreover,*

$$\int_a^b f d\alpha = \int_a^b f\beta dx.$$

For the special case β is continuous, we have $\alpha' = \beta$ and the formula becomes

$$\int_a^b f d\alpha = \int_a^b f\alpha' dx. \tag{3.3.6}$$

This provides the rule for moving α from inside d to outside d, which is consistent with the notation $d\alpha = \alpha' dx$ for the differential.

Proof. Suppose $|f| < B$ for a constant B. Since β is Riemann integrable, for any $\epsilon > 0$, there is $\delta > 0$, such that $\|P\| < \delta$ implies $\sum \omega_{[x_{i-1}, x_i]}(\beta) \Delta x_i < \epsilon$. For the same choice of P and x_i^*, the Riemann-Stieltjes sum of f with respect to α is

$$S(P, f, \alpha) = \sum f(x_i^*)(\alpha(x_i) - \alpha(x_{i-1})) = \sum f(x_i^*) \int_{x_{i-1}}^{x_i} \beta(t)dt,$$

and the Riemann sum of $f\beta$ is

$$S(P, f\beta) = \sum f(x_i^*) \beta(x_i^*) \Delta x_i.$$

Then in case $\|P\| < \delta$, we have

$$|S(P, f, \alpha) - S(P, f\beta)| \leq \sum |f(x_i^*)| \left| \int_{x_{i-1}}^{x_i} \beta(t)dt - \beta(x_i^*)\Delta x_i \right|$$

$$\leq \sum B\omega_{[x_{i-1}, x_i]}(\beta)\Delta x_i \leq B\epsilon.$$

This implies that the Riemann-Stieltjes sum of f with respect to α converges if and only if the Riemann sum of $f\beta$ converges. Moreover, the two limits are the same.

$\square$

The relation between the Riemann-Stieltjes integral and the ordinary Riemann integral suggests that the integration by parts and the change of variable formulae can be extended. The integration by parts for the Riemann-Stieltjes integral turns out to be a lot more symmetric and elegant.

Theorem 3.3.6. *Suppose f is Riemann-Stieltjes integrable with respect to α. Then α is Riemann-Stieltjes integrable with respect to f. Moreover,*

$$\int_a^b f d\alpha + \int_a^b \alpha df = f(b)\alpha(b) - f(a)\alpha(a). \tag{3.3.7}$$

Proof. Let

$$P: a = x_0 < x_1 < x_2 < \cdots < x_n = b$$

be a partition of $[a, b]$ and $x_i^* \in [x_{i-1}, x_i]$ are chosen for $1 \leq i \leq n$. Consider

$$Q: a = x_0^* \leq x_1^* \leq x_2^* \leq \cdots \leq x_n^* \leq x_{n+1}^* = b.$$

Q is almost a partition except some repetition may happen among the partition points. By choosing $x_{i-1} \in [x_{i-1}^*, x_i^*]$ for $1 \leq i \leq n+1$, we still use the notation

$$S(Q, f, \alpha) = \sum_{i=1}^{n+1} f(x_{i-1})(\alpha(x_i^*) - \alpha(x_{i-1}^*))$$

to denote the Riemann-Stieltjes sum of f with respect to α. In case a repetition $x_i^* = x_{i-1}^*$ happens, the corresponding term in the sum simply vanishes. Therefore $S(Q, f, \alpha)$ is the same as the Riemann-Stieltjes sum after removing all the repetitions and we may pretend there is no repetition in $S(Q, f, \alpha)$ in the subsequent argument. In particular, by the integrability of f with respect to α, for any $\epsilon > 0$, there is δ, such that

$$\|Q\| < \delta \implies \left| S(Q, f, \alpha) - \int_a^b f d\alpha \right| < \epsilon.$$

Since $x_i^* - x_{i-1}^* \leq x_i - x_{i-2} \leq 2\|P\|$, we have $\|Q\| \leq 2\|P\|$ and

$$\|P\| < \frac{\delta}{2} \implies \left| S(Q, f, \alpha) - \int_a^b f d\alpha \right| < \epsilon.$$

Note that (see Figure 4.1 for the geometric meaning)

$$
\begin{aligned}
S(Q, f, \alpha) &= \sum_{i=1}^{n+1} f(x_{i-1})\alpha(x_i^*) - \sum_{i=1}^{n+1} f(x_{i-1})\alpha(x_{i-1}^*) \\
&= \sum_{i=1}^{n} f(x_{i-1})\alpha(x_i^*) + f(b)\alpha(b) - \sum_{i=2}^{n+1} f(x_{i-1})\alpha(x_{i-1}^*) - f(a)\alpha(a) \\
&= \sum_{i=1}^{n} f(x_{i-1})\alpha(x_i^*) - \sum_{i=1}^{n} f(x_i)\alpha(x_i^*) + f(b)\alpha(b) - f(a)\alpha(a) \\
&= -S(P, \alpha, f) + f(b)\alpha(b) - f(a)\alpha(a).
\end{aligned}
$$

Therefore we have

$$
\|P\| < \frac{\delta}{2} \implies \left| S(P, \alpha, f) - \left(f(b)\alpha(b) - f(a)\alpha(a) - \int_a^b f d\alpha \right) \right| < \epsilon.
$$

This proves that α is Riemann-Stieltjes integrable with respect to f, and the formula (3.3.7) holds.

$\square$

Exercise 3.3.27. Suppose $f(x)$ is a non-negative and decreasing function on $[a, b]$ that is Riemann-Stieltjes integrable with respect to α. Suppose $m \le \alpha \le M$ on $[a, b]$. Prove

$$
f(a)(m - \alpha(a)) \le \int_a^b f d\alpha \le f(a)(M - \alpha(a))
$$

and derive the second integral mean value theorem in Exercises 3.2.58 and 3.2.59 without assuming the differentiability.

Exercise 3.3.28. Prove the general version of the Dirichlet and Abel tests (Propositions 3.3.3 and 3.3.4) without the differentiability assumption on $f(x)$.

In view of Theorem 3.3.5, the following extends Theorem 3.2.5 for the change of variable.

Theorem 3.3.7. *Suppose ϕ is increasing and continuous on $[a, b]$. Suppose f is Riemann-Stieltjes integrable with respect to α on $[\phi(a), \phi(b)]$. Then $f \circ \phi$ is Riemann-Stieltjes integrable with respect to $\alpha \circ \phi$ on $[a, b]$. Moreover,*

$$
\int_{\phi(a)}^{\phi(b)} f d\alpha = \int_a^b (f \circ \phi) d(\alpha \circ \phi). \tag{3.3.8}
$$

Proof. Let P be a partition of $[a, b]$. Similar to the proof of Theorem 3.2.5, we have a partition $\phi(P)$ of $[\phi(a), \phi(b)]$. Choose x_i^* for P and choose corresponding $\phi(x_i^*)$ for $\phi(P)$. Then the Riemann-Stieltjes sum of $f \circ \phi$ with respect to $\alpha \circ \phi$ is

$$
S(P, f \circ \phi, \alpha \circ \phi) = \sum f(\phi(x_i^*))(\alpha(\phi(x_i)) - \alpha(\phi(x_{i-1}))) = S(\phi(P), f, \alpha).
$$

Since f is Riemann-Stieltjes integrable with respect to α, for any $\epsilon > 0$, there is $\delta > 0$, such that

$$
\|Q\| < \delta \implies \left| S(Q, f, \alpha) - \int_{\phi(a)}^{\phi(a)} f d\alpha \right| < \epsilon.
$$

The continuity of ϕ implies the uniform continuity. Therefore there is $\delta' > 0$, such that $\|P\| < \delta'$ implies $\|\phi(P)\| < \delta$. Then $\|P\| < \delta'$ implies

$$\left| S(P, f \circ \phi, \alpha \circ \phi) - \int_{\phi(a)}^{\phi(a)} f d\alpha \right| = \left| S(\phi(P), f, \alpha) - \int_{\phi(a)}^{\phi(a)} f d\alpha \right| < \epsilon.$$

This proves that $f \circ \phi$ is Riemann-Stieltjes integrable with respect to $\alpha \circ \phi$ and the formula (3.3.8) holds.

$\square$

Exercise 3.3.29. What will happen to Theorem 3.3.7 if ϕ is decreasing? What if $\phi(x)$ is not assumed to be continuous?

Exercise 3.3.30. Suppose f and g are increasing functions satisfying $g(f(x)) = x$ and $f(0) = 0$. Prove that if f is continuous, then for any $a, b > 0$, we have the *Young inequality*

$$\int_0^a f(x) dx + \int_0^b g(y) dy \geq ab.$$

3.3.4 Bounded Variation Function

The discussion of the integrability in Section 3.1.2 critically depends on the inequality (3.1.6). The inequality may be extended to the Riemann-Stieltjes sum

$$|f(c)(\alpha(b) - \alpha(a)) - S(P, f, \alpha)| \leq \omega_{[a,b]}(f) V_P(\alpha), \qquad (3.3.9)$$

where

$$V_P(\alpha) = |\alpha(x_0) - \alpha(x_1)| + |\alpha(x_1) - \alpha(x_2)| + \cdots + |\alpha(x_{n-1}) - \alpha(x_n)|$$

is the *variation* of α with respect to the partition P. We say α has *bounded variation* if there is a constant V, such that $V_P(\alpha) \leq V$ for any partition P. If a partition Q refines the partition P, then we clearly have

$$V_Q(\alpha) \geq V_P(\alpha).$$

This suggests us to define the *variation* of α on an interval to be

$$V_{[a,b]}(\alpha) = \sup\{V_P(\alpha) \colon P \text{ is a partition of } [a, b]\}.$$

It is easy to see that monotone functions have bounded variations with $V_{[a,b]}(\alpha) = |\alpha(b) - \alpha(a)|$, and linear combinations of bounded variation functions also have bounded variations. Moreover, it is easy to establish the following properties for the variations.

Proposition 3.3.8. *Suppose α and β are bounded variation functions on $[a, b]$. Then*

$$V_{[a,b]}(\alpha) \geq |\alpha(b) - \alpha(a)|,$$
$$V_{[a,c]}(\alpha) = V_{[a,b]}(\alpha) + V_{[b,c]}(\alpha) \ \text{ for } \ a < b < c,$$
$$V_{[a,b]}(\lambda\alpha) = |\lambda| V_{[a,b]}(\alpha),$$
$$V_{[a,b]}(\alpha + \beta) \leq V_{[a,b]}(\alpha) + V_{[a,b]}(\beta).$$

Proof. The difference $|\alpha(b) - \alpha(a)|$ is the variation with respect to the partition $a = x_0 < x_1 = b$. Therefore the first inequality follows from the definition of the variation $V_{[a,b]}(\alpha)$.

For any partition P of $[a, c]$, we denote by $P \cup \{b\}$ the partition of $[a, b]$ obtained by adding c to P, and denote by $P_{[x,y]}$ the partition of $[x, y]$ obtained by combining x, y and those points in P lying in $[x, y]$. Then

$$V_{P_{[a,b]}}(\alpha) + V_{P_{[b,c]}}(\alpha) = V_{P \cup \{b\}}(\alpha) \geq V_P(\alpha).$$

This implies $V_{[a,c]}(\alpha) \leq V_{[a,b]}(\alpha) + V_{[b,c]}(\alpha)$. Conversely, for any partitions P' of $[a, b]$ and P'' of $[b, c]$, we denote by $P' \cup P''$ the partition of $[a, c]$ obtained by combining points in P' and P'' together. Then

$$V_P(\alpha) = V_{P'}(\alpha) + V_{P''}(\alpha).$$

This implies that $V_{[a,c]}(\alpha) \geq V_{[a,b]}(\alpha) + V_{[b,c]}(\alpha)$. The completes the proof of the second equality.

The third equality follows from $V_P(\lambda\alpha) = |\lambda|V_P(\alpha)$.

The forth inequality follows from $V_P(\alpha + \beta) \leq V_P(\alpha) + V_P(\beta)$.

$\square$

Example 3.3.16. Suppose β is Riemann integrable and $\alpha(x) = \displaystyle\int_a^x \beta(t)dt$. Then

$V_P(\alpha) = \sum \left| \displaystyle\int_{x_{i-1}}^{x_i} \beta(t)dt \right|$, and by (3.1.13) in Exercise 3.1.16,

$$|S(P, |\beta|) - V_P(\alpha)| \leq \sum \left| |\beta(x_i^*)|\Delta x_i - \left| \int_{x_{i-1}}^{x_i} \beta(t)dt \right| \right|$$

$$\leq \sum \left| \beta(x_i^*)\Delta x_i - \int_{x_{i-1}}^{x_i} \beta(t)dt \right|$$

$$\leq \sum \omega_{[x_{i-1},x_i]}(\beta)\Delta x_i.$$

This implies α has bounded variation, and

$$V_{[a,b]}(\alpha) = \int_a^b |\beta(t)|dt.$$

Exercise 3.3.31. Prove that any Lipschitz function has bounded variation.

Exercise 3.3.32. Prove that any bounded variation function is Riemann integrable.

Exercise 3.3.33. Suppose a function α has bounded variation on $[a, b]$. Prove that α is increasing if and only if $V_{[a,b]}(\alpha) = \alpha(b) - \alpha(a)$.

Exercise 3.3.34. Suppose f is Riemann-Stieltjes integrable with respect to α. Suppose $a \leq c \leq b$. Prove that

$$\left| \int_a^b fd\alpha \right| \leq \sup_{[a,b]} |f|V_{[a,b]}(\alpha),$$

$$\left| f(c)(\alpha(b) - \alpha(a)) - \int_a^b fd\alpha \right| \leq \omega_{[a,b]}(f)V_{[a,b]}(\alpha).$$

Exercise 3.3.35. Suppose f is Riemann-Stieltjes integrable with respect to a bounded variation function α, and $F(x) = \int_a^x f d\alpha$. Prove that if α is not continuous at x_0 and $f(x_0) \neq 0$, then F is not continuous at x_0.

For a function $\alpha(x)$ with bounded variation on $[a, b]$, define the *variation function*

$$v(x) = V_{[a,x]}(\alpha).$$

Intuitively, the change of α is positive when α is increasing and negative when α is decreasing. The variation function $v(x)$ keeps track of both as positive changes. Then we may further define the *positive variation function* and the *negative variation function*

$$v^+ = \frac{v + \alpha}{2}, \quad v^- = \frac{v - \alpha}{2}.$$

Intuitively, $v^+(x)$ keeps track of the positive changes only and becomes constant on the intervals on which α is decreasing, and $v^-(x)$ keeps track of the negative changes only. We have $\alpha = v^+ - v^-$. Moreover, for $a \leq x < y \leq b$, by Proposition 3.3.8, we have

$$v^+(y) - v^+(x) = \frac{V_{[x,y]}(\alpha) + \alpha(y) - \alpha(x)}{2} \geq \frac{V_{[x,y]}(\alpha) - |\alpha(y) - \alpha(x)|}{2} \geq 0.$$

Thus $v^+(x)$ is increasing. Similarly, $v^-(x)$ is also increasing. Therefore we proved the necessary part of the following.

Proposition 3.3.9. *A function has bounded variation if and only if it is the difference of two increasing functions.*

The sufficiency follows from the fact that monotone functions have bounded variations and linear combinations of bounded variation functions still have bounded variations.

Exercise 3.3.36. Suppose β is Riemann integrable and $\alpha(x) = \int_a^x \beta(t) dt$. Prove that $v^+(x) = \int_a^x \max\{0, \beta(t)\} dt$, $v^-(x) = -\int_a^x \min\{0, \beta(t)\} dt$.

Exercise 3.3.37. Suppose a function α has bounded variation, with positive and negative variation functions v^+ and v^-. Suppose $\alpha = u^+ - u^-$, where u^+ and u^- are increasing functions.

1. Prove that $V_{[a,b]}(\alpha) \leq (u^+(b) - u^+(a)) + (u^-(b) - u^-(a))$.

2. Prove that $v^+(y) - v^+(x) \leq u^+(y) - u^+(x)$ and $v^-(y) - v^-(x) \leq u^-(y) - u^-(x)$ for any $a \leq x < y \leq b$.

3. Prove that the equality in the first part holds if and only if $u^+ = v^+ + c$ and $u^- = v^- + c$ for some constant c.

The result shows that $\alpha = v^+ - v^-$ is the "most efficient" way of expressing a bounded variation function as the difference of increasing functions.

Next we establish properties of continuous bounded variation functions.

Proposition 3.3.10. *A bounded variation function is continuous if and only if its variation function is continuous.*

Proof. Let α be a bounded variation function on $[a, b]$ and $v(x) = V_{[a,x]}(\alpha)$.

By Proposition 3.3.8, for $a \leq x < y \leq b$, we have

$$|\alpha(y) - \alpha(x)| \leq V_{[x,y]}(\alpha) = v(y) - v(x).$$

This implies that if v is continuous, then α is also continuous.

Conversely, for any $\epsilon > 0$, there is a partition P of $[a, b]$, such that

$$V_P(\alpha) > V_{[a,b]}(\alpha) - \epsilon = v(b) - \epsilon.$$

If α is continuous, then there is $\delta > 0$, such that $a \leq x < y \leq b$ and $y - x < \delta$ implies $|\alpha(y) - \alpha(x)| < \epsilon$. Let $\delta_P = \min \Delta x_i$ be the length of the smallest interval in P. Then for $a \leq x < y \leq b$ satisfying $y - x < \delta' = \min\{\delta, \delta_P\}$, the interval (x, y) contains at most one point from P. Denote by $Q = P \cup \{x, y\}$ the partition of $[a, b]$ obtained by adding x, y to P. Denote by $Q_{[x,y]}$ the partition of $[x, y]$ obtained by combining x, y and those points in P lying in $[x, y]$ (as in the proof of Proposition 3.3.8). Then

$$V_{Q_{[a,x]}}(\alpha) + V_{Q_{[x,y]}}(\alpha) + V_{Q_{[y,b]}}(\alpha) = V_Q(\alpha) \geq V_P(\alpha) > V_{[a,b]}(\alpha) - \epsilon$$
$$= V_{[a,x]}(\alpha) + V_{[x,y]}(\alpha) + V_{[y,b]}(\alpha) - \epsilon.$$

By $V_{Q_{[a,x]}}(\alpha) \leq V_{[a,x]}(\alpha)$ and $V_{Q_{[y,b]}}(\alpha) \leq V_{[y,b]}(\alpha)$, we get

$$V_{Q_{[x,y]}}(\alpha) \geq V_{[x,y]}(\alpha) - \epsilon.$$

Since (x, y) contains at most one point from P. The partition $Q_{[x,y]}$ is either $x < y$ or $x < c < y$ for some $c \in P$. Since $y - x < \delta$, we have either

$$V_{Q_{[x,y]}}(\alpha) = |\alpha(y) - \alpha(x)| < \epsilon,$$

or

$$V_{Q_{[x,y]}}(\alpha) = |\alpha(y) - \alpha(c)| + |\alpha(c) - \alpha(x)| < 2\epsilon.$$

Therefore in either case, we get

$$v(y) - v(x) = V_{[x,y]}(\alpha) \leq V_{Q_{[x,y]}}(\alpha) + \epsilon \leq 3\epsilon.$$

Since the only condition for this to hold is $y - x < \delta'$, this proves the continuity of v.

$\square$

The following is a technical result about continuous bounded variation functions that will be useful for proving further results.

Proposition 3.3.11. *Suppose α is a continuous bounded variation function. Then for any $\epsilon > 0$, that there is $\delta > 0$, such that $\|P\| < \delta$ implies $V_P(\alpha) > V_{[a,b]}(\alpha) - \epsilon$.*

Proof. Example 3.3.16 shows that the close relation between the variation and the Riemann integral. The subsequent proof is similar to Exercises 3.1.38 and 3.1.39.

By definition, there is a partition Q, such that $V_Q(\alpha) > V_{[a,b]}(\alpha) - \epsilon$. Suppose Q contains n partition points. By the continuity of α, there is $\delta > 0$, such that

$$|x - y| < \delta \implies |\alpha(x) - \alpha(y)| < \frac{\epsilon}{2n}.$$

Now assume P is any partition satisfying $\|P\| < \delta$. Then the oscillations of α on the intervals in P are $< \frac{\epsilon}{2n}$. Denote by $P \cup Q$ the partition of $[a, b]$ obtained by combining points of P and Q together. Then

$$V_{P \cup Q}(\alpha) \geq V_Q(\alpha) > V_{[a,b]}(\alpha) - \epsilon.$$

On the other hand, $P \cup Q$ is obtained from P by adding no more than n points. These n points fall into at most n intervals in P, and the sums for $V_{P \cup Q}(\alpha)$ and $V_P(\alpha)$ differ only at these intervals. Specifically, if $[x_{i-1}, x_i]$ is one such interval that contains k points from Q, then the term $|\alpha(x_i) - \alpha(x_{i-1})|$ in $V_P(\alpha)$ is replaced by a sum of at most $k + 1$ terms in $V_{P \cup Q}(\alpha)$. Since $\omega_{[x_{i-1}, x_i]}(\alpha) < \frac{\epsilon}{2n}$, the sum of these $k+1$ terms is $< (k+1)\frac{\epsilon}{2n}$. Adding up all these differences, we get (note that $\sum (k_i + 1) \leq 2n$)

$$V_{P \cup Q}(\alpha) < V_P(\alpha) + 2n\frac{\epsilon}{2n} = V_P(\alpha) + \epsilon.$$

Combining the two inequalities together, we conclude that

$$\|P\| < \delta \implies V_P(\alpha) + \epsilon > V_{[a,b]}(\alpha) - \epsilon.$$

$\square$

Exercise 3.3.38. Given the conclusion of Proposition 3.3.11, prove that for any partition P of any interval $[c, d] \subset [a, b]$ satisfying $\|P\| < \delta$, we have $V_P(\alpha) > V_{[c,d]}(\alpha) - \epsilon$. Use this observation to give another proof of Proposition 3.3.10.

Exercise 3.3.39. Suppose f is Riemann-Stieltjes integrable with respect to α, and $F(x) = \int_a^x f d\alpha$. Prove that if α is continuous and has bounded variation, then F is continuous. Compare with the Exercise 3.3.35.

Now we are ready to extend Theorem 3.1.3 on the criterion for Riemann integrabilty to the Riemann-Stieltjes integrability.

Theorem 3.3.12. *Suppose f is bounded and α has bounded variation.*

1. *If f is Riemann-Stieltjes integrable with respect to α, then for any $\epsilon > 0$, there is $\delta > 0$, such that*

$$\|P\| < \delta \implies \sum \omega_{[x_{i-1}, x_i]}(f)|\Delta\alpha_i| < \epsilon.$$

2. *If for any $\epsilon > 0$, there is $\delta > 0$, such that*

$$\|P\| < \delta \implies \sum \omega_{[x_{i-1}, x_i]}(f) V_{[x_{i-1}, x_i]}(\alpha) < \epsilon,$$

then f is Riemann-Stieltjes integrable with respect to α.

Moreover, if α is monotone, then the two parts are inverse to each other and give a necessary and sufficient condition for the Riemann-Stieltjes integrability. If α is continuous, then the converse of the second part is also true and gives a necessary and sufficient condition for the Riemann-Stieltjes integrability.

Proof. The proof is very similar to the proof of Theorem 3.1.3. For the first part, we note that by taking $P = P'$ but choosing different x_i^* for P and P', we have

$$S(P, f, \alpha) - S(P', f, \alpha) = \sum (f(x_i^*) - f(x_i'^*)) \Delta \alpha_i.$$

Then for each i, we may choose $f(x_i^*) - f(x_i'^*)$ to have the same sign as $\Delta \alpha_i$ and $|f(x_i^*) - f(x_i'^*)|$ to be as close to the oscillation $\omega_{[x_{i-1}, x_i]}(f)$ as possible. The result is that $S(P, f, \alpha) - S(P', f, \alpha)$ is very close to $\sum \omega_{[x_{i-1}, x_i]}(f)|\Delta \alpha_i|$. The rest of the proof is the same.

For the second part, the key is the following estimation for a refinement Q of P.

$$
\begin{aligned}
|S(P, f, \alpha) - S(Q, f, \alpha)| &= \left| \sum_{i=1}^{n} f(x_i^*) \Delta \alpha_i - \sum_{i=1}^{n} S(Q_{[x_{i-1}, x_i]}, f, \alpha) \right| \\
&\leq \sum_{i=1}^{n} \left| f(x_i^*)(\alpha(x_{i+1}) - \alpha(x_i)) - S(Q_{[x_{i-1}, x_i]}, f, \alpha) \right| \\
&\leq \sum_{i=1}^{n} \omega_{[x_{i-1}, x_i]}(f) V_{Q_{[x_{i-1}, x_i]}}(\alpha) \\
&\leq \sum_{i=1}^{n} \omega_{[x_{i-1}, x_i]}(f) V_{[x_{i-1}, x_i]}(\alpha),
\end{aligned}
$$

where the second inequality follows from (3.3.9). The rest of the proof is the same.

If α is monotone, then $|\Delta \alpha_i| = V_{[x_{i-1}, x_i]}(\alpha)$, so that the two parts are inverse to each other.

If α is continuous, then by Proposition 3.3.11, for any $\epsilon > 0$, there is $\delta > 0$, such that $\|P\| < \delta$ implies $V_P(\alpha) > V_{[a,b]}(\alpha) - \epsilon$. Now for Riemann-Stieltjes integrable f, by the first part, there is $\delta' > 0$, such that $\|P\| < \delta'$ implies $\sum \omega_{[x_{i-1}, x_i]}(f)|\Delta \alpha_i| < \epsilon$. If $|f| < B$, then $\|P\| < \min\{\delta, \delta'\}$ implies

$$
\begin{aligned}
&\sum \omega_{[x_{i-1}, x_i]}(f) V_{[x_{i-1}, x_i]}(\alpha) \\
\leq &\sum \omega_{[x_{i-1}, x_i]}(f)|\Delta \alpha_i| + \sum \omega_{[x_{i-1}, x_i]}(f)(V_{[x_{i-1}, x_i]}(\alpha) - |\Delta \alpha_i|) \\
\leq &\epsilon + 2B \sum (V_{[x_{i-1}, x_i]}(\alpha) - |\Delta \alpha_i|) = \epsilon + 2B(V_{[a,b]}(\alpha) - V_P(\alpha)) < (2B+1)\epsilon.
\end{aligned}
$$

$\square$

With the help of the theorem, the results in Section 3.1.3 may be extended. The proofs are left as exercises.

Proposition 3.3.13. *Any continuous function is Riemann-Stieltjes integrable with respect to any function with bounded variation.*

By Theorem 3.3.6, we have the following result that extends Proposition 3.1.5.

Proposition 3.3.14. *Any function with bounded variation is Riemann-Stieltjes integrable with respect to any continuous function.*

If α is monotone or continuous, then Theorem 3.3.12 gives us necessary and sufficient criteria for the Riemann-Stieltjes integrability. Based on this, Proposition 3.1.6 can be proved just as before.

Proposition 3.3.15. *Suppose f is Riemann-Stieltjes integrable with respect to α, which is either monotone or continuous with bounded variation. Suppose the values of f lie in a finite union U of closed intervals, and ϕ is a continuous function on U. Then the composition $\phi \circ f$ is also Riemann-Stieltjes integrable with respect to α.*

By the same reason as Riemann integration, under the assumption about α in Proposition 3.3.15, the products of functions Riemann-Stieltjes integrable with respect to α are still Riemann-Stieltjes integrable with respect to α.

3.3.5 Additional Exercise

Trigonometric Integration

Exercise 3.3.40. By writing $a\sin x + b\cos x$ as a linear combination of $c\sin x + d\cos x$ and $(c\sin x + d\cos x)'$, compute the antiderivatives $\int \dfrac{a\sin x + b\cos x}{c\sin x + d\cos x}dx$ and $\int \dfrac{a\sin x + b\cos x}{(c\sin x + d\cos x)^2}dx$. Use the similar idea to compute the antiderivative $\int \dfrac{a\sin x + b\cos x + \lambda}{c\sin x + d\cos x + \mu}dx$.

Exercise 3.3.41. By writing $Au^2 + 2Buv + Cv^2$ in the form $(\alpha u + \beta v)(au + bv) + \gamma(u^2 + v^2)$, compute the antiderivative $\int \dfrac{A\sin^2 x + 2B\sin x\cos x + C\cos^2 x}{a\sin x + b\cos x}dx$.

Gamma Function

The *Gamma function* is

$$\Gamma(x) = \int_0^{+\infty} t^{x-1}e^{-t}dt.$$

Exercise 3.3.42. Prove that the function is defined and continuous for $x > 0$.

Exercise 3.3.43. Prove $\lim_{x\to 0^+} \Gamma(x) = \lim_{x\to +\infty} \Gamma(x) = +\infty$.

Exercise 3.3.44. Prove the other formulae for the Gamma function

$$\Gamma(x) = 2\int_0^\infty t^{2x-1}e^{-t^2}\,dt = a^x\int_0^\infty t^{x-1}e^{-at}\,dt.$$

Exercise 3.3.45. Prove the equalities for the Gamma function

$$\Gamma(x+1) = x\Gamma(x), \;\; \Gamma(n) = (n-1)!.$$

Beta Function

The *Beta function* is

$$B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,dt.$$

In Exercise 7.1.69, we will see that the Beta function can be written in terms of the Gamma function.

Exercise 3.3.46. Prove that the function is defined for $x, y > 0$.

Exercise 3.3.47. Use the change of variables $t = \dfrac{1}{1+u}$ to prove the other formulae for the Beta function

$$B(x,y) = \int_0^\infty \frac{t^{y-1}dt}{(1+t)^{x+y}} = \int_0^1 \frac{t^{x-1} + t^{y-1}}{(1+t)^{x+y}}\,dt.$$

Exercise 3.3.48. Prove

$$B(x,y) = 2\int_0^{\frac{\pi}{2}} \cos^{2x-1} t \sin^{2y-1} t\,dt.$$

Exercise 3.3.49. Prove the equalities for the Beta function

$$B(x,y) = B(x,y), \;\; B(x+1,y) = \frac{x}{x+y}B(x,y).$$

Chapter 4

Series

4.1   Series of Numbers

When n approaches infinity, the Taylor expansion becomes a series

$$T(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + \cdots .$$

This can be considered as the infinite degree approximation of $f(x)$ at x_0. The value of the series would be a limit, and the natural question is whether the value is equal to the function itself.

Another important and useful series is the Fourier[1] series

$$\frac{1}{2}a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + \cdots + a_n \cos nx + b_n \sin nx + \cdots$$

defined for an integrable periodic function $f(x)$ with period 2π, where

$$a_n = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos nx dx, \ \ b_n = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin nx dx.$$

Again the central question here is whether the limit of the series is equal to the function itself.

The Taylor series and the Fourier series are series of functions. In this part, we discuss the more elementary theory of the series of numbers. The series of functions will be discussed in the next part.

4.1.1   Sum of Series

A *series* is an infinite sum

$$\sum_{n=1}^{\infty} x_n = x_1 + x_2 + \cdots + x_n + \cdots .$$

The *partial sum* of the series is the sequence

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

Definition 4.1.1. A series $\sum_{n=1}^{\infty} x_n$ has *sum* s, and denoted $\sum_{n=1}^{\infty} x_n = s$, if $\lim_{n\to\infty} s_n = s$.

If the limit exists, then the series *converges*. Otherwise, the series *diverges*. If $\lim_{n\to\infty} s_n = \infty$, we also say the series diverges to infinity and write $\sum_{n=1}^{\infty} x_n = \infty$.

The Cauchy criterion for the convergence of sequences leads immediately to the Cauchy criterion for the convergence of series: A series $\sum_{n=1}^{\infty} x_n$ converges if and only for any $\epsilon > 0$, there is N, such that

$$n \geq m > N \implies |x_m + x_{m+1} + \cdots + x_n| < \epsilon. \tag{4.1.1}$$

---

[1]Jean Baptiste Joseph Fourier, born 1768 in Bourgogne (France), died 1830 in Paris (France).

For the special case that $m = n$, we find that a necessary condition for a sequence $\sum_{n=1}^{\infty} x_n$ to converge is

$$\lim_{n \to \infty} x_n = 0.$$

Like sequences, a series does not have to start with index 1. On the other hand, without loss of generality, we can always assume that a series starts with index 1 in theoretical studies. Moreover, modifying, adding or deleting finitely many terms in a series does not change the convergence of the series.

Example 4.1.1. The *geometric series* is $\sum_{n=0}^{\infty} a^n = 1 + a + a^2 + \cdots + a^n + \cdots$. The partial sum satisfies

$$(1 - a)s_n = (1 + a + a^2 + \cdots + a^n) - (a + a^2 + a^3 + \cdots + a^{n+1}) = 1 - a^{n+1}.$$

Therefore $s_n = \dfrac{1 - a^{n+1}}{1 - a}$, and

$$\sum_{n=0}^{\infty} a^n = \begin{cases} \dfrac{1}{1 - a} & \text{if } |a| < 1 \\ \text{diverges} & \text{if } |a| \geq 1 \end{cases}. \tag{4.1.2}$$

Example 4.1.2. If $x_n \geq 0$, then the partial sum is an increasing sequence. Therefore the convergence of the *non-negative series* $\sum x_n$ is equivalent to the boundedness of the partial sums. This is the reason behind the divergence of the *harmonic series* $\sum_{n=1}^{\infty} \dfrac{1}{n}$ and the convergence of the series $\sum_{n=1}^{\infty} \dfrac{1}{n^2}$ in Example 1.2.9. We also note that the computation in Example 1.2.9 tells us

$$\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = \frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \cdots + \frac{1}{(n-1)n} + \cdots = \lim_{n \to \infty} \left(\frac{1}{1} - \frac{1}{n} \right) = 1.$$

Example 4.1.3. In Example 2.3.16, the estimation of the remainder of the Taylor series tells us that $\sum_{n=0}^{\infty} \dfrac{x^n}{n!}$ converges to e^x for any x. In particular, we have

$$1 + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} + \cdots = e.$$

Exercise 4.1.1. Compute the sums of convergent series.

1. $\sum_{n=1}^{\infty} na^n$.

2. $\sum_{n=1}^{\infty} \dfrac{n}{2n - 1}$.

3. $\sum_{n=1}^{\infty} \dfrac{(-1)^n}{(2n)!}$.

4. $\sum_{n=1}^{\infty} \dfrac{(-1)^n}{(2n + 1)!}$.

5. $\sum_{n=2}^{\infty} \log \left(1 - \dfrac{1}{n^2} \right)$.

6. $\sum_{n=1}^{\infty} \dfrac{1}{\sqrt[n]{a}}$.

7. $\sum_{n=1}^{\infty} \dfrac{1}{(a + n)(a + n + 1)}$.

8. $\sum_{n=1}^{\infty} \dfrac{1}{n(n + 1)(n + 2)}$.

Exercise 4.1.2. Use Exercise 1.4.38 to show

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} - \frac{1}{8} + \cdots = \log 2,$$

$$1 + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} + \frac{1}{7} - \frac{1}{4} + \frac{1}{9} + \frac{1}{11} - \frac{1}{6} + \cdots = \frac{3}{2} \log 2.$$

Note that the second series is obtained from the first by rearranging the terms.

Exercise 4.1.3. Prove that if y is not an integer multiple of 2π, then

$$\sum_{k=0}^{n} \cos(x + ky) = \frac{1}{2\sin \frac{y}{2}} \left(\sin \left(x + \frac{2n+1}{2}y \right) - \sin \left(x - \frac{1}{2}y \right) \right),$$

$$\sum_{k=0}^{n} \sin(x + ky) = \frac{1}{2\sin \frac{y}{2}} \left(-\cos \left(x + \frac{2n+1}{2}y \right) + \cos \left(x - \frac{1}{2}y \right) \right).$$

Use integration between y and π to find the partial sum of the series $\sum \frac{\sin ny}{n}$ for $0 < y < 2\pi$. Then apply Riemann-Lebesgue Lemma in Exercise 3.2.44 to get $\sum_{n=1}^{\infty} \frac{\sin ny}{n} = \frac{\pi - y}{2}$ for $0 < y < 2\pi$.

Exercise 4.1.4. Suppose $x_n > 0$. Compute $\sum_{n=1}^{\infty} \frac{x_n}{(1 + x_1)(1 + x_2) \cdots (1 + x_n)}$.

Exercise 4.1.5. Prove that a sequence x_n converges if and only if the series $\sum(x_{n+1} - x_n)$ converges.

Exercise 4.1.6. Prove that if the series $\sum x_n$ and $\sum y_n$ converge, then the series $\sum(x_n + y_n)$ and $\sum cx_n$ also converge. Moreover,

$$\sum(x_n + y_n) = \sum x_n + \sum y_n, \quad \sum cx_n = c \sum x_n.$$

Exercise 4.1.7. Let n_k be a strictly increasing sequence of natural numbers. Then $\sum_{n=k}^{\infty}(x_{n_k} + x_{n_k+1} + \cdots + x_{n_{k+1}-1})$ is obtained from the series $\sum_{n=1}^{\infty} x_n$ by combining successive terms.

1. Prove that if $\sum_{n=1}^{\infty} x_n$ converges, then $\sum_{n=k}^{\infty}(x_{n_k} + x_{n_k+1} + \cdots + x_{n_{k+1}-1})$ also converges.

2. Prove that if the terms $x_{n_k}, x_{n_k+1}, \ldots, x_{n_{k+1}-1}$ that got combined have the same sign and $\sum_{n=k}^{\infty}(x_{n_k} + x_{n_k+1} + \cdots + x_{n_{k+1}-1})$ converges, then the $\sum_{n=1}^{\infty} x_n$ also converges,

Exercise 4.1.8. Prove that if $\lim_{n \to \infty} x_n = 0$, then $\sum x_n$ converges if and only if $\sum(x_{2n-1} + x_{2n})$ converges, and the two sums are the same. What about combining three consecutive terms?

Exercise 4.1.9. Suppose x_n is decreasing and positive. Prove that $\sum x_n$ converges if and only if $\sum 2^n x_{2^n}$ converges. Then study the convergence of $\sum \frac{1}{n^\alpha}$ and $\sum \frac{1}{n(\log n)^\alpha}$.

An *infinite product* is

$$\prod_{n=1}^{\infty} x_n = x_1 x_2 \cdots x_n \cdots .$$

The *partial product* is the sequence $p_n = x_1 x_2 \cdots x_n$. If $\lim_{n \to \infty} p_n = p$ and $p \neq 0$, then we say the infinite product *converges* to (or has the *product*) p.

Because the limit $\lim_{n \to \infty} p_n$ is assumed to be nonzero, we have $x_n \neq 0$, so that the convergence is not affected if we modify, add or delete finitely many nonzero terms. Moreover, for a convergent infinite product we have

$$\lim_{n \to \infty} x_n = \lim_{n \to \infty} \frac{p_n}{p_{n-1}} = \frac{\lim_{n \to \infty} p_n}{\lim_{n \to \infty} p_{n-1}} = \frac{p}{p} = 1.$$

In particular, the terms x_n will be positive for big enough n. Since the finitely many possibly negative (and nonzero) terms will not affect the convergence, we may pretend all terms are positive and use log to convert the infinite product to a series

$$\sum_{n=1}^{\infty} \log x_n = \log x_1 + \log x_2 + \cdots + \log x_n + \cdots ,$$

whose partial sum is

$$\log x_1 + \log x_2 + \cdots + \log x_n = \log p_n.$$

Thus the infinite product $\prod x_n$ converges if and only if the series $\sum \log x_n$ converges.

Exercise 4.1.10. Compute convergent infinite products.

1. $\prod_{n=1}^{\infty} \left(1 + \dfrac{1}{n}\right)$.

2. $\prod_{n=2}^{\infty} \left(1 + \dfrac{(-1)^n}{n}\right)$.

3. $\prod_{n=1}^{\infty} 2^{\frac{1}{n}}$.

4. $\prod_{n=1}^{\infty} 2^{\frac{(-1)^n}{n!}}$.

5. $\prod_{n=1}^{\infty} \cos \dfrac{x}{2^n}$.

Exercise 4.1.11. Use the Cauchy criterion for the convergence of the series $\sum \log x_n$ to get the Cauchy criterion for the convergence of the infinite product $\prod x_n$.

Exercise 4.1.12. Establish properties for infinite products similar to Exercises 4.1.5 and 4.1.6.

Exercise 4.1.13. Why do we have to consider the infinite product as divergent when $\lim_{n\to\infty} p_n = 0$? What bad things may happen if the case $\lim_{n\to\infty} p_n = 0$ is considered as convergent?

4.1.2 Comparison Test

Similar to the convergence of improper integrals, the convergence of series may be determined by comparing with other series.

Proposition 4.1.2 (Comparison Test). *Suppose $|x_n| \le y_n$. If $\sum y_n$ converges, then $\sum x_n$ converges.*

Proof. By the Cauchy criterion for the convergence of y_n, for any $\epsilon > 0$, there is N, such that (4.1.1) holds for $\sum y_n$. Then for $n \ge m > N$, we have

$$|x_m + x_{m+1} + \cdots + x_n| \le |x_m| + |x_{m+1}| + \cdots + |x_n| \le y_m + y_{m+1} + \cdots + y_n < \epsilon.$$

This verifies the Cauchy criterion for the convergence of $\sum x_n$.

$\square$

A series $\sum x_n$ *absolutely converges* if $\sum |x_n|$ converges. The series $\sum x_n$ in the proposition absolutely converges. The proposition also tells us that absolute convergence implies convergence. Moreover, by the discussion in

Example 4.1.2, a series $\sum x_n$ absolutely converges if and only if the partial sum of $\sum |x_n|$ is bounded.

A special case of the comparison test is that if both x_n and y_n are positive, and $\lim_{n \to \infty} \dfrac{x_n}{y_n} = l$ exists, then the convergence of $\sum y_n$ implies the convergence of $\sum x_n$. If $l \neq 0$, then the two convergences are equivalent. More generally, if $c_1 y_n \leq x_n \leq c_2 y_n$ for some $c_1, c_2 > 0$, then the two convergences are also equivalent.

A series *conditionally converges* if it converges but not absolutely converges.

Example 4.1.4. Since $\dfrac{1}{n^\alpha} \leq \dfrac{1}{n^2}$ for $\alpha \geq 2$, the convergence of $\sum \dfrac{1}{n^2}$ in Example 4.1.2 implies that $\sum \dfrac{1}{n^\alpha}$ converges for $\alpha \geq 2$. On the other hand, since $\dfrac{1}{n^\alpha} \geq \dfrac{1}{n}$ for $0 < \alpha \leq 1$, the divergence of the harmonic series $\sum \dfrac{1}{n}$ implies that $\sum \dfrac{1}{n^\alpha}$ diverges for $0 < \alpha \leq 1$.

Example 4.1.5. Suppose x_n satisfies $\sqrt[n]{|x_n|} \leq a$ for some constant $a < 1$. Then $|x_n| \leq a^n$ and the convergence of the geometric series $\sum a^n$ in Example 4.1.1 implies that $\sum x_n$ absolutely converges. This is called the *root test*.

For a specific example, let $p(t)$ be a nonzero polynomial. Then

$$\lim_{n \to \infty} \sqrt[n]{|p(n)a^n|} = |a| \lim_{n \to \infty} \sqrt[n]{|p(n)|} = |a|.$$

If $|a| < 1$, then the limit implies that for any $|a| < b < 1$, we have $\sqrt[n]{|p(n)a^n|} < b$ for sufficiently big n. Since the convergence of a series is independent of the choice of finitely many terms, we conclude that $\sum p(n)a^n$ converges for $|a| < 1$.

If $|a| \geq 1$, then $p(n)a^n$ does not converge to 0 as $n \to \infty$. Thus $\sum p(n)a^n$ diverges for $|a| \geq 1$.

Example 4.1.6. Suppose x_n satisfies $\left| \dfrac{x_{n+1}}{x_n} \right| \leq a$ for some constant $a < 1$. Then

$$|x_n| = |x_1| \left| \frac{x_2}{x_1} \right| \left| \frac{x_3}{x_2} \right| \cdots \left| \frac{x_n}{x_{n-1}} \right| \leq |x_1| a^{n-1}.$$

The convergence of the geometric series $|x_1| \sum a^{n-1}$ implies that $\sum x_n$ absolutely converges. This is called the *ratio test*.

For a specific example, the series $\sum \dfrac{(n!)^2 a^n}{(2n)!}$ satisfies

$$\lim_{n \to \infty} \left| \frac{\dfrac{(n!)^2 a^n}{(2n)!}}{\dfrac{((n-1)!)^2 a^{n-1}}{(2n-2)!}} \right| = \lim_{n \to \infty} \frac{n^2 |a|}{2n(2n-1)} = \frac{|a|}{4}.$$

Thus for $|a| < 4$, the ratio is $< \dfrac{|a| + 1}{5} < 1$ for sufficiently big n, and the series converges. If $|a| \geq 4$, then the absolute value of the quotient is > 1 and the individual term does not converge to 0 (by Exercise 1.1.33, the limit is in fact ∞), so that the series diverges.

Example 4.1.7. Consider the series

$$\sum \frac{(-1)^{n-1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots$$

By

$$0 < \frac{1}{2n-1} - \frac{1}{2n} = \frac{1}{2n(2n-1)} < \frac{1}{n^2}$$

and the comparison test, we know the series

$$\sum \left(\frac{1}{2n-1} - \frac{1}{2n} \right) = \left(1 - \frac{1}{2} \right) + \left(\frac{1}{3} - \frac{1}{4} \right) + \cdots$$

converges. The partial sum of $\sum \left(\frac{1}{2n-1} - \frac{1}{2n} \right)$ is the even part s_{2n} of the partial sum s_n of $\sum \frac{(-1)^{n-1}}{n}$. By the convergence of s_{2n} and $\lim_{n\to\infty}(s_{2n} - s_{2n-1}) = \lim_{n\to\infty} \frac{1}{2n} = 0$, we see that s_{2n-1} converges to the same limit, so that the series $\sum \frac{(-1)^{n-1}}{n}$ converges.

By the divergence of $\sum \frac{1}{n}$, we further know that $\sum \frac{(-1)^{n-1}}{n}$ converges conditionally.

Exercise 4.1.14. Suppose $p(x)$ and $q(x)$ are polynomials. Determine the convergence of $\sum \frac{p(n)}{q(n)}$.

Exercise 4.1.15. Is the convergence of $\sum x_n$ and $\sum y_n$ related to the convergence of $\sum \max\{x_n, y_n\}$ and $\sum \min\{x_n, y_n\}$?

Exercise 4.1.16. Suppose $\left| \frac{x_{n+1}}{x_n} \right| \le \left| \frac{y_{n+1}}{y_n} \right|$ for sufficiently big n. Prove that if $\sum y_n$ absolutely converges, then $\sum x_n$ absolutely converges.

Exercise 4.1.17 (Root Test). Prove that $\sum x_n$ absolutely converges if $\overline{\lim}_{n\to\infty} \sqrt[n]{|x_n|} < 1$ and diverges if $\overline{\lim}_{n\to\infty} \sqrt[n]{|x_n|} > 1$.

Exercise 4.1.18 (Ratio Test). Prove that $\sum x_n$ absolutely converges if $\overline{\lim}_{n\to\infty} \left| \frac{x_{n+1}}{x_n} \right| < 1$ and diverges if $\underline{\lim}_{n\to\infty} \left| \frac{x_{n+1}}{x_n} \right| > 1$.

Exercise 4.1.19. Prove that

$$\overline{\lim_{n\to\infty}} \sqrt[n]{|x_n|} \le \overline{\lim_{n\to\infty}} \left| \frac{x_{n+1}}{x_n} \right|, \quad \underline{\lim_{n\to\infty}} \sqrt[n]{|x_n|} \ge \underline{\lim_{n\to\infty}} \left| \frac{x_{n+1}}{x_n} \right|.$$

What do the inequalities tell you about the relation between the root and ratio tests?

Exercise 4.1.20. Determine the convergence of series.

1. $\sum \dfrac{1}{\sqrt{n(n-1)}}$.

2. $\sum \dfrac{\sin n}{\sqrt{(n^2-1)(n^2-2)}}$.

3. $\sum a^{n+(-1)^n}$.

4. $\sum a^{n^2}$.

5. $\sum n^2 a^{n^2}$.

6. $\sum \sqrt{a^n + b^n}$.

10. $\sum \left(\dfrac{an + b}{cn + d} \right)^n$.

13. $\sum \dfrac{1}{(\log n)^n}$.

7. $\sum \dfrac{1}{\sqrt{a^n + b^n}}$.

14. $\sum n! a^n$.

11. $\sum \dfrac{\log n}{n^3}$.

8. $\sum (a^{\frac{1}{n}} - 1)$.

15. $\sum \dfrac{na^n}{(n + 1)!}$.

9. $\sum \left(e^{\frac{1}{n}} - 1 - \dfrac{1}{n} \right)$.

12. $\sum \dfrac{1}{(\log n)^2}$.

16. $\sum \dfrac{\sqrt{(2n)!}a^n}{n!}$.

Exercise 4.1.21. Let $[x]$ be the biggest integer $\leq x$, defined in Exercise 1.3.5. Consider the series $\sum \dfrac{(-1)^{[\sqrt{n}]}}{n^\alpha}$, with $\alpha > 0$.

1. Prove that

$$x_n = \sum_{k=n^2}^{(n+1)^2 - 1} \frac{1}{k} = \frac{1}{n^{2\alpha}} + \frac{1}{(n^2 + 1)^\alpha} + \cdots + \frac{1}{((n + 1)^2 - 1)^\alpha}$$

satisfies

$$\frac{2n}{((n + 1)^2 - 1)^\alpha} < x_n < \frac{2n}{n^{2\alpha}}.$$

2. If $\alpha > \dfrac{1}{4}$, prove that $|x_{n+1} - x_n| < \dfrac{M}{n^{2\alpha}}$ for a constant M.

3. If $\alpha \leq \dfrac{1}{4}$, prove that x_n is increasing for sufficiently big n.

4. With the help of Exercise 4.1.7, prove that $\sum \dfrac{(-1)^{[\sqrt{n}]}}{n^\alpha}$ converges for $\alpha > \dfrac{1}{2}$ and diverges for $\alpha \leq \dfrac{1}{4}$.

Exercise 4.1.22. Suppose $x_n \neq -1$.

1. Prove that if $\sum x_n$ converges, then $\prod(1 + x_n)$ converges if and only if $\sum x_n^2$ converges.

2. Prove that if $\sum x_n^2$ converges, then $\prod(1 + x_n)$ converges if and only if $\sum x_n$ converges.

The convergence of series may also be determined by comparing with the convergence of improper integrals.

Proposition 4.1.3 (Integral Comparison Test)**.** *Suppose $f(x)$ is a decreasing function on $[a, +\infty)$ satisfying $\lim_{x \to +\infty} f(x) = 0$. Then the series $\sum f(n)$ converges if and only if the improper integral $\int_a^{+\infty} f(x)dx$ converges.*

Proof. Since the convergence is not changed if finitely many terms are modified or deleted, we may assume $a = 1$ without loss of generality.

Since $f(x)$ is decreasing, we have $f(k) \geq \int_k^{k+1} f(x)dx \geq f(k+1)$. Then

$$f(1) + f(2) + \cdots + f(n-1)$$

$$\geq \int_1^n f(x)dx = \int_1^2 f(x)dx + \int_2^3 f(x)dx + \cdots + \int_{n-1}^n f(x)dx$$

$$\geq f(2) + f(3) + \cdots + f(n).$$

This implies that $\int_1^n f(x)dx$ is bounded if and only if the partial sums of the series $\sum f(n)$ are bounded. Since $f(x) \geq 0$, the boundedness is equivalent to the convergence. Therefore $\int_a^{+\infty} f(x)dx$ converges if and only if $\sum f(n)$ converges.

$\square$

Example 4.1.8. The series $\sum \dfrac{1}{n^\alpha}$ converges if and only if $\int_1^{+\infty} \dfrac{dx}{x^\alpha}$ converges, which by Example 3.3.7 means $\alpha > 1$. The *Riemann zeta function*

$$\zeta(\alpha) = 1 + \frac{1}{2^\alpha} + \frac{1}{3^\alpha} + \cdots + \frac{1}{n^\alpha} + \cdots \tag{4.1.3}$$

is then defined for $\alpha > 1$.

Example 4.1.9. The series $\sum \dfrac{1}{n(\log n)^\alpha}$ converges if and only if $\int_2^{+\infty} \dfrac{dx}{x(\log x)^\alpha}$ converges. It is easy to see that the improper integral converges if and only if $\alpha > 1$. Therefore the series converges if and only if $\alpha > 1$.

Example 4.1.10. Suppose a is not a multiple of π. We study the convergence of $\sum \dfrac{|\sin na|}{n^\alpha}$.

For $\alpha > 1$, by $\dfrac{|\sin na|}{n^\alpha} \leq \dfrac{1}{n^\alpha}$, the convergence of $\sum \dfrac{1}{n^\alpha}$, and the comparison test, the series $\sum \dfrac{|\sin na|}{n^\alpha}$ converges.

For $\alpha \leq 1$, an argument similar to Example 3.3.12 shows that $\int_1^{+\infty} \dfrac{|\sin x|}{x^\alpha}dx$ diverges. This suggests that the series diverges. However, we cannot apply the integral comparison test because the function $\dfrac{|\sin x|}{x^\alpha}$ is not decreasing.

Assume $0 < a \leq \dfrac{\pi}{2}$. Then any interval of length $\dfrac{\pi}{2}$ contains an integer multiple of a. Choose natural numbers n_k, such that

$$\frac{4k+1}{4}\pi \leq n_k a \leq \frac{4k+3}{4}\pi.$$

Then $|\sin n_k a| \geq \dfrac{1}{\sqrt{2}}$, and for $\alpha \leq 1$,

$$\sum_{n=1}^\infty \frac{|\sin na|}{n^\alpha} \geq \sum_{n=1}^\infty \frac{|\sin na|}{n} \geq \sum_{k=1}^\infty \frac{|\sin n_k a|}{n_k} \geq \frac{1}{\sqrt{2}} \sum_{k=1}^\infty \frac{1}{n_k}.$$

By
$$\frac{1}{n_k} \geq \frac{4}{4k+3} \frac{a}{\pi} > \frac{a}{4k},$$

the divergence of $\sum_{k=1}^{\infty} \frac{a}{4k}$, and the comparison test, we conclude that $\sum_{k=1}^{\infty} \frac{1}{n_k}$ diverges. Consequently, the series $\sum_{n=1}^{\infty} \frac{|\sin na|}{n^{\alpha}}$ diverges for $\alpha \leq 1$.

For general a that is not an integer multiple of π, there is b, such that $0 < b \leq \frac{\pi}{2}$ and either $a+b$ or $a-b$ is an integer multiple of π. Then we have $|\sin na| = |\sin nb|$, and we still conclude that $\sum_{n=1}^{\infty} \frac{|\sin na|}{n^{\alpha}}$ diverges when $\alpha \leq 1$.

Exercise 4.1.23. Show that the number of n digit numbers that do not contain the digit 9 is $8 \cdot 9^{n-1}$. Then use the fact to prove that if we delete the terms in the harmonic series that contain the digit 9, then the series becomes convergent. What about deleting the terms that contain some other digit? What about the numbers expressed in the base other than 10? What about deleting similar terms in the series $\sum \frac{1}{n^{\alpha}}$?

Exercise 4.1.24. Determine the convergence of series.

1. $\sum \frac{\sin n}{\sqrt{n(n-1)(n-2)}}$.

2. $\sum \frac{1}{(\log n)^{\alpha \log n}}$.

3. $\sum \frac{1}{n^{\alpha} + (\log n)^{\beta}}$.

4. $\sum n^{\alpha} (\log n)^{\beta}$.

5. $\sum \left(1 - \frac{a \log n}{n}\right)^n$.

6. $\sum \int_0^{\frac{1}{n}} \frac{x^{\alpha}}{1+x^2} dx$.

Exercise 4.1.25. Prove that if $\sum x_n^2$ converges, then $\sum \frac{x_n}{n^{\alpha}}$ also converges for any $\alpha > \frac{1}{2}$. What if $\alpha = \frac{1}{2}$?

Exercise 4.1.26. Use the idea of the proof of Proposition 4.1.3 to prove $\log(n-1)! < n(\log n - 1) + 1 < \log n!$. Then derive the inequality $\frac{n^n}{e^{n-1}} < n! < \frac{(n+1)^{n+1}}{e^n}$.

4.1.3   Conditional Convergence

A series $\sum x_n$ is *alternating* if the signs of x_n and x_{n+1} are different. Another way of expressing an alternating series is $\sum (-1)^n x_n$, with $x_n \geq 0$.

Proposition 4.1.4 (Leibniz Test). *Suppose x_n is decreasing and $\lim_{n \to \infty} x_n = 0$. Then $\sum (-1)^n x_n$ converges. Moreover, the sum s and the partial sum s_n satisfy $|s_n - s| \leq x_{n+1}$.*

Proof. The partial sum satisfies

$$s_{2n+1} - s_{2n-1} = x_{2n} - x_{2n+1} \geq 0, \quad s_{2n+2} - s_{2n} = -x_{2n+1} + x_{2n+2} \leq 0.$$

Thus s_{2n+1} is increasing and s_{2n} is decreasing. Moreover, by $s_{2n} - s_{2n+1} = x_{2n+1} \geq 0$, s_{2n+1} has upper bound and s_{2n} has lower bound. Therefore the sequences converge. Then $\lim_{n \to \infty}(s_{2n} - s_{2n+1}) = 0$ implies that the limits are the same.

The estimation $|s_n - s| \leq x_{n+1}$ follows from

$$0 \leq s_{2n} - s \leq s_{2n} - s_{2n+1} = x_{2n+1}, \;\; 0 \leq s - s_{2n-1} \leq s_{2n} - s_{2n-1} = x_{2n}.$$

$\square$

Example 4.1.11. The series $\sum \dfrac{(-1)^n}{n^\alpha}$ converges for $\alpha > 0$. Note that by Example 4.1.8, the series absolutely converges only for $\alpha > 1$. Therefore the series conditionally converges for $0 < \alpha \leq 1$.

Exercise 4.1.27. Suppose $b \neq 0$. Find all the combinations of a and b such that $\sum n^a b^n$ converge.

Exercise 4.1.28. Find all $a \neq e^{-1}$ such that $\sum \dfrac{(na)^n}{n!}$ converges (the case $a = e^{-1}$ will be dealt with in Exercise 4.1.48).

Exercise 4.1.29. Construct a convergent series $\sum x_n$ such that the series $\sum x_n^2$ diverges. By Exercise 4.1.22, we see that the convergence of $\sum x_n$ does not necessarily imply the convergence of $\prod(1 + x_n)$.

Exercise 4.1.30. Suppose x_n is decreasing and $\lim_{n\to\infty} x_n = 1$. Prove that the "alternating infinite product" $\prod x_n^{(-1)^n}$ converges. Then find suitable x_n, such that the series $\sum y_n$ defined by $1 + y_n = x_n^{(-1)^n}$ diverges. This shows that the convergence of $\prod(1 + y_n)$ does not necessarily imply the convergence of $\sum y_n$.

We have Dirichlet and Abel tests for the convergence of improper integrals in Propositions 3.3.3 and 3.3.4, especially in the conditionally convergent cases. The analogue of the tests for the convergence of the series is the following.

Proposition 4.1.5 (Dirichlet Test). *Suppose x_n is monotone and $\lim_{n\to\infty} x_n = 0$. Suppose the partial sums of the series $\sum y_n$ are bounded. Then $\sum x_n y_n$ converges.*

Proposition 4.1.6 (Abel Test). *Suppose x_n is monotone and bounded. Suppose the series $\sum y_n$ converges. Then $\sum x_n y_n$ converges.*

Proof. The proof is the discrete version of the proof for the convergence of improper integrals. First we need the discrete version of the integration by parts, which already appeared in Exercise 3.2.55 and the proof of Theorem 3.3.6.

Let s_n be the partial sum of $\sum y_n$. Then for $n \geq m$,

$$\begin{aligned}
\sum_{k=m}^{n} x_k y_k &= x_m(s_m - s_{m-1}) + x_{m+1}(s_{m+1} - s_m) + \cdots + x_n(s_n - s_{n-1}) \\
&= -x_m s_{m-1} - (x_{m+1} - x_m)s_m - (x_{m+2} - x_{m+1})s_{m+1} \\
&\quad - \cdots - (x_n - x_{n-1})s_{n-1} + x_n s_n \\
&= x_n s_n - x_m s_{m-1} - \sum_{k=m}^{n-1} (x_{k+1} - x_k)s_k.
\end{aligned}$$
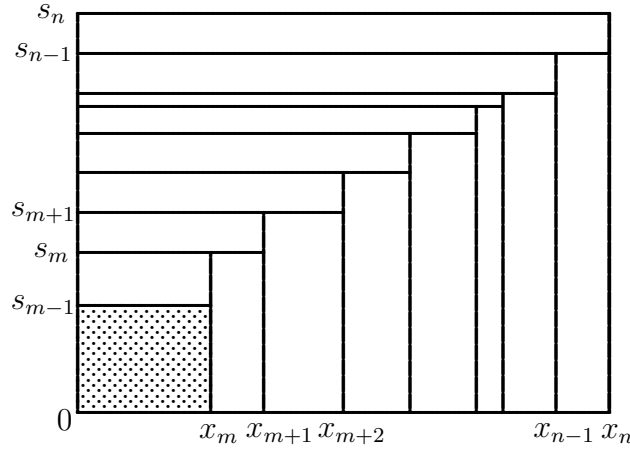
Figure 4.1: the big square is equal to shaded small square, plus vertical strips, plus horizontal strips

The geometric meaning of the equality is illustrated in Figure 4.1. Under the assumption of either test, we have $|s_k| \leq M$ for a constant M and all k. Moreover, the increments $x_{k+1} - x_k$ do not change sign. Therefore

$$\left| \sum_{k=m}^{n-1} (x_{k+1} - x_k)s_k \right| \leq \sum_{k=m}^{n-1} M|x_{k+1} - x_k| = M \left| \sum_{k=m}^{n-1} (x_{k+1} - x_k) \right| = M|x_n - x_m|.$$

Assuming the condition of the Dirichlet test, for any $\epsilon > 0$, there is N, such that $k > N$ implies $|x_k| < \epsilon$. Then $n \geq m > N$ implies

$$\left| \sum_{k=m}^{n} x_k y_k \right| \leq \epsilon|s_n| + \epsilon|s_{m-1}| + 2M\epsilon \leq 4M\epsilon.$$

This verifies the Cauchy criterion for the convergence of $\sum x_n y_n$.

Assuming the condition of the Abel test, we know both $\lim_{n \to \infty} x_n$ and $\lim_{n \to \infty} x_n s_n$ converge. The Cauchy criterion for both limits tell us that for any $\epsilon > 0$, there is N, such that $m, n > N$ implies $|x_n - x_m| < \epsilon$ and $|x_n s_n - x_m s_m| < \epsilon$. Then $n \geq m > N$ implies

$$\left| \sum_{k=m}^{n} x_k y_k \right| \leq \epsilon + M\epsilon = (M + 1)\epsilon.$$

This also verifies the Cauchy criterion for the convergence of $\sum x_n y_n$.   □

Example 4.1.12. Let us consider the discrete version of Example 3.3.12. We are concerned with the series

$$\sum \frac{\sin na}{n^\alpha} = \sum \frac{1}{n^\alpha} \sin na.$$

If a is not a multiple of π, then the partial sum of $\sum \sin na$ is

$$\sum_{k=1}^{n} \sin ka = \frac{\cos\left(n + \frac{1}{2}\right)a - \cos\frac{1}{2}a}{2\sin\frac{1}{2}a},$$

which is bounded by $\dfrac{1}{\left|\sin\frac{1}{2}a\right|}$. Moreover, for $\alpha > 0$, the sequence $\dfrac{1}{n^\alpha}$ is decreasing and converges to 0. By the Dirichlet test, the series converges.

Combined with the discussion in Example 4.1.10, we see that the series $\sum \dfrac{\sin na}{n^\alpha}$ absolutely converges for $\alpha > 1$ and conditionally converges for $0 < \alpha \le 1$.

Exercise 4.1.31. Determine the absolute and conditional convergence of series

1. $\sum (-1)^n n^\alpha \log^\beta n$.

2. $\sum \dfrac{(-1)^n}{(\log n)^{\alpha \log n}}$.

3. $\sum \dfrac{(-1)^{\frac{n(n-1)}{2}}}{n^\alpha}$.

4. $\sum \dfrac{(-1)^{\frac{n(n-1)}{2}} \sin n}{n^\alpha}$.

5. $\sum \dfrac{\cos na}{n^\alpha}$.

6. $\sum \dfrac{\sin^3 na}{n^\alpha}$.

7. $\sum \sin \sqrt{n}$.

8. $\sum \dfrac{\sin \sqrt{n}}{n^\alpha}$.

Exercise 4.1.32. Derive the Leibniz test and the Abel test from the Dirichlet test.

Exercise 4.1.33. Prove that if $\beta > \alpha$, then the convergence of $\sum \dfrac{a_n}{n^\alpha}$ implies the convergence of $\sum \dfrac{a_n}{n^\beta}$.

4.1.4 Rearrangement of Series

A *rearrangement* of a series $\sum x_n$ is $\sum x_{k_n}$, where $n \to k_n$ is a one-to-one correspondence from the index set to itself (i.e., a rearrangement of the indices).

Proposition 4.1.7. *Any rearrangement of an absolutely convergent series is still absolutely convergent. Moreover, the sum is the same.*

Proof. Let $s = \sum x_n$. For any $\epsilon > 0$, there is a natural number N, such that $\sum_{i=N+1}^\infty |x_i| < \epsilon$. Let $N' = \max\{i \colon k_i \le N\}$. Then $\sum_{i=1}^{N'} x_{k_i}$ contains all the terms $x_1, x_2, \ldots, x_N$. Therefore for $n > N'$, the difference $\sum_{i=1}^n x_{k_i} - \sum_{i=1}^N x_i$ is a sum of some non-repeating terms in $\sum_{i=N+1}^\infty x_i$, and we have

$$\left| \sum_{i=1}^n x_{k_i} - s \right| \le \left| \sum_{i=1}^n x_{k_i} - \sum_{i=1}^N x_i \right| + \left| \sum_{i=1}^N x_i - s \right| \le 2 \sum_{i=N+1}^\infty |x_i| < 2\epsilon.$$

The absolute convergence of the rearrangement may be obtained by applying what was just proved to $\sum |x_n|$.

$\square$

In contrast to absolutely convergent series, Exercise 4.1.2 shows that the rearrangement of conditionally convergent series may change the sum. In fact, the rearrangement can produce any behavior we wish to have.

Proposition 4.1.8. *A conditionally convergent series may be rearranged to have any number as the sum, or to become divergent.*

Proof. Suppose $\sum x_n$ conditionally converges. Let $\sum x_n'$ and $\sum x_n''$ be the series obtained by respectively taking only the non-negative terms and the negative terms. If $\sum x_n'$ converges, then $\sum x_n'' = \sum x_n - \sum x_n'$ also converges. Therefore $\sum |x_n| = \sum x_n' - \sum x_n''$ converges. Since $\sum |x_n|$ is assumed to diverge, the contradiction shows that $\sum x_n'$ diverges. Because $x_n' \geq 0$, we conclude $\sum x_n' = +\infty$. Similarly, we have $\sum x_n'' = -\infty$.

Let s be any number. By $\sum x_n' = +\infty$, there is m_1, such that

$$\sum_{k=1}^{m_1-1} x_k' \leq s < \sum_{k=1}^{m_1} x_k'.$$

Then by $\sum x_n'' = -\infty$, there is n_1, such that

$$\sum_{k=1}^{n_1-1} x_k'' \geq s - \sum_{k=1}^{m_1} x_k' > \sum_{k=1}^{n_1} x_k''.$$

Then by $\sum_{n>n_1} x_n' = +\infty$, there is m_2, such that

$$\sum_{k=m_1+1}^{m_2-1} x_k' \leq s - \sum_{k=1}^{m_1} x_k' - \sum_{k=1}^{n_1} x_k'' < \sum_{k=m_1+1}^{m_2} x_k'.$$

Keep going, we get

$$-x_{n_p}'' \geq s - \sum_{k=1}^{m_p} x_k' - \sum_{k=1}^{n_p} x_k'' > 0, \quad -x_{m_p}' \leq s - \sum_{k=1}^{m_p} x_k' - \sum_{k=1}^{n_{p-1}} x_k'' < 0.$$

Then for $m_{p-1} < j \leq m_p$, by $x_k' > 0$, we have

$$-x_{m_p}' \leq s - \sum_{k=1}^{m_p} x_k' - \sum_{k=1}^{n_{p-1}} x_k'' \leq s - \sum_{k=1}^{j} x_k' - \sum_{k=1}^{n_{p-1}} x_k'' \leq s - \sum_{k=1}^{m_{p-1}} x_k' - \sum_{k=1}^{n_{p-1}} x_k'' \leq -x_{n_{p-1}}'',$$

and for $n_{p-1} < j \leq n_p$, by $x_k'' < 0$, we have

$$-x_{n_p}'' \geq s - \sum_{k=1}^{m_p} x_k' - \sum_{k=1}^{n_p} x_k'' \geq s - \sum_{k=1}^{m_p} x_k' - \sum_{k=1}^{j} x_k'' \geq s - \sum_{k=1}^{m_p} x_k' - \sum_{k=1}^{n_{p-1}} x_k'' \geq -x_{m_p}'.$$

The convergence of $\sum x_n$ implies $\lim_{n\to\infty} x_n = 0$. Then the two estimations above show that the rearranged series

$$x_1' + \cdots + x_{m_1}' + x_1'' + \cdots + x_{n_1}'' + x_{m_1+1}' + \cdots + x_{m_2}' + x_{n_1+1}'' + \cdots + x_{n_2}'' + \cdots$$

converges to s.

$\square$

Exercise 4.1.34. Rearrange the series $1 - \dfrac{1}{2} + \dfrac{1}{3} - \dfrac{1}{4} + \cdots$ so that p positive terms are followed by q negative terms and the pattern repeated. Show that the sum of the new series is $\log 2 + \dfrac{1}{2} \log \dfrac{p}{q}$. For any real number, expand the idea to construct a rearrangement to have the given number as the limit.

Exercise 4.1.35. Prove that if the rearrangement satisfies $|k_n - n| < M$ for a constant M, then $\sum x_{k_n}$ converges if and only if $\sum x_n$ converges, and the sums are the same.

Exercise 4.1.36. Suppose $\sum x_n$ conditionally converges. Prove that for any s and t satisfying $s < t$, there is a rearrangement with the partial sum s'_n satisfying $\underline{\lim}_{n\to\infty} s'_n = s$, $\overline{\lim}_{n\to\infty} s'_n = t$. Moreover, prove that any number between s and t is the limit of a convergent subsequence of $\{s'_n\}$.

The product of two series $\sum x_n$ and $\sum y_n$ involves the product $x_m y_n$ for all m and n. In general, the *product series* $\sum x_m y_n$ makes sense only after we arrange all the terms into a linear sequence $\sum (xy)_k = \sum x_{m_k} y_{n_k}$, which is given by a one-to-one correspondence $(m_k, n_k) \colon \mathbb{N} \to \mathbb{N} \times \mathbb{N}$. For example, the following is the "diagonal arrangement"

$$\sum (xy)_k = x_1 y_1 + x_1 y_2 + x_2 y_1 + \cdots \tag{4.1.4}$$
$$+ x_1 y_{n-1} + x_2 y_{n-2} + \cdots + x_{n-1} y_1 + \cdots ,$$

and the following is the "square arrangement"

$$\sum (xy)_k = x_1 y_1 + x_1 y_2 + x_2 y_2 + x_2 y_1 + \cdots \tag{4.1.5}$$
$$+ x_1 y_n + x_2 y_n + \cdots + x_n y_{n-1} + x_n y_n + x_n y_{n-1} + \cdots + x_n y_1 + \cdots$$
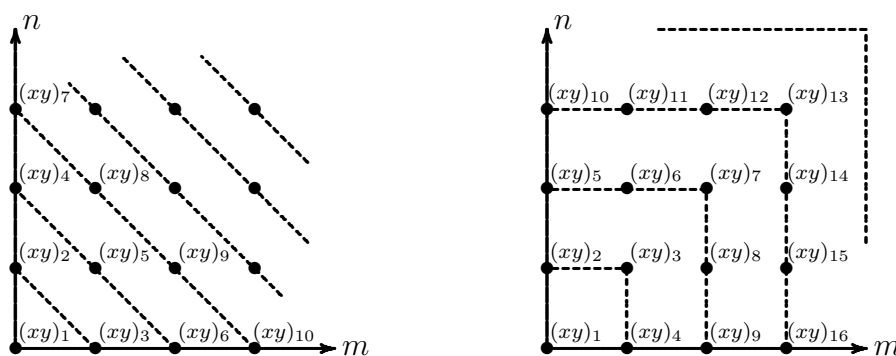
Figure 4.2: diagonal and square arrangements

In view of Proposition 4.1.7, the following result shows that the arrangement does not matter if the series absolutely converge.

Proposition 4.1.9. *Suppose $\sum x_n$ and $\sum y_n$ absolutely converge. Then $\sum x_m y_n$ also absolutely converges, and $\sum x_m y_n = (\sum x_m)(\sum y_n)$.*

Proof. Let

$$s = \sum x_m, \quad t = \sum y_n, \quad S = \sum |x_m|, \quad T = \sum |y_n|$$

For any $\epsilon > 0$, there is a natural number N, such that

$$\sum_{m>N} |x_m| < \epsilon, \quad \sum_{n>N} |y_n| < \epsilon.$$

Then

$$\left| \sum_{m=1}^{N} x_m - s \right| < \epsilon, \quad \left| \sum_{n=1}^{N} y_n - t \right| < \epsilon, \quad |s| \leq S, \quad \left| \sum_{n=1}^{N} y_n \right| \leq T.$$

Let $K = \max\{k \colon m_k \leq N \text{ and } n_k \leq N\}$. Then $\sum_{i=1}^{K}(xy)_i$ contains all the terms $x_m y_n$ with $1 \leq m, n \leq N$. Therefore for $k > K$, the difference $\sum_{i=1}^{k}(xy)_i - \left(\sum_{m=1}^{N} x_m \right) \left(\sum_{n=1}^{N} y_n \right)$ is a sum of some non-repeating terms $x_m y_n$ with either $m > N$ or $n > N$, and we have

$$\left| \sum_{i=1}^{k}(xy)_i - \left(\sum_{m=1}^{N} x_m \right) \left(\sum_{n=1}^{N} y_n \right) \right| \leq \sum_{\text{either } m>N \text{ or } n>N} |x_m y_n|$$

$$\leq \sum_{m>N} |x_m y_n| + \sum_{n>N} |x_m y_n|$$

$$= \sum_{m>N} |x_m| \sum |y_n| + \sum |x_m| \sum_{n>N} |y_n|$$

$$\leq \epsilon(T + S).$$

Combined with

$$\left| \left(\sum_{m=1}^{N} x_m \right) \left(\sum_{n=1}^{N} y_n \right) - st \right| \leq \left| \sum_{m=1}^{N} x_m - s \right| \left| \sum_{n=1}^{N} y_n \right| + |s| \left| \sum_{n=1}^{N} y_n - t \right|$$

$$\leq \epsilon \left| \sum_{n=1}^{N} y_n \right| + |s|\epsilon \leq \epsilon(T + S),$$

we get

$$\left| \sum_{i=1}^{k}(xy)_i - st \right| \leq 2\epsilon(T + S).$$

The absolute convergence of the product series may be obtained by applying what was just proved to $\sum |x_m|$ and $\sum |y_n|$. $\qquad\square$

Example 4.1.13. The geometric series absolutely converges to $\dfrac{1}{1-a}$ for $|a| < 1$. The product of two copies of the geometric series is

$$\sum_{i,j \geq 0} a^i a^j = \sum_{n=1}^{\infty} \sum_{i+j=n} a^n = \sum_{n=1}^{\infty}(n+1)a^n.$$

Thus we conclude

$$1 + 2a + 3a^2 + \cdots + (n+1)a^n + \cdots = \frac{1}{(1-a)^2}.$$

Exercise 4.1.37. Use the Taylor series of e^x to verify $e^x e^y = e^{x+y}$.

Exercise 4.1.38. Suppose $\sum x_n$ and $\sum y_n$ converge (not necessarily absolutely). Does the square arrangement (4.1.5) converge to $(\sum x_n)(\sum y_n)$?

Exercise 4.1.39. Suppose $\sum x_n$ absolutely converges and $\sum y_n$ converges. Prove that the diagonal arrangement (4.1.4) converges. Moreover, show that the condition of absolute convergence is necessary by considering the product of $\sum \frac{(-1)^n}{\sqrt{n}}$ with itself.

Exercise 4.1.40. Let p_n, $n = 1, 2, \ldots$, be prime numbers in increasing order. Let S_n be the set of natural numbers whose prime factors are among $p_1, p_2, \ldots, p_n$. For example, $20 \notin S_2$ and $20 \in S_3$ because the prime factors of 20 are 2 and 5.

1. Prove that $\prod_{i=1}^k \left(1 - \frac{1}{p_i^\alpha}\right)^{-1} = \sum_{n \in S_k} \frac{1}{n^\alpha}$.

2. When does the infinite product $\prod \left(1 - \frac{1}{p_n^\alpha}\right)$ converge?

3. When does the series $\sum \frac{1}{p_n^\alpha}$ converge?

Note that the zeta function (see Example 4.1.8) $\zeta(\alpha) = \prod_{i=1}^\infty \left(1 - \frac{1}{p_i^\alpha}\right)^{-1}$ when the right side converges. The equality relates the function to the number theory.

4.1.5 Additional Exercise

Convergence of Series

Exercise 4.1.41. Suppose $x_n > 0$ and x_n is increasing. Prove that $\sum \frac{1}{x_n}$ converges if and only if $\sum \frac{n}{x_1 + x_2 + \cdots + x_n}$ converges.

Exercise 4.1.42. Suppose $x_n > 0$. Prove that $\sum x_n$ converges if and only if $\sum \frac{x_n}{x_1 + x_2 + \cdots + x_n}$ converges.

Exercise 4.1.43. Suppose $x_n \geq 0$. Prove that $\sum \frac{x_n}{(x_1 + x_2 + \cdots + x_n)^2}$ converges.

Exercise 4.1.44. Suppose x_n is a non-negative and decreasing sequence. Prove that if $\lim_{n \to \infty} x_n = 0$ and $\sum_{i=1}^n (x_i - x_n) = \sum_{i=1}^n x_i - n x_n \leq B$ for a fixed bound B, then $\sum x_n$ converges.

Raabe[2] and Bertrand[3] Tests

Exercise 4.1.16 is the mother of ratio tests. By applying the exercise to the power series, we get the ratio test in Example 4.1.8 and Exercise 4.1.18. By applying the exercise to the series in Examples 4.1.8 and 4.1.9, we get the Rabbe and Bertrand tests.

Exercise 4.1.45. Prove that if $\left|\frac{x_{n+1}}{x_n}\right| \leq 1 - \frac{a}{n}$ for some $a > 1$ and big n, then $\sum x_n$ absolutely converges.

---

[2]Joseph Ludwig Raabe, born 1801 in Brody (now Ukrain), died 1859 in Zürich (Switzerland).

[3]Joseph Louis François Bertrand, born 1822 and died 1900 in Paris (France).

Exercise 4.1.46. Prove that if $\left|\dfrac{x_{n+1}}{x_n}\right| \geq 1 - \dfrac{1}{n-a}$ for some constant a and big n, then $\sum x_n$ does not absolutely converge.

Exercise 4.1.47. Rephrase the Raabe test in Exercises 4.1.45 and 4.1.46 in terms of the quotient $\left|\dfrac{x_n}{x_{n+1}}\right|$.

Exercise 4.1.48. Determine the convergence of the series $\sum \dfrac{\alpha(\alpha+1)\cdots(\alpha+n)}{\beta(\beta+1)\cdots(\beta+n)}$ and $\sum \dfrac{n^n}{e^n n!}$.

Exercise 4.1.49. Derive the condition for the absolute convergence of $\sum x_n$ in terms of the comparison between $\left|\dfrac{x_{n+1}}{x_n}\right|$ and $1 - \dfrac{a}{n\log n}$, or the comparison between $\left|\dfrac{x_n}{x_{n+1}}\right|$ and $1 + \dfrac{a}{n\log n}$.

Kummer Test

Exercise 4.1.50. Prove that if there are $c_n > 0$ and $\delta > 0$, such that $c_n - c_{n+1}\left|\dfrac{x_{n+1}}{x_n}\right| \geq \delta$ for sufficiently big n, then $\sum x_n$ absolutely converges.

Exercise 4.1.51. Prove that if $x_n > 0$, and there are $c_n > 0$, such that $c_n - c_{n+1}\dfrac{x_{n+1}}{x_n} \leq 0$ and $\sum \dfrac{1}{c_n}$ diverges, then $\sum x_n$ diverges.

Exercise 4.1.52. Prove that if $\sum x_n$ absolutely converges, then there are $c_n > 0$, such that $c_n - c_{n+1}\left|\dfrac{x_{n+1}}{x_n}\right| \geq 1$ for all n.

Exercise 4.1.53. Derive ratio, Raabe and Bertrand tests from the Kummer test.

Absolutely Convergence of Infinite Product

Suppose $x_n \neq -1$. An infinite product $\prod(1 + x_n)$ *absolutely converges* if $\prod(1 + |x_n|)$ converges.

Exercise 4.1.54. Prove that $\prod(1 + x_n)$ absolutely converges if and only if the series $\sum x_n$ absolutely converges. Moreover, $\prod(1 + |x_n|) = +\infty$ if and only if $\sum x_n = +\infty$.

Exercise 4.1.55. Prove that if the infinite product absolutely converges, then the infinite product converges.

Exercise 4.1.56. Suppose $0 < x_n < 1$. Prove that $\prod(1 + x_n)$ converges if and only if $\prod(1 - x_n)$ converges. Moreover, $\prod(1 + x_n) = +\infty$ if and only if $\prod(1 - x_n) = 0$.

Exercise 4.1.57. Suppose $\prod(1+x_n)$ converges to a positive number but $\prod(1+|x_n|)$ diverges. Prove that by rearranging the terms in $\prod(1 + x_n)$, the infinite product can have any positive number as the limit, or to diverge.

Ratio Rule

By relating x_n to the partial product of $\prod \dfrac{x_{n+1}}{x_n}$, the limit of the sequence x_n can be studied by considering the ratio $\dfrac{x_{n+1}}{x_n}$. This leads to the extension of the ratio rules in Exercises 1.1.32, 2.2.29, 2.2.32.

Exercise 4.1.58. Suppose $\left|\dfrac{x_{n+1}}{x_n}\right| \le 1 - y_n$ and $0 < y_n < 1$. Use Exercises 4.1.54 and 4.1.56 to prove that if $\sum y_n = +\infty$, then $\lim_{n \to \infty} x_n = 0$.

Exercise 4.1.59. Suppose $\left|\dfrac{x_{n+1}}{x_n}\right| \ge 1 + y_n$ and $0 < y_n < 1$. Prove that if $\sum y_n = +\infty$, then $\lim_{n \to \infty} x_n = \infty$.

Exercise 4.1.60. Suppose $1 - y_n \le \dfrac{x_{n+1}}{x_n} \le 1 + z_n$ and $0 < y_n, z_n < 1$. Prove that if $\sum y_n$ and $\sum z_n$ converge, then $\lim_{n \to \infty} x_n$ converges to a nonzero limit.

Exercise 4.1.61. Study $\lim_{n \to \infty} \dfrac{(n+a)^{n+\frac{1}{2}}}{(\pm e)^n n!}$, the case not yet settled in Exercise 2.2.30.

An Example by Borel

Suppose $a_n > 0$ and $\sum \sqrt{a_n}$ converges. Suppose $\{r_n\}$ is all the rational numbers in $[0, 1]$. We study the convergence of the series $\sum \dfrac{a_n}{|x - r_n|}$ for $x \in [0, 1]$?

Exercise 4.1.62. Prove that if $x \notin \cup_n (r_n - c\sqrt{a_n}, r_n + c\sqrt{a_n})$, then the series converges.

Exercise 4.1.63. Use Heine-Borel theorem to prove that if $\sum \sqrt{a_n} < \dfrac{1}{2c}$, then $[0, 1] \not\subset \cup_n (r_n - c\sqrt{a_n}, r_n + c\sqrt{a_n})$. By Exercise 4.1.62, this implies that the series converges for some $x \in [0, 1]$.

Approximate Partial Sum by Integral

The proof of Proposition 4.1.3 gives an estimation of the partial sum for $\sum f(n)$ by the integral of $f(x)$ on suitable intervals. The idea leads to an estimation of $n!$ in Exercise 4.1.26. In what follows, we study the approximation in general.

Suppose $f(x)$ is a decreasing function on $[1, +\infty)$ satisfying $\lim_{x \to +\infty} f(x) = 0$. Denote $d_n = f(1) + f(2) + \cdots + f(n-1) - \displaystyle\int_1^n f(x)dx$.

Exercise 4.1.64. Prove that d_n is increasing and $0 \le d_n \le f(1) - f(n)$. This implies that d_n converge to a limit γ.

Exercise 4.1.65. Prove that if $f(x)$ is convex, then $d_n \ge \dfrac{1}{2}(f(1) - f(n))$.

Exercise 4.1.66. By using Exercise 3.1.53, prove that if $f(x)$ is convex and differentiable, then for $m > n$, we have

$$\left| d_n - d_m + \frac{1}{2}(f(n) - f(m)) \right| \le \frac{1}{8}(f'(m) - f'(n)).$$

Exercise 4.1.67. For convex and differentiable $f(x)$, prove

$$\left| d_n - \gamma + \frac{1}{2}f(n) \right| \le -\frac{1}{8}f'(n).$$

Exercise 4.1.68. By using Exercise 3.2.34, prove that if $f(x)$ has second order derivative, then for $m > n$, we have

$$\frac{1}{24} \sum_{k=n}^{m-1} \inf_{[k,k+1]} f'' \le d_n - d_m + \frac{1}{2}(f(n) - f(m)) - \frac{1}{8}(f'(n) - f'(m)) \le \frac{1}{24} \sum_{k=n}^{m-1} \sup_{[k,k+1]} f''.$$

Exercise 4.1.69. Let γ be the Euler-Mascheroni constant in Exercise 1.4.38. Prove

$$\frac{1}{24(n+1)^2} \le 1 + \frac{1}{2} + \cdots + \frac{1}{n-1} - \log n - \gamma + \frac{1}{2n} + \frac{1}{8n^2} \le \frac{1}{24(n-1)^2}.$$

Exercise 4.1.70. Estimate $1 + \dfrac{1}{\sqrt{2}} + \dfrac{1}{\sqrt{3}} + \cdots + \dfrac{1}{\sqrt{n-1}} - 2\sqrt{n}$ (see Exercise 1.2.14).

4.2   Series of Functions

We study of the series of functions, especially the Taylor and Fourier series. To preserve the properties such as limit, differentiation, integration of the functions under the infinite sum, some condition on the uniformity on the convergence is needed. We first introduce such uniformity condition for sequences and series of functions. Then we will study Taylor and Fourier series in detail.

4.2.1   Uniform Convergence

A sequence of functions $f_n(x)$ *converges* to a function $f(x)$ if $\lim_{n \to \infty} f_n(x) = f(x)$ for each x. The following are some examples.

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

$$\lim_{n \to \infty} x^n = \begin{cases} 0 & \text{if } |x| < 1 \\ 1 & \text{if } x = 1 \end{cases}. \qquad\qquad x \in (-1, 1]$$

$$\lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n = e^x, \qquad\qquad\qquad x \in (-\infty, +\infty)$$

The second example indicates that the continuity of $f_n(x)$ does not necessarily imply the continuity of the limit function. In other words, the equality $\lim_{n \to \infty} \lim_{x \to a} f_n(x) = \lim_{x \to a} f(x) = \lim_{x \to a} \lim_{n \to \infty} f_n(x)$ does not necessarily hold. The next example shows that the limit and the integration may not commute.

Example 4.2.1. Consider the function $f_n(x)$ in Figure 4.3. We have $\lim_{n \to \infty} n f_n(x) = 0$ and $\displaystyle\int_0^1 n f_n(x) dx = \frac{1}{2}$. In particular,

$$\lim_{n \to \infty} \int_0^1 n f_n(x) dx = \frac{1}{2} \ne \int_0^1 \lim_{n \to \infty} n f_n(x) dx = 0.$$
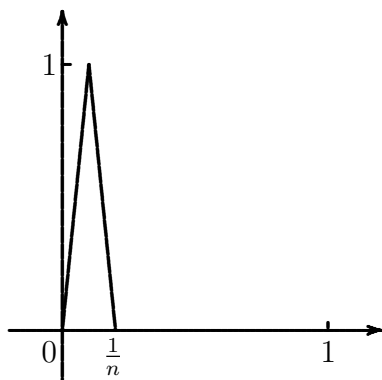
Figure 4.3: a non-uniform convergent sequence

Given $\lim_{n\to\infty} f_n(x) = f(x)$ on $[a, b]$, how can we prove $\lim_{n\to\infty} \int_a^b f_n(x)dx = \int_a^b f(x)dx$? We have

$$\left| \int_a^b f_n(x)dx - \int_a^b f(x)dx \right| \leq \int_a^b |f_n(x) - f(x)|dx.$$

If for any $\epsilon > 0$, there is N, such that $|f_n(x) - f(x)| < \epsilon$ for all $n > N$ and $x \in [a, b]$, then we have $\left| \int_a^b f_n(x)dx - \int_a^b f(x)dx \right| \leq \epsilon(b - a)$ for $n > N$. The proof leads to the following definition.

Definition 4.2.1. A sequence of functions $f_n(x)$ *uniformly converges* to a function $f(x)$ on $[a, b]$ if for any $\epsilon > 0$, there is N, such that

$$n > N, x \in [a, b] \implies |f_n(x) - f(x)| < \epsilon. \tag{4.2.1}$$

The key point of the definition is that the inequality holds for all x in the defining domain. In other words, N is independent of the choice of x. Of course, the defining domain does not have to be closed interval only. The interval $[a, b]$ in the definition can be replaced by any set of numbers.
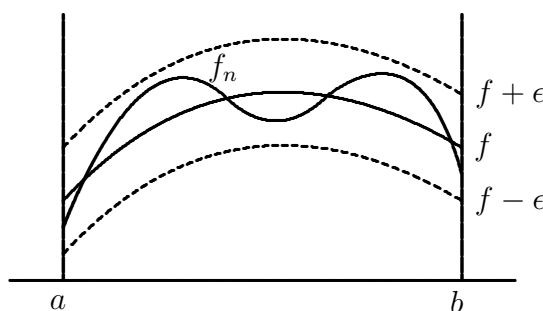
Figure 4.4: uniform convergence

It is not difficult to derive the Cauchy criterion for the uniform convergence (the defining domain is omitted in the statement).

Proposition 4.2.2 (Cauchy Criterion). *A sequence $f_n(x)$ uniformly converges if and only if for any $\epsilon > 0$, there is N, such that*

$$m, n > N \implies |f_m(x) - f_n(x)| < \epsilon. \tag{4.2.2}$$

Example 4.2.2. For any fixed R and $\epsilon > 0$, we have

$$x \geq R, n > \frac{1}{\epsilon} - R \implies n + x > \frac{1}{\epsilon} \implies \left| \frac{1}{n+x} \right| < \epsilon.$$

Therefore $\dfrac{1}{n+x}$ uniformly converges to 0 on $[R, +\infty)$ for any R.

On the other hand, the sequence is defined on the set X of all real numbers except negative integers. However, the sequence is not uniformly convergent on X. Specifically, for $\epsilon = 1$ and any N, pick a natural number $n > N$. Then for $x = -n + \dfrac{1}{2} \in X$, we have $\left| \dfrac{1}{n+x} - 0 \right| = 2 > \epsilon$.

Example 4.2.3. Let $0 < R < 1$. For any $\epsilon > 0$, we have

$$|x| \leq R, n > \log_R \epsilon = \frac{\log \epsilon}{\log R} \implies |x^n| \leq R^n < R^{\log_R \epsilon} = \epsilon.$$

Therefore x^n uniformly converges to 0 on $[-R, R]$ for any $0 < R < 1$.

On the other hand, for any $0 < \epsilon < 1$ and any N, pick a natural number $n > N$. Then we have $0 < x = \sqrt[n]{\epsilon} < 1$ and $x^n = \epsilon$. This shows the convergence is not uniform on $(0, 1)$.

Example 4.2.4. We have $\lim_{n \to \infty} n \log \left(1 + \dfrac{x}{n} \right) = x$ for any x. Moreover, for any fixed R and $\epsilon > 0$, by Taylor expansion we have

$$|x| \leq R, n > \frac{R^2}{2\epsilon} \implies \left| n \log \left(1 + \frac{x}{n} \right) - x \right| = \left| n \left(\frac{x}{n} - \frac{c^2}{2n^2} \right) - x \right| \leq \frac{R^2}{2n} < \epsilon,$$

where $|c| < |x|$. Therefore the convergence is uniform on $[-R, R]$ for any $R > 0$.

Taking the exponential, we get $\lim_{n \to \infty} \left(1 + \dfrac{x}{n} \right)^n = e^x$. To see that the convergence is also uniform on $[-R, R]$, we use the uniform continuity of e^x on $[-R - 1, R + 1]$. For any $\epsilon > 0$, there is $0 < \delta < 1$, such that

$$|x| \leq R + 1, |y| \leq R + 1, |x - y| < \delta \implies |e^x - e^y| < \epsilon.$$

Then by the estimation above, we get

$$|x| \leq R, n > \frac{R^2}{2\delta} \implies \left| n \log \left(1 + \frac{x}{n} \right) - x \right| \leq \delta < 1$$
$$\implies \left| \left(1 + \frac{x}{n} \right)^n - e^x \right| = \left| e^{n \log \left(1 + \frac{x}{n} \right)} - e^x \right| < \epsilon.$$

On the other hand, for any fixed n, we have $\lim_{x \to +\infty} \left(\left(1 + \dfrac{x}{n} \right)^n - e^x \right) = \infty$. This implies that for $\epsilon = 1$ and any natural number n, there is x, such that $\left| \left(1 + \dfrac{x}{n} \right)^n - e^x \right| > \epsilon$. This shows that the convergence is not uniform on $(-\infty, +\infty)$.

Exercise 4.2.1. Determine the intervals on which the sequences uniformly converge.

1. $x^{\frac{1}{n}}$.

2. $n(x^{\frac{1}{n}} - 1)$.

3. $\dfrac{1}{nx + 1}$.

4. $\dfrac{\sin nx}{n}$.

5. $\sin \dfrac{x}{n}$.

6. $\sqrt[n]{1 + x^n}$.

7. $\left(x + \dfrac{1}{n}\right)^{\alpha}$.

8. $\left(1 + \dfrac{x}{n}\right)^{\alpha}$.

9. $\log\left(1 + \dfrac{x}{n}\right)$.

Exercise 4.2.2. Suppose $\lim_{n \to \infty} f_n(x) = f(x)$ uniformly on $[a, b]$ and uniformly on $[b, c]$. Prove that $\lim_{n \to \infty} f_n(x) = f(x)$ uniformly on $[a, c]$. What about other combinations of intervals?

Exercise 4.2.3. Prove that $\lim_{n \to \infty} f_n(x) = f(x)$ uniformly if and only if $\lim_{n \to \infty} \sup |f_n(x) - f(x)| = 0$, where the supremum is taken over all x in the defining domain.

Exercise 4.2.4. Suppose $\lim_{n \to \infty} f_n(x) = f(x)$ uniformly on X. Suppose $g(y)$ is a function on Y such that $g(y) \in X$ for any $y \in Y$. Prove that $\lim_{n \to \infty} f_n(g(y)) = f(g(y))$ uniformly on Y. Is it also true that $\lim_{n \to \infty} f_n(x) = f(x)$ not uniformly implies $\lim_{n \to \infty} f_n(g(y)) = f(g(y))$ not uniformly?

Exercise 4.2.5. Suppose $\lim_{n \to \infty} f_n(x) = f(x)$ uniformly on X. Suppose $f_n(x) \in Y$ for all n and all $x \in X$. Prove that if $g(y)$ is uniformly continuous on Y, then $\lim_{n \to \infty} g(f_n(x)) = g(f(x))$ uniformly on X.

Exercise 4.2.6. Are the sum, product, composition, maximum, etc, of two uniformly convergent sequences of functions still uniformly convergent?

Exercise 4.2.7. Suppose $f(x)$ is integrable on $[a, b + 1]$. Prove that the sequence $f_n(x) = \dfrac{1}{n} \sum_{i=0}^{n-1} f\left(x + \dfrac{i}{n}\right)$ converges uniformly to $\displaystyle\int_x^{x+1} f(t)dt$ on $[a, b]$.

A series $\sum u_n(x)$ of functions uniformly converges if the sequence of partial sum functions uniformly converges. Applying the Cauchy criterion, we find $\sum u_n(x)$ uniformly converges if and only if for any $\epsilon > 0$, there is N, such that

$$n \geq m > N \implies |u_m(x) + u_{m+1}(x) + \cdots + u_n(x)| < \epsilon. \qquad (4.2.3)$$

Based on this, it is possible to extend various results on the convergence of series of numbers to the uniform convergence of series of functions. For example, the uniform convergence of a series $\sum u_n(x)$ would imply that the sequence $u_n(x)$ uniformly converges to 0. The comparison test can also be extended.

Proposition 4.2.3 (Comparison Test). *Suppose $|u_n(x)| \leq v_n(x)$. If $\sum v_n(x)$ uniformly converges, then $\sum u_n(x)$ uniformly converges.*

It is possible for $\sum u_n(x)$ to uniformly converge but $\sum |u_n(x)|$ not uniformly converge (either diverge, or converge but not uniformly). So we also need to extend the Leibniz test, the Dirichlet test, and the Abel test. The proof is the same as before and makes the additional use of the following condition: A sequence of functions $f_n(x)$ is *uniformly bounded* if there is M, such that $|f_n(x)| < M$ for all n and x.

Proposition 4.2.4 (Leibniz Test). *Suppose $u_n(x)$ is a monotone sequence for each x, and uniformly converges to 0. Then $\sum (-1)^n u_n(x)$ uniformly converges.*

Proposition 4.2.5 (Dirichlet Test). *Suppose $u_n(x)$ is a monotone sequence for each x, and uniformly converges to 0. Suppose the partial sums of $\sum v_n(x)$ are uniformly bounded. Then $\sum u_n(x)v_n(x)$ uniformly converges.*

Proposition 4.2.6 (Abel Test). *Suppose $u_n(x)$ is a monotone sequence for each x, and is uniformly bounded. Suppose $\sum v_n(x)$ uniformly converges. Then $\sum u_n(x)v_n(x)$ uniformly converges.*

Note that in the monotone condition, it is allowed to have $u_n(x)$ increasing for some x and decreasing for some other x.

Example 4.2.5. We have $\left| \dfrac{(-1)^n}{n^2 + x^2} \right| \leq \dfrac{1}{n^2}$. By considering $\dfrac{1}{n^2}$ as constant functions, the series $\sum \dfrac{1}{n^2}$ uniformly converges. Therefore by the comparison test, the series $\sum \dfrac{(-1)^n}{n^2 + x^2}$ also uniformly converges.

Example 4.2.6. The partial sum of the geometric series $1 + x + x^2 + \cdots + x^n + \cdots$ is $\dfrac{1 - x^{n+1}}{1 - x}$, and the sum is $\dfrac{1}{1 - x}$ for $|x| < 1$. By

$$\left| \frac{1 - x^{n+1}}{1 - x} - \frac{1}{1 - x} \right| = \frac{|x^{n+1}|}{1 - x},$$

and an argument similar to the one in Example 4.2.3, we find the series uniformly converges on $[-R, R]$ for any $0 < R < 1$, but does not uniformly converge on $(-1, 1)$.

Example 4.2.7. The series $\sum \dfrac{(-1)^n}{n} x^n$ converges to a function $f(x)$ on $(-1, 1]$. We expect $f(x) = -\log(1 + x)$ but the fact is yet to be established. Where does it converge uniformly?

For any $0 < R < 1$, we have $\left| \dfrac{(-1)^n}{n} x^n \right| \leq R^n$ on $[-R, R]$. By considering each R^n as a constant function, the series $\sum R^n$ converges uniformly. Therefore by the comparison test, the series $\sum \dfrac{(-1)^n}{n} x^n$ converges uniformly on $[-R, R]$.

The convergence is not uniform on $(-1, 0]$. For any $0 < x < 1$, we have

$$\left| \sum_{i=n+1}^{2n} \frac{(-1)^i}{i} (-x)^i \right| = \frac{x^{n+1}}{n+1} + \frac{x^{n+2}}{n+2} + \cdots + \frac{x^{2n}}{2n} \geq \frac{x^{2n}}{2}.$$

For $\epsilon = \dfrac{1}{3}$ and any n, we can find x close to 1, such that $\dfrac{x^{2n}}{2} > \epsilon$. Therefore the Cauchy criterion for the uniform convergence fails.

The convergence is uniform on $[0, 1]$. For each $x \in [0, 1]$, the series is an alternating series and $\dfrac{1}{n} x^n$ is decreasing. By the estimation in Proposition 4.1.4, we get

$$\left| \sum_{i=0}^{n} \frac{(-1)^i}{i} x^i - f(x) \right| \leq \frac{1}{n+1} x^{n+1} \leq \frac{1}{n+1}$$

for all $x \in [0, 1]$. From this it is easy to see the series uniformly converges on $[0, 1]$.

The uniform convergence on $[0, 1]$ will be extended in a later Proposition 4.2.13.

Example 4.2.8. By Example 2.3.16, the Taylor series $1 + \dfrac{1}{1!}x + \dfrac{1}{2!}x^2 + \cdots + \dfrac{1}{n!}x^n + \cdots$ of e^x converges to e^x. By the remainder formula $(2.3.6)$ in Proposition 2.3.3, the difference between the sum and the partial sum is

$$\left| \sum_{k=0}^{n} \frac{1}{k!}x^k - e^x \right| < \frac{e^{|x|}|x|^{n+1}}{(n+1)!}.$$

Then it is easy to see that for any R, the Taylor series uniformly converges for $|x| \le R$.

Alternatively, for $|x| \le R$, we have $\left| \dfrac{x^n}{n!} \right| \le \dfrac{R^n}{n!}$. Then the series $\sum \dfrac{R^n}{n!}$, in which each term is considered as a constant function, uniformly converges. Therefore by the comparison test, the Taylor series uniformly converges.

On the other hand, the Taylor series of e^x does not uniformly converge on $(-\infty, +\infty)$ because the sequence $\dfrac{x^n}{n!}$ does not uniformly converge to 0.

Example 4.2.9. By Example 4.1.12, we know the series $\sum \dfrac{\sin nx}{n^\alpha}$ converges for all x and $\alpha > 0$.

If $\alpha > 1$, then by $\left| \dfrac{\sin nx}{n^\alpha} \right| \le \dfrac{1}{n^\alpha}$ and the convergence of $\sum \dfrac{1}{n^\alpha}$, we conclude that $\sum \dfrac{\sin nx}{n^\alpha}$ uniformly converges on the whole real line.

Suppose $0 < \alpha \le 1$. Then in the argument in Example 4.1.12, the partial sum of the series $\sum \sin nx$ is bounded by $\dfrac{1}{\left| \sin \dfrac{1}{2}x \right|}$. For any $\delta > 0$, the bounded is uniform on the interval $[\delta, 2\pi - \delta]$. Since $\dfrac{1}{n^\alpha}$ can be considered as a monotone sequence uniformly converging to 0, we may use the Dirichlet test to conclude that the series $\sum \dfrac{\sin nx}{n^\alpha}$ uniformly converges on $[\delta, 2\pi - \delta]$.

Exercise 4.2.8. Determine the intervals on which the series uniformly converge.

1. $\sum x^n e^{-nx}$.

2. $\sum \dfrac{x^n}{\sqrt{n}}$.

3. $\sum \left(\dfrac{x(x+n)}{n} \right)^n$.

4. $\sum \dfrac{1}{n^\alpha + x^\alpha}$.

5. $\sum \dfrac{1}{x + a^n}$.

6. $\sum \dfrac{x^n}{1 - x^n}$.

7. $\sum \dfrac{\sin nx}{n^\alpha (\log n)^\beta}$.

8. $\sum \dfrac{\cos nx}{n^\alpha}$.

9. $\sum \dfrac{\sin^3 nx}{n^\alpha}$.

Exercise 4.2.9. Show that $\sum (-1)^n x^n (1 - x)$ uniformly converges on $[0, 1]$ and absolutely converges for each $x \in [0, 1]$. However, the series $\sum |(-1)^n x^n (1 - x)|$ does not uniformly converge.

Exercise 4.2.10. Regarding Exercise 4.1.25, if $\sum f_n(x)^2$ uniformly converges, can you conclude that $\sum \dfrac{f_n(x)}{n^\alpha}$ also uniformly converges for any $\alpha > \dfrac{1}{2}$?

Exercise 4.2.11. Can you establish the uniform convergence version of Exercise 4.1.16?

4.2.2   Properties of Uniform Convergence

Proposition 4.2.7. *Suppose $f_n(x)$ uniformly converges to $f(x)$ for $x \neq a$. If $\lim_{x \to a} f_n(x) = l_n$ exists, then both $\lim_{n \to \infty} l_n$ and $\lim_{x \to a} f(x)$ converge and are equal.*

The conclusion is

$$\lim_{x \to a} \lim_{n \to \infty} f_n(x) = \lim_{n \to \infty} \lim_{x \to a} f_n(x). \tag{4.2.4}$$

Therefore uniform convergence implies that the two limits $\lim_{x \to a}$ and $\lim_{n \to \infty}$ commute.

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

$$m, n > N, x \neq a \implies |f_m(x) - f_n(x)| < \epsilon. \tag{4.2.5}$$

Taking $\lim_{x \to a}$ of the right side of (4.2.5), we get $|l_m - l_n| \leq \epsilon$. Therefore $\lim_{n \to \infty} l_n$ converges by the Cauchy criterion. Let $l = \lim_{n \to \infty} l_n$. Then

$$n > N \implies |l - l_n| \leq \epsilon.$$

On the other hand, taking $\lim_{m \to \infty}$ of the right side of (4.2.5), we get

$$n > N, x \neq a \implies |f(x) - f_n(x)| \leq \epsilon.$$

Fix one $n > N$. Since $\lim_{x \to a} f_n(x) = l_n$, there is $\delta > 0$, such that

$$0 < |x - a| < \delta \implies |f_n(x) - l_n| < \epsilon.$$

Then for the fixed choice of n, we have

$$0 < |x - a| < \delta \implies |f(x) - l| \leq |f(x) - f_n(x)| + |f_n(x) - l_n| + |l - l_n| < 3\epsilon.$$

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

$\square$

An immediate consequence of Proposition 4.2.7 is that the uniform convergence preserves the continuity.

Proposition 4.2.8. *Suppose $f_n(x)$ is continuous and the sequence $f_n(x)$ converges to $f(x)$.*

1. *If the convergence is uniform, then $f(x)$ is also continuous.*

2. *(Dini Theorem) If $f_n(x)$ is monotone in n and $f(x)$ is continuous, then the convergence is uniform on any bounded and closed interval.*

Proof. By taking $l_n = \lim_{x \to a} f_n(x) = f_n(a)$, the first part is a consequence of Proposition 4.2.7.

Now suppose $f_n(x)$ is monotone in n and $f(x)$ is continuous. If the limit $\lim f_n(x) = f(x)$ is not uniform on $[a, b]$, then there is $\epsilon > 0$, a subsequence $f_{n_k}(x)$, and a sequence $x_k \in [a, b]$, such that $|f_{n_k}(x_k) - f(x_k)| \geq \epsilon$. In the

bounded and closed interval $[a, b]$, x_k has a convergent subsequence. Thus without loss of generality, we may assume x_k converges to $c \in [a, b]$. By the monotone assumption, we have

$$n \leq m \implies |f_n(x) - f(x)| \geq |f_m(x) - f(x)| \geq \epsilon.$$

Therefore

$$n \leq n_k \implies |f_n(x_k) - f(x_k)| \geq |f_{n_k}(x_k) - f(x_k)| \geq \epsilon.$$

For each fixed n, by taking $k \to \infty$ and the continuity of $f(x)$ and $f_n(x)$, we get $|f_n(c) - f(c)| \geq \epsilon$. This contradicts with the assumption that $f_n(c)$ converges to $f(c)$. Thus we proved that $f_n(x)$ must converge to $f(x)$ uniformly on $[a, b]$.

$\square$

Note that in the proof of the second part, we only need $f_n(x)$ to be monotone for each x. It is not required that the sequence should be increasing for all x or decreasing for all x.

Example 4.2.10. Since x^n are continuous and the limit

$$\lim_{n \to \infty} x^n = \begin{cases} 0 & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x = 1 \end{cases}$$

is not continuous on $[0, 1]$, the convergence is not uniform on $[0, 1]$. On the other hand, for any $0 < R < 1$, the sequence x^n is decreasing in n on $[0, R]$ and the limit function 0 is continuous on the interval. By Dini Theorem, the limit $\lim_{n \to \infty} x^n = 0$ is uniform on $[0, R]$.

Example 4.2.11. By Example 4.2.4, we know the function e^x is the uniform limit of the polynomials $\left(1 + \dfrac{x}{n}\right)^n$ on $[-R, R]$ for any $R > 0$. Since polynomials are continuous, we conclude that e^x is continuous on any $[-R, R]$. In other words, e^x is continuous on $(-\infty, +\infty)$.

Instead of deriving the continuity of e^x from the uniform convergence, we may also use the continuity of e^x and the fact that the sequence $\left(1 + \dfrac{x}{n}\right)^n$ is monotone to derive that the convergence is uniform on $[-R, R]$ for any $R > 0$.

Example 4.2.12. The function $f_n(x)$ given in Figure 4.3 is continuous, and the limit $\lim_{n \to \infty} f_n(x) = 0$ is also continuous. However, the convergence is not uniform. Note that the sequence $f_n(x)$ is not monotone, so that Dini Theorem cannot be applied.

Exercise 4.2.12. Suppose $f_n(x)$ uniformly converges to $f(x)$. Prove that if $f_n(x)$ are uniformly continuous, then $f(x)$ is uniformly continuous.

The discussion leading to the definition of uniform convergence makes us expect the following.

Proposition 4.2.9. *Suppose $f_n(x)$ is integrable on a bounded interval $[a, b]$ and the sequence $f_n(x)$ uniformly converges. Then $\lim_{n \to \infty} f_n(x)$ is integrable, with*

$$\int_a^b \lim_{n \to \infty} f_n(x) dx = \lim_{n \to \infty} \int_a^b f_n(x) dx. \qquad (4.2.6)$$

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

$$m, n > N, x \in [a, b] \implies |f_m(x) - f_n(x)| < \epsilon.$$

This further implies

$$m, n > N \implies \left| \int_a^b f_m(x)dx - \int_a^b f_n(x)dx \right| \leq \epsilon(b - a).$$

By Cauchy criterion, the limit $\lim_{n \to \infty} \int_a^b f_n(x)dx$ converges. Denote the limit by I.

Let $f(x) = \lim_{n \to \infty} f_n(x)$. For any $\epsilon > 0$, there is N, such that

$$n > N, x \in [a, b] \implies |f_n(x) - f(x)| < \epsilon.$$

This implies that for any partition P of $[a, b]$ and the same choice of x_i^* for all functions, we have

$$n > N \implies |S(P, f_n(x)) - S(P, f(x))| < \epsilon(b - a).$$

Now we fix one $n > N$ satisfying

$$\left| \int_a^b f_n(x)dx - I \right| < \epsilon.$$

For the fixed integrable function $f_n(x)$, there is $\delta > 0$, such that

$$\|P\| < \delta \implies \left| S(P, f_n(x)) - \int_a^b f_n(x)dx \right| < \epsilon.$$

Combining everything together, we find $\|P\| < \delta$ implies

$$
\begin{aligned}
|S(P, f(x)) - I| \leq & |S(P, f_n(x)) - S(P, f(x))| \\
& + \left| S(P, f_n(x)) - \int_a^b f_n(x)dx \right| + \left| \int_a^b f_n(x)dx - I \right| \\
< & \epsilon(b - a + 2).
\end{aligned}
$$

This shows that $f(x)$ is integrable, with $\int_a^b f(x)dx = I$.

$\square$

Example 4.2.13. Let $r_n, n = 1, 2, \ldots$, be all the rational numbers in $[0, 1]$. Then the functions

$$f_n(x) = \begin{cases} 1 & \text{if } x = r_1, r_2, \ldots, r_n \\ 0 & \text{otherwise} \end{cases}$$

are integrable. However, the limit $\lim_{n \to \infty} f_n(x) = D(x)$ is the Dirichlet function, which is not Riemann integrable. Of course the limit is not uniform.

The example shows that the limit of Riemann integrable functions are not necessarily Riemann integrable. Thus we often need to attach the uniform convergence condition if we want to consider the limit of Riemann integral. The annoyance will be resolved by the introduction of Lebesgue integral. The Lebesgue integral is an extension of the Riemann integral that allows more functions (such as the Dirichlet function) to be integrable, and has the property that the limit of Lebesgue integrable functions are (almost always) Lebesgue integrable.

Example 4.2.14. By Example 4.2.3, the limit

$$\lim_{n\to\infty} x^n = f(x) = \begin{cases} 0 & \text{if } |x| < 1 \\ 1 & \text{if } x = 1 \end{cases}$$

is not uniform on $[0, 1]$. However, we still have

$$\lim_{n\to\infty} \int_0^1 x^n dx = \lim_{n\to\infty} \frac{1}{n+1} = 0 = \int_0^1 f(x)dx.$$

In general, we have the *dominant convergence theorem* in the theory of Lebesgue integral: If $f_n(x)$ are uniformly bounded and Lebesgue integrable, then $f(x) = \lim_{n\to\infty} f_n(x)$ is Lebesgue integrable, and the equality (4.2.6) holds.

Exercise 4.2.13. Suppose $f_n(x)$ is integrable on a bounded interval $[a, b]$ and $\lim_{n\to\infty} f_n(x) = f(x)$ uniformly. Prove that the convergence $\lim_{n\to\infty} \int_a^x f_n(t)dt = \int_a^x \lim_{n\to\infty} f_n(t)dt$ is uniform for $x \in [a, b]$.

Exercise 4.2.14. Extend Proposition 4.2.9 to the Riemann-Stieltjes integral.

Exercise 4.2.15. Suppose f is Riemann-Stieltjes integrable with respect to each α_n. Will the uniform convergence of α_n tell you something about the limit of $\int_a^b f d\alpha_n$.

Proposition 4.2.10. *Suppose $f_n(x)$ are differentiable on an interval, such that $f_n(x_0)$ converges at some point x_0 and $f_n'(x)$ uniformly converges to $g(x)$. Then $f_n(x)$ uniformly converges and*

$$\left(\lim_{n\to\infty} f_n(x)\right)' = \lim_{n\to\infty} f_n'(x). \tag{4.2.7}$$

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

$$m, n > N \implies |f_m'(x) - f_n'(x)| < \epsilon, \ \ |f_m(x_0) - f_n(x_0)| < \epsilon.$$

Applying the mean value theorem to $f_m(x) - f_n(x)$, we get

$$|(f_m(y) - f_n(y)) - (f_m(x) - f_n(x))| = |f_m'(c) - f_n'(c)||y - x| \le \epsilon|x - y| \tag{4.2.8}$$

for $m, n > N$ and any x, y in the interval. In particular, we get

$$|f_m(x) - f_n(x)| \le \epsilon|x_0 - x| + |f_m(x_0) - f_n(x_0)| \le \epsilon(|x_0 - x| + 1)$$

for $m, n > N$ and any x. This implies that the sequence $f_n(x)$ uniformly converges on any bounded interval.

Let $f(x) = \lim_{n\to\infty} f_n(x)$. By taking $\lim_{m\to\infty}$ in (4.2.8), we get

$$|(f(y) - f(x)) - (f_n(y) - f_n(x))| = |(f(y) - f_n(y)) - (f(x) - f_n(x))| \le \epsilon|y - x|.$$

This means that for fixed x, the function $g_n(y) = \dfrac{f_n(y) - f_n(x)}{y - x}$ uniformly converges to $g(y) = \dfrac{f(y) - f(x)}{y - x}$ for all y not equal to x. By Proposition 4.2.7, we conclude that $\lim_{y\to x} g(y)$ converges and

$$f'(x) = \lim_{y\to x} g(y) = \lim_{n\to\infty}\lim_{y\to x} g_n(y) = \lim_{n\to\infty} f_n'(x).$$

$\square$

Exercise 4.2.16. Derive the derivative of e^x from $e^x = \lim_{n \to \infty} \left(1 + \dfrac{x}{n}\right)^n$ or $e^x = \lim_{n \to \infty} \left(1 + \dfrac{1}{1!}x + \dfrac{1}{2!}x^2 + \cdots + \dfrac{1}{n!}x^n\right)$ and the uniform convergence.

The properties of uniform convergence can be rephrased for series.

1. Suppose $\sum u_n(x)$ uniformly converges. Then taking the limit commutes with the sum:
$$\sum \lim_{x \to a} u_n(x) = \lim_{x \to a} \sum u_n(x).$$

2. Suppose $u_n(x)$ is continuous and $\sum u_n(x)$ uniformly converges. Then $\sum u_n(x)$ is also continuous.

3. (Dini Theorem) Suppose $u_n(x)$ is continuous, $u_n(x) \geq 0$ and $\sum u_n(x)$ converges to a continuous function. Then $\sum u_n(x)$ uniformly converges.

4. Suppose $u_n(x)$ is integrable and $\sum u_n(x)$ uniformly converges. Then $\sum u_n(x)$ is integrable and the integration commutes with the sum:
$$\int_a^b \sum u_n(x)dx = \sum \int_a^b u_n(x)dx.$$

5. Suppose $\sum u_n(x_0)$ converges and $\sum u_n'(x)$ uniformly converges. Then $\sum u_n(x)$ uniformly converges and is differentiable, and the derivative commutes with the sum:
$$\left(\sum u_n(x)\right)' = \sum u_n'(x).$$

Example 4.2.15. By Example 4.1.8, the Riemann zeta function
$$\zeta(x) = 1 + \frac{1}{2^x} + \frac{1}{3^x} + \cdots + \frac{1}{n^x} + \cdots$$
is defined on $(1, +\infty)$. For any $R > 1$, we have
$$x \geq R \implies 0 < \frac{1}{n^x} \leq \frac{1}{n^R}.$$

Since the series $\sum \dfrac{1}{n^R}$ of numbers (considered as constant functions) converges, the series $\sum \dfrac{1}{n^x}$ uniformly converges on $[R, +\infty)$. Therefore $\zeta(x)$ is continuous on $[R, +\infty)$ for any $R > 1$. Since $R > 1$ is arbitrary, we conclude that $\zeta(x)$ is continuous on $(1, +\infty)$.

Consider the series $-\sum \dfrac{\log n}{n^x}$ obtained by differentiating $\zeta(x)$ term by term. For any $R > 1$, choose R' satisfying $R > R' > 1$. Then $0 < \dfrac{\log n}{n^x} \leq \dfrac{\log n}{n^R} < \dfrac{1}{n^{R'}}$ for $x > R$ and sufficiently big n. The convergence of the series $\sum \dfrac{1}{n^{R'}}$ of numbers implies the series $-\sum \dfrac{\log n}{n^x}$ uniformly convergent on $[R, +\infty)$ for any $R > 1$. This implies that $\zeta(x)$ is differentiable and $\zeta'(x) = -\sum \dfrac{\log n}{n^x}$. Further argument shows that $\zeta(x)$ has derivative of any order.

Example 4.2.16. Let $h(x)$ be given by $h(x) = |x|$ on $[-1, 1]$ and $h(x + 2) = h(x)$ for any x. The function is continuous and satisfies $0 \leq h(x) \leq 1$. Therefore

$$f(x) = \sum_{n=0}^{\infty} \left(\frac{3}{4}\right)^n h(4^n x)$$

uniformly converges and is a continuous function. However, we will show that the function is not differentiable anywhere.

Let $\delta_k = \pm \dfrac{1}{2 \cdot 4^k}$. Then for any n, by $|h(x) - h(y)| \leq |x - y|$, we have

$$\left| \frac{h(4^n(a + \delta_k)) - h(4^n a)}{\delta_k} \right| \leq \frac{4^n \delta_k}{\delta_k} = 4^n.$$

For $n > k$, $4^n \delta_k$ is a multiple of 2, and we have

$$\frac{h(4^n(a + \delta_k)) - h(4^n a)}{\delta_k} = 0.$$

For $n = k$, we have $4^k \delta_k = \pm \dfrac{1}{2}$. By choosing $\pm$ sign so that there is no integer between $4^k a$ and $4^k a \pm \dfrac{1}{2}$, we can make sure that $|h(4^k(a + \delta_k)) - h(4^k a)| = |4^k(a + \delta_k) - 4^k a| = \dfrac{1}{2}$. Then

$$\left| \frac{h(4^k(a + \delta_k)) - h(4^k a)}{\delta_k} \right| = 4^k.$$

Thus for any fixed a, by choosing a sequence δ_k with suitable $\pm$ sign, we get

$$\left| \frac{f(a + \delta_k) - f(a)}{\delta_k} \right| \geq \left(\frac{3}{4}\right)^k 4^k - \sum_{n=0}^{k-1} \left(\frac{3}{4}\right)^n 4^n = 3^k - \frac{3^k - 1}{3 - 1} = \frac{3^k + 1}{2}.$$

This implies that $\lim_{\delta \to 0} \dfrac{f(a + \delta) - f(a)}{\delta}$ diverges.

Exercise 4.2.17. Justify the equalities.

1. $\displaystyle \int_0^1 x^{ax} dx = \sum_{n=1}^{\infty} \frac{(-a)^{n-1}}{n^n}$.

2. $\displaystyle \int_{\lambda a}^a \left(\sum_{n=0}^{\infty} \lambda^n \tan \lambda^n x \right) dx = -\log|\cos a|$ for $|a| < \dfrac{\pi}{2}$, $|\lambda| < 1$.

3. $\displaystyle \int_x^{+\infty} (\zeta(t) - 1) dt = \sum_{n=2}^{\infty} \frac{1}{n^x \log n}$ for $x > 1$.

Exercise 4.2.18. Find the places where the series converge and have derivatives. Also find the highest order of the derivative.

1. $\displaystyle \sum \frac{1}{n(\log n)^x}$.

2. $\displaystyle \sum \left(x + \frac{1}{n} \right)^n$.

3. $\sum_{n=-\infty}^{+\infty} \dfrac{1}{|n-x|^\alpha}$.

4. $\sum \dfrac{(-1)^n}{n^x}$.

4.2.3   Power Series

A *power series* is a series of the form

$$\sum_{n=0}^{\infty} a_n x^n = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n + \cdots,$$

or more generally, of the form $\sum_{n=0}^{\infty} a_n (x - x_0)^n$. Taylor series are power series.

Theorem 4.2.11. *Suppose*

$$R = \frac{1}{\overline{\lim}_{n\to\infty} \sqrt[n]{|a_n|}}. \qquad (4.2.9)$$

Then the power series absolutely converges for $|x| < R$ and diverges for $|x| > R$. Moreover, for any $0 < r < R$, the power series uniformly converges for $|x| \le r$.

Proof. For any $0 < r < R$, choose r' satisfying $r < r' < R$. Then we have $\dfrac{1}{r'} > \overline{\lim}_{n\to\infty} \sqrt[n]{|a_n|}$. By the second part of Proposition 1.2.10, this means that $\dfrac{1}{r'} > \sqrt[n]{|a_n|}$ for all but finitely many n. Thus for $|x| \le r$, we have

$$|a_n x^n| = \left(\sqrt[n]{|a_n|}|x| \right)^n < \left(\frac{|x|}{r'} \right)^n \le \left(\frac{r}{r'} \right)^n$$

for all but finitely many n. Since $0 < r < r'$, the series $\sum \left(\dfrac{r}{r'} \right)^n$ of numbers converges. Then by the comparison test, the series $\sum a_n x^n$ of functions uniformly converges for $|x| \le r$.

Since the series converges for $|x| \le r$, where r can be any number satisfying $0 < r < R$, we conclude that the series converges for $|x| < R$. On the other hand, if $|x| > R$, then $\dfrac{1}{|x|} < \overline{\lim}_{n\to\infty} \sqrt[n]{|a_n|}$. By the first part of Proposition 1.2.10, there are infinitely many n satisfying $\dfrac{1}{|x|} < \sqrt[n]{|a_n|}$, or $|a_n x^n| > 1$. Thus the terms of the series $\sum a_n x^n$ do not converge to 0, and the series diverges.

$\square$

The number R in Theorem 4.2.11 is called the *radius of convergence* for the power series. By the theorem, the radius is characterized by the property that the power series converges for $|x| < R$ and diverges for $|x| > R$.

Example 4.2.17. By Example 4.2.6, the geometrical series $\sum x^n$ converges for $|x| < 1$ and diverges for $|x| \ge 1$. Thus the radius of convergence of the geometric series

is 1. Alternatively, the radius of convergence can be obtained by $\overline{\lim}_{n\to\infty} \sqrt[n]{|a_n|} = \lim_{n\to\infty} \sqrt[n]{1} = 1$.

By Example 4.2.8 (or Example 2.3.16), the Taylor series $\sum \dfrac{x^n}{n!}$ of e^x converges for any x. Therefore the radius of convergence is ∞. By the similar reason, the radius of convergence for the Taylor series of $\sin x$ and $\cos x$ is also ∞.

The Taylor series of $\log(1 + x)$ is $\sum \dfrac{(-1)^{n+1}}{n} x^n$. Since $\overline{\lim}_{n\to\infty} \sqrt[n]{|a_n|} = \lim_{n\to\infty} \sqrt[n]{\dfrac{1}{n}} = 1$, the radius of convergence is $\dfrac{1}{1} = 1$.

Example 4.2.18. By $\overline{\lim}_{n\to\infty} \sqrt[n]{n!} = \infty$, the radius of convergence for the series $\sum n! x^n$ is 0. In other words, the series diverges for all $x \neq 0$.

By $\overline{\lim}_{n\to\infty} \sqrt[n]{2^n + 3^n} = 3$, the radius of convergence for the series $\sum (2^n + 3^n) x^n$ is $\dfrac{1}{3}$.

By $\overline{\lim}_{n\to\infty} \sqrt[n]{\dfrac{1}{n^n}} = 0$, the radius of convergence for the series $\sum \dfrac{(-1)^n}{n^n} x^n$ is ∞. In other words, the series converges for all x.

Exercise 4.2.19. Prove the radius of convergence R of a power series $\sum a_n x^n$ satisfies

$$\lim_{n\to\infty} \left| \frac{a_n}{a_{n+1}} \right| \leq R \leq \overline{\lim}_{n\to\infty} \left| \frac{a_n}{a_{n+1}} \right|.$$

Then use the result to show that the radius of convergence for the Taylor series of $(1 + x)^\alpha$ and $\log(1 + x)$ is 1.

Exercise 4.2.20. Suppose the radii of convergence for the power series $\sum a_n x^n$ and $\sum b_n x^n$ are R and R'.

1. Prove that the radius of convergence for the sum power series $\sum (a_n + b_n) x^n$ is at least $\min\{R, R'\}$.

2. Prove that if $R \neq R'$, then the radius of convergence for the sum power series is equal to $\min\{R, R'\}$.

What about the radius of convergence for the product power series $\sum (a_0 b_n + a_1 b_{n-1} + \cdots + a_n b_0) x^n$?

Exercise 4.2.21. Find the radius of convergence.

1. $\sum x^{n^2}$.

2. $\sum \dfrac{2^n}{n} x^n$.

3. $\sum 2^n x^{n^2 - 1}$.

4. $\sum \dfrac{(2n)!}{(n!)^2} x^n$.

5. $\sum a^{n^2} x^n$.

6. $\sum \left(\dfrac{n+1}{n} \right)^n x^n$.

7. $\sum (-1)^n \left(\dfrac{n+1}{n} \right)^{n^2} x^n$.

8. $\sum (-1)^n \left(\dfrac{n+1}{n} \right)^{n^2} x^{n^2}$.

9. $\sum (-1)^n \left(\dfrac{n+1}{2n} \right)^n x^n$.

Due to the uniform convergence, the calculus of the power series can be done term by term.

Proposition 4.2.12. *Suppose R is the radius of convergence of a power series $\sum a_n x^n$. Then the sum has derivative of any order for $|x| < R$. Moreover, the derivative and integral can be taken term by term for $|x| < R$:*

$$(a_0 + a_1 x + \cdots + a_n x^n + \cdots)' = a_1 + 2a_2 x + \cdots + n a_n x^{n-1} + \cdots ,$$

$$\int_0^x (a_0 + a_1 t + \cdots + a_n t^n + \cdots) \, dt = a_0 x + \frac{a_1}{2} x^2 + \frac{a_2}{3} x^3 + \cdots + \frac{a_n}{n} x^{n+1} + \cdots .$$

Note that by the formula (4.2.9), the radius of convergence for the derivative and integral power series remains the same.

Example 4.2.19. The power series $\sum n x^n = x + 2x^2 + 3x^3 + \cdots$ has radius of convergence 1. To find the sum, we note that

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

Therefore

$$\sum n x^n = x(1 + 2x + 3x^2 + \cdots) = \frac{x}{(1-x)^2}.$$

Example 4.2.20. By the equality

$$\frac{1}{1-x} = 1 + x + x^2 + \cdots + x^n + \cdots , \quad |x| < 1,$$

we get

$$\frac{1}{1+x} = 1 - x + x^2 + \cdots + (-1)^n x^n + \cdots , \quad |x| < 1.$$

By integration, we get

$$\log(1 + x) = x - \frac{1}{2} x^2 + \frac{1}{3} x^3 + \cdots + \frac{(-1)^n}{n+1} x^{n+1} + \cdots , \quad |x| < 1.$$

Exercise 4.2.22. Find the sum of power series.

1. $\sum_{n=1}^{\infty} n^2 x^n$.

2. $\sum_{n=0}^{\infty} \frac{x^{2n+1}}{2n+1}$.

3. $\sum_{n=1}^{\infty} \frac{x^n}{n(n+1)}$.

4. $\sum_{n=1}^{\infty} \frac{x^n}{n(n+1)(n+2)}$.

Exercise 4.2.23. Find the Taylor series of the function and the radius of convergence. Then explain why the sum of the Taylor series is the given function.

1. $\arcsin x$.

2. $\int_0^x \frac{\sin t}{t} \, dt$.

3. $\arctan x$.

4. $\int_0^x e^{-t^2} \, dt$.

5. $\log(x + \sqrt{1 + x^2})$.

6. $\int_0^x \frac{dt}{\sqrt{1 - t^4}} \, dt$.

Exercise 4.2.24. Verify that the functions defined by the power series are the solutions of the differential equations.

1. $f(x) = \sum_{n=0}^{\infty} \frac{x^{4n}}{(4n)!}$, $f^{(4)}(x) = f(x)$.

2. $f(x) = \sum_{n=0}^{\infty} \frac{x^n}{(n!)^2}$, $x f''(x) + f'(x) - f(x) = 0$.

Exercise 4.2.25. Show that $f(x) = \sum_{n=1}^{\infty} \frac{x^n}{n^2} = -\int_0^x \frac{\log(1-t)}{t} dt$ and then verify that $f(x) + f(1-x) + \log x \log(1-x) = f(1)$.

Exercise 4.2.26 (Euler). The vibration of a circular drumhead is described by the differential equation

$$f''(x) + \frac{1}{x} f'(x) + \left(a^2 - \frac{b^2}{x^2} \right) f(x) = 0.$$

Verify that

$$f(x) = x^b \left[1 - \frac{1}{b+1} \left(\frac{ax}{2} \right)^2 + \frac{1}{2!(b+1)(b+2)} \left(\frac{ax}{2} \right)^4 - \frac{1}{3!(b+1)(b+2)(b+3)} \left(\frac{ax}{2} \right)^6 + \cdots \right]$$

is a solution of the equation.

Exercise 4.2.27. Derive the derivative of e^x, $\sin x$ and $\cos x$ from the uniform convergence of their Taylor series.

Exercise 4.2.28. Prove that if the radius of convergence for the power series $\sum a_n x^n$ is nonzero, then the sum function $f(x)$ satisfies $f^{(n)}(0) = n! a_n$.

The results so far concerns only with what happens *within* the radius of convergence. The following describes what happens *at* the radius of convergence.

Proposition 4.2.13 (Abel Theorem). *Suppose R is the radius of convergence of a power series. If the power series converges at $x = R$, then the series uniformly converges on $[0, R]$. If the power series converges at $x = -R$, then the series uniformly converges on $[-R, 0]$.*

By Proposition 4.2.7, a consequence of Abel theorem is that if $\sum a_n x^n$ converges at $x = R > 0$, then

$$\lim_{x \to R^-} \sum a_n x^n = \sum a_n R^n.$$

Similar equality holds when the power series converges at $x = -R$.

Proof. Suppose $\sum a_n R^n$ converges. Then the Abel test in Proposition 4.2.6 may be applied to $u_n(x) = \frac{x^n}{R^n}$ and $v_n(x) = a_n R^n$ on $[0, R]$. The functions $u_n(x)$ are all bounded by 1, and the sequence $u_n(x)$ is increasing for each $x \in [0, R]$. The convergence of series $\sum v_n(x)$ is uniform because the series is independent of x. The conclusion is that $\sum u_n(x) v_n(x) = \sum a_n x^n$ converges uniformly on $[0, R]$. $\square$

Example 4.2.21. By Example 4.2.17, the Taylor series of $\log(1+x)$ has radius of convergence 1. The series also converges at $x = 1$ because it is alternating. By Example 4.2.20 and the remark made after Proposition 4.2.13, we have

$$1 - \frac{1}{2} + \frac{1}{3} + \cdots + \frac{(-1)^{n+1}}{n} + \cdots$$
$$= \lim_{x \to 1^-} \left(x - \frac{1}{2} x^2 + \frac{1}{3} x^3 + \cdots + \frac{(-1)^n}{n+1} x^{n+1} + \cdots \right)$$
$$= \lim_{x \to 1^-} \log(1+x) = \log 2.$$

Exercise 4.2.29. Use the Taylor series of $\arctan x$ to show that

$$1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \cdots = \frac{\pi}{4}.$$

Then show

$$1 + \frac{1}{2} - \frac{1}{3} - \frac{1}{4} + \frac{1}{5} + \frac{1}{6} - \frac{1}{7} - \frac{1}{8} + \cdots = \frac{\pi}{4} + \frac{\log 2}{2}.$$

Exercise 4.2.30. Discuss the convergence of the Taylor series of $\arcsin x$ at the radius of convergence.

Exercise 4.2.31. By making use of the Taylor series of $(1 + x)^\alpha$, prove that

$$\sum_{n=0}^{\infty} \frac{\alpha(\alpha - 1) \cdots (\alpha - n + 1)}{n!} = 2^\alpha, \quad \alpha > -1,$$

and

$$\sum_{n=0}^{\infty} \frac{\alpha(\alpha + 1) \cdots (\alpha + n - 1)}{n!} = 0, \quad \alpha < 0.$$

Note that the convergence can be obtained by the Leibniz test and the Raabe test (see Example 4.1.45).

Exercise 4.2.32. The series $\sum c_n$ with $c_n = a_0 b_n + a_1 b_{n-1} + \cdots + a_n b_0$ is obtained by combining terms in the diagonal arrangement of the product of the series $\sum a_n$ and $\sum b_n$. By considering the power series $\sum a_n x^n$, $\sum b_n x^n$, $\sum c_n x^n$ at $x = 1$, prove that if $\sum a_n$, $\sum b_n$ and $\sum c_n$ converge, then $\sum c_n = (\sum a_n)(\sum b_n)$. Compare with Exercise 4.1.39.

4.2.4 Fourier Series

A *trigonometric series* is an infinite sum of trigonometric functions

$$\frac{a_0}{2} + \sum_{n=1}^{\infty}(a_n \cos nx + b_n \sin nx). \qquad (4.2.10)$$

If the series uniformly converges to a continuous function $f(x)$ (this happens, for example, when $\sum(|a_n| + |b_n|)$ converges), then the sum $f(x)$ is necessarily a periodic function with period 2π:

$$f(x + 2\pi) = f(x).$$

The *inner product* of two periodic integrable *real* functions with period 2π is

$$\langle f, g \rangle = \int_0^{2\pi} f(x)g(x)dx. \qquad (4.2.11)$$

Because of the periodicity, the integral can be taken over any interval of length 2π. The two functions are said to be *orthogonal* if $\langle f, g \rangle = 0$. A collection of such functions is called an *orthogonal system* if they are pairwise orthogonal.

The inner product is linear with respect to f:

$$\langle f_1 + f_2, g \rangle = \langle f_1, g \rangle + \langle f_2, g \rangle, \quad \langle cf, g \rangle = c \langle f, g \rangle.$$

It is also *symmetric*:

$$\langle f, g \rangle = \langle g, f \rangle.$$

Therefore the inner product is also linear in g, and we say the inner product is *bilinear*. Moreover, the inner product has the *positivity* property:

$$\langle f, f \rangle \geq 0.$$

By Exercise 3.2.3, the functions

$$\phi_1(x) = \frac{1}{2}, \quad \phi_2(x) = \cos x, \quad \phi_3(x) = \sin x, \quad \phi_4(x) = \cos 2x, \quad \phi_5(x) = \sin 2x, \quad \ldots$$

used in the construction of the trigonometric series form an orthogonal system. Denote the corresponding coefficients by $c_1 = a_0$, $c_2 = a_1$, $c_3 = b_1$, $\ldots$. Suppose $\sum |c_n|$ converges. Then by the uniform convergence and the bilinear property of the inner product, we have

$$\langle f, \phi_n \rangle = \int_0^{2\pi} f(x) \phi_n(x) dx$$
$$= c_0 \int_0^{2\pi} \phi_0(x) \phi_n(x) dx + c_1 \int_0^{2\pi} \phi_1(x) \phi_n(x) dx + c_2 \int_0^{2\pi} \phi_2(x) \phi_n(x) dx + \cdots$$
$$= c_0 \langle \phi_0, \phi_n \rangle + c_1 \langle \phi_1, \phi_n \rangle + c_2 \langle \phi_2, \phi_n \rangle + \cdots$$
$$= c_n \langle \phi_n, \phi_n \rangle.$$

Therefore

$$c_n = \frac{\langle f, \phi_n \rangle}{\langle \phi_n, \phi_n \rangle}. \tag{4.2.12}$$

In our case (again by the computation in Exercise 3.2.3), we get

$$a_n = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos nx\, dx, \quad b_n = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin nx\, dx. \tag{4.2.13}$$

These are the *Fourier coefficients*.

Conversely, for any periodic integrable function with period 2π, we may compute the Fourier coefficients as above and construct the *Fourier series*

$$f(x) \sim \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx).$$

The notation $\sim$ only means that the function and the series are related. The two will be equal only under certain conditions.

Example 4.2.22. Let $0 \leq a \leq 2\pi$. Then

$$f(x) = \begin{cases} 0 & \text{if } 2n\pi < x < a + 2n\pi \\ 1 & \text{if } a + 2n\pi < x < 2(n+1)\pi \end{cases}, \quad n \in \mathbb{Z}$$

240   CHAPTER 4. SERIES

is the periodic function with period 2π that uniquely extends the function

$$f(x) = \begin{cases} 0 & \text{if } 0 < x < a \\ 1 & \text{if } a < x < 2\pi \end{cases}$$

on $(0, 2\pi)$. Note that since the Fourier coefficients are defined as integrals, the values of the function at 0 and a are not important. The Fourier series is

$$1 - \frac{a}{2\pi} - \sum \frac{1}{n\pi}(\sin na \cos nx + (1 - \cos na)\sin nx).$$

Example 4.2.23. Let $f(x)$ be the periodic function with period 2π given by $f(x) = x^2$ on $(0, 2\pi)$. Then

$$a_0 = \frac{1}{\pi}\int_0^{2\pi} x^2 dx = \frac{8\pi^2}{3},$$

$$a_n = \frac{1}{\pi}\int_0^{2\pi} x^2\cos nx dx = -\frac{2}{n\pi}\int_0^{2\pi} x\sin nx dx = \frac{2}{n^2\pi}\left(2\pi - \int_0^{2\pi}\cos nx dx\right) = \frac{4}{n^2},$$

$$b_n = \frac{1}{\pi}\int_0^{2\pi} x^2\sin nx dx = -\frac{1}{n\pi}\left(4\pi^2 - 2\int_0^{2\pi} x\cos nx dx\right)$$

$$= -\frac{4\pi}{n} - \frac{2}{n^2\pi}\int_0^{2\pi}\sin nx dx = -\frac{4\pi}{n}.$$

Therefore the Fourier series is

$$\frac{4\pi^2}{3} + \sum\left(\frac{4}{n^2}\cos nx - \frac{4\pi}{n}\sin nx\right).$$

Exercise 4.2.33. Compute the Fourier series for the periodic functions with period 2π and satisfy the given formulae.

1. $f(x) = x$ on $(0, 2\pi)$.

2. $f(x) = a^x$ on $(0, 2\pi)$.

3. $f(x) = a^x$ on $(-\pi, \pi)$.

4. $f(x) = |\sin x|$.

5. $f(x) = \begin{cases} 0 & \text{if } 0 < x < a \\ x - a & \text{if } a < x < 2\pi \end{cases}$.

6. $f(x) = \begin{cases} x^2 & \text{if } 0 < x < \pi \\ -x^2 & \text{if } \pi < x < 2\pi \end{cases}$.

Exercise 4.2.34. Suppose $l > 0$ and $f(x)$ is a periodic function with period l: $f(x + l) = f(x)$. Then $f\left(\frac{lx}{2\pi}\right)$ is a periodic function with period 2π. Use the observation to derive the formula for the coefficients of the Fourier series

$$\frac{a_0}{2} + \sum_{n=1}^{\infty}\left(a_n\cos\frac{2n\pi x}{l} + b_n\sin\frac{2n\pi x}{l}\right).$$

Exercise 4.2.35. What can you say about the Fourier series of odd periodic functions? What about even periodic functions?

Exercise 4.2.36. How can you extend a function $f(x)$ on $\left(0, \frac{\pi}{2}\right)$ so that its Fourier series is of the form $\sum b_{2n-1}\sin(2n-1)x$?

Exercise 4.2.37. Suppose k is a integer. How are the Fourier series of $f(x)$ and $f(kx + a)$ related?

Exercise 4.2.38. Suppose $f(x)$ is a periodic differentiable function with period 2π, such that $f'(x)$ is integrable. How are the Fourier series of $f(x)$ and $f'(x)$ related?

Exercise 4.2.39. Show that $\{\cos nx\}$ is an orthogonal system on $(0, \pi)$ with respect to the inner product $\langle f, g \rangle = \int_0^\pi f(x)g(x)dx$. Show that $\{\sin nx\}$ is also an orthogonal system on $(0, \pi)$. However, the combined system is not an orthogonal system on $(0, \pi)$.

Similar to the Taylor series, we are interested in whether the Fourier series of $f(x)$ converges to $f(x)$. We start the discussion with the following result.

Proposition 4.2.14 (Bessel Inequality). *For any periodic integrable function* $f(x)$ *of period* 2π, *the Fourier coefficients satisfy*

$$\frac{a_0^2}{2} + \sum(a_n^2 + b_n^2) \le \frac{1}{\pi} \int_0^{2\pi} f(x)^2 dx. \qquad (4.2.14)$$

Proof. Let $s_N = \sum_{n=1}^N c_n \phi_n$. Then

$$\langle f, s_N \rangle = \sum_{n=1}^N c_n \langle f, \phi_n \rangle \qquad \text{(inner product is bilinear)}$$

$$= \sum_{n=1}^N c_n^2 \langle \phi_n, \phi_n \rangle \qquad \text{(definition of } c_n)$$

$$= \sum_{m=1}^N \sum_{n=1}^N c_m c_n \langle \phi_m, \phi_n \rangle \qquad (\phi_n \text{ is orthogonal})$$

$$= \left\langle \sum_{m=1}^N \phi_m, \sum_{n=1}^N \phi_n \right\rangle = \langle s_N, s_N \rangle. \qquad \text{(inner product is bilinear)}$$

Thus by the bilinearity and the symmetric property of the inner product, we have

$$0 \le \langle f - s_N, f - s_N \rangle = \langle f, f \rangle - \langle f, s_N \rangle - \langle s_N, f \rangle + \langle s_N, s_N \rangle$$

$$= \langle f, f \rangle - \langle f, s_N \rangle = \int_0^{2\pi} f(x)^2 dx - \sum_{n=1}^N c_n^2 \langle \phi_n, \phi_n \rangle.$$

In our case, we have

$$\left\langle \frac{1}{2}, \frac{1}{2} \right\rangle = \frac{\pi}{2}, \quad \langle \cos nx, \cos nx \rangle = \pi, \quad \langle \sin nx, \sin nx \rangle = \pi,$$

and the inequality (4.2.14) follows.

$\square$

Note that the Bessel inequality applies to any orthogonal system $\{\phi_n\}$:

$$\langle f, f \rangle \ge \sum_{n=1}^\infty c_n^2 \langle \phi_n, \phi_n \rangle, \quad \text{where } c_n = \frac{\langle f, \phi_n \rangle}{\langle \phi_n, \phi_n \rangle}. \qquad (4.2.15)$$

As a matter of fact, the inequality becomes an equality for the trigonometric series (the fact is related to the *completeness* of the trigonometric system).

Proposition 4.2.15 (Riemann-Lebesgue Lemma). *Suppose $f(x)$ is an integrable function on a bounded interval $[a, b]$. Then*

$$\lim_{n \to \infty} \int_a^b f(x) \cos nx \, dx = 0, \quad \lim_{n \to \infty} \int_a^b f(x) \sin nx \, dx = 0. \qquad (4.2.16)$$

Proof. For a periodic integrable function $f(x)$ with period 2π, Proposition 4.2.14 tells us that the Fourier coefficients converge to 0, which means

$$\lim_{n \to \infty} \int_0^{2\pi} f(x) \cos nx \, dx = 0, \quad \lim_{n \to \infty} \int_0^{2\pi} f(x) \sin nx \, dx = 0.$$

If $|b - a| \leq 2\pi$, then $f(x)$ may be extended to a periodic integrable function $F(x)$ with period 2π satisfying

$$F(x) = \begin{cases} f(x) & \text{if } a < x < b \\ 0 & \text{if } b < x < a + 2\pi \end{cases}.$$

Then

$$\int_a^b f(x) \cos nx \, dx = \int_a^{a+2\pi} F(x) \cos nx \, dx = \int_0^{2\pi} F(x) \cos nx \, dx.$$

We just proved that this approaches 0 as $n \to \infty$.

In general, for bounded $[a, b]$, the integration $\int_a^b$ may be divided into finitely many integrations on intervals of length $\leq 2\pi$. Then the limit also approaches 0.

$\square$

A more general version of Riemann-Lebesgue Lemma can be found in Exercise 3.2.44.

Note that by $\cos(n + \lambda)x = \cos \lambda x \sin nx - \sin \lambda x \cos nx$ and the similar formula for $\sin(n + \lambda)x$, we also have

$$\lim_{n \to \infty} \int_a^b f(x) \cos(n + \lambda)x \, dx = 0, \quad \lim_{n \to \infty} \int_a^b f(x) \sin(n + \lambda)x \, dx = 0. \quad (4.2.17)$$

Exercise 4.2.40. Suppose $f(x)$ is a periodic function with period 2π that is monotone on $(a, a + 2\pi)$ for some a. Prove that there is a constant M, such that the Fourier coefficients satisfy $|a_n| \leq \dfrac{M}{n}$ and $|b_n| \leq \dfrac{M}{n}$. Can you draw stronger conclusion if $f'(x)$ is monotone?

Exercise 4.2.41. Suppose $f(x)$ is decreasing and $\lim_{x \to +\infty} f(x) = 0$. Prove that $\int_a^{+\infty} f(x) \cos nx \, dx$ converges and $\lim_{n \to \infty} \int_a^{+\infty} f(x) \cos nx \, dx = 0$.

Now we are ready for a simple convergence criterion for the Fourier series.

Theorem 4.2.16. *Suppose $f(x)$ is a periodic integrable function with period 2π. Suppose at a, $f(x)$ has the left limit $f(a^-)$ and the right limit $f(a^+)$, and there are M and $\delta > 0$, such that*

$$0 < t < \delta \implies |f(a+t) - f(a^+)| \le Mt, \;\; |f(a-t) - f(a^-)| \le Mt.$$

Then the Fourier series converges to $\dfrac{f(a^+) + f(a^-)}{2}$ at a.

The condition of the theorem is satisfied everywhere if the function is piecewise Lipschitz.

Proof. The partial sum of the Fourier series is

$$s_N(a) = \frac{a_0}{2} + \sum_{n=1}^{N}(a_n \cos na + b_n \sin na)$$

$$= \frac{1}{\pi}\int_0^{2\pi} f(t)\left(\frac{1}{2} + \sum_{n=1}^{N}(\cos nt \cos na + \sin nt \sin na)\right) dt$$

$$= \frac{1}{\pi}\int_0^{2\pi} f(t)\left(\frac{1}{2} + \sum_{n=1}^{N}\cos n(t-a)\right) dt$$

$$= \frac{1}{\pi}\int_0^{2\pi} f(t)D_N(t-a)dt = \frac{1}{\pi}\int_{-a}^{2\pi-a} f(t)D_N(t)dt$$

$$= \frac{1}{\pi}\int_{-\pi}^{\pi} f(a+t)D_N(t)dt,$$

where the last equality is due to the fact that the integrand is periodic. The *Dirichlet kernel function*

$$D_N(t) = \frac{1}{2} + \sum_{n=1}^{N}\cos nt = \frac{\sin\left(N+\dfrac{1}{2}\right)t}{2\sin\dfrac{t}{2}}. \qquad (4.2.18)$$

satisfies

$$\frac{1}{\pi}\int_{-\pi}^{0} D_N(t)dt = \frac{1}{\pi}\int_0^{\pi} D_N(t)dt = \frac{1}{2}.$$

Therefore

$$s_N(a) - \frac{f(a^+) + f(a^-)}{2}$$

$$= \frac{1}{\pi}\int_{-\pi}^{\pi} f(a+t)D_N(t)dt - \frac{1}{\pi}\int_{-\pi}^{0} f(a^-)D_N(t)dt - \frac{1}{\pi}\int_0^{\pi} f(a^+)D_N(t)dt$$

$$= \frac{1}{\pi}\int_{-\pi}^{0}(f(a+t) - f(a^-))D_N(t)dt + \frac{1}{\pi}\int_0^{\pi}(f(a+t) - f(a^+))D_N(t)dt.$$

Note that for any fixed a, the function $\dfrac{f(a+t) - f(a^-)}{2\sin\dfrac{t}{2}}$ is integrable on $[-\pi, -\epsilon]$ for any $\epsilon > 0$. Moreover, the assumption tells us that the function

is bounded on $[-\pi, 0]$. Thus by Exercise 3.1.8, the function is integrable on $[-\pi, 0]$. Then by (4.2.17),

$$\int_{-\pi}^0 (f(a+t) - f(a^-))D_N(t)dt = \int_{-\pi}^0 \frac{f(a+t) - f(a^-)}{2\sin\dfrac{t}{2}} \sin\left(N + \frac{1}{2}\right) t dt$$

approaches 0 as $N \to \infty$. By the similar reason, $\displaystyle\int_0^\pi (f(a+t) - f(a^+))D_N(t)dt$ also converges to 0.

$\square$

Example 4.2.24. The function in Example 4.2.23 satisfies the condition of Theorem 4.2.16. By evaluating the Fourier series at $x = 0$, we get

$$\frac{(2\pi)^2 + 0^2}{2} = \frac{4\pi^2}{3} + \sum \frac{4}{n^2}.$$

From this we get

$$\frac{1}{1^2} + \frac{1}{2^2} + \cdots + \frac{1}{n^2} + \cdots = \frac{\pi^2}{6}.$$

Exercise 4.2.42. By evaluating the Fourier series in Example 4.2.23, show that $\displaystyle\sum \frac{\sin nx}{n} = \frac{\pi - x}{2}$ on $(0, 2\pi)$. Then derive the sum of the following infinite series.

1. $1 - \dfrac{1}{3} + \dfrac{1}{5} - \dfrac{1}{7} + \dfrac{1}{9} - \dfrac{1}{11} + \cdots$.

2. $1 + \dfrac{1}{5} - \dfrac{1}{7} - \dfrac{1}{11} + \dfrac{1}{13} + \dfrac{1}{17} + \cdots$.

3. $1 + \dfrac{1}{2} - \dfrac{1}{4} - \dfrac{1}{5} + \dfrac{1}{7} + \dfrac{1}{8} - \dfrac{1}{10} - \dfrac{1}{11} + \cdots$.

4. $1 + \dfrac{1}{3} - \dfrac{1}{5} - \dfrac{1}{7} + \dfrac{1}{9} + \dfrac{1}{11} - \dfrac{1}{13} - \dfrac{1}{15} + \cdots$.

Exercise 4.2.43. By studying the Fourier series of x^4, compute $\displaystyle\sum \frac{1}{n^4}$.

Exercise 4.2.44. Prove that if $f(x)$ has continuous second order derivative, then the Fourier series uniformly converges to $f(x)$.

4.2.5   Additional Exercise

Uniform Convergence of Double Sequence

A double sequence $x_{m,n}$ is indexed by natural numbers m and n. We write $\lim_{n\to\infty} x_{m,n} = l_m$ if the limit holds for any fixed m. We say the convergence is uniform if for any $\epsilon > 0$, there is N, such that

$$n > N \implies |x_{m,n} - l_m| < \epsilon.$$

The key point here is that N depends on ϵ only, and is independent of m.

The uniform convergence of series $\sum_n x_{m,n}$ can also be defined accordingly.

Exercise 4.2.45. Determine uniform convergence.

1. $\lim_{n\to\infty} \dfrac{1}{m+n} = 0.$

4. $\lim_{n\to\infty} \dfrac{1}{n^m} = 0 \ (n > 1).$

2. $\lim_{n\to\infty} \dfrac{m}{m+n} = 0.$

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

3. $\lim_{n\to\infty} \dfrac{1}{mn} = 0.$

6. $\lim_{n\to\infty} \dfrac{m}{n} = 0.$

Exercise 4.2.46. Construct a double sequence $x_{m,n}$ satisfying $\lim_{m\to\infty} x_{m,n} = 0$ and $\lim_{n\to\infty} x_{m,n} = 1$. In particular, the double limits $\lim_{m\to\infty} \lim_{n\to\infty} x_{m,n}$ and $\lim_{n\to\infty} \lim_{m\to\infty} x_{m,n}$ are different.

Exercise 4.2.47. Are the sum, product, maximum, etc, of two uniformly convergent double sequences still uniformly convergent?

Exercise 4.2.48. Suppose $\lim_{n\to\infty} x_{m,n} = l_m$ uniformly and $\lim_{m\to\infty} x_{m,n} = k_n$ exists. Prove that both $\lim_{m\to\infty} l_m$ and $\lim_{n\to\infty} k_n$ converge and are equal.

Exercise 4.2.49. State the Cauchy criterion for the uniform convergence of the series $\sum_n x_{m,n}$. State the corresponding comparison test.

Exercise 4.2.50. Determine uniform convergence $(\alpha, \beta > 0)$.

1. $\sum_n a^{mn}.$

3. $\sum_n \dfrac{1}{n^{\alpha m}}.$

2. $\sum_n \dfrac{1}{m^\alpha n^\beta}.$

4. $\sum_n \dfrac{(m+n)!}{m!n!} a^n.$

Double Series

A *double series* $\sum_{m,n\geq 1} x_{m,n}$ may converge in several different ways.

First, the double series *converges to sum s* if for any $\epsilon > 0$, there is N, such that

$$m, n > N \implies \left| \sum_{i=1}^{m} \sum_{j=1}^{n} x_{m,n} - s \right| < \epsilon.$$

Second, the double series has *repeated sum* $\sum_m \sum_n x_{m,n}$ if $\sum_n x_{m,n}$ converges for each m and the series $\sum_m (\sum_n x_{m,n})$ again converges. Of course, there is another repeated sum $\sum_n \sum_m x_{m,n}$.

Third, a one-to-one correspondence $k \in \mathbb{N} \mapsto (m(k), n(k)) \in \mathbb{N}^2$ arranges the double series into a single series $\sum_k x_{m(k),n(k)}$, and we may consider the sum of the single series.

Fourth, for any finite subset $A \subset \mathbb{N}^2$, we may define the partial sum

$$s_A = \sum_{(m,n)\in A} x_{m,n}.$$

Then for any sequence A_k of finite subsets satisfying $A_k \subset A_{k+1}$ and $\cup A_k = \mathbb{N}^2$, we say the double series converges to s with respect to the sequence A_k if $\lim_{k\to\infty} s_{A_k} = s$. For example, we have the *spherical sum* by considering $A_k = \{(m,n) \in \mathbb{N}^2 : m^2 + n^2 \leq k^2\}$ and the *triangular sum* by considering $A_k = \{(m,n) \in \mathbb{N}^2 : m + n \leq k\}$.

Finally, the double series *absolutely converges* (see Exercise 4.2.52 for the reason for the terminology) to s if for any $\epsilon > 0$, there is N, such that $|s_A - s| < \epsilon$ for any A containing all (m,n) satisfying $m \leq N$ and $n \leq N$.

Exercise 4.2.51. State the Cauchy criterion for the convergence of the double series $\sum_{m,n\geq 1} x_{m,n}$ (in the first sense). State the corresponding comparison test.

Exercise 4.2.52. Prove that a double series $\sum_{m,n\geq 1} x_{m,n}$ absolutely converges (in the final sense) if and only if $\sum_{m,n\geq 1} |x_{m,n}|$ converges (in the first sense).

Exercise 4.2.53. Prove that a double series $\sum_{m,n\geq 1} x_{m,n}$ absolutely converges if and only if all the arrangement series $\sum_k x_{m(k),n(k)}$ converge. Moreover, the arrangement series have the same sum.

Exercise 4.2.54. Prove that a double series $\sum_{m,n\geq 1} x_{m,n}$ absolutely converges if and only if the double series converge with respect to all the sequences A_n. Moreover, the sums with respect to all the sequences A_n are the same.

Exercise 4.2.55. Prove that if a double series $\sum_{m,n\geq 1} x_{m,n}$ absolutely converges, then the two repeated sums converge and have the same value.

Exercise 4.2.56. If a double series does not converge absolutely, what can happen to various sums?

Exercise 4.2.57. Study the convergence and values of the following double series.

1. $\sum_{m,n\geq 1} a^{mn}$.

2. $\sum_{m,n\geq 1} \dfrac{1}{(m+n)^\alpha}$.

3. $\sum_{m,n\geq 1} \dfrac{(-1)^{m+n}}{(m+n)^\alpha}$.

4. $\sum_{m,n\geq 2} \dfrac{1}{n^m}$.

Uniform Convergence of Two Variable Function

Let $f(x,t)$ be a two variable function. If $\lim_{t\to a} f(x,t) = l(x)$, then we say $f(x,t)$ converges to $l(x)$ as $t \to a$. The convergence is *uniform* if for any $\epsilon > 0$, there is $\delta > 0$, such that

$$0 < |t-a| < \delta \implies |f(x,t) - l(x)| < \epsilon.$$

Similar definition can be made for the one-side limit and the limit at infinity.

Exercise 4.2.58. Determine the intervals on which the limits are uniformly convergent.

1. $\lim_{t\to\infty} \dfrac{1}{xt} = 0$.

2. $\lim_{t\to 0} x^t = 1$.

3. $\lim_{t\to 0} t(x^t - 1) = 1$.

4. $\lim_{t\to\infty} \dfrac{1}{tx+1} = 0$.

5. $\lim_{t\to 0} \sqrt{x+t} = \sqrt{x}$.

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

Exercise 4.2.59. State the Cauchy criterion for $f(x,t)$ to uniformly converges to $l(x)$ as $t \to a$.

Exercise 4.2.60. Prove that $f(x,t)$ uniformly converges to $l(x)$ as $x \to a$ if and only if $f(x,t_n)$ uniformly converges to $l(x)$ for any sequence t_n satisfying $t_n \neq a$ and $\lim_{n\to\infty} t_n \to a$.

Exercise 4.2.61. Suppose $f(x)$ is continuous on an open interval containing $[a,b]$. Prove that $f(x+t)$ uniformly converges to $f(x)$ on $[a,b]$ as $t \to 0$.

Exercise 4.2.62. Suppose $f(x)$ has continuous an open interval containing $[a, b]$. Prove that $\dfrac{f(x+t) - f(x)}{t}$ uniformly converges to $f'(x)$ on $[a, b]$ as $t \to 0$.

Exercise 4.2.63. Suppose $f(x)$ is continuous on an open interval containing $[a, b]$. Prove that the fundamental theorem of calculus

$$\lim_{t \to 0} \frac{1}{t} \int_x^{x+t} f(t)dt = f(x)$$

converges uniformly.

Exercise 4.2.64. What properties of the uniform convergence of $f_n(x)$ still hold for the uniform convergence of $f(x, t)$?

The Series $\sum_{n=1}^{\infty} \dfrac{a_n}{n^x}$

The series $\sum_{n=1}^{\infty} \dfrac{a_n}{n^x}$ has many properties similar to the power series.

Exercise 4.2.65. Use the Abel test in Proposition 4.2.6 to show that if $\sum \dfrac{a_n}{n^r}$ converges, then $\sum \dfrac{a_n}{n^x}$ uniformly converges on $[r, +\infty)$, and $\sum \dfrac{a_n}{n^r} = \lim_{x \to r^+} \sum \dfrac{a_n}{n^x}$.

Exercise 4.2.66. Prove that there is R, such that $\sum \dfrac{a_n}{n^x}$ converges on $(R, +\infty)$ and diverges on $(-\infty, R)$. Moreover, prove that $R \geq \overline{\lim}_{n \to \infty} \dfrac{\log|a_n|}{\log n}$.

Exercise 4.2.67. Prove that we can take terms wise integration and derivative of any order of the series on $(R, +\infty)$.

Exercise 4.2.68. Prove that there is R', such that the series absolutely converges on $(R', +\infty)$ and absolutely diverges on $(R', +\infty)$. Moreover, prove that $\overline{\lim}_{n \to \infty} \dfrac{\log|a_n|}{\log n} + 1 \geq R' \geq \underline{\lim}_{n \to \infty} \dfrac{\log|a_n|}{\log n} + 1$.

Exercise 4.2.69. Give an example such that the inequalities in Exercises 4.2.66 and 4.2.68 are strict.

Continuous But Not Differentiable Function

Exercise 4.2.70. Let a_n be any sequence of points. Let $|b| < 1$. Prove that the function $f(x) = \sum b^n |x - a_n|$ is continuous and is not differentiable precisely at $\{a_n\}$.

Exercise 4.2.71 (Riemann). Let $((x))$ be the periodic function with period 1, determined by $((x)) = x$ for $-\dfrac{1}{2} < x < \dfrac{1}{2}$ and $((\dfrac{1}{2})) = 0$. Let $f(x) = \sum_{n=1}^{\infty} \dfrac{((nx))}{n^2}$.

1. Prove that $f(x)$ is not continuous precisely at rational numbers with even denominators, i.e., number of the form $r = \dfrac{a}{2b}$, where a and b are odd integers.

2. Compute $f(r^+) - f(r)$ and $f(r^-) - f(r)$ at discontinuous points (you may need the conclusion of Exercise 4.2.24 for the precise value).

3. Prove that $f(x)$ is integrable, and $F(x) = \int_0^x f(x)dx$ is not differentiable precisely at rational numbers with even denominators.

Exercise 4.2.72 (Weierstrass). Suppose $0 < b < 1$ and a is an odd integer satisfying $ab > 1 + \dfrac{3\pi}{2}$. Prove that $\sum_{n=0}^{\infty} b^n \cos(a^n \pi x)$ is continuous but nowhere differentiable.

Exercise 4.2.73 (Weierstrass). Let $\{a_n\}$ be a bounded countable set of numbers. Let $h(x) = x + \dfrac{x}{2} \sin \log |x|$, $h(0) = 0$, and $0 < b < 1$. Prove that $\sum_{n=1}^{\infty} b^n h(x - a_n)$ is continuous, strictly increasing, and is not differentiable precisely at $\{a_n\}$.

Chapter 5

Multivariable Function

5.1   Limit and Continuity

The extension of the analysis from single variable to multivariable means that we are going from the real line $\mathbb{R}$ to the Euclidean space $\mathbb{R}^n$. While many concepts and results may be extended, the Euclidean space is more complicated in various aspects. The first complication is the many possible choices of the distance in a Euclidean space, which we will show to be all equivalent as far as the mathematical analysis is concerned. The second complication is that it is not sufficient to just do analysis on the rectangles, which are the obvious generalizations of the intervals on the real lines. For example, to analyze the temperature around the globe, we need to deal with a function defined on the 2-dimensional sphere inside the 3-dimensional Euclidean space. Thus we need to set up proper topological concepts that extend closed interval (closed subsets), bounded closed interval (compact subsets), and open interval (open subsets). Once the concepts are established, we find that most of the discussion about the limits and continuity of single variable functions may be extended to multivariable functions (the only exceptions are those dealing with orders among real numbers, such as monotone properties).

5.1.1   Limit in Euclidean Space

The n-dimensional (real) *Euclidean space* $\mathbb{R}^n$ is the collection of n-tuples of real numbers

$$\vec{x} = (x_1, x_2, \ldots, x_n), \ \ x_i \in \mathbb{R}.$$

Geometrically, $\mathbb{R}^1$ is the usual real line and $\mathbb{R}^2$ is a plane with origin. An element of $\mathbb{R}^n$ can be considered as a *point*, or as an arrow starting from the origin and ending at the point, called a *vector*.

The operation of *addition* and *scalar multiplication* may be applied to Euclidean vectors

$$\vec{x} + \vec{y} = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n), \ \ c\vec{x} = (cx_1, cx_2, \ldots, cx_n). \quad (5.1.1)$$

The operations satisfy the usual properties such as the commutativity and the associativity. In general, a *vector space* is a set with two operations satisfying these usual properties.

The operation

$$\vec{x} \cdot \vec{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n. \quad (5.1.2)$$

is called the *dot product* and satisfies the usual properties such as the bilinearity and the positivity. In general, an *inner product* on a vector space is an operation with numerical value satisfying these usual properties.

The dot product, or the inner product in general, induces the *length* (or the *Euclidean norm*)

$$\|\vec{x}\|_2 = \sqrt{\vec{x} \cdot \vec{x}} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}. \quad (5.1.3)$$

It also induces the *angle* θ between two nonzero vectors by the formula

$$\cos\theta = \frac{\vec{x}\cdot\vec{y}}{\|\vec{x}\|_2\|\vec{y}\|_2}. \tag{5.1.4}$$

The definition of the angle is justified by the *Schwarz inequality*

$$|\vec{x}\cdot\vec{y}| \le \|\vec{x}\|_2\|\vec{y}\|_2. \tag{5.1.5}$$

By the angle formula, two vectors are *orthogonal* and denoted $\vec{x}\perp\vec{y}$, if $\vec{x}\cdot\vec{y}=0$. Moreover, the *orthogonal projection* of a vector $\vec{x}$ on another vector $\vec{y}$ is

$$\mathrm{proj}_{\vec{y}}\vec{x} = \|\vec{x}\|_2\cos\theta\frac{\vec{y}}{\|\vec{y}\|_2} = \frac{\vec{x}\cdot\vec{y}}{\vec{y}\cdot\vec{y}}\vec{y}. \tag{5.1.6}$$

Finally, the area of the parallelogram formed by two vectors is

$$A(\vec{x},\vec{y}) = \|\vec{x}\|_2\|\vec{y}\|_2|\sin\theta| = \sqrt{(\vec{x}\cdot\vec{x})(\vec{y}\cdot\vec{y})-(\vec{x}\cdot\vec{y})^2}. \tag{5.1.7}$$

A *norm* on a vector space is a function $\|\vec{x}\|$ satisfying

1. Positivity: $\|\vec{x}\|\ge 0$, and $\|\vec{x}\|=0$ if and only if $\vec{x}=\vec{0}=(0,0,\dots,0)$.

2. Scalar Property: $\|c\vec{x}\|=|c|\|\vec{x}\|$.

3. Triangle Inequality: $\|\vec{x}+\vec{y}\|\le\|\vec{x}\|+\|\vec{y}\|$.

The norm induces the distance $\|\vec{x}-\vec{y}\|$ between two vectors. The most often used norms are the Euclidean norm $\|\vec{x}\|_2$ and

$$\|\vec{x}\|_1 = |x_1|+|x_2|+\dots+|x_n|, \quad \|\vec{x}\|_\infty = \max\{|x_1|,|x_2|,\dots,|x_n|\}.$$

In general, for any $p\ge 1$, the L_p-*norm* is defined in Exercise 5.1.3.

$$B_{L_2}(\vec{x},\epsilon) \qquad B_{L_\infty}(\vec{x},\epsilon) \qquad B_{L_1}(\vec{x},\epsilon)$$
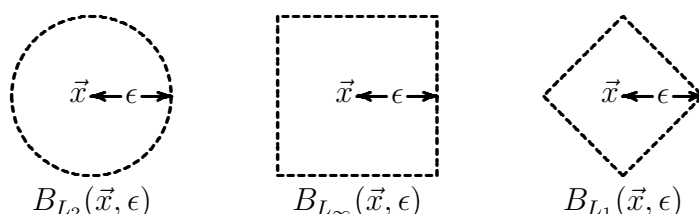
Figure 5.1: balls with respect to different norms

Given any norm, we have the *(open) ball* and the *closed ball*

$$B(\vec{x},\epsilon) = \{\vec{y}\colon \|\vec{y}-\vec{x}\|<\epsilon\},$$
$$\bar{B}(\vec{x},\epsilon) = \{\vec{y}\colon \|\vec{y}-\vec{x}\|\le\epsilon\}$$

of radius ϵ and centered at $\vec{x}$. Moreover, for the Euclidean norm, we have the (closed) Euclidean ball and the sphere of radius R

$$B_R^n = \{\vec{x}\colon \|\vec{x}\|_2\le R\} = \{(x_1,x_2,\dots,x_n)\colon x_1^2+x_2^2+\dots+x_n^2\le R^2\},$$
$$S_R^{n-1} = \{\vec{x}\colon \|\vec{x}\|_2 = R\} = \{(x_1,x_2,\dots,x_n)\colon x_1^2+x_2^2+\dots+x_n^2 = R^2\}.$$

For the radius $R=1$, we have the *unit ball* $B^n = B_1^n$ and the *unit sphere* $S^{n-1} = S_1^{n-1}$.

Exercise 5.1.1. Find all the norms on $\mathbb{R}$.

Exercise 5.1.2. Directly verify the Schwarz inequality for the dot product

$$|x_1y_1 + x_2y_2 + \cdots + x_ny_n| \leq \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}\sqrt{y_1^2 + y_2^2 + \cdots + y_n^2},$$

and find the condition for the equality to hold. Moreover, derive the triangle inequality for the norm $\|\vec{x}\|_2$ from the Schwarz inequality.

Note that the Hölder inequality in Exercise 2.2.41 generalizes the Schwarz inequality.

Exercise 5.1.3. Prove that for any $p \geq 1$, the L_p-norm

$$\|\vec{x}\|_p = \sqrt[p]{|x_1|^p + |x_2|^p + \cdots + |x_n|^p} \tag{5.1.8}$$

satisfies the three conditions for the norm. Then prove that the L_p-norms satisfy

$$\|\vec{x}\|_\infty \leq \|\vec{x}\|_p \leq \sqrt[p]{n}\|\vec{x}\|_\infty.$$

Exercise 5.1.4. Prove that for any positive numbers a_1, a_2, ..., a_n, and $p \geq 1$,

$$\|\vec{x}\| = \sqrt[p]{a_1|x_1|^p + a_2|x_2|^p + \cdots + a_n|x_n|^p}$$

is a norm.

Exercise 5.1.5. Suppose $\|\vec{x}\|$ and $\|\|\vec{y}\|\|$ are norms on $\mathbb{R}^m$ and $\mathbb{R}^n$. Prove that $\|(\vec{x}, \vec{y})\| = \max\{\|\vec{x}\|, \|\|\vec{y}\|\|\}$ is a norm on $\mathbb{R}^m \times \mathbb{R}^n = \mathbb{R}^{m+n}$. Prove that if $m = n$, then $\|\vec{x}\| + \|\|\vec{x}\|\|$ is a norm on $\mathbb{R}^n$.

Exercise 5.1.6. Prove that for any norm and any vector $\vec{x}$, there is a number $r \geq 0$ and a vector $\vec{u}$, such that $\vec{x} = r\vec{u}$ and $\|\vec{u}\| = 1$. The expression $\vec{x} = r\vec{u}$ is the *polar decomposition* that describes a vector as characterized by the length r and the direction $\vec{u}$.

Exercise 5.1.7. Prove that any norm satisfies $\|\|\vec{x}\| - \|\vec{y}\|\| \leq \|\vec{x} - \vec{y}\|$.

Exercise 5.1.8. Prove that if $\vec{y} \in B(\vec{x}, \epsilon)$, then $B(\vec{y}, \delta) \subset B(\vec{x}, \epsilon)$ for some radius $\delta > 0$. In fact, we can take $\delta = \epsilon - \|\vec{y} - \vec{x}\|$.

Comparing the single variable (represented by $\mathbb{R}$) with the multivariable (represented by $\mathbb{R}^n$), we find that the addition is generalized in a unique way, and the multiplication is generalized to the scalar multiplication and the dot product. Moreover, the absolute value?and therefore the distance) is generalized to many possible choices of norms.

A sequence of vectors $\{\vec{x}_n\}$ converges to a vector $\vec{l}$ *with respect to a norm* $\|\vec{x}\|$, denoted $\lim_{n\to\infty} \vec{x}_n = \vec{l}$, if for any $\epsilon > 0$, there is N, such that

$$n > N \implies \|\vec{x}_n - \vec{l}\| < \epsilon. \tag{5.1.9}$$

The right side is the same as $\vec{x}_n \in B(\vec{l}, \epsilon)$. The definition appears to depend on the choice of norm. For example, $\lim_{n\to\infty}(x_n, y_n) = (l, k)$ with respect to the Euclidean norm means

$$\lim_{n\to\infty} \sqrt{|x_n - l|^2 + |y_n - k|^2} = 0,$$

and the same limit with respect to the L_∞-norm means each coordinate converges

$$\lim_{n\to\infty} x_n = l, \quad \lim_{n\to\infty} y_n = k.$$

On the other hand, if two norms $\|\vec{x}\|$ and $\|\|\vec{x}\|\|$ satisfy

$$c_1\|\vec{x}\| \leq \|\|\vec{x}\|\| \leq c_2\|\vec{x}\| \tag{5.1.10}$$

for some $c_1, c_2 > 0$, then it is easy to see that the convergence with respect to $\|\vec{x}\|$ is equivalent to the convergence with respect to $\|\|\vec{x}\|\|$. Because of the remark, two norms related as in (5.1.10) are said to be *equivalent*. For example, since all L_p-norms are equivalent by Exercise 5.1.3, the L_p-convergence in $\mathbb{R}^n$ means exactly that each coordinate converges.

In Theorem 5.1.13, we will prove that all norms on $\mathbb{R}^n$ (and on finite dimensional vector spaces in general) are equivalent, so that the convergence is independent of the choice of the norm. For the moment, however, all the subsequent discussions can be rigorously verified with regard to the L_p-norm. After Theorem 5.1.13 is established, the discussion becomes valid for any norm.

Because the L_∞-convergence is the same as the convergence of each coordinate, by applying the Cauchy criterion to each coordinate, we get the Cauchy criterion for the convergence of vectors: For any $\epsilon > 0$, there is N, such that $m, n > N$ implies $\|\vec{x}_m - \vec{x}_n\| < \epsilon$. This extends Theorems 1.2.2 and 1.2.12.

By considering individual coordinates, Proposition 1.2.1, Theorem 1.2.8 and Proposition 1.2.9 can be extended to vectors. Because of the lack of order among vectors, Proposition 1.2.7 cannot be extended directly to vectors, but can still be applied to individual coordinates. For the same reason, the concept of upper and lower limits cannot be extended.

Exercise 5.1.9. Prove that if the first norm is equivalent to the second norm, and the second norm is equivalent to the third norm, then the first norm is equivalent to the third norm.

Exercise 5.1.10. Prove that if $\lim_{n\to\infty} \vec{x}_n = \vec{l}$, then $\lim_{n\to\infty} \|\vec{x}_n\| = \|\vec{l}\|$. Prove the converse is true if $\vec{l} = \vec{0}$. In particular, convergent sequences of vectors are bounded.

Exercise 5.1.11. Prove that if $\lim_{n\to\infty} \vec{x}_n = \vec{l}$ and $\lim_{n\to\infty} \vec{y}_n = \vec{k}$, then $\lim_{n\to\infty} (\vec{x}_n + \vec{y}_n) = \vec{l} + \vec{k}$ and $\lim_{n\to\infty} c\vec{x}_n = c\vec{l}$.

Exercise 5.1.12. Prove that if $\lim_{n\to\infty} c_n = c$ and $\lim_{n\to\infty} \vec{x}_n = \vec{x}$, then $\lim_{n\to\infty} c_n\vec{x}_n = c\vec{x}$.

Exercise 5.1.13. Prove that if $\lim_{n\to\infty} \vec{x}_n = \vec{x}$ and $\lim_{n\to\infty} \vec{y}_n = \vec{y}$ with respect to the Euclidean norm $\|\vec{x}\|_2$, then $\lim_{n\to\infty} \vec{x}_n \cdot \vec{y}_n = \vec{x} \cdot \vec{y}$. Of course the Euclidean norm can be replaced by any norm after Theorem 5.1.13 is established.

5.1.2 Topology in Euclidean Space

The theory of one variable functions are often discussed over intervals. For multivariable functions, it is no longer sufficient to only consider rectangles.

We need some basic concepts and results for the most useful type of subsets of the Euclidean space.

The subsequent definitions are made with respect to any norm. The results are stated for any norm. The discussion and proofs are mostly based on the L_∞-norm. After Theorem 5.1.13 is established, the discussion and results are valid for any norm.

Many important results on one-variable functions are stated on a bounded and closed interval. Since the key property used for proving the results is that any sequence in the interval has a subsequence converging to a number in the interval, we introduce the following concept.

Definition 5.1.1. A subset of a Euclidean space is *compact* if any sequence in the subset has a convergent subsequence with the limit still in the subset.

A subset K is *bounded* if there is a constant M such that $\|\vec{x}\| < M$ for any $\vec{x} \in K$. By applying Theorem 1.2.8 to each coordinate (for L_∞-norm, with the help of Exercise 1.2.24), we find that any sequence in a bounded subset K has a convergent subsequence. For K to be compact (with respect to the L_∞-norm for the moment), it remains to make sure the limit is still inside K. This leads to the following concept.

Definition 5.1.2. A subset of a Euclidean space is *closed* if the limit of any convergent sequence in the subset is still in the subset.

In other words, a subset is closed if it contains all its limits.

Proposition 5.1.3. *A subset of the Euclidean space is compact if and only if it is bounded and closed.*

For the special case of the L_∞-norm, we have argued that bounded and closed subsets are compact. The proof of the converse is given below, for any norm.

Proof. Suppose K is not bounded. Then there is a sequence $\{\vec{x}_n\}$ in K satisfying $\lim_{n\to\infty}\|\vec{x}_n\| = \infty$. This implies that $\lim_{n\to\infty}\|\vec{x}_{n_k}\| = \infty$ for any subsequence $\{\vec{x}_{n_k}\}$, so that any subsequence is not bounded and must diverge (see Exercise 5.1.10). Thus K is not compact.

Suppose K is not closed. Then there is a convergent sequence $\{\vec{x}_n\}$ in K such that $\lim_{n\to\infty}\vec{x}_n = \vec{y} \notin K$. In particular, any subsequence of $\{\vec{x}_n\}$ converges to $\vec{y} \notin K$. Thus no subsequence has limit in K, and K is not compact. $\qquad\square$

Exercise 5.1.14. Prove that if two norms are equivalent, then a subset is bounded, compact, or closed with respect to one norm if and only if it has the same property with respect to the other norm.

Exercise 5.1.15. Prove that the subsets $\{\vec{x}\colon \|\vec{x}\|_\infty = 1\}$ and $\{\vec{x}\colon \|\vec{x}\|_\infty \le 1\}$ are compact with respect to the L_∞-norm.

Exercise 5.1.16. Suppose K is a compact subset. Prove that there are $\vec{a}, \vec{b} \in K$, such that $\|\vec{a} - \vec{b}\| = \max_{\vec{x},\vec{y}\in K} \|\vec{x} - \vec{y}\|$.

Exercise 5.1.17. Prove that the closed ball $\bar{B}(\vec{x}, \epsilon)$ is a closed subset (so the name is appropriate).

Exercise 5.1.18. Prove that the intersection of closed subsets is closed, the union of finitely many closed subsets is also closed, and the product of closed subsets is closed (use the product norm in Exercise 5.1.5). What about compact subsets?

The definition of closed subsets suggests that, to make any subset closed, we should add all the limits of convergent subsequences in the subset.

Definition 5.1.4. The *closure* $\bar{A}$ of a subset A consists of the limits of all convergent sequences in A.

Clearly, a subset A is closed if and only if $\bar{A} \subset A$. The closure is also characterized by the following result.

Proposition 5.1.5. *The closure of A is closed. It consists of points $\vec{y}$ satisfying the property that for any $\epsilon > 0$, we have $\|\vec{y} - \vec{x}\| < \epsilon$ for some $\vec{x} \in A$. Moreover, it is the smallest closed subset containing A.*

The property in the proposition means $B(\vec{y}, \epsilon) \cap A \neq \emptyset$ for any $\epsilon > 0$. The geometric meaning is illustrated in Figure 5.2.
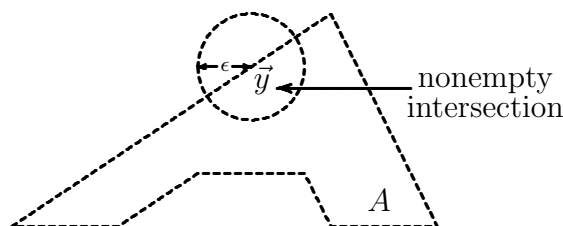
Figure 5.2: characterization of points in the closure

If a subset A is bounded, then it is contained in the closed cube $[-M, M]^n$ for some big M. The proposition implies that the closure $\bar{A}$ is also contained in $[-M, M]^n$ and is therefore bounded. Then by Proposition 5.1.3, the closure is compact.

Proof. Any $\vec{y} \in \bar{A}$ is the limit of a sequence $\{\vec{x}_n\}$ in A. Thus for any $\epsilon > 0$, we have $\|\vec{y} - \vec{x}_n\| < \epsilon$ for big n, where we also note that $\vec{x}_n \in A$.

Conversely, suppose $\vec{y}$ has the property described in the proposition. Then for $\epsilon = \dfrac{1}{n}$, we have $\vec{x}_n \in A$ satisfying $\|\vec{y} - \vec{x}_n\| < \dfrac{1}{n}$. This implies $\vec{y}$ is the limit of the sequence $\{\vec{x}_n\}$ in A.

Next we prove the closure is closed. By the definition of closure, this is the same as showing $\bar{\bar{A}} \subset \bar{A}$. By the characterization just proved, if $\vec{z} \in \bar{\bar{A}}$, then for any $\epsilon > 0$, we have $\|\vec{z} - \vec{y}\| < \epsilon$ for some $\vec{y} \in \bar{A}$. Using the characterization of the closure again, we have $\|\vec{y} - \vec{x}\| < \epsilon$ for some $\vec{x} \in A$. Therefore we get

$$\|\vec{z} - \vec{x}\| \leq \|\vec{z} - \vec{y}\| + \|\vec{y} - \vec{x}\| < 2\epsilon.$$

By the characterization of the closure, this shows that $\vec{z} \in \bar{A}$ and completes the proof that $\bar{A}$ is closed.

Finally, we prove that the closure is the smallest closed subset containing A. Since any $\vec{x} \in A$ is the limit of the constant sequence $\{\vec{x}\}$ in A, we have $A \subset \bar{A}$. Therefore the closure is a closed subset containing A. Moreover, suppose C is a closed subset containing A. Then any $\vec{y} \in \bar{A}$ is the limit of a convergent sequence in A. Since $A \subset C$, the sequence also lies in C. Since C is closed, it contains the limit $\vec{z}$ of the convergence sequence. This proves that $\bar{A} \subset C$. Therefore the closure is the smallest closed subset containing A.

<div align="right">□</div>

Definition 5.1.6. The *boundary* ∂A of a subset A consists of points that are simultaneously limits of A and $\mathbb{R}^n - A$.

The definition means $\partial A = \bar{A} \cap \overline{\mathbb{R}^n - A}$. Equivalently, $\vec{y} \in \partial A$ means $B(\vec{y}, \epsilon) \cap A \neq \emptyset$ and $B(\vec{y}, \epsilon) \cap (\mathbb{R}^n - A) \neq \emptyset$ for any $\epsilon > 0$.

Exercise 5.1.19. Prove that for any norm, the closed ball $\bar{B}(\vec{x}, \epsilon)$ is the closure of $B(\vec{x}, \epsilon)$, and the sphere $\bar{B}(\vec{x}, \epsilon) - B(\vec{x}, \epsilon)$ is the boundary of $B(\vec{x}, \epsilon)$.

Exercise 5.1.20. Prove that A is closed if and only if $\bar{A} = A$.

Exercise 5.1.21. Prove that if two norms are equivalent, then the closure and the boundary of a subset with respect to one norm is the same as the the closure and the boundary with respect to the other norm.

Exercise 5.1.22. Prove properties of the closure and the boundary (the norm on the product of Euclidean space is given in Exercise 5.1.5).

1. $A \subset B \implies \bar{A} \subset \bar{B}$.

2. $\overline{A \cup B} = \bar{A} \cup \bar{B}$.

3. $\overline{A \cap B} \subset \bar{A} \cap \bar{B}$.

4. $\overline{A \times B} = \bar{A} \times \bar{B}$.

5. $\bar{A} = \partial A \cup A$.

6. $\partial A = \partial(\mathbb{R}^n - A)$.

7. $\partial(A \cup B) \subset \partial A \cup \partial B$.

8. $\partial(A \cap B) \subset \partial A \cup \partial B$.

9. $\partial(A - B) \subset \partial A \cup \partial B$.

10. $\partial(A \times B) = (\partial A \times \bar{B}) \cup (\bar{A} \times \partial B)$.

Exercise 5.1.23. Suppose $\vec{x} \in A$ and $\vec{y} \notin A$. Suppose $\phi(t) = (1 - t)\vec{x} + t\vec{y}$ and $\tau = \sup\{t \colon \phi(t) \in A\}$. Prove that $\phi(\tau) \in \partial A$.

In the analysis of one-variable functions, such as differentiability, we often require functions to be defined at all the points near a given point. For multivariable functions, the requirement becomes the following concept.

Definition 5.1.7. A subset of a Euclidean space is *open* if for any point in the subset, all the points near the point are also in the subset.

Thus a subset U is open if

$$\vec{x} \in U \implies B(\vec{x}, \epsilon) \subset U \text{ for some } \epsilon > 0. \qquad (5.1.11)$$

The geometrical meaning is illustrated in Figure 5.3.

Proposition 5.1.8. *A subset U of $\mathbb{R}^n$ is open if and only if the complement $\mathbb{R}^n - U$ is closed.*
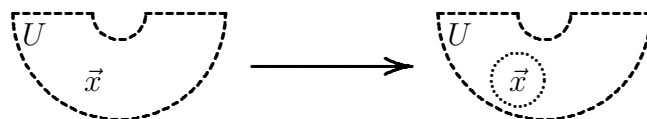
Figure 5.3: definition of open subset

Proof. The subset U is open if and only if (5.1.11) holds. Since the right side of (5.1.11) is the same as $B(\vec{x}, \epsilon) \cap (\mathbb{R}^n - U) = \emptyset$ for some $\epsilon > 0$, by Proposition 5.1.5, we see that U is open if and only if

$$\vec{x} \in U \implies \vec{x} \notin \overline{\mathbb{R}^n - U}.$$

This is logically the same as

$$\vec{x} \in \overline{\mathbb{R}^n - U} \implies \vec{x} \notin U.$$

Since the right side is the same as $\vec{x} \in \mathbb{R}^n - U$, the implication means exactly $\overline{\mathbb{R}^n - U} \subset \mathbb{R}^n - U$, which by the definition means the complement $\mathbb{R}^n - U$ is closed.

$\square$

The following result says that if an open subset U contains a compact subset K, then U contains an ϵ-neighborhood of K.

Proposition 5.1.9. *Suppose a compact subset K is contained in an open subset U. Then there is $\epsilon > 0$, such that $\|\vec{y} - \vec{x}\| < \epsilon$ and $\vec{x} \in K$ implies $\vec{y} \in U$.*

Proof. Suppose the conclusion is not true. Then for any n, there are $\vec{x}_n \in K$ and $\vec{y}_n \notin U$, such that $\|\vec{y}_n - \vec{x}_n\| < \dfrac{1}{n}$. Since K is compact, there is a subsequence $\{\vec{x}_{n_k}\}$ converging to $\vec{z} \in K$. By $\|\vec{y}_n - \vec{x}_n\| < \dfrac{1}{n}$, we know the subsequence $\{\vec{y}_{n_k}\}$ also converges to $\vec{z}$. Since $\vec{y}_{n_k} \in \mathbb{R}^n - U$ and $\mathbb{R}^n - U$ is closed by Proposition 5.1.8, we get $\vec{z} \in \mathbb{R}^n - U$. This contradicts with $\vec{z} \in K \subset U$.

$\square$

Finally, the Heine-Borel Theorem in Section 1.2.6 can be extended to Euclidean space.

Theorem 5.1.10. *Suppose K is a compact subset. Suppose $\{U_i\}$ is a collection of open subsets such that $K \subset \cup U_i$. Then $K \subset U_{i_1} \cup U_{i_2} \cup \cdots \cup U_{i_k}$ for finitely open subsets in the collection.*

The extension can be proved for the L_∞-norm by adopting the original Heine-Borel Theorem with little modification. The compact subset K is contained in a bounded rectangle $I = [\alpha_1, \beta_1] \times [\alpha_2, \beta_2] \times \cdots \times [\alpha_n, \beta_n]$. By replacing each component interval $[\alpha_i, \beta_i]$ with either $\left[\alpha_i, \dfrac{\alpha_i + \beta_i}{2}\right]$ or

$\left[\dfrac{\alpha_i + \beta_i}{2}, \beta_i\right]$, the rectangle can be divided into 2^n rectangles. If K cannot be covered by finitely many open subsets from $\{U_i\}$, then for one of the 2^n rectangles, denoted I_1, the intersection $K_1 = K \cap I_1$ cannot be covered by finitely many open subsets from $\{U_i\}$. The rest of the construction and the argument can proceed as before.

Exercise 5.1.24. Prove that if two norms are equivalent, then a subset is open with respect to one norm if and only if it is open with respect to the other norm.

Exercise 5.1.25. Prove that for any norm, the ball $B(\vec{x}, \epsilon)$ is open.

Exercise 5.1.26. Prove that the intersection of finitely many open subsets is open, the union of open subsets is also open, and the product of open subsets is open (use the product norm in Exercise 5.1.5).

Exercise 5.1.27. Suppose $f(x)$ is a continuous function on $\mathbb{R}$. Prove that the subset $\{(x, y) \colon y < f(x)\}$ is open, with the subset $\{(x, y) \colon y = f(x)\}$ as the boundary. What if $f(x)$ is not continuous?

5.1.3   Multivariable Function

Multivariable functions may be defined on all kinds of subsets of the Euclidean space. Compared with the single variable case, we often need to pay more attention to the subset on which the function is defined.

A multivariable function f on a subset A may be visualized either by the graph $\{(\vec{x}, f(\vec{x})) \colon \vec{x} \in A\}$ or the levels $\{\vec{x} \in A \colon f(\vec{x}) = c\}$.
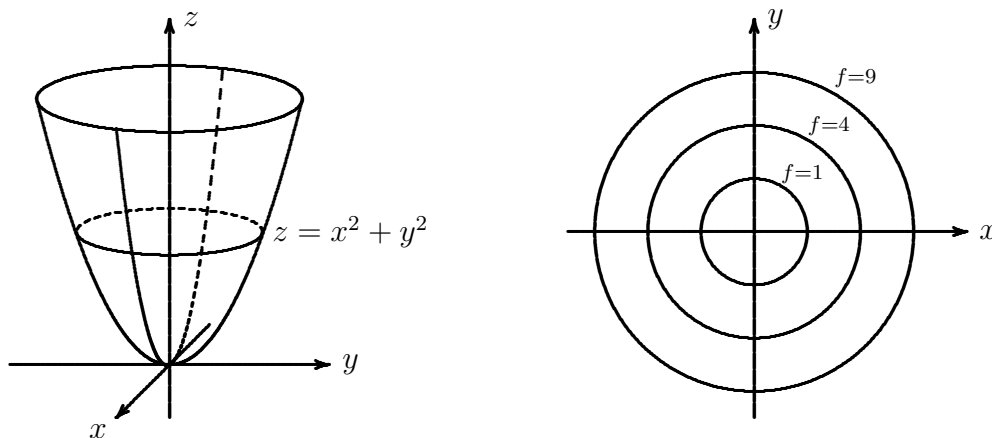
Figure 5.4: graph and level of $x^2 + y^2$

Given a norm, a multivariable function $f(\vec{x}) = f(x_1, x_2, \ldots, x_n)$ has limit l at $\vec{a} = (a_1, a_2, \ldots, a_n)$ if for any $\epsilon > 0$, there is $\delta > 0$, such that

$$0 < \|\vec{x} - \vec{a}\| < \delta \implies |f(\vec{x}) - l| < \epsilon. \qquad (5.1.12)$$

For the L_∞-norm, this means

$$|x_i - a_i| < \delta \text{ for all } i \text{ and } x_i \neq a_i \text{ for some } i \implies |f(x_1, x_2, \ldots, x_n) - l| < \epsilon.$$

After we show that all norms are equivalent in Theorem 5.1.13, we will know that the continuity is independent of the choice of the norm.

In general, if a function is defined on a subset A, then the limit is denoted as $\lim_{\vec{x}\in A,\vec{x}\to\vec{a}} f(\vec{x}) = l$. If a function is defined for all $\vec{x}$ satisfying $0 < \|\vec{x} - \vec{a}\| < \delta$ for some δ (i.e., $\vec{x}$ is in the *punctured ball* $B(\vec{a},\delta) - \vec{a}$), then we simply denote $\lim_{\vec{x}\to\vec{a}} f(\vec{x}) = l$ or $\lim_{x_1\to a_1,\ldots,x_n\to a_n} f(x_1, x_2, \ldots, x_n) = l$.

If $B \subset A$, then by restricting the definition of the limit from $\vec{x} \in A$ to $\vec{x} \in B$, we find

$$\lim_{\vec{x}\in A,\vec{x}\to\vec{a}} f(\vec{x}) = l \implies \lim_{\vec{x}\in B,\vec{x}\to\vec{a}} f(\vec{x}) = l.$$

In particular, if the restrictions of a function on different subsets give different limits, then the limit diverges.

The limits of multivariable functions have all the usual properties as before. We may also define various variations at ∞ by replacing $\|\vec{x} - \vec{a}\| < \delta$ with $\|\vec{x}\| > N$ or replacing $|f(\vec{x}) - l| < \epsilon$ by $|f(\vec{x})| > b$.

Example 5.1.1. Consider the function $f(x,y) = \dfrac{xy(x^2 - y^2)}{x^2 + y^2}$ defined for $(x,y) \neq (0,0)$. Since $|f(x,y)| \leq |xy|$, we find $\|(x,y)\|_\infty < \delta = \sqrt{\epsilon}$ implies $|f(x,y)| \leq \epsilon$. Therefore we have $\lim_{x,y\to 0} f(x,y) = 0$ in the L_∞-norm. By the equivalence between the L_p-norms, the limit also holds for other L_p-norms.

Example 5.1.2. Consider the function $f(x,y) = \dfrac{xy}{x^2 + y^2}$ defined for $(x,y) \neq (0,0)$. We have $f(x,cx) = \dfrac{c}{1 + c^2}$, so that $\lim_{y=cx,(x,y)\to(0,0)} f(x,y) = \dfrac{c}{1 + c^2}$. Thus the restriction of the function to straight lines of different slopes gives different limits. We conclude that the function diverges at $(0,0)$.

Exercise 5.1.28. Describe the graphs and the levels of functions.

1. $ax + by$.

2. $ax + by + cz$.

3. $\dfrac{(x - x_0)^2}{a^2} + \dfrac{(y - y_0)^2}{b^2}$.

4. $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c^2}$.

5. $\dfrac{(x - x_0)^2}{a^2} - \dfrac{(y - x_0)^2}{b^2}$.

6. $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} - \dfrac{z^2}{c^2}$.

7. $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} - \dfrac{z^2}{c^2}$.

8. $(x^2 + y^2)^2$.

9. $|x|^p + |y|^p$.

10. xy.

Exercise 5.1.29. What is the relation between $\lim_{\vec{x}\in A,\vec{x}\to\vec{a}} f(\vec{x})$, $\lim_{\vec{x}\in B,\vec{x}\to\vec{a}} f(\vec{x})$, $\lim_{\vec{x}\in A\cup B,\vec{x}\to\vec{a}} f(\vec{x})$, and $\lim_{\vec{x}\in A\cap B,\vec{x}\to\vec{a}} f(\vec{x})$.

Exercise 5.1.30. Prove that $\lim_{\vec{x}\to\vec{a}} f(\vec{x}) = l$ if and only if $\lim_{n\to\infty} f(\vec{x}_n) = l$ for any sequence $\{\vec{x}_n\}$ satisfying $\vec{x}_n \neq \vec{a}$ and $\lim_{n\to\infty} \vec{x}_n = \vec{a}$.

Exercise 5.1.31. For a subset $A \subset \mathbb{R}^n$, the *characteristic function* is

$$\chi_A(\vec{x}) = \begin{cases} 1 & \text{if } \vec{x} \in A \\ 0 & \text{if } \vec{x} \notin A \end{cases}$$

Find the condition for $\lim_{\vec{x}\to\vec{a}} \chi_A(\vec{x})$ to exist.

Exercise 5.1.32. Find the condition for $\lim_{x,y\to 0^+} \dfrac{x^p y^q}{(x^m + y^n)^k} = 0$, where all the parameters are positive. Extend the discussion to more variables.

Exercise 5.1.33. Compute the convergent limits. All parameters are positive.

1. $\lim_{x \to 1, y \to 1} \dfrac{1}{x-y}$.

2. $\lim_{x \to 0, y \to 0, z \to 0} \dfrac{xyz}{x^2+y^2+z^2}$.

3. $\lim_{x \to 0, y \to 0}(x-y)\sin \dfrac{1}{x^2+y^2}$.

4. $\lim_{x \to \infty, y \to \infty}(x-y)\sin \dfrac{1}{x^2+y^2}$.

5. $\lim_{x \to \infty, y \to \infty} \dfrac{(x^2+y^2)^p}{(x^4+y^4)^q}$.

6. $\lim_{x \to 0, y \to 0, 0 < x < y^2} \dfrac{x^p y}{x^2+y^2}$.

7. $\lim_{x \to \infty, y \to \infty, ax \le y \le bx} \dfrac{1}{xy}$.

8. $\lim_{x \to 0^+, y \to 0^+}(x+y)^{xy}$.

9. $\lim_{x \to \infty, y \to 0}\left(1+\dfrac{1}{x}\right)^{\frac{x^2}{x+y}}$.

10. $\lim_{x \to +\infty, y \to +\infty}\left(1+\dfrac{1}{x}\right)^{\frac{x^2}{x+y}}$.

11. $\lim_{x \to +\infty, y \to +\infty} \dfrac{x^2+y^2}{e^{x+y}}$.

For multivariable functions, we may first take the limit in some variables and then take the limit in the other variables. In this way, we get repeated limits such as $\lim_{\vec{x} \to \vec{a}} \lim_{\vec{y} \to \vec{b}} f(\vec{x}, \vec{y})$. The following result tells us the relation between the repeated limits and the usual limit $\lim_{\vec{x} \to \vec{a}, \vec{y} \to \vec{b}} f(\vec{x}, \vec{y}) = \lim_{(\vec{x}, \vec{y}) \to (\vec{a}, \vec{b})} f(\vec{x}, \vec{y})$. The proof is given for the the case $\|(\vec{x}, \vec{y})\| = \|\vec{x}\| + \|\vec{y}\|$.

Proposition 5.1.11. *Suppose $f(\vec{x}, \vec{y})$ is defined near $(\vec{a}, \vec{b})$. Suppose*

1. *The limit $\lim_{(\vec{x}, \vec{y}) \to (\vec{a}, \vec{b})} f(\vec{x}, \vec{y})$ converges to l.*

2. *For each $\vec{x}$ near $\vec{a}$, the limit $\lim_{\vec{y} \to \vec{b}} f(\vec{x}, \vec{y})$ converges to $g(\vec{x})$.*

Then the repeated limit $\lim_{\vec{x} \to \vec{a}} \lim_{\vec{y} \to \vec{b}} f(\vec{x}, \vec{y}) = \lim_{\vec{x} \to \vec{a}} g(\vec{x})$ exists and is equal to l.

Proof. For any $\epsilon > 0$, there is $\delta > 0$, such that

$$0 < \max\{\|\vec{x}-\vec{a}\|, \|\vec{y}-\vec{b}\|\} < \delta \implies |f(\vec{x}, \vec{y}) - l| < \epsilon.$$

Then for each fixed $\vec{x}$ near $\vec{a}$, taking $\vec{y} \to \vec{b}$ in the inequality above gives us

$$0 < \|\vec{x}-\vec{a}\| < \delta \implies |g(\vec{x}) - l| = \lim_{\vec{y} \to \vec{b}} |f(\vec{x}, \vec{y}) - l| \le \epsilon.$$

This implies that $\lim_{\vec{x} \to \vec{a}} g(\vec{x})$ converges to l.

$\square$

Example 5.1.3. The function $f(x, y) = \dfrac{xy}{x^2+y^2}$ in Example 5.1.2 satisfies $\lim_{x \to 0} f(x, y) = 0$ for each y and $\lim_{y \to 0} f(x, y) = 0$ for each x. Therefore both repeated limits $\lim_{y \to 0} \lim_{x \to 0} f(x, y)$ and $\lim_{x \to 0} \lim_{y \to 0} f(x, y)$ exist and are equal to 0. However, the limit $\lim_{(x,y) \to (0,0)} f(x, y)$ diverges.

Exercise 5.1.34. Study the usual limit $\lim_{x \to 0, y \to 0}$ and the repeated limits $\lim_{x \to 0} \lim_{y \to 0}$, $\lim_{y \to 0} \lim_{x \to 0}$.

1. $\dfrac{x - y + x^2 + y^2}{x + y}$.

4. $\dfrac{x^2 y^2}{|x|^3 + |y|^3}$.

2. $x \sin \dfrac{1}{x} + y \cos \dfrac{1}{x}$.

5. $(x + y) \sin \dfrac{1}{x} \sin \dfrac{1}{y}$.

3. $\dfrac{|x|^\alpha |y|^\beta}{(x^2 + y^2)^\gamma}$, $\alpha, \beta, \gamma > 0$.

6. $\dfrac{e^x - e^y}{\sin xy}$.

Exercise 5.1.35. Establish a concept of the uniform convergence and find the condition for the commutativity of the repeated limits such as $\lim_{x \to a} \lim_{y \to b} f(x, y) = \lim_{y \to b} \lim_{x \to a} f(x, y)$, similar to Proposition 4.2.7.

Exercise 5.1.36. Construct a function $f(x, y)$, such that the limit $\lim_{x \to a, y \to b} f(x, y)$ converges, but the two repeated limits do not converge.

5.1.4 Continuous Function

A multivariable function $f(\vec{x})$ is *continuous* at $\vec{a}$ if it is defined at $\vec{a}$ and $\lim_{\vec{x} \to \vec{a}} f(\vec{x}) = f(\vec{a})$. For the L_∞-norm, this means that for any $\epsilon > 0$, there is $\delta > 0$, such that

$$|x_i - a_i| < \delta \text{ for all } i \implies |f(x_1, x_2, \ldots, x_n) - f(a_1, a_2, \ldots, a_n)| < \epsilon.$$

Like the single variable case, it is easy to see that the arithmetic combinations of continuous functions are continuous, and the exponential of continuous functions are continuous.

A function $f(\vec{x})$ on A is *uniformly continuous* if for any $\epsilon > 0$, there is $\delta > 0$, such that

$$\vec{x}, \vec{y} \in A, \ \|\vec{x} - \vec{y}\| < \delta \implies |f(\vec{x}) - f(\vec{y})| < \epsilon. \tag{5.1.13}$$

The following extension of Theorem 1.4.5 is proved for any norm.

Theorem 5.1.12. *A continuous function on a compact subset is bounded, uniformly continuous, and reaches its maximum and minimum.*

Proof. Suppose $f(\vec{x})$ is continuous on a compact subset K. If $f(\vec{x})$ is not bounded, then there is a sequence $\{\vec{x}_n\}$ in K such that $\lim_{n \to \infty} f(\vec{x}_n) = \infty$. Since K is compact, there is a subsequence $\{\vec{x}_{n_k}\}$ converging to $\vec{a} \in K$. By the continuity of $f(\vec{x})$ on K, we have $\lim_{k \to \infty} f(\vec{x}_{n_k}) = f(\vec{a})$. This contradicts with $\lim_{n \to \infty} f(\vec{x}_n) = \infty$.

The function is bounded and we may let $\beta = \sup_{\vec{x} \in K} f(\vec{x})$. There is a sequence $\{\vec{x}_n\}$ in K such that $\lim_{n \to \infty} f(\vec{x}_n) = \beta$. By the similar argument as before, there is a subsequence converging to $\vec{a} \in K$, such that $f(\vec{a}) = \lim_{k \to \infty} f(\vec{x}_{n_k}) = \beta$. Thus the maximum is reached at $\vec{a}$.

The proof of Theorem 1.4.4 can be adopted to prove the uniform continuity.

$\square$

Exercise 5.1.37. Prove that any norm is a continuous function with respect to itself. Is the continuity uniform?

Exercise 5.1.38. Prove that if $f(\vec{x})$ and $g(\vec{x})$ are continuous at $\vec{x}_0$, and $h(u, v)$ is continuous at $(f(\vec{x}_0), g(\vec{x}_0))$ with respect to the L_∞-norm on $\mathbb{R}^2$, then $h(f(\vec{x}), g(\vec{x}))$ is continuous at $\vec{x}_0$.

Exercise 5.1.39. Prove that if $f(x, y)$ is monotone and continuous in x and in y, then $f(x, y)$ is continuous with respect to the L_∞-norm. What if the function is monotone and continuous in x but is only continuous in y?

Exercise 5.1.40. Study the continuity.

1. $\begin{cases} \dfrac{\sin xy}{x} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$.

2. $\begin{cases} \dfrac{\sin xy}{|x| + |y|} & \text{if} (x, y) \neq (0, 0) \\ 0 & \text{if} (x, y) = (0, 0) \end{cases}$.

3. $\begin{cases} y & \text{if } x \text{ is rational} \\ -y & \text{if } x \text{ is irrational} \end{cases}$.

4. $\begin{cases} x(x^2 + y^2)^p & \text{if } y \neq 0 \\ 0 & \text{if } y = 0 \end{cases}$.

5. $\begin{cases} y \log(x^2 + y^2) & \text{if} (x, y) \neq (0, 0) \\ 0 & \text{if} (x, y) = (0, 0) \end{cases}$.

6. $\begin{cases} e^{\frac{x}{y}} & \text{if } y \neq 0 \\ 0 & \text{if } y = 0 \end{cases}$.

Exercise 5.1.41. Suppose $\phi(x)$ is uniformly continuous. Is $f(x, y) = \phi(x)$ uniformly continuous with respect to the L_∞-norm?

Exercise 5.1.42. Prove that with respect to the L_∞-norm, any uniformly continuous function on a bounded subset is bounded.

Exercise 5.1.43. Prove that the sum of uniformly continuous functions is uniformly continuous. What about the product?

By using Theorem 5.1.12, we can establish the following result, which allows us to freely use any norm in the discussion of vectors and multivariable functions. For example, a subset is compact with respect to any norm if and only if it is closed and bounded.

Theorem 5.1.13. *All norms on a finite dimensional vector space are equivalent.*

Proof. We prove the claim for norms on a Euclidean space. Since linear algebra tells us that any finite dimensional vectors space is *isomorphic* to a Euclidean space, the result applies to any vector space.

We compare any norm $\|\vec{x}\|$ with the L_∞-norm $\|\vec{x}\|_\infty$. Let $\vec{e}_1 = (1, 0, \dots, 0)$, $\vec{e}_2 = (0, 1, \dots, 0)$, ..., $\vec{e}_n = (0, 0, \dots, 1)$ be the *standard basis* of $\mathbb{R}^n$. By the conditions for norm, we have

$$\begin{aligned}
\|\vec{x}\| &= \|x_1 \vec{e}_1 + x_2 \vec{e}_2 + \cdots + x_n \vec{e}_n\| \\
&\leq |x_1| \|\vec{e}_1\| + |x_2| \|\vec{e}_2\| + \cdots + |x_n| \|\vec{e}_n\| \\
&\leq (\|\vec{e}_1\| + \|\vec{e}_2\| + \cdots + \|\vec{e}_n\|) \max\{|x_1|, |x_2|, \dots, |x_n|\} \\
&= (\|\vec{e}_1\| + \|\vec{e}_2\| + \cdots + \|\vec{e}_n\|) \|\vec{x}\|_\infty = c_1 \|\vec{x}\|_\infty.
\end{aligned}$$

We remark that this implies (see Exercise 5.1.7)

$$\|\vec{x} - \vec{a}\|_\infty \leq \delta \implies \big| \|\vec{x}\| - \|\vec{a}\| \big| \leq \|\vec{x} - \vec{a}\| \leq c_1 \delta.$$

Based on this, it is easy to see that the function $\|x\|$ is continuous with respect to the L_∞-norm.

It remains to prove that $\|x\| \geq c_2 \|\vec{x}\|_\infty$ for some $c_2 > 0$. The inequality implies

$$\|\vec{x}\|_\infty = 1 \implies \|\vec{x}\| \geq c_2. \tag{5.1.14}$$

Conversely, suppose the implication (5.1.14) holds. For any $\vec{x}$, we write $\vec{x} = r\vec{u}$ with $r = \|\vec{x}\|_\infty$ and $\|\vec{u}\|_\infty = 1$ (see Exercise 5.1.6). Then

$$\|\vec{x}\| = \|r\vec{u}\| = r\|\vec{u}\| \geq rc_2 = c_2 \|\vec{x}\|_\infty.$$

Therefore we conclude the inequality $\|\vec{x}\| \geq c_2 \|\vec{x}\|_\infty$ is equivalent to the implication (5.1.14).

The subset $K = \{\vec{x} \colon \|\vec{x}\|_\infty = 1\}$ is bounded and closed with respect to the L_∞-norm (see Exercise 5.1.15). By Proposition 5.1.3, which was fully proved for the L_∞-norm, K is compact with respect to the L_∞-norm. Then by Theorem 5.1.12, the continuity of the function $\|\vec{x}\|$ on K with respect to the L_∞-norm implies that $\|\vec{x}\|$ reaches its minimum on K at $\vec{a} \in K$. Then we conclude the implication (5.1.14) holds for $c_2 = \|\vec{a}\|$, with

$$\vec{a} \in K \implies \|\vec{a}\|_\infty = 1 \implies \vec{a} \neq \vec{0} \implies c_2 = \|\vec{a}\| > 0.$$

$\square$

5.1.5 Multivariable Map

Multivariable analysis is not restricted to multivariable functions. We may also consider maps from one multivariable to another multivariable. These are the maps between Euclidean spaces.

A map from $A \subset \mathbb{R}^n$ to $\mathbb{R}^m$ has m-dimensional vectors as values. Instead of using many arrows, we denote the map by

$$F(\vec{x}) = (f_1(\vec{x}), f_2(\vec{x}), \dots, f_m(\vec{x})) \colon A \subset \mathbb{R}^n \to \mathbb{R}^m.$$

Two maps into the same Euclidean space may be added. A number can be multiplied to a map into a Euclidean space. If $F \colon A \subset \mathbb{R}^n \to \mathbb{R}^m$ and $G \colon B \subset \mathbb{R}^k \to \mathbb{R}^n$ are maps such that $G(\vec{x}) \in A$ for any $\vec{x} \in B$, then we have the composition $(F \circ G)(\vec{x}) = F(G(\vec{x})) \colon B \subset \mathbb{R}^k \to \mathbb{R}^m$.

The map converges to $\vec{l} \in \mathbb{R}^m$ at $\vec{a} \in \mathbb{R}^n$, denoted $\lim_{\vec{x} \to \vec{a}} F(\vec{x}) = \vec{l}$, if for any $\epsilon > 0$, there is $\delta > 0$, such that

$$0 < \|\vec{x} - \vec{a}\| < \delta \implies \|F(\vec{x}) - \vec{l}\| < \epsilon. \tag{5.1.15}$$

By using the L_∞-norm, it is easy to see that

$$\lim_{\vec{x} \to \vec{a}} F(\vec{x}) = \vec{l} \iff \lim_{\vec{x} \to \vec{a}} f_i(\vec{x}) = l_i \text{ for all } i. \tag{5.1.16}$$

The property also holds for any norm because any norm is equivalent to the L_∞-norm.

The map is continuous at $\vec{a}$ if $\lim_{\vec{x} \to \vec{a}} F(\vec{x}) = F(\vec{a})$. By (5.1.16), a map is continuous if and only if its coordinate functions are continuous.

Many properties can be extended from functions to maps. For example, the sum and the composition of continuous maps are still continuous. The scalar product of a continuous function and a continuous map is a continuous map. The dot product of two continuous maps is a continuous function.

Some special cases of maps can be visualized in various ways. For example, a function is a map to $\mathbb{R}$ and can be visualized by its graph or its levels. On the other hand, a *(parametrized) curve* (or a *path*) in $\mathbb{R}^n$ is a continuous map

$$\phi(t) = (x_1(t), x_2(t), \ldots, x_n(t)) \colon [a, b] \to \mathbb{R}^n.$$

The continuity means that each coordinate function $x_i(t)$ is continuous. For example, the straight line passing through $\vec{a}$ and $\vec{b}$ is

$$\phi(t) = (1 - t)\vec{a} + t\vec{b} = \vec{a} + t(\vec{b} - \vec{a}), \tag{5.1.17}$$

and the unit circle on the plane is

$$\phi(t) = (\cos t, \sin t) \colon [0, 2\pi] \to \mathbb{R}^2. \tag{5.1.18}$$

We say the path connects $\phi(a)$ to $\phi(b)$. A subset $A \subset \mathbb{R}^n$ is *path connected* if any two points in A are connected by a path in A (i.e., $\phi(t) \in A$ for any $t \in [a, b]$).

Similar to curves, a map $\mathbb{R}^2 \to \mathbb{R}^n$ may be considered as a *(parametrized) surface*. For example, the sphere in $\mathbb{R}^3$ may be parametrized by

$$\sigma(\phi, \theta) = (\sin \phi \cos \theta, \sin \phi \sin \theta, \cos \phi) \colon [0, \pi] \times [0, 2\pi] \to \mathbb{R}^3, \tag{5.1.19}$$

and the torus by $(a > b > 0)$

$$\sigma(\phi, \theta) = ((a + b \cos \phi) \cos \theta, (a + b \cos \phi) \sin \theta, b \sin \phi) \colon [0, 2\pi] \times [0, 2\pi] \to \mathbb{R}^3. \tag{5.1.20}$$

A map $\mathbb{R}^n \to \mathbb{R}^n$ may be considered as a *change of variable*, or a *transform*, or a *vector field*. For example,

$$(x, y) = (r \cos \theta, r \sin \theta) \colon [0, \infty) \times \mathbb{R} \to \mathbb{R}^2$$

is the polar coordinate. Moreover,

$$R_\theta(x, y) = (x \cos \theta - y \sin \theta, x \sin \theta + y \cos \theta) \colon \mathbb{R}^2 \to \mathbb{R}^2$$

transforms the plane by rotation of angle θ, and

$$\vec{x} \to \vec{a} + \vec{x} \colon \mathbb{R}^n \to \mathbb{R}^n$$

shifts all the Euclidean space by $\vec{a}$. A vector field assigns an arrow to a point. For example, the vector field

$$F(\vec{x}) = \vec{x} \colon \mathbb{R}^n \to \mathbb{R}^n$$

is a vector field in the radial direction, while

$$F(x, y) = (y, -x) \colon \mathbb{R}^2 \to \mathbb{R}^2$$

is a counterclockwise rotating vector field, just like the water flow in the sink.
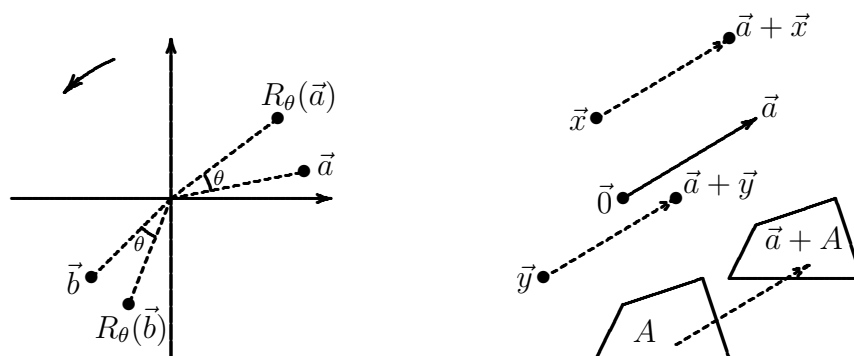
Figure 5.5: rotation and shifting
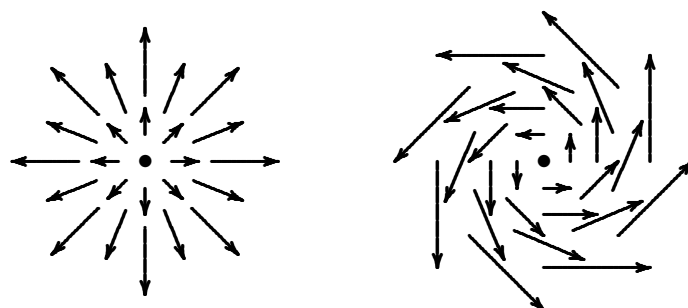
Figure 5.6: radial and rotational vector fields

Example 5.1.4. By direct argument, it is easy to see that the product function $\mu(x, y) = xy \colon \mathbb{R}^2 \to \mathbb{R}$ is continuous. Suppose $f(\vec{x})$ and $g(\vec{x})$ are continuous. Then $F(\vec{x}) = (f(\vec{x}), g(\vec{x})) \colon \mathbb{R}^n \to \mathbb{R}^2$ is also continuous because each coordinate function is continuous. The composition $(\mu \circ F)(\vec{x}) = f(\vec{x})g(\vec{x})$ of the two continuous maps is also continuous. Therefore we conclude that the product of continuous functions is continuous.

Exercise 5.1.44. Describe the maps in suitable ways.

1. $F(x) = (\cos x, \sin x, x)$.

2. $F(x, y) = (x, y, x + y)$.

3. $F(x, y, z) = (y, z, x)$.

4. $F(x, y, z) = (y, -x, z)$.

5. $F(x, y) = (x^2, y^2)$.

6. $F(x, y) = (x^2 - y^2, 2xy)$.

7. $F(\vec{x}) = 2\vec{x}$.

8. $F(\vec{x}) = \vec{a} - \vec{x}$.

The meaning of the sixth map can be seen through the polar coordinate.

Exercise 5.1.45. Prove that the composition of two continuous maps is continuous.

Exercise 5.1.46. Prove that the addition $(\vec{x}, \vec{y}) \to \vec{x} + \vec{y} \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ and the scalar multiplication $(c, \vec{x}) \to c\vec{x} \colon \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ are continuous.

The following extends the similar Propositions 1.3.8 and 1.4.3 for single variable functions. The same proof applies here.

Proposition 5.1.14. *For a map $F(\vec{x})$, $\lim_{\vec{x} \to \vec{a}} F(\vec{x}) = \vec{l}$ if and only if $\lim_{n \to \infty} F(\vec{x}_n) = \vec{l}$ for any sequence $\{\vec{x}_n\}$ satisfying $\vec{x}_n \neq \vec{a}$ and $\lim_{n \to \infty} \vec{x}_n = \vec{a}$. In particular, F is continuous at $\vec{a}$ if and only if $\lim_{n \to \infty} F(\vec{x}_n) = F(\vec{a})$ for any sequence $\{\vec{x}_n\}$ satisfying $\lim_{n \to \infty} \vec{x}_n = \vec{a}$.*

Theorem 5.1.12 on the boundedness and the uniform continuity can also be extended. The extreme is meaningless for maps.

Theorem 5.1.15. *Suppose $F(\vec{x})$ is a continuous map on a compact subset K. Then $F(\vec{x})$ is bounded and uniformly continuous: For any $\epsilon > 0$, there is $\delta > 0$, such that*

$$\vec{x}, \vec{y} \in K, \ \ \|\vec{x} - \vec{y}\| < \delta \implies \|F(\vec{x}) - F(\vec{y})\| < \epsilon. \tag{5.1.21}$$

We do not expect the intermediate value theorem to extend to multivariable functions in general because the theorem makes critical use of the interval. For the multivariable case, the interval may be replaced by the path connected condition.

Theorem 5.1.16. *Suppose $f(\vec{x})$ is a continuous function on a path connected subset A. Then for any $\vec{a}, \vec{b} \in A$ and y between $f(\vec{a})$ and $f(\vec{b})$, there is $\vec{c} \in A$, such that $f(\vec{c}) = y$.*

Proof. Since A is path connected, there is a continuous path $\phi(t) \colon [a, b] \to A$ such that $\phi(a) = \vec{a}$ and $\phi(b) = \vec{b}$. The composition $g(t) = f(\phi(t))$ is then a continuous function for $t \in [a, b]$, and y is between $g(a) = f(\vec{a})$ and $g(b) = f(\vec{b})$. By Theorem 1.4.6, there is $c \in [a, b]$, such that $g(c) = y$. The conclusion is the same as $f(\vec{c}) = y$ for $\vec{c} = \phi(c)$. $\qquad\square$

We cannot talk about multivariable monotone maps. Thus the only part of Theorem 1.4.8 that can be extended is the continuity.

Theorem 5.1.17. *Suppose $F \colon K \subset \mathbb{R}^n \to \mathbb{R}^m$ is a one-to-one and continuous map on a compact set K. Then the inverse map $F^{-1} \colon F(K) \subset \mathbb{R}^m \to \mathbb{R}^n$ is continuous.*

Proof. We claim that F^{-1} is in fact uniformly continuous. Suppose it is not uniformly continuous, then there is $\epsilon > 0$ and $\vec{\xi}_k = F(\vec{x}_k), \vec{\eta}_k = F(\vec{y}_k) \in F(K)$, such that $\|\vec{\xi}_k - \vec{\eta}_k\| \to 0$ as $k \to \infty$ and $\|\vec{x}_k - \vec{y}_k\| = \|F^{-1}(\vec{\xi}_k) - F^{-1}(\vec{\eta}_k)\| \geq \epsilon$. By the compactness of K, we can find k_p, such that both $\lim_{p \to \infty} \vec{x}_{k_p} = \vec{a}$ and $\lim_{p \to \infty} \vec{y}_{k_p} = \vec{b}$ exist. Then by the continuity of F, we have $\lim_{p \to \infty} \vec{\xi}_{k_p} = \lim_{p \to \infty} F(\vec{x}_{k_p}) = F(\vec{a})$ and $\lim_{p \to \infty} \vec{\eta}_{k_p} = \lim_{p \to \infty} F(\vec{y}_{k_p}) = F(\vec{b})$. We conclude that $\|F(\vec{a}) - F(\vec{b})\| = \lim_{p \to \infty} \|\vec{\xi}_{k_p} - \vec{\eta}_{k_p}\| = 0$, so that $F(\vec{a}) = F(\vec{b})$. Since F is one-to-one, we get $\vec{a} = \vec{b}$. This is in contradiction with $\lim_{p \to \infty} \vec{x}_{k_p} = \vec{a}$, $\lim_{p \to \infty} \vec{y}_{k_p} = \vec{b}$, and $\|\vec{x}_k - \vec{y}_k\| \geq \epsilon$. $\qquad\square$

Exercise 5.1.47. Prove the following are equivalent for a map F.

1. The map is continuous.

2. The preimage $F^{-1}(C) = \{\vec{x} \colon F(\vec{x}) \in C\} \subset \mathbb{R}^n$ of any closed subset $C \subset \mathbb{R}^m$ is closed.

3. The preimage $F^{-1}(U)$ of any open subset $U \subset \mathbb{R}^m$ is open.

4. The preimage $F^{-1}(B(\vec{b}, \epsilon))$ of any open ball $B(\vec{b}, \epsilon)$ is open.

Exercise 5.1.48. Prove the following are equivalent for a map F.

1. The map is continuous.

2. $F(\bar{A}) \subset \overline{F(A)}$ for any subset A.

3. $F^{-1}(\bar{B}) \supset \overline{F^{-1}(B)}$ for any subset B.

Exercise 5.1.49. Suppose F is a continuous map on a compact subset K. Prove that the image subset $F(K) = \{F(\vec{x}) \colon \vec{x} \in K\}$ is compact.

Exercise 5.1.50. A map $F(\vec{x})$ is Lipschitz if $\|F(\vec{x}) - F(\vec{y})\| \leq L\|\vec{x} - \vec{y}\|$ for some constant L. Prove that Lipschitz maps are uniformly continuous.

5.1.6 Additional Exercise

Exercise 5.1.51. Suppose $f(x)$ has continuous derivative. Prove that

$$\lim_{(x,y) \to (a,a)} \frac{f(x) - f(y)}{x - y} = f'(a).$$

What if the continuity condition is dropped? Is there a similar conclusion for the second order derivative?

Exercise 5.1.52. Prove that $F \colon \mathbb{R}^n \to \mathbb{R}^m$ is continuous if and only if $F \cdot \vec{a}$ is a continuous function for any $\vec{a} \in \mathbb{R}^m$.

Repeated Extreme

Exercise 5.1.53. For a function $f(\vec{x}, \vec{y})$ on $A \times B$. Prove that

$$\inf_{\vec{y} \in B} \sup_{\vec{x} \in A} f(\vec{x}, \vec{y}) \geq \sup_{\vec{x} \in A} \inf_{\vec{y} \in B} f(\vec{x}, \vec{y}) \geq \inf_{\vec{x} \in A} \inf_{\vec{y} \in B} f(\vec{x}, \vec{y}) = \inf_{\vec{x} \in A, \vec{y} \in B} f(\vec{x}, \vec{y}).$$

Exercise 5.1.54. For any $a \geq b_1 \geq c_1 \geq d$ and $a \geq b_2 \geq c_2 \geq d$, can you construct a function $f(x, y)$ on $[0, 1] \times [0, 1]$ such that

$$\sup_{\vec{x} \in A, \vec{y} \in B} f = a, \qquad \inf_{\vec{x} \in A, \vec{y} \in B} f = d,$$

$$\inf_{\vec{y} \in B} \sup_{\vec{x} \in A} f = b_1. \qquad \sup_{\vec{x} \in A} \inf_{\vec{y} \in B} f = c_1,$$

$$\inf_{\vec{x} \in A} \sup_{\vec{y} \in B} f = b_2, \qquad \sup_{\vec{y} \in B} \inf_{\vec{x} \in A} f = c_2.$$

Exercise 5.1.55. Suppose $f(x, y)$ is a function on $[0, 1] \times [0, 1]$. Prove that if $f(x, y)$ is increasing in x, then $\inf_{y \in [0,1]} \sup_{x \in [0,1]} f(x, y) = \sup_{x \in [0,1]} \inf_{y \in [0,1]} f(x, y)$. Can the interval $[0, 1]$ be changed to $(0, 1)$?

Exercise 5.1.56. Extend the discussion to three or more repeated extrema. For example, can you give a simple criterion for comparing two strings of repeated extrema?

Limit Along Any Path

Exercise 5.1.57. Prove that if $\lim_{\vec{x} \to \vec{a}} f(\vec{x})$ converges along any continuous path leading to $\vec{a}$, then the limit exists.

Exercise 5.1.58. Show that $\lim_{x\to 0, y\to 0} \dfrac{xy^2}{x^2 + y^4}$ does not exist, although the limit converges zero along any straight line leading to $(0,0)$.

Homogeneous and Multihomogeneous Function

A function $f(\vec{x})$ is *homogeneous* of degree α if $f(c\vec{x}) = c^{\alpha} f(\vec{x})$ for any $c > 0$. More generally, a function is *multihomogeneous* if

$$cf(x_1, x_2, \ldots, x_n) = f(c^{\beta_1} x_1, c^{\beta_2} x_2, \ldots, c^{\beta_n} x_n).$$

The later part of the proof of Theorem 5.1.13 makes use of the fact that any norm is a homogeneous function of degree 1 (also see Exercise 5.1.61).

Exercise 5.1.59. Prove that two homogeneous functions of the same degree are equal away from $\vec{0}$ if and only if their restrictions on the unit sphere S^{n-1} are equal.

Exercise 5.1.60. Prove that a homogeneous function is bigger than another homogeneous function of the same degree away from $\vec{0}$ if and only if the inequality holds on S^{n-1}.

Exercise 5.1.61. Suppose $f(\vec{x})$ is a continuous homogeneous function of degree α satisfying $f(\vec{x}) > 0$ for $\vec{x} \neq \vec{0}$. Prove that there is $c > 0$, such that $f(\vec{x}) \geq c\|\vec{x}\|^{\alpha}$ for any $\vec{x}$.

Exercise 5.1.62. Prove that a homogeneous function is continuous away from $\vec{0}$ if and only if its restriction on S^{n-1} is continuous. Then find the condition for a homogeneous function to be continuous at $\vec{0}$.

Continuous Map and Function on Compact Set

Exercise 5.1.63. Suppose $F(\vec{x}, \vec{y})$ is a continuous map on $A \times K \subset \mathbb{R}^m \times \mathbb{R}^n$. Prove that if K is compact, then for any $\vec{a} \in A$ and $\epsilon > 0$, there is $\delta > 0$, such that $\vec{x} \in A$, $\vec{y} \in K$, and $\|\vec{x} - \vec{a}\| < \delta$ imply $\|F(\vec{x}, \vec{y}) - F(\vec{a}, \vec{y})\| < \epsilon$.

Exercise 5.1.64. Suppose $f(\vec{x}, \vec{y})$ is a continuous function on $A \times K$. Prove that if K is compact, then $g(\vec{x}) = \max_{\vec{y} \in K} f(\vec{x}, \vec{y})$ is a continuous function on A.

Exercise 5.1.65. Suppose $f(\vec{x}, y)$ is a continuous function on $A \times [0, 1]$. Prove that $g(\vec{x}) = \displaystyle\int_0^1 f(\vec{x}, y) dy$ is a continuous function on A.

Continuity in Coordinates

A function $f(\vec{x}, \vec{y})$ is continuous in $\vec{x}$ if $\lim_{\vec{x}\to\vec{a}} f(\vec{x}, \vec{y}) = f(\vec{a}, \vec{y})$. It is uniformly continuous in $\vec{x}$ if the limit is uniform in $\vec{y}$: For any $\epsilon > 0$, there is $\delta > 0$, such that $\|\vec{x} - \vec{x}'\| < \delta$ implies $\|F(\vec{x}, \vec{y}) - F(\vec{x}', \vec{y})\| \leq \epsilon$ for any $\vec{y}$.

Exercise 5.1.66. Prove that if $f(\vec{x}, \vec{y})$ is continuous, then $f(\vec{x}, \vec{y})$ is continuous in $\vec{x}$. Moreover, show that the function $f(x, y) = \begin{cases} \dfrac{xy}{x^2 + y^2} & \text{if} (x, y) \neq (0, 0) \\ 0 & \text{if} (x, y) = (0, 0) \end{cases}$ is continuous in both x and y, but is not a continuous function.

Exercise 5.1.67. Prove that if $f(\vec{x}, \vec{y})$ is continuous in $\vec{x}$ and is uniformly continuous in $\vec{y}$, then $f(\vec{x}, \vec{y})$ is continuous.

Exercise 5.1.68. Suppose $f(\vec{x}, y)$ is continuous in $\vec{x}$ and in y. Prove that if $f(\vec{x}, y)$ is monotone in y, then $f(\vec{x}, y)$ is continuous.

5.2 Multivariable Algebra

To extend the differentiation from single to multivariable, we need to consider multivariable linear and polynomial maps. We introduce the necessary linear and multilinear algebras for multivariable analysis, including linear transform, linear functional, bilinear form, multilinear form, and exterior algebra. The discussion will be made on Euclidean spaces and can be easily extended to general finite dimensional vector spaces.

5.2.1 Linear Transform

A map $L \colon \mathbb{R}^n \to \mathbb{R}^m$ is a *linear transform* if it preserves the addition and scalar multiplication

$$L(\vec{x} + \vec{y}) = L(\vec{x}) + L(\vec{y}), \ \ L(c\vec{x}) = cL(\vec{x}). \tag{5.2.1}$$

As a consequence, a linear transform also preserves the linear combinations

$$L(c_1\vec{x}_1 + c_2\vec{x}_2 + \cdots + c_k\vec{x}_k) = c_1L(\vec{x}_1) + c_2L(\vec{x}_2) + \cdots + c_kL(\vec{x}_k).$$

In particular, for the standard basis $\vec{e}_1 = (1, 0, \ldots, 0)$, $\vec{e}_2 = (0, 1, \ldots, 0)$, $\ldots$, $\vec{e}_n = (0, 0, \ldots, 1)$ of $\mathbb{R}^n$ and the image $\vec{a}_i = L(\vec{e}_i)$ of the standard basis, we get

$$L(\vec{x}) = x_1\vec{a}_1 + x_2\vec{a}_2 + \cdots + x_n\vec{a}_n. \tag{5.2.2}$$

Conversely, any map given by the formula (5.2.2) is a linear transform.

Denote $\vec{a}_j = (a_{1j}, a_{2j}, \ldots, a_{mj})$. Then the i-th coordinate of $L(\vec{x})$ is

$$l_i(\vec{x}) = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n.$$

By expressing vectors as *vertical* lists of coordinates, the linear transform becomes

$$L(\vec{x}) = \begin{pmatrix} l_1(\vec{x}) \\ l_2(\vec{x}) \\ \vdots \\ l_m(\vec{x}) \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix}$$

$$= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = A\vec{x},$$

where A is an $m \times n$ matrix. Therefore matrices are really the explicit presentations of linear transforms. A linear transform L produces a matrix $A = \begin{pmatrix} L(\vec{e}_1) & L(\vec{e}_2) & \cdots & L(\vec{e}_n) \end{pmatrix}$. Conversely, a matrix A produces a linear transform $L(\vec{x}) = A\vec{x}$. Under the correspondence, the addition, scalar multiplication and composition of linear transforms become the addition, scalar multiplication and product of matrices.

A number valued linear transform $l\colon \mathbb{R}^n \to \mathbb{R}$ is called a *linear functional*. A linear functional can be written as

$$l(\vec{x}) = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = \vec{a} \cdot \vec{x} \qquad (5.2.3)$$

for a unique vector $\vec{a} = (a_1, a_2, \ldots, a_n)$. The levels $l(\vec{x}) = c$ of the linear functional are $(n-1)$-dimensional hyperplanes in $\mathbb{R}^n$ orthogonal to $\vec{a}$.

Exercise 5.2.1. Verify that a map given by the formula (5.2.2) is a linear transform. In other words, the map satisfies (5.2.1).

Exercise 5.2.2. Suppose $L, K\colon \mathbb{R}^n \to \mathbb{R}^m$ are linear transforms and c is a number. Verify that the maps $K + L$ and cK defined by

$$(K + L)(\vec{x}) = K(\vec{x}) + L(\vec{x}), \;\; (cK)(\vec{x}) = c(K(\vec{x}))$$

are linear transforms. Moreover, prove that if K and L correspond to $m \times n$ matrices A and B, then $K + L$ and cK correspond to the matrices $A + B$ and cA (the right way is to study the column vectors of $K + L$ and cK).

Exercise 5.2.3. Suppose $L\colon \mathbb{R}^n \to \mathbb{R}^m$ and $K\colon \mathbb{R}^k \to \mathbb{R}^n$ are linear transforms. Verify that the composition $L \circ K\colon \mathbb{R}^k \to \mathbb{R}^m$ is a linear transform. Moreover, prove that if K and L correspond to matrices A and B, then $L \circ K$ correspond to the product matrix BA (again, study the column vectors of the matrix corresponding to the composition).

Exercise 5.2.4. Suppose $L\colon \mathbb{R}^n \to \mathbb{R}^m$ is a linear transform. Verify that for any $\vec{y} \in \mathbb{R}^m$, $\vec{x} \mapsto L(\vec{x}) \cdot \vec{y}$ is a linear functional on $\mathbb{R}^n$. The linear functional (of $\vec{x}$) can be expressed as $L(\vec{x}) \cdot \vec{y} = \vec{x} \cdot \vec{z}$ for a unique $\vec{z} \in \mathbb{R}^n$.

1. Verify that $L^*(\vec{y}) = \vec{z}\colon \mathbb{R}^m \to \mathbb{R}^n$ is a linear transform. Note that L^* is the unique linear transform satisfying

$$L(\vec{x}) \cdot \vec{y} = \vec{x} \cdot L^*(\vec{y}) \text{ for all } \vec{x} \in \mathbb{R}^n \text{ and } \vec{y} \in \mathbb{R}^m. \qquad (5.2.4)$$

2. Prove that $(L + K)^* = L^* + K^*$, $(cK)^* = cK^*$, $(L \circ K)^* = K^* \circ L^*$.

3. Prove that if L corresponds to the $m \times n$ matrix A, then L^* corresponds to the transpose A^T of A. In particular, this implies $(BA)^T = A^T B^T$.

The linear transform L^* is called the *adjoint* of L.

Given a linear transform $L\colon \mathbb{R}^n \to \mathbb{R}^m$ and norms on $\mathbb{R}^n$ and $\mathbb{R}^m$, we have

$$\|L(\vec{x})\| \le |x_1| \|\vec{a}_1\| + |x_2| \|\vec{a}_2\| + \cdots + |x_n| \|\vec{a}_n\| \le (\|\vec{a}_1\| + \|\vec{a}_2\| + \cdots + \|\vec{a}_n\|) \|\vec{x}\|_\infty.$$

By the equivalence of norms on $\mathbb{R}^n$, we can find λ, such that $\|L(\vec{x})\| \le \lambda \|\vec{x}\|$. By $\|c\vec{x}\| = |c| \|\vec{x}\|$ and $L(c\vec{x}) = cL(\vec{x})$, the smallest such λ is

$$\begin{aligned}
\|L\| &= \inf\{\lambda\colon \|L(\vec{x})\| \le \lambda \|\vec{x}\| \text{ for any } \vec{x}\} \\
&= \inf\{\lambda\colon \|L(\vec{x})\| \le \lambda \text{ for any } \vec{x} \text{ satisfying } \|\vec{x}\| = 1\} \\
&= \sup\{\|L(\vec{x})\|\colon \|\vec{x}\| = 1\}. \qquad (5.2.5)
\end{aligned}$$

The number $\|L\|$ is the *norm of the linear transform* (with respect to the given norms on the Euclidean spaces). The norm $\|A\|$ of a matrix A is the norm of the corresponding linear transform. The inequality

$$\|L(\vec{x}) - L(\vec{y})\| = \|L(\vec{x} - \vec{y})\| \le \lambda \|\vec{x} - \vec{y}\|$$

also implies that linear transforms from finitely dimensional vector spaces are continuous.

The linear functions used in the differentiation of single variable functions will be extended to maps of the form $\vec{a} + L(\vec{x})$. We will call such maps *linear maps* (the formal name is *affine maps*) to distinguish from linear transforms (for which $\vec{a} = \vec{0}$).

Theorem 5.2.1. *The norm of linear transform satisfies the three axioms for the norm. Moreover, the norm of composition satisfies $\|L \circ K\| \leq \|L\| \|K\|$.*

Proof. By the definition, we have $\|L\| \geq 0$. If $\|L\| = 0$, then $\|L(\vec{x})\| = 0$ for any $\vec{x}$. By the positivity of norm, we get $L(\vec{x}) = \vec{0}$ for any $\vec{x}$. In other words, L is the zero transform. This verifies the positivity of $\|L\|$.

By the definition of $\|L\|$ and $\|K\|$, we have

$$\|L(\vec{x}) + K(\vec{x})\| \leq \|L(\vec{x})\| + \|K(\vec{x})\| \leq \|L\| \|\vec{x}\| + \|K\| \|\vec{x}\| \leq (\|L\| + \|K\|) \|\vec{x}\|.$$

Then by the definition of $\|L + K\|$, we get $\|L + K\| \leq \|L\| + \|K\|$. This verifies the triangle inequality.

The scalar property can be similarly proved.

Finally, we have

$$\|(L \circ K)(\vec{x})\| = \|L(K(\vec{x}))\| \leq \|L\| \|K(\vec{x})\| \leq \|L\| \|K\| \|\vec{x}\|.$$

Then by the definition of $\|L \circ K\|$, we have $\|L \circ K\| \leq \|L\| \|K\|$.

$\square$

Exercise 5.2.5. Prove that with respect to the Euclidean norm on $\mathbb{R}^n$ and the absolute value on $\mathbb{R}$, the norm of the linear functional $l(\vec{x}) = \vec{a} \cdot \vec{x} \colon \mathbb{R}^n \to \mathbb{R}$ is $\|\vec{a}\|_2$. Then extend the conclusion to linear transforms $\mathbb{R}^n \to \mathbb{R}^m$ with the Euclidean norm on $\mathbb{R}^n$ and the L_∞-norm on $\mathbb{R}^m$.

Exercise 5.2.6. Prove that if the Euclidean spaces are given the Euclidean norms, then $\|L\| = \sup_{\|\vec{x}\| = \|\vec{y}\| = 1} L(\vec{x}) \cdot \vec{y}$. This implies that the adjoint linear transform in Exercise 5.2.4 satisfies $\|L^*\| = \|L\|$ with respect to the Euclidean norms.

Exercise 5.2.7. Suppose $A = (a_{ij})$ is the matrix of a linear transform. Prove that $\|A\| \leq \sqrt{\sum a_{ij}^2}$ with respect to the Euclidean norms.

Exercise 5.2.8. Prove that the supremum in the definition of the norm of linear transform is in fact maximum. In other words, the supremum is reached at some vector of unit length.

Exercise 5.2.9. Suppose a continuous map $F \colon \mathbb{R}^n \to \mathbb{R}^m$ is *additive*: $F(\vec{x} + \vec{y}) = F(\vec{x}) + F(\vec{y})$. Prove that F is a linear map.

By using matrices, the collection of all the linear transforms from $\mathbb{R}^n$ to $\mathbb{R}^m$ is easily identified with the Euclidean space R^{mn}. On the space R^{mn}, the preferred norm is the norm of linear transform, although the choice is often not important because all the norms are equivalent.

Example 5.2.1. For an $m \times n$ matrix A, the left multiplication $L_A(X) = AX \colon \mathbb{R}^{nk} \to \mathbb{R}^{mk}$ is a linear transform. By $\|AX\| \leq \|A\| \|X\|$, we have $\|L_A\| \leq \|A\|$. Similarly, the right multiplication $R_A(X) = XA \colon \mathbb{R}^{km} \to \mathbb{R}^{kn}$ is a linear transform with $\|R_A\| \leq \|A\|$.

Example 5.2.2. Consider the map of taking squares $F(L) = L^2 \colon R^{n^2} \to R^{n^2}$. If H is a linear transform satisfying $\|H\| < \epsilon$, then by Proposition 5.2.1, we have

$$\|(L+H)^2 - L^2\| = \|LH + HL + H^2\| \leq \|LH\| + \|HL\| + \|H^2\|$$
$$\leq \|L\|\|H\| + \|H\|\|L\| + \|H\|^2 < 2\epsilon\|L\| + \epsilon^2.$$

From this it is easy to deduce that the square map is continuous.

Exercise 5.2.10. Prove that with respect to the norm of linear transform, the addition, scalar multiplication and composition of linear transforms are continuous maps.

Exercise 5.2.11. Is it possible to have $\|L_A\| = \|A\|$ for some choices of norms on the Euclidean space? Does the equality always hold for the special choice?

Exercise 5.2.12. Suppose $\mathbb{R}^n$ is given a norm, which induces a norm for the collection $\mathbb{R}^{n^2}$ of linear transforms from $\mathbb{R}^n$ to itself. Study the inverse of linear transforms.

1. Prove that if $\|L\| < 1$, then $\sum_{n=0}^{\infty} L^n = 1 + L + L^2 + \cdots$ converges in $\mathbb{R}^{n^2}$ and is the inverse of $I - L$ (I is the identity transform).

2. Prove that if L is invertible, then for any linear transform K satisfying $\|K - L\| < \dfrac{1}{\|L^{-1}\|}$, K is also invertible and $\|K^{-1} - L^{-1}\| \leq \dfrac{\|K - L\|\|L^{-1}\|^2}{1 - \|K - L\|\|L^{-1}\|}$. This implies that the collection $GL(n)$ of invertible linear transforms on $\mathbb{R}^n$ is an open subset of $\mathbb{R}^{n^2}$.

3. Prove that $L \mapsto L^{-1}$ is a continuous map on $GL(n)$.

5.2.2   Bilinear and Quadratic Form

A map $B \colon \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^k$ is *bilinear* if it is linear in the first vector

$$B(\vec{x} + \vec{x}', \vec{y}) = B(\vec{x}, \vec{y}) + B(\vec{x}', \vec{y}), \quad B(c\vec{x}, \vec{y}) = cB(\vec{x}, \vec{y}), \qquad (5.2.6)$$

and is also linear in the second vector

$$B(\vec{x}, \vec{y} + \vec{y}') = B(\vec{x}, \vec{y}) + B(\vec{x}, \vec{y}'), \quad B(\vec{x}, c\vec{y}) = cB(\vec{x}, \vec{y}). \qquad (5.2.7)$$

Bilinear maps can be considered as generalized products because the scalar product

$$c\vec{x} \colon \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n,$$

the dot product

$$\vec{x} \cdot \vec{y} \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R},$$

the 3-dimensional *cross product*

$$(x_1, x_2, x_3) \times (y_1, y_2, y_3) = (x_2 y_3 - x_3 y_2, x_3 y_1 - x_1 y_3, x_1 y_2 - x_2 y_1) \colon \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3,$$

and the matrix product

$$AB \colon \mathbb{R}^{mn} \times \mathbb{R}^{nk} \to \mathbb{R}^{mk}$$

are all bilinear.

The addition and scalar multiplication of bilinear maps are still bilinear. Moreover, if B is bilinear and L and K are linear, then $B(L(\vec{u}), K(\vec{v}))$ is bilinear (note that $(L, K)(\vec{u}, \vec{v}) = (L(\vec{u}), K(\vec{v}))$ is a linear transform). On the other hand, if L is linear, then $L(B(\vec{x}, \vec{y}))$ is bilinear.

For the standard bases $\vec{e}_i$ and $\vec{f}_j$ of $\mathbb{R}^m$ and $\mathbb{R}^n$, we get

$$B(\vec{x}, \vec{y}) = B\left(\sum_i x_i \vec{e}_i, \sum_j y_j \vec{f}_j\right) = \sum_{i,j} x_i y_j B(\vec{e}_i, \vec{f}_j) = \sum_{i,j} x_i y_j \vec{a}_{ij}. \quad (5.2.8)$$

Conversely, any map given by the formula (5.2.8) is bilinear. Moreover, we have

$$\|B(\vec{x}, \vec{y})\| \le \sum_{i,j} |x_i||y_j|\|\vec{a}_{ij}\| \le \left(\sum_{i,j} \|\vec{a}_{ij}\|\right) \|\vec{x}\|_\infty \|\vec{y}\|_\infty.$$

This implies that the bilinear map is continuous, and we may define the norm of the bilinear map to be

$$\|B\| = \inf\{\lambda \colon \|B(\vec{x}, \vec{y})\| \le \lambda \|\vec{x}\|\|\vec{y}\| \text{ for any } \vec{x}\}$$
$$= \sup\{\|B(\vec{x}, \vec{y})\| \colon \|\vec{x}\| = \|\vec{y}\| = 1\}. \quad (5.2.9)$$

The norm $\|B\|$ depends on the choices of norms on the three spaces $\mathbb{R}^m$, $\mathbb{R}^n$ and $\mathbb{R}^k$.

A map is bilinear if and only if its coordinate functions are *bilinear functions*. Bilinear functions $b(\vec{x}, \vec{y})$ for $\vec{x} \in \mathbb{R}^m$ and $\vec{y} \in \mathbb{R}^n$ are equivalent to matrices $A = (a_{ij})$ via the formula

$$b(\vec{x}, \vec{y}) = \sum_{i,j} a_{ij} x_i y_j = A\vec{x} \cdot \vec{y}. \quad (5.2.10)$$

The relation can also be written in terms of the linear transform as $b(\vec{x}, \vec{y}) = L(\vec{x}) \cdot \vec{y}$. The addition and scalar multiplication of bilinear functions correspond to the addition and scalar multiplication of matrices or linear transforms.

A bilinear function on $\mathbb{R}^n \times \mathbb{R}^n$ is called a *bilinear form* on $\mathbb{R}^n$. A bilinear form is *symmetric* if $b(\vec{x}, \vec{y}) = b(\vec{y}, \vec{x})$, and is *skew-symmetric* if $b(\vec{x}, \vec{y}) = -b(\vec{y}, \vec{x})$. We also call the corresponding matrices symmetric and skew-symmetric.

Exercise 5.2.13. Find the norms of the scalar product, dot product and matrix product with respect to the Euclidean norms on the Euclidean spaces.

Exercise 5.2.14. Prove that the norm of bilinear map satisfies the three conditions for norms.

Exercise 5.2.15. Find the norm of a bilinear function $b(\vec{x}, \vec{y}) = L(\vec{x}) \cdot \vec{y}$ with respect to the Euclidean norms.

Exercise 5.2.16. Prove that a bilinear form is skew-symmetric if and only if $b(\vec{x}, \vec{x}) = 0$ for any $\vec{x}$.

Exercise 5.2.17. Prove that any bilinear form is the unique sum of a symmetric form and a skew-symmetric form.

Exercise 5.2.18. Suppose $B\colon \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^k$ is a continuous map satisfying $B(\vec{x} + \vec{x}', \vec{y}) = B(\vec{x}, \vec{y}) + B(\vec{x}', \vec{y})$ and $B(\vec{x}, \vec{y} + \vec{y}') = B(\vec{x}, \vec{y}) + B(\vec{x}, \vec{y}')$. Prove that B is a bilinear map.

A skew-symmetric bilinear form on $\mathbb{R}^2$ is

$$b(\vec{x}, \vec{y}) = a_{12} x_1 y_2 + a_{21} x_2 y_1 = a_{12} x_1 y_2 - a_{12} x_2 y_1 = a_{12} \det \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \end{pmatrix}.$$

A skew-symmetric bilinear form on $\mathbb{R}^3$ is

$$\begin{aligned}
b(\vec{x}, \vec{y}) &= a_{12} x_1 y_2 + a_{21} x_2 y_1 + a_{13} x_1 y_3 + a_{31} x_3 y_1 + a_{23} x_2 y_3 + a_{23} x_3 y_2 \\
&= a_{12}(x_1 y_2 - x_2 y_1) + a_{31}(x_3 y_1 - x_1 y_3) + a_{23}(x_2 y_3 - x_3 y_2) \\
&= a_{23} \det \begin{pmatrix} x_2 & y_2 \\ x_3 & y_3 \end{pmatrix} + a_{31} \det \begin{pmatrix} x_3 & y_3 \\ x_1 & y_1 \end{pmatrix} + a_{12} \det \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \end{pmatrix} \\
&= \vec{a} \cdot (\vec{x} \times \vec{y}).
\end{aligned}$$

In general, a skew-symmetric bilinear form on $\mathbb{R}^n$ is

$$b(\vec{x}, \vec{y}) = \sum_{i \neq j} a_{ij} x_i y_j = \sum_{i < j} a_{ij}(x_i y_j - x_j y_i) = \sum_{i < j} a_{ij} \det \begin{pmatrix} x_i & y_i \\ x_j & y_j \end{pmatrix}.$$

Motivated by the formula on $\mathbb{R}^3$, we may define *n-dimensional cross product*

$$\vec{x} \times \vec{y} = \left((-1)^{i+j+1}(x_i y_j - x_j y_i) \right)_{i<j} \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^{\frac{n(n-1)}{2}}. \qquad (5.2.11)$$

Then we still have

$$b(\vec{x}, \vec{y}) = \vec{a} \cdot (\vec{x} \times \vec{y}),$$

where the ij-coordinate of $\vec{a} \in \mathbb{R}^{\frac{n(n-1)}{2}}$ is a_{ij} when $i + j$ is odd and is a_{ji} when $i + j$ is even. The formula for skew-symmetric bilinear forms may be compared with the formula (5.2.3) for linear functionals.

The general cross product $\vec{x} \times \vec{y}$ is still bilinear and satisfies $\vec{x} \times \vec{y} = -\vec{y} \times \vec{x}$. Therefore

$$\begin{aligned}
&(x_1 \vec{a}_1 + x_2 \vec{a}_2) \times (y_1 \vec{a}_1 + y_2 \vec{a}_2) \\
={}&x_1 y_1 \vec{a}_1 \times \vec{a}_1 + x_1 y_2 \vec{a}_1 \times \vec{a}_2 + x_2 y_1 \vec{a}_2 \times \vec{a}_1 + x_2 y_2 \vec{a}_2 \times \vec{a}_2 \\
={}&x_1 y_2 \vec{a}_1 \times \vec{a}_2 - x_2 y_1 \vec{a}_1 \times \vec{a}_2 = \det \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \end{pmatrix} \vec{a}_1 \times \vec{a}_2. \qquad (5.2.12)
\end{aligned}$$

In $\mathbb{R}^3$, the cross product $\vec{x} \times \vec{y}$ is a vector orthogonal to the two vectors, with the direction determined by the right hand rule from $\vec{x}$ to $\vec{y}$. Moreover, the Euclidean norm of $\vec{x} \times \vec{y}$ is given by the area of the parallelogram spanned by the two vectors

$$\|\vec{x} \times \vec{y}\|_2 = \|\vec{x}\|_2 \|\vec{y}\|_2 |\sin \theta|,$$

where θ is the angle between the two vectors. In other dimensions, we cannot talk about the direction because $\vec{x} \times \vec{y}$ is in a different Euclidean space.

However, the length of the cross product is still the area of the parallelogram spanned by the two vectors

$$\text{Area}(\vec{x}, \vec{y}) = \sqrt{(\vec{x} \cdot \vec{x})(\vec{y} \cdot \vec{y}) - (\vec{x} \cdot \vec{y})^2} = \sqrt{\sum_{i,j} x_i^2 y_j^2 - \sum_{i,j} x_i y_i x_j y_j}$$

$$= \sqrt{\sum_{i<j}(x_i^2 y_j^2 + x_j^2 y_i^2 - 2x_i y_i x_j y_j)} = \sqrt{\sum_{i<j}(x_i y_j - x_j y_i)^2}$$

$$= \|\vec{x} \times \vec{y}\|_2. \tag{5.2.13}$$

In $\mathbb{R}^2$, the cross product $\vec{x} \times \vec{y} \in \mathbb{R}$ is the determinant of the matrix formed by the two vectors. The Euclidean norm of the cross product is $|\det \begin{pmatrix} \vec{x} & \vec{y} \end{pmatrix}|$, which is the area of the parallelogram spanned by $\vec{x}$ and $\vec{y}$. A linear transform

$$L(\vec{x}) = A\vec{x} = x_1 \vec{a}_1 + x_2 \vec{a}_2 \colon \mathbb{R}^2 \to \mathbb{R}^n$$

takes the parallelogram spanned by $\vec{x}, \vec{y} \in \mathbb{R}^2$ to the parallelogram spanned by $L(\vec{x}), L(\vec{y}) \in \mathbb{R}^n$. By (5.2.12), the areas of the parallelograms are related by

$$\text{Area}(L(\vec{x}), L(\vec{y})) = |\det \begin{pmatrix} \vec{x} & \vec{y} \end{pmatrix}| \|\vec{a}_1 \times \vec{a}_2\|_2 = \|\vec{a}_1 \times \vec{a}_2\|_2 \text{Area}(\vec{x}, \vec{y}).$$

Exercise 5.2.19 (Archimedes[1]). Find the area of the region A bounded by $x = 0$, $y = 4$, and $y = x^2$ in the following way.

1. Show that the area of the triangle with vertices (a, a^2), $(a+h, (a+h)^2)$ and $(a+2h, (a+2h)^2)$ is $|h|^3$.

2. Let P_n be the polygon bounded by the line connecting $(0,0)$ to $(0,4)$, the line connecting $(0,4)$ to $(2,4)$, and the line segments connecting points on the parabola $y = x^2$ with $x = 0, \dfrac{1}{2^{n-1}}, \dfrac{2}{2^{n-1}}, \dots, \dfrac{2^n}{2^{n-1}}$.

3. Prove that the area of the polygon $P_n - P_{n-1}$ is $\dfrac{1}{4^{n-1}}$.

4. Prove that the area of the region A is $\sum_{n=0}^{\infty} \dfrac{1}{4^{n-1}} = \dfrac{16}{3}$.

A *quadratic form*

$$q(\vec{x}) = b(\vec{x}, \vec{x}) = \sum_{i,j} a_{ij} x_i x_j = A\vec{x} \cdot \vec{x} \tag{5.2.14}$$

is obtained by taking two vectors in a bilinear form to be the same. Since skew-symmetric bilinear forms induce the zero quadratic form, usually only symmetric bilinear forms are used to induce quadratic forms. Therefore quadratic forms are in one-to-one correspondence with symmetric bilinear

---

[1]Archimedes of Syracuse, born 287 BC and died 212 BC.

forms (and symmetric matrices, and *self-adjoint linear transforms*) by (note that $a_{ij} = a_{ji}$)

$$q(x_1, x_2, \ldots, x_n)$$
$$= \sum_{1 \le i \le n} a_{ii} x_i^2 + 2 \sum_{1 \le i < j \le n} a_{ij} x_i x_j \qquad (5.2.15)$$
$$= a_{11} x_1^2 + a_{22} x_2^2 + \cdots + a_{nn} x_n^2 + 2a_{12} x_1 x_2 + 2a_{13} x_1 x_3 + \cdots + 2a_{(n-1)n} x_{n-1} x_n.$$

Exercise 5.2.20. Prove that a symmetric bilinear form can be recovered from the corresponding quadratic form by

$$b(\vec{x}, \vec{y}) = \frac{1}{4}(q(\vec{x} + \vec{y}) - q(\vec{x} - \vec{y})) = \frac{1}{2}(q(\vec{x} + \vec{y}) - q(\vec{x}) - q(\vec{y})). \qquad (5.2.16)$$

Exercise 5.2.21. Prove that a quadratic form is homogeneous of second order

$$q(c\vec{x}) = c^2 q(\vec{x}),$$

and satisfies the *parellelogram law*

$$q(\vec{x} + \vec{y}) + q(\vec{x} - \vec{y}) = 2q(\vec{x}) + 2q(\vec{y}). \qquad (5.2.17)$$

Exercise 5.2.22. Suppose a function q satisfies the parellelogram law (5.2.17). Define a function $b(\vec{x}, \vec{y})$ by (5.2.16).

1. Prove that $q(\vec{0}) = 0$, $q(-\vec{x}) = q(\vec{x})$.

2. Prove that b is symmetric.

3. Prove that b satisfies $b(\vec{x} + \vec{y}, \vec{z}) + b(\vec{x} - \vec{y}, \vec{z}) = 2b(\vec{x}, \vec{z})$.

4. Suppose $f(\vec{x})$ is a function satisfying $f(\vec{0}) = 0$ and $f(\vec{x}+\vec{y})+f(\vec{x}-\vec{y}) = 2f(\vec{x})$. Prove that f is additive (see Exercise 5.2.9): $f(\vec{x} + \vec{y}) = f(\vec{x}) + f(\vec{y})$.

5. Prove that if q is continuous, then b is a bilinear form.

A quadratic form is *positive definite* if

$$\vec{x} \ne 0 \implies q(\vec{x}) > 0.$$

It is *negative definite* if

$$\vec{x} \ne 0 \implies q(\vec{x}) < 0.$$

It is *indefinite* if it takes both positive and negative values. In addition to the three possibilities, a quadratic form is *semi-positive definite* if $q(\vec{x}) \ge 0$ for all $\vec{x}$ and $q(\vec{x}) = 0$ for some $\vec{x} \ne \vec{0}$. It is *semi-negative definite* if $q(\vec{x}) \le 0$ for all $\vec{x}$ and $q(\vec{x}) = 0$ for some $\vec{x} \ne \vec{0}$.

The terms of the form $a_{ij} x_i x_j$, $i \ne j$ are *cross terms*. A quadratic form without cross terms is $q = a_{11} x_1^2 + a_{22} x_2^2 + \cdots + a_{nn} x_n^2$. It is positive definite if and only if all $a_{ii} > 0$. It is negative definite if and only if all $a_{ii} < 0$. It is indefinite if and only if some $a_{ii} > 0$ and some $a_{jj} < 0$. In general, the cross terms may be eliminated by the technique of completing the squares. Then the nature of the quadratic form can be determined.

Example 5.2.3. Consider the quadratic from $q(x, y, z) = x^2 + 13y^2 + 14z^2 + 6xy + 2xz + 18yz$. Putting together all the terms involving x and completing a square, we get

$$\begin{aligned} q &= x^2 + 6xy + 2xz + 13y^2 + 14z^2 + 18yz \\ &= [x^2 + 2x(3y + z) + (3y + z)^2] + 13y^2 + 14z^2 + 18yz - (3y + z)^2 \\ &= (x + 3y + z)^2 + 4y^2 + 13z^2 + 12yz. \end{aligned}$$

The remaining terms involve only y and z. Putting together all the terms involving y and completing a square, we get

$$4y^2 + 13z^2 + 12yz = (2y + 3z)^2 + 4z^2.$$

Thus $q = (x + 3y + z)^2 + (2y + 3z)^2 + (2z)^2 = u^2 + v^2 + w^2$ is positive definite.

Example 5.2.4. The cross terms in the quadratic form $q = 4x_1^2 + 19x_2^2 - 4x_4^2 - 4x_1x_2 + 4x_1x_3 - 8x_1x_4 + 10x_2x_3 + 16x_2x_4 + 12x_3x_4$ can be eliminated as follows.

$$q = 4[x_1^2 - x_1x_2 + x_1x_3 - 2x_1x_4] + 19x_2^2 - 4x_4^2 + 10x_2x_3 + 16x_2x_4 + 12x_3x_4$$

$$= 4\left[x_1^2 + 2x_1\left(-\frac{1}{2}x_2 + \frac{1}{2}x_3 - x_4\right) + \left(-\frac{1}{2}x_2 + \frac{1}{2}x_3 - x_4\right)^2\right]$$

$$+ 19x_2^2 - 4x_4^2 + 10x_2x_3 + 16x_2x_4 + 12x_3x_4 - 4\left(-\frac{1}{2}x_2 + \frac{1}{2}x_3 - x_4\right)^2$$

$$= 4\left(x_1 - \frac{1}{2}x_2 + \frac{1}{2}x_3 - x_4\right)^2 + 18\left[x_2^2 + \frac{2}{3}x_2x_3 + \frac{2}{3}x_2x_4\right] - x_3^2 - 8x_4^2 + 16x_3x_4$$

$$= (2x_1 - x_2 + x_3 - 2x_4)^2 + 18\left[x_2^2 + 2x_2\left(\frac{1}{3}x_3 + \frac{1}{3}x_4\right) + \left(\frac{1}{3}x_3 + \frac{1}{3}x_4\right)^2\right]$$

$$- x_3^2 - 8x_4^2 + 16x_3x_4 - 18\left(\frac{1}{3}x_3 + \frac{1}{3}x_4\right)^2$$

$$= (2x_1 - x_2 + x_3 - 2x_4)^2 + 18\left(x_2 + \frac{1}{3}x_3 + \frac{1}{3}x_4\right)^2 - 3(x_3^2 - 4x_3x_4) - 10x_4^2$$

$$= (2x_1 - x_2 + x_3 - 2x_4)^2 + 2(3x_2 + x_3 + x_4)^2$$

$$- 3[x_3^2 + 2x_3(-2x_4) + (-2x_4)^2] - 10x_4^2 + 3(-2x_4)^2$$

$$= (2x_1 - x_2 + x_3 - 2x_4)^2 + 2(3x_2 + x_3 + x_4)^2 - 3(x_3 - 2x_4)^2 + 2x_4^2$$

$$= y_1^2 + 2y_2^2 - 3y_3^3 + 2y_4^2.$$

The result tells us that q is indefinite.

Example 5.2.5. The quadratic form $q = 4xy + y^2$ has no x^2 term. We may complete the square by using the y^2 term and get $q = (y + 2x)^2 - 4x^2 = u^2 - 4v^2$, which is indefinite.

The quadratic form $q = xy + yz$ has no square terms. We may eliminate the cross terms by introducing $x = x_1 + y_1$, $y = x_1 - y_1$, so that $q = x_1^2 - y_1^2 + x_1z - y_1z$. Then we complete the square and get $q = \left(x_1 - \frac{1}{2}z\right)^2 - \left(y_1 + \frac{1}{2}z\right)^2 = \frac{1}{4}(x + y - z)^2 - \frac{1}{4}(x - y + z)^2$. The quadratic form is also indefinite.

Exercise 5.2.23. Eliminate the cross terms in the following quadratic forms. Then determine the nature of the quadratic forms.

1. $x^2 + 4xy - 5y^2$.

2. $2x^2 + 4xy$.

3. $4x_1^2 + 4x_1x_2 + 5x_2^2$.

4. $x^2 + 2y^2 + z^2 + 2xy - 2xz$.

5. $-2u^2 - v^2 - 6w^2 - 4uw + 2vw$.

6. $x_1^2 + x_3^2 + 2x_1x_2 + 2x_1x_3 + 2x_1x_4 + 2x_3x_4$.

Exercise 5.2.24. Eliminate the cross terms in the quadratic form $x^2 + 2y^2 + z^2 + 2xy - 2xz$ by first completing a square for terms involving z, then completing for terms involving y.

Exercise 5.2.25. Prove that if a quadratic form is positive definite, then all the coefficients of the square terms must be positive.

Exercise 5.2.26. Prove that if a quadratic form $q(\vec{x})$ is positive definite, then there is $\lambda > 0$, such that $q(\vec{x}) \geq \lambda \|\vec{x}\|^2$ for any $\vec{x}$.

Now let us study the technique of completing the squares in general.

The k-th *principal minor* of a matrix A is the determinant of the $k \times k$ matrix formed by the entries in the first k rows and first k columns of A. Consider a quadratic form $q(\vec{x}) = A\vec{x} \cdot \vec{x}$, where A is a symmetric matrix. Let

$$d_1 = a_{11}, \ \ d_2 = \det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \ \ \ldots, \ \ d_n = \det A$$

be the principal minors of the coefficient matrix.

If $d_1 \neq 0$, then eliminating all the cross terms involving x_1 gives us

$$q(\vec{x}) = a_{11}\left(x_1^2 + 2x_1\frac{1}{a_{11}}(a_{12}x_2 + \cdots + a_{1n}x_n) + \frac{1}{a_{11}^2}(a_{12}x_2 + \cdots + a_{1n}x_n)^2\right)$$
$$+ a_{22}x_2^2 + \cdots + a_{nn}x_n^2 + 2a_{23}x_2x_3 + 2a_{24}x_2x_4 + \cdots + 2a_{(n-1)n}x_{n-1}x_n$$
$$- \frac{1}{a_{11}}(a_{12}x_2 + \cdots + a_{1n}x_n)^2$$
$$= d_1\left(x_1 + \frac{a_{12}}{d_1}x_2 + \cdots + \frac{a_{1n}}{d_1}x_n\right)^2 + q_2(\vec{x}_2),$$

where q_2 is a quadratic form of the truncated vector $\vec{x}_2 = (x_2, x_3, \ldots, x_n)$. The coefficient matrix A_2 for q_2 is obtained as follows. For each $2 \leq i \leq n$, adding $-\dfrac{a_{1i}}{a_{11}}$ multiple of the first column of A to the i-th column will make the i-th term in the first row to become zero. Then we get a matrix $\begin{pmatrix} d_1 & \vec{0} \\ * & A_2 \end{pmatrix}$. Since the column operation does not change the determinant of the matrix (and all the principal minors), the principal minors $d_1^{(2)}, d_2^{(2)}, \ldots, d_{n-1}^{(2)}$ of A_2 are related to the principal minors $d_1^{(1)} = d_1$, $d_2^{(1)} = d_2$, ..., $d_n^{(1)} = d_n$ of A_1 by $d_{k+1}^{(1)} = d_1 d_k^{(2)}$.

The discussion sets up an inductive argument. Assume $d_1, d_2, \ldots, d_k$ are all nonzero. Then we may complete the squares in k steps and obtain

$$q(\vec{x}) = d_1^{(1)}(x_1 + b_{12}x_2 + \cdots + b_{1n}x_n)^2 + d_1^{(2)}(x_2 + b_{23}x_3 + \cdots + b_{2n}x_n)^2$$
$$+ \cdots + d_1^{(k)}(x_k + a_{k(k+1)}x_{k+1} + \cdots + a_{kn}x_n)^2 + q_{k+1}(\vec{x}_{k+1}),$$

with

$$d_1^{(i)} = \frac{d_2^{(i-1)}}{d_1^{(i-1)}} = \frac{d_3^{(i-2)}}{d_2^{(i-2)}} = \cdots = \frac{d_i^{(1)}}{d_{i-1}^{(1)}} = \frac{d_i}{d_{i-1}},$$

and the coefficient of x_{k+1}^2 in q_{k+1} is $d_1^{(k+1)} = \dfrac{d_{k+1}}{d_k}$.

Thus we have shown that if d_1, d_2, ..., d_n are all nonzero, then there is an "upper triangular" change of variables

$$\begin{aligned}
y_1 &= & x_1 &+b_{12}x_2 &+b_{13}x_3 &+\cdots &+b_{1n}x_n, \\
y_2 &= & & x_2 &+b_{23}x_3 &+\cdots &+b_{2n}x_n, \\
&\vdots & \\
y_n &= & & & & & x_n,
\end{aligned}$$

such that $q = d_1 y_1^2 + \dfrac{d_2}{d_1} y_2^2 + \cdots + \dfrac{d_n}{d_{n-1}} y_n^2$. Consequently, we get Sylvester's criterion

1. If $d_1 > 0$, $d_2 > 0$, ..., $d_n > 0$, then $q(\vec{x})$ is positive definite.

2. If $-d_1 > 0$, $d_2 > 0$, ..., $(-1)^n d_n > 0$, then $q(\vec{x})$ is negative definite.

3. If $d_1 > 0$, $d_2 > 0$, ..., $d_k > 0$, $d_{k+1} < 0$, or $-d_1 > 0$, $d_2 > 0$, ..., $(-1)^k d_k > 0$, $(-1)^{k+1} d_{k+1} < 0$, then $q(\vec{x})$ is indefinite.

5.2.3 Multilinear Map and Polynomial

A map $F(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k)\colon \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \cdots \times \mathbb{R}^{n_k} \to \mathbb{R}^m$ is *multilinear* if it is linear in each of its k variables. For example, if $B_1(\vec{x}, \vec{y})$ and $B_2(\vec{u}, \vec{v})$ are bilinear, then $B_1(\vec{x}, B_2(\vec{u}, \vec{v}))$ is a trilinear map in $\vec{x}$, $\vec{u}$, $\vec{v}$. The addition, scalar multiplication and composition of multilinear maps of matching types are still multilinear.

Similar to the discussion of linear and bilinear maps, a map is multilinear if and only if it is given by

$$F = \sum_{i_1, i_2, \cdots, i_k} x_{1i_1} x_{2i_2} \cdots x_{ki_k} \vec{a}_{i_1 i_2 \cdots i_k}, \quad \vec{a}_{i_1 i_2 \cdots i_k} = F(\vec{e}_{i_1}, \vec{e}_{i_2}, \ldots, \vec{e}_{i_k}), \quad (5.2.18)$$

in terms of the coordinates of the variables. Moreover, we have

$$\|F(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k)\| \leq \left(\sum_{i_1, i_2, \cdots, i_k} \|\vec{a}_{i_1 i_2 \cdots i_k}\| \right) \|\vec{x}_1\|_\infty \|\vec{x}_2\|_\infty \cdots \|\vec{x}_k\|_\infty.$$

Thus the norm of multilinear maps can be defined similar to the linear and bilinear maps.

A map is multilinear if and only if its coordinate functions are multilinear (of the same type). Multilinear functions are equivalent to the collection $(a_{i_1 i_2 \cdots i_k})$ of its coefficients, which can be imagined as a "k-dimensional matrix". Multilinear functions on $\mathbb{R}^n \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n$ are called *multilinear forms* on $\mathbb{R}^n$.

A multilinear map F on $\mathbb{R}^n \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n$ is *symmetric* if switching any two variables does not change the value

$$F(\vec{x}_1, \ldots, \vec{x}_i, \ldots, \vec{x}_j, \ldots, \vec{x}_k) = F(\vec{x}_1, \ldots, \vec{x}_j, \ldots, \vec{x}_i, \ldots, \vec{x}_k).$$

This is equivalent to the coefficients $\vec{a}_{i_1 i_2 \cdots i_k}$ being independent of the order of indices.

A *k-th order form*

$$\phi(\vec{x}) = f(\vec{x}, \vec{x}, \ldots, \vec{x}) = \sum_{i_1, i_2, \ldots, i_k} a_{i_1 i_2 \cdots i_k} x_{i_1} x_{i_2} \cdots x_{i_k} \qquad (5.2.19)$$

is obtained by taking all vectors in a multilinear form of k vectors to be the same. Similar to the quadratic forms, usually only symmetric multilinear forms are used here, and this gives an equivalence between symmetric multilinear forms of k vectors and k-th order forms. Write $a^{k_1 k_2 \cdots k_n} = a_{i_1 i_2 \cdots i_k}$ when the collection $\{i_1, i_2, \ldots, i_k\}$ consists of k_i copies of i for any $1 \le i \le n$. For example,

$$a_{244} x_2 x_4 x_4 = a_{424} x_4 x_2 x_4 = a_{442} x_4 x_4 x_2 = a^{0102} x_1^0 x_2^1 x_3^0 x_4^2.$$

Then we have

$$\phi(x_1, x_2, \ldots, x_n) = \sum_{k_1 + k_2 + \cdots + k_n = k, k_i \ge 0} \frac{k!}{k_1! k_2! \cdots k_n!} a^{k_1 k_2 \cdots k_n} x_1^{k_1} x_2^{k_2} \cdots x_n^{k_n}.$$
$$(5.2.20)$$

The k-th order forms satisfy $\phi(c\vec{x}) = c^k \phi(\vec{x})$. They are homogeneous functions of order k and are multivariable analogues of the monomial x^k for single variable polynomials. Thus a k-th order multivariable polynomial on $\mathbb{R}^n$ is a linear combination of j-th order forms with $j \le k$

$$p(\vec{x}) = \phi_0(\vec{x}) + \phi_1(\vec{x}) + \phi_2(\vec{x}) + \cdots + \phi_k(\vec{x})$$
$$= \sum_{k_1 + k_2 + \cdots + k_n \le k, k_i \ge 0} b^{k_1 k_2 \cdots k_n} x_1^{k_1} x_2^{k_2} \cdots x_n^{k_n}. \qquad (5.2.21)$$

By defining a polynomial as the sum of reductions of multilinear forms, the concept extends to general vector spaces and to maps between vector spaces. In particular, a map $F \colon \mathbb{R}^n \to \mathbb{R}^m$ is a polynomial map if and only if all its coordinate functions are polynomials.

A multilinear map F on $\mathbb{R}^n \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n$ is *alternating* if switching any two variables changes the sign

$$F(\vec{x}_1, \ldots, \vec{x}_i, \ldots, \vec{x}_j, \ldots, \vec{x}_k) = -F(\vec{x}_1, \ldots, \vec{x}_j, \ldots, \vec{x}_i, \ldots, \vec{x}_k).$$

This is equivalent to the coefficients $\vec{a}_{i_1 i_2 \cdots i_k}$ changing signs when two indices are exchanged. The alternating property is also equivalent to the value being zero when two vectors are equal (see Exercise 5.2.16)

$$F(\vec{x}_1, \ldots, \vec{y}, \ldots, \vec{y}, \ldots, \vec{x}_k) = 0.$$

In particular, if $k > n$, then at least two indices in $\vec{a}_{i_1 i_2 \cdots i_k}$ must be the same, and the coefficient vector is zero. Therefore any alternating multilinear map of k-variables in $\mathbb{R}^n$ is zero when $k > n$.

Suppose $k = n$. Then we have

$$
\begin{aligned}
& F(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n) \\
= & \sum_{i_1, i_2, \ldots, i_n} x_{1 i_1} x_{2 i_2} \cdots x_{n i_n} F(\vec{e}_{i_1}, \vec{e}_{i_2}, \ldots, \vec{e}_{i_n}) \\
= & \sum_{i_1, i_2, \ldots, i_n} \operatorname{sign}(i_1, i_2, \ldots, i_n) x_{1 i_1} x_{2 i_2} \cdots x_{n i_n} F(\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n) \\
= & \det \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{nn} \end{pmatrix} F(\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n),
\end{aligned}
\tag{5.2.22}
$$

where $(i_1, i_2, \ldots, i_n)$ is a rearrangement of $(1, 2, \ldots, n)$ (called a *permutation*), and $\operatorname{sign}(i_1, i_2, \ldots, i_n)$ is 1 if it takes even number of steps to recover $(1, 2, \ldots, n)$ from $(i_1, i_2, \ldots, i_n)$ by exchanging pairs of numbers, and $\operatorname{sign}(i_1, i_2, \ldots, i_n)$ is -1 if it takes odd number of steps. Thus the determinant of $n \times n$ matrices is the unique multilinear alternating function of the n column vectors, such that the value at the identity matrix (corresponding to the columns $\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n$) is 1.

Suppose $k \leq n$. Then we have

$$
F(\vec{e}_{i_1}, \vec{e}_{i_2}, \ldots, \vec{e}_{i_k}) = \pm F(\vec{e}_{j_1}, \vec{e}_{j_2}, \ldots, \vec{e}_{j_k}),
$$

where $j_1 < j_2 < \cdots < j_k$ is the rearrangement of $i_1, i_2, \ldots, i_k$ in increasing order. A computation similar to (5.2.22) tells us

$$
\begin{aligned}
& F(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k) \\
= & \sum_{i_1, i_2, \ldots, i_k} x_{1 i_1} x_{2 i_2} \cdots x_{k i_k} F(\vec{e}_{i_1}, \vec{e}_{i_2}, \ldots, \vec{e}_{i_k}) \\
= & \sum_{j_1 < j_2 < \cdots < j_k} \det \begin{pmatrix} x_{1 j_1} & x_{2 j_1} & \cdots & x_{k j_1} \\ x_{1 j_2} & x_{2 j_2} & \cdots & x_{k j_2} \\ \vdots & \vdots & & \vdots \\ x_{1 j_k} & x_{2 j_k} & \cdots & x_{k j_k} \end{pmatrix} F(\vec{e}_{j_1}, \vec{e}_{j_2}, \ldots, \vec{e}_{j_k}).
\end{aligned}
\tag{5.2.23}
$$

The skew-symmetric bilinear forms can be described by the generalized cross product $\vec{x} \times \vec{y} \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^{\frac{n(n-1)}{2}}$, where the coordinates of $\mathbb{R}^{\frac{n(n-1)}{2}}$ are indexed by $1 \leq i < j \leq n$. For alternating multilinear k-forms, we need to consider cross product of k vectors with value in a Euclidean space in which the coordinates are indexed by $1 \leq i_1 < i_2 < \cdots < i_k \leq n$. Therefore for any subset $I = \{i_1, i_2, \ldots, i_k\} \subset [n] = \{1, 2, \ldots, n\}$, arrange the indices in I in increasing order and introduce the symbol

$$
\vec{e}_{\wedge I} = \vec{e}_{i_1} \wedge \vec{e}_{i_2} \wedge \cdots \wedge \vec{e}_{i_k}.
$$

Define the k-th *exterior product* $\Lambda^k \mathbb{R}^n \cong \mathbb{R}^{\frac{n!}{k!(n-k)!}}$ of $\mathbb{R}^n$ to be the vector space with symbols $\vec{e}_{\wedge I}$ as basis (called the *standard basis* of $\Lambda^k \mathbb{R}^n$). Then define the k-th order *wedge product* to be the map

$$\vec{x}_1 \wedge \vec{x}_2 \wedge \cdots \wedge \vec{x}_k \colon \mathbb{R}^n \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \Lambda^k \mathbb{R}^n$$

given by

$$\vec{x}_1 \wedge \vec{x}_2 \wedge \cdots \wedge \vec{x}_k = \sum_{i_1 < i_2 < \cdots < i_k} \det \begin{pmatrix} x_{1i_1} & x_{2i_1} & \cdots & x_{ki_1} \\ x_{1i_2} & x_{2i_2} & \cdots & x_{ki_2} \\ \vdots & \vdots & & \vdots \\ x_{1i_k} & x_{2i_k} & \cdots & x_{ki_k} \end{pmatrix} \vec{e}_{i_1} \wedge \vec{e}_{i_2} \wedge \cdots \wedge \vec{e}_{i_k}.$$
$$(5.2.24)$$

The wedge product is multilinear and alternating. The notation is consistent in the sense that if $1 \le i_1 < i_2 < \cdots < i_k \le n$ and $\vec{x}_j = \vec{e}_{i_j}$, then the right side of (5.2.24) is indeed equal to $\vec{e}_{i_1} \wedge \vec{e}_{i_2} \wedge \cdots \wedge \vec{e}_{i_k}$. Moreover, if $I = \{i_1, i_2, \ldots, i_k\}$ but i_j is not necessarily increasing, then $\vec{e}_{i_1} \wedge \vec{e}_{i_2} \wedge \cdots \wedge \vec{e}_{i_k} = \pm \vec{e}_{\wedge I}$, where the sign is determined by the number of exchanges needed in order to rearrange the indices in increasing order.

Suppose

$$\vec{y}_1 = a_{11}\vec{x}_1 + a_{12}\vec{x}_2 + \cdots + a_{1k}\vec{x}_k,$$
$$\vec{y}_2 = a_{21}\vec{x}_1 + a_{22}\vec{x}_2 + \cdots + a_{2k}\vec{x}_k,$$
$$\vdots$$
$$\vec{y}_k = a_{k1}\vec{x}_1 + a_{k2}\vec{x}_2 + \cdots + a_{kk}\vec{x}_k,$$

and denote the coefficient matrix $A = (a_{ij})$. Then we have

$$\vec{y}_1 \wedge \vec{y}_2 \wedge \cdots \wedge \vec{y}_k = (\det A)\vec{x}_1 \wedge \vec{x}_2 \wedge \cdots \wedge \vec{x}_k. \qquad (5.2.25)$$

The equality can be derived similar to (5.2.22). It can also be derived by the fact that both sides are alternating multilinear maps of the row vectors of A, and both sides have the same values when the rows of A are standard basis vectors.

For the special case that $\vec{x}_i$ form the standard basis of the Euclidean space, the equality (5.2.25) becomes

$$\vec{x}_1 \wedge \vec{x}_2 \wedge \cdots \wedge \vec{x}_n = \det \begin{pmatrix} \vec{x}_1 & \vec{x}_2 & \cdots & \vec{x}_n \end{pmatrix} \vec{e}_{[n]} \ \text{ for } \ \vec{x}_i \in \mathbb{R}^n. \qquad (5.2.26)$$

Taking the map F in the equality (5.2.22) to be the wedge product will also lead to (5.2.26). The equality also generalizes (5.2.12) for the case $n = 2$.

If $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k$ are linearly independent, then by extending the vectors to a basis of $\mathbb{R}^n$, the equality (5.2.26) implies $\vec{x}_1 \wedge \vec{x}_2 \wedge \cdots \wedge \vec{x}_k \ne 0$. Conversely, if the vectors are linearly dependent, then one vector, say $\vec{x}_1$, can be written as a linear combination of the others

$$\vec{x}_1 = c_2 \vec{x}_2 + \cdots + c_k \vec{x}_k,$$

and by the alternating property of the wedge product,

$$\vec{x}_1 \wedge \vec{x}_2 \wedge \cdots \wedge \vec{x}_k = c_2 \vec{x}_2 \wedge \vec{x}_2 \wedge \cdots \wedge \vec{x}_k + \cdots + c_k \vec{x}_k \wedge \vec{x}_2 \wedge \cdots \wedge \vec{x}_k = 0.$$

Therefore we conclude that

$$\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k \text{ are linearly independent} \iff \vec{x}_1 \wedge \vec{x}_2 \wedge \cdots \wedge \vec{x}_k \neq 0. \quad (5.2.27)$$

Given a linear transform $L \colon \mathbb{R}^n \to \mathbb{R}^m$, we define a linear transform

$$\Lambda L = \Lambda^k L \colon \Lambda^k \mathbb{R}^n \to \Lambda^k \mathbb{R}^m$$

by linearly extending the values on the basis vectors

$$\Lambda L(\vec{e}_{i_1} \wedge \vec{e}_{i_2} \wedge \cdots \wedge \vec{e}_{i_k}) = L(\vec{e}_{i_1}) \wedge L(\vec{e}_{i_2}) \wedge \cdots \wedge L(\vec{e}_{i_k}).$$

Then we have

$$\Lambda L(\vec{x}_1 \wedge \vec{x}_2 \wedge \cdots \wedge \vec{x}_k) = L(\vec{x}_1) \wedge L(\vec{x}_2) \wedge \cdots \wedge L(\vec{x}_k) \quad (5.2.28)$$

because both sides are multilinear maps from $(\mathbb{R}^n)^k$ to $\Lambda \mathbb{R}^m$ with the same values at the standard basis vectors. The equality (5.2.28) implies that

$$\Lambda(L \circ K) = \Lambda L \circ \Lambda K.$$

Therefore if L is invertible, then ΛL is also invertible. Since any basis $\vec{b}_1$, $\vec{b}_2$, ..., $\vec{b}_n$ of $\mathbb{R}^n$ is the image of the standard basis $\vec{e}_1$, $\vec{e}_2$, ..., $\vec{e}_n$ under an invertible linear map, we conclude from (5.2.28) that $\vec{b}_{\wedge I} = \vec{b}_{i_1} \wedge \vec{b}_{i_2} \wedge \cdots \wedge \vec{b}_{i_k}$ also form a basis of the exterior product $\Lambda^k \mathbb{R}^n$. This suggests that the exterior product ΛV can be defined for any vector space V.

Because the dimension of $\Lambda^n \mathbb{R}^n$ is 1, the linear transform $\Lambda^n L \colon \Lambda^n \mathbb{R}^n \to \Lambda^n \mathbb{R}^n$ induced by a linear transform $L \colon \mathbb{R}^n \to \mathbb{R}^n$ is multiplying a number

$$L(\vec{x}_1) \wedge L(\vec{x}_2) \wedge \cdots \wedge L(\vec{x}_n) = d\vec{x}_1 \wedge \vec{x}_2 \wedge \cdots \wedge \vec{x}_n. \quad (5.2.29)$$

By comparing the special case

$$L(\vec{e}_1) \wedge L(\vec{e}_2) \wedge \cdots \wedge L(\vec{e}_n) = d\vec{e}_{\wedge[n]}$$

with (5.2.26), we find the number d to be the determinant of the matrix corresponding to L. Therefore the equation 5.2.29 gives a definition

$$L(\vec{x}_1) \wedge L(\vec{x}_2) \wedge \cdots \wedge L(\vec{x}_n) = (\det L)\vec{x}_1 \wedge \vec{x}_2 \wedge \cdots \wedge \vec{x}_n \quad (5.2.30)$$

of the determinant of a linear transform from $\mathbb{R}^n$ to itself without explicitly referring to the standard basis. Such a definition can be directly extended to general vector spaces. We also note that the equality $\Lambda^n(L \circ K) = \Lambda^n L \circ \Lambda^n K$ simply means

$$\det(L \circ K) = \det L \det K. \quad (5.2.31)$$

This gives a conceptual proof of the similar equality for matrices.

We have $\Lambda^0 \mathbb{R}^n = \mathbb{R}$, $\Lambda^1 \mathbb{R}^n = \mathbb{R}^n$, $\Lambda^n \mathbb{R}^n = \mathbb{R}\vec{e}_{\wedge[n]}$, and $\Lambda^k \mathbb{R}^n = 0$ for $k > n$. The total exterior product space

$$\Lambda \mathbb{R}^n = \Lambda^0 \mathbb{R}^n \oplus \Lambda^1 \mathbb{R}^n \oplus \Lambda^2 \mathbb{R}^n \oplus \cdots \oplus \Lambda^n \mathbb{R}^n$$

has dimension 2^n. Define the wedge product

$$\wedge \colon \Lambda^k \mathbb{R}^n \times \Lambda^l \mathbb{R}^n \to \Lambda^{k+l} \mathbb{R}^n, \ \ \Lambda \mathbb{R}^n \times \Lambda \mathbb{R}^n \to \Lambda \mathbb{R}^n$$

to be the bilinear map that extends the obvious values at the standard basis vectors. Then we have

$$(\vec{x}_1 \wedge \vec{x}_2 \wedge \cdots \wedge \vec{x}_k) \wedge (\vec{y}_1 \wedge \vec{y}_2 \wedge \cdots \wedge \vec{y}_l) = \vec{x}_1 \wedge \vec{x}_2 \wedge \cdots \wedge \vec{x}_k \wedge \vec{y}_1 \wedge \vec{y}_2 \wedge \cdots \wedge \vec{y}_l \quad (5.2.32)$$

because both sides are multilinear maps from $(\mathbb{R}^n)^{k+l}$ to $\Lambda^{k+l} \mathbb{R}^n$ with the same values at the standard basis vectors $\vec{e}_{\wedge I}$. The *associativity*

$$(\vec{\lambda} \wedge \vec{\mu}) \wedge \vec{\nu} = \vec{\lambda} \wedge (\vec{\mu} \wedge \vec{\nu}) \quad (5.2.33)$$

and the *graded commutativity*

$$\vec{\lambda} \wedge \vec{\mu} = (-1)^{kl} \vec{\mu} \wedge \vec{\lambda} \ \text{ for } \ \vec{\lambda} \in \Lambda^k \mathbb{R}^n, \vec{\mu} \in \Lambda^l \mathbb{R}^n \quad (5.2.34)$$

can also be derived by considering both sides as multilinear maps that coincide at the standard basis vectors. This makes $\Lambda \mathbb{R}^n$ into an algebra with unit $1 \in \Lambda^0 \mathbb{R}^n$, called the *exterior algebra*. The exterior algebra can be generated from the unit 1 and the vectors in $\mathbb{R}^n$ by the wedge product.

For a linear transform $L \colon \mathbb{R}^n \to \mathbb{R}^m$, the induced transform ΛL on the exterior algebra is an algebra homomorphism

$$\Lambda L(1) = 1,$$
$$\Lambda L(a\vec{\lambda} + b\vec{\mu}) = a\Lambda L(\vec{\lambda}) + b\Lambda L(\vec{\mu}),$$
$$\Lambda L(\vec{\lambda} \wedge \vec{\mu}) = \Lambda L(\vec{\lambda}) \wedge \Lambda L(\vec{\mu}).$$

By taking the standard basis $\wedge \vec{e}_I$ as an orthonormal basis, the exterior algebra $\Lambda \mathbb{R}^n$ becomes an inner product space. Then we have the equality

$$(\vec{x}_1 \wedge \vec{x}_2 \wedge \cdots \wedge \vec{x}_k) \cdot (\vec{y}_1 \wedge \vec{y}_2 \wedge \cdots \wedge \vec{y}_k) = \det(\vec{x}_i \cdot \vec{y}_j)_{1 \le i,j \le k} = \det X^T Y, \quad (5.2.35)$$

where

$$X = \begin{pmatrix} \vec{x}_1 & \vec{x}_2 & \cdots & \vec{x}_k \end{pmatrix}, \ \ Y = \begin{pmatrix} \vec{y}_1 & \vec{y}_2 & \cdots & \vec{y}_k \end{pmatrix}.$$

The equality is a consequence of the fact that both sides are multilinear functions on $(\mathbb{R}^n)^{2k}$ and are equal at the standard basis. As a result, if $U \colon \mathbb{R}^n \to \mathbb{R}^n$ is an orthogonal transform, then $\Lambda U \colon \Lambda \mathbb{R}^n \to \Lambda \mathbb{R}^n$ is also an orthogonal transform. In other words, if $\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_n$ is an orthogonal basis of $\mathbb{R}^n$, then $\vec{b}_{\wedge I}$ also form an orthogonal basis of $\Lambda \mathbb{R}^n$.

The equality (5.2.35) also tells us that if $\vec{\lambda} = \vec{x}_1 \wedge \vec{x}_2 \wedge \cdots \wedge \vec{x}_k$, $\vec{\mu} = \vec{y}_1 \wedge \vec{y}_2 \wedge \cdots \wedge \vec{y}_k$, $\vec{\xi} = \vec{z}_1 \wedge \vec{z}_2 \wedge \cdots \wedge \vec{z}_l$, $\vec{\eta} = \vec{w}_1 \wedge \vec{w}_2 \wedge \cdots \wedge \vec{w}_l$, such that $\vec{x}_i \cdot \vec{w}_j = 0$, then

$$(\vec{\lambda} \wedge \vec{\xi}) \cdot (\vec{\mu} \wedge \vec{\eta}) = (\vec{\lambda} \cdot \vec{\mu})(\vec{\eta} \cdot \vec{\xi}). \quad (5.2.36)$$

In general, for a subspace $V \subset \mathbb{R}^n$, we let ΛV to be the subspace of $\Lambda \mathbb{R}^n$ consisting of linear combination of $\vec{x}_1 \wedge \vec{x}_2 \wedge \cdots \wedge \vec{x}_k$ with $\vec{x}_i \in V$. Now if V and W are subspaces orthogonal to each other, then the equality (5.2.36) holds for $\vec{\lambda} \in \Lambda V$ and $\vec{\eta} \in \Lambda W$.

As suggested by the cross product and the extension to higher dimension, the Euclidean norm

$$\|\vec{x}_1 \wedge \vec{x}_2 \wedge \cdots \wedge \vec{x}_k\|_2 = \sqrt{\det(\vec{x}_i \cdot \vec{x}_j)_{1 \le i,j \le k}} = \sqrt{\det X^T X} \qquad (5.2.37)$$

with respect to the inner product in $\Lambda^k \mathbb{R}^n$ is the k-dimensional volume of the parallelepiped spanned by the k vectors in $\mathbb{R}^n$. Strictly speaking, the concept of volume is yet to be defined, and the formula needs to be justified. The justification will be carried out in Section 7.1.5. Note that by (5.2.36), we have

$$\vec{\lambda} \in \Lambda V, \ \vec{\mu} \in \Lambda W, \ V \perp W \implies \|\vec{\lambda} \wedge \vec{\mu}\|_2 = \|\vec{\lambda}\|_2 \|\vec{\mu}\|_2. \qquad (5.2.38)$$

Define the *duality map* (or *Hodge star operator*) on the exterior algebra

$$\vec{\lambda} \mapsto \vec{\lambda}^\star \colon \Lambda^k \mathbb{R}^n \to \Lambda^{n-k} \mathbb{R}^n,$$

as a linear map with the value on the standard basis given by

$$\vec{e}_{\wedge I}^\star = \pm \vec{e}_{\wedge([n]-I)},$$

where the sign is chosen so that

$$\vec{e}_{\wedge I} \wedge \vec{e}_{\wedge I}^\star = \vec{e}_{\wedge[n]} = \vec{e}_1 \wedge \vec{e}_2 \wedge \cdots \wedge \vec{e}_n. \qquad (5.2.39)$$

For example, we have

$$\vec{e}_i^\star = (-1)^{i-1} \vec{e}_{\wedge([n]-i)},$$
$$\vec{e}_{i \wedge j}^\star = (-1)^{i+j-1} \vec{e}_{\wedge([n]-\{i,j\})}, \quad \text{for } i < j,$$
$$\vec{e}_{\wedge([n]-i)}^\star = (-1)^{n-i} \vec{e}_i.$$

Moreover, the dual in $\mathbb{R}^2$ is the counterclockwise rotation by 90 degrees

$$(x_1, x_2)^\star = x_1 \vec{e}_1^\star + x_2 \vec{e}_2^\star = x_1 \vec{e}_2 - x_2 \vec{e}_1 = (-x_2, x_1), \qquad (5.2.40)$$

and the cross product in $\mathbb{R}^3$ is the dual of the wedge product

$$
\begin{aligned}
&((x_1, x_2, x_3) \wedge (y_1, y_2, y_3))^\star \\
=& \det \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \end{pmatrix} (\vec{e}_1 \wedge \vec{e}_2)^* + \det \begin{pmatrix} x_1 & y_1 \\ x_3 & y_3 \end{pmatrix} (\vec{e}_1 \wedge \vec{e}_3)^* + \det \begin{pmatrix} x_2 & y_2 \\ x_3 & y_3 \end{pmatrix} (\vec{e}_2 \wedge \vec{e}_3)^* \\
=& \det \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \end{pmatrix} \vec{e}_3 - \det \begin{pmatrix} x_1 & y_1 \\ x_3 & y_3 \end{pmatrix} \vec{e}_2 + \det \begin{pmatrix} x_2 & y_2 \\ x_3 & y_3 \end{pmatrix} \vec{e}_1 \\
=& (x_1, x_2, x_3) \times (y_1, y_2, y_3).
\end{aligned} \qquad (5.2.41)
$$

Suppose I consists of k indices. Then by $\vec{e}_{\wedge I} \wedge \vec{e}_{\wedge I}^\star = \vec{e}_{\wedge[n]} = \vec{e}_{\wedge I}^\star \wedge \vec{e}_{\wedge I}^{\star\star} = (-1)^{k(n-k)}\vec{e}_{\wedge I}^{\star\star} \wedge \vec{e}_{\wedge I}^\star$, we have

$$\vec{\lambda}^{\star\star} = (-1)^{k(n-k)}\vec{\lambda} \ \text{ for } \ \vec{\lambda} \in \Lambda^k\mathbb{R}^n. \tag{5.2.42}$$

We also have

$$\vec{\lambda} \wedge \vec{\mu}^\star = (\vec{\lambda} \cdot \vec{\mu})\vec{e}_{\wedge[n]} \tag{5.2.43}$$

because both sides are bilinear maps from $(\Lambda^k\mathbb{R}^n)^2$ to $\Lambda^n\mathbb{R}^n$ that are equal at the standard basis vectors. By $(\vec{\lambda}^\star \cdot \vec{\mu}^\star)\vec{e}_{\wedge[n]} = \vec{\lambda}^\star \wedge \vec{\mu}^{\star\star} = (-1)^{k(n-k)}\vec{\lambda}^\star \wedge \vec{\mu} = \vec{\mu} \wedge \vec{\lambda}^\star = (\vec{\mu} \cdot \vec{\lambda})\vec{e}_{\wedge[n]} = (\vec{\lambda} \cdot \vec{\mu})\vec{e}_{\wedge[n]}$, we find the duality map preserves the inner product

$$\vec{\lambda}^\star \cdot \vec{\mu}^\star = \vec{\lambda} \cdot \vec{\mu}, \tag{5.2.44}$$

and therefore also preserves the Euclidean norm

$$\|\vec{\lambda}^\star\|_2 = \|\vec{\lambda}\|_2.$$

The equalities (5.2.42) and (5.2.43) imply

$$\vec{\mu} \wedge \vec{\lambda} = (\vec{\mu} \cdot \vec{\lambda}^*)\vec{e}_{\wedge[n]}. \tag{5.2.45}$$

Since a vector $\vec{\xi} \in \Lambda^k\mathbb{R}^n$ is uniquely determined by the dot product $\vec{\mu} \cdot \vec{\xi}$ for all $\vec{\mu} \in \Lambda^k\mathbb{R}^n$, the dual $\vec{\lambda}^*$ may be characterized as the unique exterior vector satisfying (5.2.45) for all $\vec{\mu}$.

Suppose $U \colon \mathbb{R}^n \to \mathbb{R}^n$ is an orthogonal transform. Then

$$\Lambda U(\vec{\mu}) \wedge \Lambda U(\vec{\lambda}) = \Lambda U(\vec{\mu} \wedge \vec{\lambda}) = (\vec{\mu} \cdot \vec{\lambda}^*)\Lambda U(\vec{e}_{\wedge[n]})$$
$$= (\Lambda U(\vec{\mu}) \cdot \Lambda U(\vec{\lambda}^*))(\det U)\vec{e}_{\wedge[n]}.$$

Since $\Lambda U(\vec{\mu})$ can be any vector in $\Lambda^k\mathbb{R}^n$, the equality tells us

$$\Lambda U(\vec{\lambda})^* = (\det U)\Lambda U(\vec{\lambda}^*) \ \text{ for } \ \vec{\lambda} \in \Lambda^k\mathbb{R}^n \text{ and orthogonal } U. \tag{5.2.46}$$

A basis $\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_n$ of $\mathbb{R}^n$ is *positively oriented* if

$$\det \begin{pmatrix} \vec{b}_1 & \vec{b}_2 & \cdots & \vec{b}_n \end{pmatrix} > 0.$$

By (5.2.26), this means

$$\vec{b}_1 \wedge \vec{b}_2 \wedge \cdots \wedge \vec{b}_n = d\vec{e}_1 \wedge \vec{e}_2 \wedge \cdots \wedge \vec{e}_n, \ \ d > 0.$$

Suppose $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k$ and $\vec{x}_{k+1}, \vec{x}_{k+2}, \ldots, \vec{x}_n$ are two collections of linearly independent vectors, such that any vector in the first collection is orthogonal to any vector in the second. Then the union $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n$ of two collections is a basis in $\mathbb{R}^n$. By modifying $\vec{x}_n$ by a sign if necessary, we assume the basis $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n$ is positively oriented. Then we claim that

$$(\vec{x}_1 \wedge \vec{x}_2 \wedge \cdots \wedge \vec{x}_k)^* = c\vec{x}_k \wedge \vec{x}_{k+1} \wedge \cdots \wedge \vec{x}_n \tag{5.2.47}$$

with

$$c = \frac{\|\vec{x}_1 \wedge \vec{x}_2 \wedge \cdots \wedge \vec{x}_k\|_2}{\|\vec{x}_k \wedge \vec{x}_{k+1} \wedge \cdots \wedge \vec{x}_n\|_2} > 0.$$

The equality (5.2.47) suggests that the duality can be defined in the exterior algebra of any inner product space.

To derive the equality (5.2.47), we choose an orthonormal bases $\vec{b}_1$, $\vec{b}_2$, ..., $\vec{b}_k$ for the span of $\vec{x}_1$, $\vec{x}_2$, ..., $\vec{x}_k$ and an orthonormal bases $\vec{b}_{k+1}$, $\vec{b}_{k+2}$, ..., $\vec{b}_n$ for the span of $\vec{x}_{k+1}$, $\vec{x}_{k+2}$, ..., $\vec{x}_n$, such that

$$\vec{x}_1 \wedge \vec{x}_2 \wedge \cdots \wedge \vec{x}_k = \alpha \vec{b}_1 \wedge \vec{b}_2 \wedge \cdots \wedge \vec{b}_k,$$
$$\vec{x}_k \wedge \vec{x}_{k+1} \wedge \cdots \wedge \vec{x}_n = \beta \vec{b}_{k+1} \wedge \vec{b}_{k+2} \wedge \cdots \wedge \vec{b}_n,$$

for some $\alpha, \beta > 0$. Then $\vec{b}_1$, $\vec{b}_2$, ..., $\vec{b}_n$ is also a positively oriented orthonormal basis of $\mathbb{R}^n$. Let U be the orthogonal transform given by the matrix $\begin{pmatrix} \vec{b}_1 & \vec{b}_2 & \cdots & \vec{b}_n \end{pmatrix}$. Then $\det U > 0$ and by (5.2.46)

$$\begin{aligned}
\vec{b}_1 \wedge \vec{b}_2 \wedge \cdots \wedge \vec{b}_k^* &= (\Lambda U(\vec{e}_1 \wedge \vec{e}_2 \wedge \cdots \wedge \vec{e}_k))^* \\
&= \Lambda U(\vec{e}_{k+1} \wedge \vec{e}_{k+2} \wedge \cdots \wedge \vec{e}_n) \\
&= \vec{b}_{k+1} \wedge \vec{b}_{k+2} \wedge \cdots \wedge \vec{b}_n,
\end{aligned}$$

Therefore we conclude $(\vec{x}_1 \wedge \vec{x}_2 \wedge \cdots \wedge \vec{x}_k)^* = \dfrac{\alpha}{\beta} \vec{x}_k \wedge \vec{x}_{k+1} \wedge \cdots \wedge \vec{x}_n$.

In the special case $k = n - 1$, we find $\vec{y} = (\vec{x}_1 \wedge \vec{x}_2 \wedge \cdots \wedge \vec{x}_{n-1})^*$ is a vector $\mathbb{R}^n$ such that

$$\vec{x}_i \cdot \vec{y} = 0, \ \ \vec{x}_1 \wedge \vec{x}_2 \wedge \cdots \wedge \vec{x}_{n-1} \wedge \vec{y} = \|\vec{x}_1 \wedge \vec{x}_2 \wedge \cdots \wedge \vec{x}_{n-1}\|_2^2 \vec{e}_{[n]}.$$

Therefore $\vec{y}$ is a vector orthogonal to $\vec{x}_1$, $\vec{x}_2$, ..., $\vec{x}_{n-1}$ and has "compatible" direction. The geometrical meaning and the equality (5.2.41) shows that the map

$$(\vec{x}_1 \wedge \vec{x}_2 \wedge \cdots \wedge \vec{x}_{n-1})^* \colon (\mathbb{R}^n)^{n-1} \to \mathbb{R}^n$$

is the true generalization of the cross product in $\mathbb{R}^3$.

5.2.4 Additional Exercises

Isometry between Normed Spaces

Let $\|\vec{x}\|$ and $\|\|\vec{x}\|\|$ be norms on $\mathbb{R}^n$ and $\mathbb{R}^m$. A map $F \colon \mathbb{R}^n \to \mathbb{R}^m$ is an *isometry* if it satisfies $\|\|F(\vec{x}) - F(\vec{y})\|\| = \|\vec{x} - \vec{y}\|$. Exercise 7.2.5 says that isometries preserve the length of curves.

A map $F \colon \mathbb{R}^n \to \mathbb{R}^m$ is *affine* if the following equivalent conditions are satisfied.

1. $F((1-t)\vec{x} + t\vec{y}) = (1-t)F(\vec{x}) + tF(\vec{y})$ for any $\vec{x}, \vec{y} \in \mathbb{R}^n$ and any $0 \le t \le 1$.

2. $F((1-t)\vec{x} + t\vec{y}) = (1-t)F(\vec{x}) + tF(\vec{y})$ for any $\vec{x}, \vec{y} \in \mathbb{R}^n$ and any $t \in \mathbb{R}$.

3. $F(\vec{x}) = \vec{a} + L(\vec{x})$ for a fixed vector $\vec{a}$ (necessarily equal to $F(\vec{0})$) and a linear transform L.

Exercise 5.2.27. Suppose the norm $\|\|\vec{x}\|\|$ on $\mathbb{R}^m$ is *strictly convex*, in the sense $\|\|\vec{x} + \vec{y}\|\| = \|\|\vec{x}\|\| + \|\|\vec{y}\|\|$ implies $\vec{x}$ and $\vec{y}$ are parallel. Prove that for any isometry F and $\vec{x}, \vec{y}, \vec{z} = (1 - t)\vec{x} + t\vec{y} \in \mathbb{R}^n$, $0 \leq t \leq 1$, the vectors $F(\vec{z}) - F(\vec{x})$ and $F(\vec{y}) - F(\vec{z})$ must be parallel. Then prove that the isometry F is affine.

Exercise 5.2.28. Show that the L_p-norm is strictly convex for $1 < p < \infty$. Then show that an isometry between Euclidean spaces with the Euclidean norms must be of the form $\vec{a} + L(\vec{x})$, where L is a linear transform with its matrix A satisfying $A^T A = I$.

Exercise 5.2.29. For any $\vec{u} \in \mathbb{R}^n$, the map $\phi(\vec{x}) = 2\vec{u} - \vec{x}$ is the reflection with respect to $\vec{u}$. A subset $K \subset \mathbb{R}^n$ is symmetric with respect to $\vec{u}$ if $\vec{x} \in K$ implies $\phi(\vec{x}) \in K$. The subset has radius $r(K) = \sup_{\vec{x} \in K} \|\vec{x} - \vec{u}\|$.

1. Prove that ϕ is an isometry from $(\mathbb{R}^n, \|\vec{x}\|)$ to itself, $\|\phi(\vec{x}) - \vec{x}\| = 2\|\vec{x} - \vec{u}\|$, and $\vec{u}$ is the only point fixed by ϕ (i.e., satisfying $\phi(\vec{x}) = \vec{x}$).

2. For a subset K symmetric with respect to $\vec{u}$, prove that the subset has diameter $\sup_{\vec{x}, \vec{y} \in K} \|\vec{x} - \vec{y}\| = 2r(K)$. Then prove that the subset $K' = \{\vec{x} \colon K \subset B(\vec{x}, r(K))\}$ has radius $r(K') \leq \dfrac{1}{2} r(K)$.

3. For any $\vec{a}, \vec{b} \in \mathbb{R}^n$, denote $\vec{u} = \dfrac{\vec{a} + \vec{b}}{2}$. Prove that

$$K_0 = \{\vec{x} \colon \|\vec{x} - \vec{a}\| = \|\vec{x} - \vec{b}\| = \frac{1}{2}\|\vec{a} - \vec{b}\|\}$$

is symmetric with respect to $\vec{u}$. Then prove that the sequence K_n defined by $K_{n+1} = K_n'$ satisfies $\cap K_n = \{\vec{u}\}$.

The last part gives a characterization of the middle point $\vec{u}$ of two points $\vec{a}$ and $\vec{b}$ purely in terms of the norm.

Exercise 5.2.30 (Mazur-Ulam Theorem). Prove that an invertible isometry is necessarily affine. Specifically, suppose $F \colon (\mathbb{R}^n, \|\vec{x}\|) \to (\mathbb{R}^n, \|\|\vec{x}\|\|)$ is an invertible isometry. By using the characterization of the middle point in Exercise 5.2.28, prove that the map preserves the middle point

$$F\left(\frac{\vec{a} + \vec{b}}{2}\right) = \frac{F(\vec{a}) + F(\vec{b})}{2}.$$

Then further prove that the property implies F is affine.

Exercise 5.2.31. Let $\phi(t)$ be a real function and consider $F(t) = (t, \phi(t)) \colon \mathbb{R} \to \mathbb{R}^2$. For the absolute value on $\mathbb{R}$ and the L_∞-norm on $\mathbb{R}^2$, find suitable condition on ϕ to make sure F is an isometry. The exercise shows that an isometry is not necessarily affine in general.

Chapter 6

Multivariable Differentiation

6.1   Differentiation

The differentiation can be extended from single to multivariable by considering the linear approximation. The coefficients in the linear approximation are given by partial derivatives, and most of the properties of differentiation can be extended. However, the multivariable differentiability is no longer equivalent to the existence of partial derivatives. Moreover, we can consider derivatives in directions other than the coordinate directions (which give the partial derivatives).

6.1.1   Differentiability and Derivative

The differentiation of maps between Euclidean spaces can be defined by directly generalizing the definition for single variable functions. Denote $\Delta \vec{x} = \vec{x} - \vec{x}_0$.

Definition 6.1.1. A map $F(\vec{x})$ defined on a ball around $\vec{x}_0$ is *differentiable* at $\vec{x}_0$ if there is a linear map $\vec{a} + L(\Delta \vec{x})$, such that for any $\epsilon > 0$, there is $\delta > 0$, such that

$$\|\Delta \vec{x}\| = \|\vec{x} - \vec{x}_0\| < \delta \implies \|F(\vec{x}) - \vec{a} - L(\Delta \vec{x})\| \le \epsilon \|\Delta \vec{x}\|. \qquad (6.1.1)$$

The linear transform L is the *derivative* of F at $\vec{x}_0$ and is denoted $F'(\vec{x}_0)$.

The condition (6.1.1) can be rephrased as

$$\vec{a} = F(\vec{x}_0), \quad \lim_{\Delta \vec{x} \to \vec{0}} \frac{\|F(\vec{x}_0 + \Delta \vec{x}) - F(\vec{x}_0) - L(\Delta \vec{x})\|}{\|\Delta \vec{x}\|} = 0. \qquad (6.1.2)$$

The differentiability at a point requires the map to be defined everywhere near the point. Thus for single variable functions, the differentiability is defined only for functions on open intervals or unions of open intervals. For multivariable maps, the differentiability is defined only for maps on open subsets. In the future, the differentiability will be extended to maps defined on "differentiable subsets" (called submanifolds).

Similar to the single variable case, we denote the *differential $dF = L(d\vec{x})$* of a map F. Again at the moment it is only a symbolic notation.

Example 6.1.1. For the map $F(x, y) = (x^2 + y^2, xy) \colon \mathbb{R}^2 \to \mathbb{R}^2$, we have

$$F(x, y) - F(x_0, y_0) = (2x_0 \Delta x + \Delta x^2 + 2y_0 \Delta y + \Delta y^2, y_0 \Delta x + x_0 \Delta y + \Delta x \Delta y)$$
$$= (2x_0 \Delta x + 2y_0 \Delta y, y_0 \Delta x + x_0 \Delta y) + (\Delta x^2 + \Delta y^2, \Delta x \Delta y).$$

Since $\|(\Delta x^2 + \Delta y^2, \Delta x \Delta y)\|_\infty \le 2\|(\Delta x, \Delta y)\|_\infty^2$, F is differentiable at (x_0, y_0) and the derivative $F'(x_0, y_0)$ is the linear transform $(u, v) \mapsto (2x_0 u + 2y_0 v, y_0 u + x_0 v)$. The differential is $d_{(x_0, y_0)}F = (2x_0 dx + 2y_0 dy, y_0 dx + x_0 dy)$.

Example 6.1.2. For the function $f(\vec{x}) = \vec{x} \cdot \vec{x} = \|\vec{x}\|_2^2$, we have

$$f(\vec{x}_0 + \Delta \vec{x}) - f(\vec{x}_0) = 2\vec{x}_0 \cdot \Delta \vec{x} + \|\Delta \vec{x}\|_2^2.$$

Since $2\vec{x}_0 \cdot \Delta \vec{x}$ is a linear functional of $\Delta \vec{x}$, f is differentiable, with the derivative $f'(\vec{x}_0)(\vec{v}) = 2\vec{x}_0 \cdot \vec{v}$ and the differential $d_{\vec{x}_0} f = 2\vec{x}_0 \cdot d\vec{x}$.

Example 6.1.3. The space of $n \times n$ matrices form the Euclidean space $\mathbb{R}^{n^2}$. For the map $F(X) = X^2 \colon \mathbb{R}^{n^2} \to \mathbb{R}^{n^2}$ of taking squares of matrices, we have

$$F(A + H) - F(A) = (A^2 + AH + HA + H^2) - A^2 = AH + HA + H^2.$$

The map $H \mapsto AH + HA$ is a linear transform, and by Proposition 5.2.1, $\|H^2\| \le \|H\|^2 \le \epsilon \|H\|$ when $\|H\| < \epsilon$. Therefore the map is differentiable, with the derivative $F'(A)(H) = AH + HA$ and the differential $d_A F = A(dX) + (dX)A$.

Exercise 6.1.1. Use the definition to show the differentiability of the function $f(x, y) = ax^2 + 2bxy + cy^2$ and find the derivative.

Exercise 6.1.2. Prove that if a map is differentiable at $\vec{x}_0$, then the map is continuous at $\vec{x}_0$. Then show that the Euclidean norm $\|\vec{x}\|_2$ is continuous but not differentiable at $\vec{0}$.

Exercise 6.1.3. Suppose F is differentiable at $\vec{x}_0$, with $F(\vec{x}_0) = \vec{0}$. Suppose a function $\lambda(\vec{x})$ is continuous at $\vec{x}_0$. Prove that $\lambda(\vec{x})F(\vec{x})$ is differentiable at $\vec{x}_0$.

Exercise 6.1.4. Find the condition for a homogeneous function to be differentiable at $\vec{0}$. What about a multihomogeneous function?

Exercise 6.1.5. Suppose A is an $n \times n$ matrix. Find the derivative of the function $A\vec{x} \cdot \vec{x}$.

Exercise 6.1.6. Let X^T be the transpose of a matrix X. Prove that the derivative of $F(X) = X^T X \colon \mathbb{R}^{n^2} \to \mathbb{R}^{n^2}$ is $F'(A)(H) = A^T H + H^T A$.

Exercise 6.1.7. For any natural number k, find the derivative of the map of taking the k-th power of matrices.

Exercise 6.1.8. Use the equality $(I + H)^{-1} = I - H + H^2(I + H)^{-1}$ and Exercise 5.2.12 to find the derivative of the inverse matrix map at the identity matrix I.

6.1.2 Partial Derivative

By considering the L_∞-norm, we see that a map is differentiable if and only if each coordinate function is differentiable. Moreover, the linear approximation is obtained by putting together the linear approximations of the coordinate functions.

A function $f(\vec{x})$ on $\mathbb{R}^n$ is approximated at $\vec{x}_0$ by a linear function

$$a + b_1 \Delta x_1 + b_2 \Delta x_2 + \cdots + b_n \Delta x_n$$

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

$$\begin{aligned}
&|\Delta x_i| = |x_i - x_{i0}| < \delta, \text{ for all } i \\
\implies &|f(x_1, x_2, \ldots, x_n) - a - b_1 \Delta x_1 - b_2 \Delta x_2 - \cdots - b_n \Delta x_n| \\
&\le \epsilon \max\{|\Delta x_1|, |\Delta x_2|, \ldots, |\Delta x_n|\}.
\end{aligned}$$

If we fix $x_2 = x_{20}, \ldots, x_n = x_{n0}$, and let only x_1 change, then the above says that $f(x_1, x_{20}, \ldots, x_{n0})$ is approximated by the linear function $a + b_1 \Delta x_1$ at $x_1 = x_{10}$. Thus $a = f(\vec{x}_0)$ and

$$b_1 = \lim_{\Delta x_1 \to 0} \frac{f(x_{10} + \Delta x_1, x_{20}, \ldots, x_{n0}) - f(x_{10}, x_{20}, \ldots, x_{n0})}{\Delta x_1}$$

is the derivative of $f(\vec{x})$ in x_1 with all the other coordinates fixed. The coefficient is called the *partial derivative* in the first variable. The other coefficients are the similar partial derivatives and denoted

$$b_i = \frac{\partial f}{\partial x_i} = D_{x_i} f = f_{x_i}. \tag{6.1.3}$$

Using the notation, the derivative $f'(\vec{x})$ is the linear transform

$$(v_1, v_2, \ldots, v_n) \mapsto \frac{\partial f}{\partial x_1} v_1 + \frac{\partial f}{\partial x_2} v_2 + \cdots + \frac{\partial f}{\partial x_n} v_n \colon \mathbb{R}^n \to \mathbb{R},$$

and the differential of the function is

$$df = \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2 + \cdots + \frac{\partial f}{\partial x_n} dx_n. \tag{6.1.4}$$

Moreover, introduce the *gradient*

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n} \right). \tag{6.1.5}$$

Then the linear approximation is $f(\vec{x}_0) + \nabla f(\vec{x}_0) \cdot \Delta \vec{x}$, the derivative is $f'(\vec{x})(\vec{v}) = \nabla f(\vec{x}_0) \cdot \vec{v}$, and the differential is $df = \nabla f \cdot d\vec{x}$.

Example 6.1.4. The function $f(x, y) = 1 + 2x + xy^2$ has partial derivatives at $(0,0)$

$$f(0,0) = 1, \quad f_x(0,0) = (2 + y^2)|_{x=0, y=0} = 2, \quad f_y(0,0) = 2xy|_{x=0, y=0} = 0.$$

So the candidate for the linear approximation is $1 + 2(x - 0) + 0(y - 0) = 1 + 2x$. Then we verify that (note that $\Delta x = x$, $\Delta y = y$)

$$|f(x, y) - 1 - 2x| = |xy^2| \le \|(x, y)\|_\infty^3.$$

Therefore f is differentiable at $(0, 0)$, with $d_{(0,0)} f = 2dx$.

Example 6.1.5. The function $f(x, y) = \sqrt{|xy|}$ satisfies $f(0,0) = 0$, $f_x(0,0) = 0$, $f_y(0,0) = 0$. So the candidate for the linear approximation at $(0,0)$ is the zero function. However, by Example 5.1.2, the limit

$$\lim_{(x,y)\to(0,0)} \frac{|f(x, y)|}{\|(x, y)\|_2} = \lim_{x\to 0, y\to 0} \sqrt{\frac{|xy|}{x^2 + y^2}}$$

does not exist. Therefore f is not differentiable at $(0, 0)$, despite the existence of the partial derivatives.

Exercise 6.1.9. Compute the partial derivatives.

1. $4xy^3 + 5x^3 y - 3y^2$.
2. $\arctan \dfrac{y}{x}$.
3. $e^{xyz} \sin(x + 2y + 3z)$.
4. $\log(x^2 + y^2)$.
5. x^{y^z}.
6. $(xy)^z$.

Exercise 6.1.10. Discuss the continuity, the existence of partial derivatives and the differentiability at $(0,0)$. $p, q > 0$.

1. $f(x, y) = \begin{cases} \dfrac{|x|^p y}{x^2 + y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0) \end{cases}$.

2. $f(x, y) = \begin{cases} \dfrac{xy}{(x^2 + y^2)^p} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0) \end{cases}$.

3. $f(x, y) = \begin{cases} (x^2 + y^2)^p \sin \dfrac{1}{x^2 + y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0) \end{cases}$.

4. $f(x, y) = \begin{cases} |x|^p |y|^q \sin \dfrac{1}{x^2 + y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0) \end{cases}$.

Exercise 6.1.11. Prove the properties of the gradient.

1. $\nabla c = \vec{0}$.

2. $\nabla(f + g) = \nabla f + \nabla g$.

3. $\nabla(fg) = f\nabla g + g\nabla f$.

4. $\nabla \phi(f) = \phi'(f)\nabla f$.

Putting the linear approximations of coordinate functions together, we get the linear approximation of a differentiable map. In particular, the derivative linear transform $F'(\vec{x}_0)$ is given by the *Jacobian matrix*

$$\frac{\partial F}{\partial \vec{x}} = \frac{\partial(f_1, f_2, \ldots, f_m)}{\partial(x_1, x_2, \ldots, x_n)} = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2} & \cdots & \dfrac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial f_m}{\partial x_1} & \dfrac{\partial f_m}{\partial x_2} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{pmatrix}. \tag{6.1.6}$$

For a differentiable multivariable function $f(\vec{x}) \colon \mathbb{R}^n \to \mathbb{R}$, the Jacobian matrix is the gradient ∇f considered as a $1 \times n$ matrix.

It is rather tedious to first use the partial derivatives to find the candidate linear approximation and then verify the condition for differentiability. Fortunately, there is a simple sufficient (but not necessary) condition for the differentiability.

Proposition 6.1.2. *Suppose a map is defined near $\vec{x}_0$. If all the partial derivatives exist near $\vec{x}_0$ and the partial derivatives are continuous at $\vec{x}_0$, then the map is differentiable at $\vec{x}_0$.*

A differentiable map $F \colon \mathbb{R}^n \to \mathbb{R}^m$ is *continuously differentiable* if the map $\vec{x} \mapsto F'(\vec{x}) \colon \mathbb{R}^n \to \mathbb{R}^{mn}$ is continuous. This is equivalent to all the partial derivatives are continuous.

Proof. Assume the partial derivatives $f_x(x, y)$ and $f_y(x, y)$ exist near (x_0, y_0). Applying the mean value theorem to $f(t, y)$ for $t \in [x_0, x]$ and to $f(x_0, s)$ for

$s \in [y_0, y]$, we get

$$\begin{aligned} f(x, y) - f(x_0, y_0) &= (f(x, y) - f(x_0, y)) + (f(x_0, y) - f(x_0, y_0)) \\ &= f_x(c, y)(x - x_0) + f_y(x_0, d)(y - y_0) \\ &= f_x(c, y)\Delta x + f_y(x_0, d)\Delta y \end{aligned}$$

for some $c \in [x_0, x]$ and $d \in [y_0, y]$. If f_x and f_y are continuous at (x_0, y_0), then for any $\epsilon > 0$, there is $\delta > 0$, such that

$$|\Delta x| < \delta, |\Delta y| < \delta \implies |f_x(x, y) - f_x(x_0, y_0)| < \epsilon, |f_y(x, y) - f_y(x_0, y_0)| < \epsilon.$$

By $|c - x_0| < |\Delta x|$ and $|d - y_0| < |\Delta y|$, we find $|\Delta x| < \delta$ and $|\Delta y| < \delta$ implies $|f_x(c, y) - f_x(x_0, y_0)| < \epsilon$, $|f_y(x_0, d) - f_y(x_0, y_0)| < \epsilon$, so that

$$|f(x, y) - f(x_0, y_0) - f_x(x_0, y_0)\Delta x - f_y(x_0, y_0)\Delta y| \le \epsilon|\Delta x| + \epsilon|\Delta y|.$$

This shows that $f(x, y)$ is differentiable at (x_0, y_0).

The proof for the general case is similar.

$\square$

Example 6.1.6. The function $f(x, y) = 1 + 2x + xy^2$ in Example 6.1.4 has continuous partial derivatives $f_x = 2 + y^2$, $f_y = 2xy$. Therefore the function is differentiable with $df = (2 + y^2)dx + 2xydy$.

Example 6.1.7. Express the map in Example 6.1.1 as $u(x, y) = x^2 + y^2$, $v(x, y) = xy$. The partial derivatives $u_x = 2x$, $u_y = 2y$, $v_x = y$, $v_y = x$ are continuous. Therefore the map is differentiable, with the differential $\begin{pmatrix} du \\ dv \end{pmatrix} = \begin{pmatrix} 2xdx + 2ydy \\ ydx + xdy \end{pmatrix} = \begin{pmatrix} 2x & 2y \\ y & x \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix}$.

Example 6.1.8. On the plane, the polar coordinate (r, θ) and the cartesian coordinate (x, y) are related by $x = r\cos\theta$, $y = r\sin\theta$. The relation is differentiable because the partial derivatives are continuous. The Jacobian matrix $\dfrac{\partial(x, y)}{\partial(r, \theta)} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$ and the differential $\begin{pmatrix} dx \\ dy \end{pmatrix} = \begin{pmatrix} \cos\theta dr - r\sin\theta d\theta \\ \sin\theta dr + r\cos\theta d\theta \end{pmatrix} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} \begin{pmatrix} dr \\ d\theta \end{pmatrix}$.

Example 6.1.9. By Proposition 2.1.3, the differentiability of a parametrized curve

$$\phi(t) = (x_1(t), x_2(t), \dots, x_n(t)) \colon (a, b) \to \mathbb{R}^n$$

is equivalent to the existence of the derivatives x_i'. The Jacobian matrix is the vertical version of the *tangent vector* $\phi' = (x_1', x_2', \dots, x_n')$. As a linear map, the derivative is $u \mapsto u\phi'$.

The definition of parametrized curve allows the constant curve and allows the tangent vector to become zero. It also allows continuously differentiable curves to appear to have sharp corners, such as this example

$$\phi(t) = \begin{cases} (t^2, 0) & \text{if } t \le 0 \\ (0, t^2) & \text{if } t > 0 \end{cases}.$$

To avoid such undesirable situations, we say a differentiable parametrized curve is *regular* if the tangent vector ϕ' is never zero. We will see that by the inverse function theorem (see Theorem 6.2.1), this is equivalent to the possibility of local reparametrization by one coordinate (i.e., $t = x_i$ for some i).

Example 6.1.10. The parametrized sphere (5.1.19) and the parametrized torus (5.1.20) are differentiable surfaces in $\mathbb{R}^3$. In general, if a parametrized surface $\sigma(u, v)\colon \mathbb{R}^2 \to \mathbb{R}^n$ is differentiable, then the tangent vectors σ_u and σ_v span the *tangent plane* $T_{(u_0, v_0)}S$. As a linear map, the derivative σ' is $(s, t) \mapsto s\sigma_u + t\sigma_v$.

The undesirable situation that may happen for parametrized curves may also happen for parametrized surfaces. We say a differentiable parametrized surface is *regular* if the tangent vectors σ_u and σ_v are always linearly independent. We will see that by the inverse function theorem (see Theorem 6.2.1), this is equivalent to the possibility of local reparametrization by two coordinates (i.e., $u = x_i$ and $v = x_j$ for some i and j).

Exercise 6.1.12. Compute the Jacobian matrices and the differentials.

1. $r = \sqrt{x^2 + y^2}$, $\theta = \arctan \dfrac{y}{x}$.

2. $u_1 = x_1 + x_2 + x_3$, $u_2 = x_1 x_2 + x_2 x_3 + x_3 x_1$, $u_3 = x_1 x_2 x_3$.

3. $x = r \sin \phi \cos \theta$, $y = r \sin \phi \sin \theta$, $z = r \cos \phi$.

Exercise 6.1.13. Suppose $F(\vec{x})$ is a bounded map near $\vec{0}$. Prove that $\|\vec{x}\|_2^2 F(\vec{x})$ is differentiable at $\vec{0}$. Then use this to construct a function that is differentiable at $\vec{0}$ but has no partial derivatives away from $\vec{0}$.

Exercise 6.1.14. Construct a function that is differentiable everywhere but the partial derivatives are not continuous at $\vec{0}$.

Exercise 6.1.15. Prove that if $f_x(x_0, y_0)$ exists and $f_y(x, y)$ is continuous at (x_0, y_0), then $f(x, y)$ is differentiable at (x_0, y_0). Extend the result to three or more variables.

Exercise 6.1.16. Prove that a function $f(\vec{x})$ is differentiable at $\vec{x}_0$ if and only if $f(\vec{x}) = f(\vec{x}_0) + J(\vec{x}) \cdot \Delta\vec{x}$, where $J\colon \mathbb{R}^n \to \mathbb{R}^n$ is continuous at $\vec{x}_0$. Extend the fact to differentiable maps.

6.1.3 Rules of Differentiation

The rules of computation of the derivatives can be extended from single to multivariable since the underlying principle still holds. If maps have linear approximations, then their addition, scalar multiplication and composition are still approximated by the similar combinations of the linear approximations.

Suppose $F, G\colon \mathbb{R}^n \to \mathbb{R}^m$ are differentiable at $\vec{x}_0$. Then the sum $F + G\colon \mathbb{R}^n \to \mathbb{R}^m$ is also differentiable at $\vec{x}_0$, with the derivative

$$(F + G)' = F' + G'.$$

Therefore the Jacobian matrix for $F + G$ is the sum of the Jacobian matrices of F and G. In terms of the individual entries of the Jacobian matrix, the equality means $\dfrac{\partial(f + g)}{\partial x_i} = \dfrac{\partial f}{\partial x_i} + \dfrac{\partial g}{\partial x_i}$.

Suppose a function $\lambda\colon \mathbb{R}^n \to \mathbb{R}$ and a map $F\colon \mathbb{R}^n \to \mathbb{R}^m$ are differentiable at $\vec{x}_0$. Then the scalar multiplication $\lambda F\colon \mathbb{R}^n \to \mathbb{R}^m$ is also differentiable at $\vec{x}_0$, with the derivative given by the *Leibniz formula*

$$(\lambda F)' = \lambda' F + \lambda F'.$$

Note that both sides are supposed to be linear transforms from $\mathbb{R}^n$ to $\mathbb{R}^m$. At $\vec{x}_0$, we have $\lambda(\vec{x}_0) \in \mathbb{R}$, $\lambda'(\vec{x}_0) \colon \mathbb{R}^n \to \mathbb{R}$, $F(\vec{x}_0) \in \mathbb{R}^m$ and $F'(\vec{x}_0) \colon \mathbb{R}^n \to \mathbb{R}^m$. Then we see that the right side takes a vector in $\mathbb{R}^n$ to a vector in $\mathbb{R}^m$. Of course, in terms of the individual entries of the Jacobian matrix, the equality means $\dfrac{\partial(\lambda f)}{\partial x_i} = \dfrac{\partial \lambda}{\partial x_i} f + \lambda \dfrac{\partial f}{\partial x_i}$.

The most general form of the Leibniz formula can be found in Exercise 6.1.17.

For the composition, we get multivariable version of the chain rule. Suppose $F \colon \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\vec{x}_0$ and $G \colon \mathbb{R}^m \to \mathbb{R}^k$ is differentiable at $\vec{y}_0 = F(\vec{x}_0)$. Then the composition $G \circ F \colon \mathbb{R}^n \to \mathbb{R}^k$ is also differentiable at $\vec{x}_0$, with the derivative given by the *chain rule*

$$(G \circ F)' = G' \circ F'.$$

The right side is a composition of the linear transforms $F'(\vec{x}_0) \colon \mathbb{R}^n \to \mathbb{R}^m$ and $G'(\vec{y}_0) = G'(F(\vec{x}_0)) \colon \mathbb{R}^m \to \mathbb{R}^k$. Therefore the Jacobian matrix of the composition is the product of the Jacobian matrices of individual maps. In terms of the individual entries of the Jacobian matrix, the chain rule means

$$\frac{\partial(g \circ F)}{\partial x_i} = \frac{\partial g}{\partial f_1}\frac{\partial f_1}{\partial x_i} + \frac{\partial g}{\partial f_2}\frac{\partial f_2}{\partial x_i} + \cdots + \frac{\partial g}{\partial f_m}\frac{\partial f_m}{\partial x_i}.$$

Example 6.1.11. Suppose $u = e^x \sin(y + 2z)$, $x = t^2$, $y = t^3 - 2t$, $z = t^2 + t$. The relations can be considered as a composition $t \mapsto (x, y, z) \mapsto u$. To compute the derivative of $u(t) = u(x(t), y(t), z(t))$, we differentiate the relations

$$du = e^x \sin(y + 2z)dx + e^x \cos(y + 2z)dy + 2e^x \cos(y + 2z)dz,$$
$$dx = 2tdt,$$
$$dy = (3t^2 - 2)dt,$$
$$dz = (2t + 1)dt.$$

Substituting dx, dy, dz into du, we get

$$du = (e^x \sin(y + 2z))2tdt + (e^x \cos(y + 2z))(3t^2 - 2)dt + (2e^x \cos(y + 2z))(2t + 1)dt$$
$$= e^{t^2}(2t \sin(t^3 + 2t^2) + (3t^2 - 2 + 4t + 2) \cos(t^3 + 2t^2))dt.$$

Thus

$$\frac{du}{dt} = e^{t^2}(2t \sin(t^3 + 2t^2) + (3t^2 + 4t) \cos(t^3 + 2t^2)).$$

We note that the coefficients in the relation between the differentials form the Jacobian matrix. The substitution of the differential relation besically computes the product of the Jacobian matrix.

Example 6.1.12. Suppose $u = e^x \sin(y + 2z)$, $x = s^2 t$, $y = s + 2t$, $z = s^2 - t$. Then

$$du = e^x \sin(y + 2z)dx + e^x \cos(y + 2z)dy + 2e^x \cos(y + 2z)dz,$$
$$dx = 2stds + s^2 dt,$$
$$dy = ds + 2dt,$$
$$dz = 2sds - dt.$$

Substituting dx, dy, dz into du, we get

$$du = e^x \sin(y + 2z)(2st\,ds + s^2\,dt) + e^x \cos(y + 2z)(ds + 2dt) + 2e^x \cos(y + 2z)(2s\,ds - dt)$$
$$= e^{s^2 t}(2st \sin(s + 2s^2) + (1 + 4s) \cos(s + 2s^2))ds + e^{s^2 t}s^2 \sin(s + 2s^2)dt.$$

In other words,

$$u_s = e^{s^2 t}(2st \sin(s + 2s^2) + (1 + 4s) \cos(s + 2s^2)),$$
$$u_t = e^{s^2 t}s^2 \sin(s + 2s^2).$$

Exercise 6.1.17. Suppose $F\colon \mathbb{R}^p \to \mathbb{R}^m$ and $G\colon \mathbb{R}^p \to \mathbb{R}^n$ are differentiable at $\vec{z}_0 \in \mathbb{R}^p$, and $B\colon \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^k$ is a bilinear map. Prove that $B(F, G)\colon \mathbb{R}^p \to \mathbb{R}^k$ is also differentiable at $\vec{z}_0$, with the derivative given by the Leibniz formula

$$B(F, G)' = B(F', G) + B(F, G').$$

Extend the Leibniz formula to multilinear maps.

Exercise 6.1.18. Prove the chain rule for multivariable maps. Does the chain rule formula $\dfrac{dg(x(t), y(t))}{dt} = g_x(x(t), y(t))x'(t) + g_y(x(t), y(t))y'(t)$ still hold if g is not assumed to be differentiable?

Exercise 6.1.19. Suppose $F, G\colon \mathbb{R}^n \to \mathbb{R}^n$ are maps that are invertible to each other. Suppose F is differentiable at $\vec{x}_0$ and G is differentiable at $\vec{y}_0 = F(\vec{x}_0)$. Prove that the linear transforms $F'(\vec{x}_0)$ and $G'(\vec{y}_0)$ are inverse to each other. In particular, the derivative of a differentiable invertible map is an invertible linear transform.

Exercise 6.1.20. Compute partial derivatives.

1. $z = uv + \sin t$, $u = e^t$, $v = \cos t$, find z_t.

2. $z = x^2 \log y$, $x = \dfrac{u}{v}$, $y = u - v$, find z_u, z_v.

3. $u = f(x + y, x - y)$, find u_x, u_y.

4. $u = f(r \cos \theta, r \sin \theta)$, find u_r, u_θ.

5. $u = f\left(\dfrac{x}{y}, \dfrac{y}{z}, \dfrac{z}{x}\right)$, find u_x, u_y, u_z.

Exercise 6.1.21. Compute the derivative of x^x by considering $f = u^v$, $u = x$, $v = x$.

Exercise 6.1.22. Prove that a differentiable function $f(x, y)$ depends only on the angle in the polar coordinate if and only if $xf_x + yf_y = 0$. Can you make a similar claim for a differentiable function that depends only on the length.

Exercise 6.1.23. Suppose $f(x, y)$ is a differentiable function satisfying $f(x, x^2) = 1$, $f_x(x, x^2) = x$. Find $f_y(x, x^2)$.

Exercise 6.1.24. For a differentiable function f, show that $u = yf(x^2 - y^2)$ satisfies $y^2 u_x + xyu_y = u$.

Exercise 6.1.25. Suppose differentiable functions f and g satisfy $f(u, v, w) = g(x, y, z)$ for $u = \sqrt{yz}$, $v = \sqrt{zx}$, $w = \sqrt{xy}$. Prove that $uf_u + vf_v + wf_w = xg_x + yg_y + zg_z$. Exercise 6.3.47 is a vast generalization.

Exercise 6.1.26. Find the partial differential equation characterizing differentiable functions $f(x, y)$ of the form $f(x, y) = \phi(xy)$. What about functions of the form $f(x, y) = \phi\left(\dfrac{y}{x}\right)$?

Exercise 6.1.27. Discuss the differentiability of a function $f(\|\vec{x}\|_2)$ of the Euclidean norm.

Exercise 6.1.28. For any invertible matrix A, the inverse map is the composition of the following three maps:

$$X \mapsto A^{-1}X, \ \ X \mapsto X^{-1}, \ \ X \mapsto XA^{-1}.$$

Use this and Exercise 6.1.8 to find the derivative of the inverse matrix map at a matrix A.

6.1.4   Directional Derivative

The restriction of a function f on a parametrized curve $\phi(t)\colon (a, b) \to \mathbb{R}^n$ is the composition $f(\phi(t))$. If both f and ϕ are differentiable, then the change of the function f along the curve ϕ is measured by

$$f(\phi(t))' = \frac{\partial f}{\partial x_1}\frac{dx_1}{dt} + \frac{\partial f}{\partial x_2}\frac{dx_2}{dt} + \cdots + \frac{\partial f}{\partial x_n}\frac{dx_n}{dt} = \nabla f \cdot \phi'. \qquad (6.1.7)$$

The formula shows that the change depends on the tangent vector instead of the curve itself. Define the derivative of f along any vector $\vec{v}$ to be

$$D_{\vec{v}}f = \lim_{t \to 0} \frac{f(\vec{x} + t\vec{v}) - f(\vec{x})}{t} = f'(\vec{x})(\vec{v}) = \nabla f \cdot \vec{v}. \qquad (6.1.8)$$

Then we get $f(\phi(t))' = D_{\phi'(t)}f$. Note that the derivative along $\vec{v}$ is the value of the derivative linear functional $f'(\vec{x})$ at $\vec{v}$. Thus $D_{\vec{v}}f$ is a linear function of $\vec{v}$ and the partial derivatives are simply the values of the linear function at the standard basis vectors.

If $\vec{v}$ has unit Euclidean length (i.e., $\|\vec{v}\|_2 = 1$), then only the direction of the vector matters and $D_{\vec{v}}f$ is called the *directional derivative*. Let θ be the angle between ∇f and $\vec{v}$. Then $D_{\vec{v}}f = \|\nabla f\|_2 \cos\theta$. The formula shows that the function "does not change" in directions orthogonal to its gradient and increases the most (at the rate of $\|\nabla f\|_2$) in the direction of its gradient (and decreases the most in the opposite direction of its gradient).

Geometrically, the restriction of the function f on a parametrized curve ϕ does not change if and only if the curve lies in a level

$$S_c = \{\vec{x}\colon f(\vec{x}) = c\}$$

of the function, which is typically an $(n-1)$-dimensional hypersurface in $\mathbb{R}^n$. The *tangent space* of the level hypersurface consists of the tangent vectors of all the curves inside the level

$$\begin{aligned} TS_c &= \{\phi'(t)\colon \phi(t) \in S_c\} = \{\phi'(t)\colon f(\phi(t)) = c\} \\ &= \{\phi'(t)\colon D_{\phi'(t)}f = 0\} = \{\vec{v}\colon \nabla f \cdot \vec{v} = 0\}. \end{aligned}$$

Therefore the tangent space of the level is the vector subspace orthogonal to the gradient.

On the other hand, the function changes only if one jumps from one level S_c to a different level $S_{c'}$. The change is fastest (i.e., the directional derivative is biggest) when one moves in the direction orthogonal to the levels, which is the direction of ∇f. Moreover, the magnitude $\|\nabla f\|_2$ of the gradient measures how fast this fastest jump is. Therefore $\|\nabla f\|_2$ measures how much the levels are "squeezed" together.
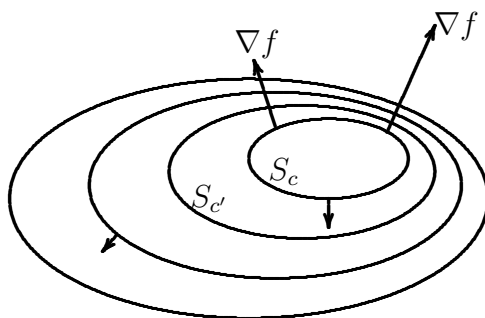
Figure 6.1: gradient and level

Example 6.1.13. The gradient of $f = x^2 + y^2 + z^2$ is $\nabla f = (2x, 2y, 2z)$. The derivative at $(1, 1, 1)$ in the direction $(2, 1, 2)$ is

$$D_{\frac{1}{3}(2,1,2)}f(1,1,1) = \nabla f(1,1,1) \cdot \frac{1}{3}(2,1,2) = \frac{1}{3}(2,2,2) \cdot (2,1,2) = \frac{10}{3}.$$

Note the vector $(2, 1, 2)$ is divided by its length, so that only the direction counts.

Example 6.1.14. Suppose the derivative of f in the directions $(1, 1)$ and $(1, -1)$ are respectively 2 and -3. Then

$$D_{(1,1)}f = \sqrt{2}D_{\frac{1}{\sqrt{2}}(1,1)}f = 2\sqrt{2},$$

$$D_{(1,-1)}f = \sqrt{2}D_{\frac{1}{\sqrt{2}}(1,-1)}f = -3\sqrt{2},$$

and

$$f_x = D_{(1,0)}f = \frac{1}{2}(D_{(1,1)}f + D_{(1,-1)}f) = \frac{-1}{\sqrt{2}},$$

$$f_y = D_{(0,1)}f = \frac{1}{2}(D_{(1,1)}f - D_{(1,-1)}f) = \frac{5}{\sqrt{2}}.$$

Example 6.1.15. We try to characterize continuously differentiable functions $f(x, y)$ satisfying $xf_x = yf_y$. Note that the condition is the same as $\nabla f \perp (x, -y)$, which is equivalent to ∇f being parallel to $(y, x) = \nabla(xy)$. This means that the levels of f and the levels of xy are tangential everywhere. Therefore we expect the levels of $f(x, y)$ and xy to be the same, and we should have $f(x, y) = h(xy)$ for some continuously differentiable $h(t)$.

For the rigorous argument, let $f(x, y) = g(x, xy)$, or $g(x, z) = f\left(x, \frac{z}{x}\right)$ (we define such g in case $x \neq 0$, and we define $f(x, y) = g(xy, y)$ in case $y \neq 0$). Then

$$g_x = f_x - \frac{z}{x^2}f_y = \frac{1}{x}\left(xf_x - \frac{z}{x}f_y\right) = 0.$$

This shows that g is independent of the first variable, and we have $f(x, y) = h(xy)$.

A vast extension of the discussion is the theory of functional dependence. See Exercises 6.2.20 through 6.2.25.

Exercise 6.1.29. Compute directional derivatives.

1. $4xy^3 + 5x^3y - 3y^2$ at $(0, 1)$, in direction $(1, 2)$.

2. $\arctan \dfrac{y}{x}$ at $(1, 1)$, in direction $(1, -1)$.

3. $xy + yz + zx$ at $(1, 2, 3)$, in direction $(2, 2, 1)$.

4. $a + b_1x_1 + b_2x_2 + \cdots + b_nx_n$ at $(1, 1, \ldots, 1)$, in direction $(v_1, v_2, \ldots, v_n)$.

5. $x_1^2 + x_2^2 + \cdots + x_n^2$ at $(1, -1, \ldots, (-1)^n)$, in direction $(1, 1, \ldots, 1)$.

Exercise 6.1.30. Suppose $Df = 1$ in direction $(1, 2, 2)$, $Df = \sqrt{2}$ in direction $(0, 1, -1)$, $Df = 3$ in direction $(0, 0, 1)$. Find the gradient of f.

Exercise 6.1.31. The curves $\phi(t) = (t, t^2)$ and $\psi(t) = (t^2, t)$ intersect at $(1, 1)$. Suppose the derivatives of $f(x, y)$ along the two curves are respectively 2 and 3 at $(1, 1)$, what is the gradient of f at $(1, 1)$.

Exercise 6.1.32. Suppose $\vec{u}_1$, $\vec{u}_2$, $\ldots$, $\vec{u}_n$ is an orthonormal basis. Prove that

$$\nabla f = (D_{\vec{u}_1} f)\vec{u}_1 + (D_{\vec{u}_2} f)\vec{u}_2 + \cdots + (D_{\vec{u}_n} f)\vec{u}_n.$$

Exercise 6.1.33. Express the gradient of $f(x, y)$ in terms of the partial derivatives f_r, f_θ and the directions $\vec{e}_r = \dfrac{\vec{x}_r}{\|\vec{x}_r\|_2} = (\cos\theta, \sin\theta)$, $\vec{e}_\theta = \dfrac{\vec{x}_\theta}{\|\vec{x}_\theta\|_2} = (-\sin\theta, \cos\theta)$ in the polar coordinates. Exercise 6.2.16 is a vast generalization.

Exercise 6.1.34. Find continuously differentiable functions satisfying the equations.

1. $af_x = bf_y$.

2. $yf_x = xf_y$.

3. $xf_x + yf_y = 0$.

4. $(x + y)f_x + (x - y)f_y = 0$.

5. $f_x = f_y = f_z$.

6. $xf_x = yf_y = zf_z$.

Exercise 6.1.35. Suppose $\vec{a}$ is a nonzero vector in $\mathbb{R}^n$. Suppose $\vec{b}_1$, $\vec{b}_2$, $\ldots$, $\vec{b}_{n-1}$ are linearly independent vectors orthogonal to $\vec{a}$. Prove that a function f on whole $\mathbb{R}^n$ (or on any open convex subset) satisfies $D_{\vec{a}} f = 0$ if and only if $f(\vec{x}) = g(\vec{b}_1 \cdot \vec{x}, \vec{b}_2 \cdot \vec{x}, \ldots, \vec{b}_{n-1} \cdot \vec{x})$ for some function g on $\mathbb{R}^{n-1}$.

The idea underlying the discussion of the directional derivative can be used to derive the following partial extension of the mean value theorem.

Proposition 6.1.3. *Suppose f is a function differentiable along the straight line connecting $\vec{a}$ and $\vec{b}$. Then there is $\vec{c}$ on the straight line, such that*

$$f(\vec{b}) - f(\vec{a}) = \nabla f(\vec{c}) \cdot (\vec{b} - \vec{a}).$$

Proof. The straight line connecting $\vec{a}$ to $\vec{b}$ is $\phi(t) = (1 - t)\vec{a} + t\vec{b}$ for $t \in [0, 1]$. The single variable function $f(\phi(t))$ is differentiable with

$$(f(\phi(t)))' = \nabla f(\phi(t)) \cdot \phi'(t) = \nabla f(\phi(t)) \cdot (\vec{b} - \vec{a}).$$

By applying the main value theorem to $f(\phi(t))$, we find $0 < c < 1$, such that

$$f(\vec{b}) - f(\vec{a}) = f(\phi(1)) - f(\phi(0)) = (f(\phi(t)))'_{t=c}(1 - 0) = \nabla f(\phi(c)) \cdot (\vec{b} - \vec{a}).$$

Let $\vec{c} = \phi(c)$. The proposition is proved.

□

Although the proposition may be extended to individual coordinates of a multivariable map, the choice of $\vec{c}$ may be different for different coordinates. The following is a more unified extension.

Proposition 6.1.4. *Suppose F is a map differentiable along the straight line connecting $\vec{a}$ and $\vec{b}$. Then there is $\vec{c}$ on the straight line, such that*

$$\|F(\vec{b}) - F(\vec{a})\|_2 \le \|F'(\vec{c})\| \|\vec{b} - \vec{a}\|_2.$$

where the norm $\|F'(\vec{c})\|$ is with respect to the Euclidean norms.

Proof. Fix any vector $\vec{v}$ and consider the function $f(\vec{x}) = F(\vec{x}) \cdot \vec{v}$. Then $\nabla f(\vec{x}) \cdot \vec{u} = F'(\vec{x})(\vec{u}) \cdot \vec{v}$. By Proposition 6.1.3, we have

$$|(F(\vec{b}) - F(\vec{a})) \cdot \vec{v}| = |\nabla f(\vec{c}) \cdot (\vec{b} - \vec{a})| = |F'(\vec{c})(\vec{b} - \vec{a}) \cdot \vec{v}|$$
$$\le \|F'(\vec{c})(\vec{b} - \vec{a})\|_2 \|\vec{v}\|_2 \le \|F'(\vec{c})\| \|\vec{b} - \vec{a}\|_2 \|\vec{v}\|_2,$$

By taking $\vec{v} = F(\vec{b}) - F(\vec{a})$ at the beginning, we get $\|F(\vec{b}) - F(\vec{a})\|_2 \le \|F'(\vec{c})\| \|\vec{b} - \vec{a}\|_2$.

□

As a consequence of the mean value theorem, Theorem 2.2.3 can be extended from single to multivariable.

Proposition 6.1.5. *Suppose a differentiable map on a path connected open subset has zero derivative everywhere. Then the map is a constant map.*

Proof. For any $\vec{x} \in U$, there is $\epsilon > 0$, such that $\|\vec{y} - \vec{x}\| < \epsilon$ implies $\vec{y} \in U$. Then the straight line connecting $\vec{x}$ and $\vec{y}$ lies in U. By Proposition 6.1.4 and $\nabla F'(\vec{c}) = 0$, we have $F(\vec{y}) = F(\vec{x})$. Thus F is constant near $\vec{x}$.

For any two points $\vec{x}, \vec{y} \in U$, there is a path $\phi(t)$, $t \in [a, b]$, connecting $\vec{x}$ to $\vec{y}$. For any $t_0 \in [a, b]$, there is $\epsilon > 0$, such that $\|\vec{y} - \phi(t_0)\| < \epsilon$ implies $\vec{y} \in U$. Since ϕ is continuous, there is $\delta > 0$, such that $|t - t_0| < \delta$ implies $\|\phi(t) - \phi(t_0)\| < \epsilon$. Then $F(\phi(t))$ is constant on $(t_0 - \mu, t_0 + \mu)$. In particular, we have $(F(\phi(t)))' = 0$ for any $t \in [a, b]$. For the map $F \circ \phi \colon [a, b] \to \mathbb{R}^n$, the vanishing of the derivative on (a, b) implies that $F(\phi(a)) = F(\phi(b))$, which is $F(\vec{x}) = F(\vec{y})$. This completes the proof that F is constant on U.

□

Exercise 6.1.36. What can you say about Proposition 6.1.4 if the norms are different from the Euclidean norm?

Exercise 6.1.37. What can you say about Proposition 6.1.5 if only some partial derivatives are constantly zero? The answer makes Example 6.1.15 more rigorous.

6.2 Inverse and Implicit Function

The inverse and implicit function theorems basically say that if the similar inverse and implicit problems can be solved for the linear approximation, then the problems can be solved for the multivariable map itself. The result has geometrical interpretation related to the definition of hypersurface.

6.2.1 Inverse Differentiation

Suppose a single variable function $f(x)$ has continuous derivative near x_0. If $f'(x_0) \neq 0$, then $f'(x)$ is either positive for all x near x_0 or negative for all x near x_0. Thus $f(x)$ is monotone and is therefore invertible near x_0. The multivariable extension is the following.

Theorem 6.2.1 (Inverse Function Theorem). *Suppose $F \colon \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable near $\vec{x}_0$. Suppose the derivative $F'(\vec{x}_0)$ is an invertible linear map. Then there is an open subset U containing $\vec{x}_0$, such that $F(U)$ is also open, $F \colon U \to F(U)$ is invertible, $F^{-1} \colon F(U) \to U$ is differentiable, and $(F^{-1})'(\vec{y}) = (F'(\vec{x}))^{-1}$ when $\vec{y} = F(\vec{x})$.*

The theorem basically says that if the linear approximation of a map is invertible, then the map is locally also invertible. Moreover, the differential $d\vec{x} = (F^{-1})'d\vec{y}$ is the solution of the equation $d\vec{y} = F'd\vec{x}$. Exercise 2.1.24 shows that the continuity assumption cannot be dropped from the theorem.

Proof. For any $\vec{y}$ near $\vec{y}_0 = F(\vec{x}_0)$, we wish to find $\vec{x}$ near $\vec{x}_0$ satisfying $F(\vec{x}) = \vec{y}$. To solve the problem, we approximate the map $F(\vec{x})$ by the linear map $L_0(\vec{x}) = F(\vec{x}_0) + F'(\vec{x}_0)(\vec{x} - \vec{x}_0)$ and solve the similar equation $L_0(\vec{x}) = \vec{y}$. The solution $\vec{x}_1$ of the approximate linear equation satisfies

$$\vec{y} = F(\vec{x}_0) + F'(\vec{x}_0)(\vec{x}_1 - \vec{x}_0).$$

Although not exactly equal to the solution $\vec{x}$ that we are looking for, $\vec{x}_1$ is often closer to $\vec{x}$ than $\vec{x}_0$ (see Figure 6.2). So we repeat the process by using $L_1(\vec{x}) = F(\vec{x}_1) + F'(\vec{x}_0)(\vec{x} - \vec{x}_1)$ to approximate F near $\vec{x}_1$ and solve the similar approximate linear equation $L_1(\vec{x}) = \vec{y}$ to get solution $\vec{x}_2$. Continuing the process with the linear approximation $L_2(\vec{x}) = F(\vec{x}_2) + F'(\vec{x}_0)(\vec{x} - \vec{x}_2)$ of F near $\vec{x}_2$ and so on, we get an inductively defined sequence $\{\vec{x}_k\}$ by

$$\vec{y} = L_k(\vec{x}_{k+1}) = F(\vec{x}_k) + F'(\vec{x}_0)(\vec{x}_{k+1} - \vec{x}_k). \tag{6.2.1}$$

To prove the expectation that the sequence $\{\vec{x}_k\}$ converges to the solution $\vec{x}$ of the equation $F(\vec{x}) = \vec{y}$, we substitute $F(\vec{x}_k)$ in (6.2.1) by using the approximation

$$F(\vec{x}) = F(\vec{x}_0) + F'(\vec{x}_0)(\vec{x} - \vec{x}_0) + R(\vec{x}) \tag{6.2.2}$$

of F near $\vec{x}_0$. We get

$$\begin{aligned} \vec{y} &= F(\vec{x}_0) + F'(\vec{x}_0)(\vec{x}_k - \vec{x}_0) + R(\vec{x}_k) + F'(\vec{x}_0)(\vec{x}_{k+1} - \vec{x}_k) \\ &= F(\vec{x}_0) + F'(\vec{x}_0)(\vec{x}_{k+1} - \vec{x}_0) + R(\vec{x}_k). \end{aligned} \tag{6.2.3}$$
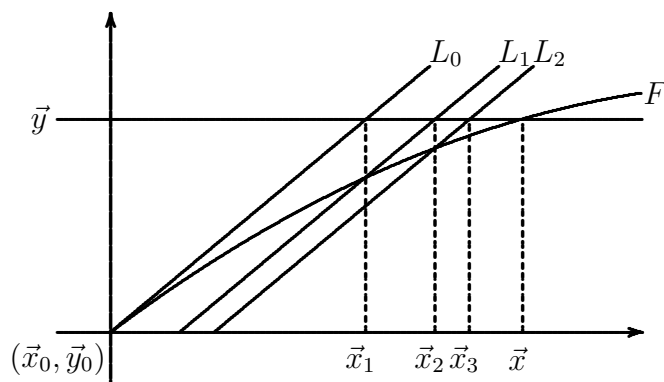
Figure 6.2: find $\vec{x}$ satisfying $F(\vec{x}) = \vec{y}$

Write the formula (6.2.3) with $k - 1$ in place of k and taking the difference with (6.2.3), we get

$$F'(\vec{x}_0)(\vec{x}_{k+1} - \vec{x}_k) + R(\vec{x}_k) - R(\vec{x}_{k-1}) = \vec{0}. \qquad (6.2.4)$$

This gives us a relation between $\vec{x}_{k+1} - \vec{x}_k$ and $\vec{x}_k - \vec{x}_{k-1}$.

By the continuity of the differentiation at $\vec{x}_0$, for any $\epsilon > 0$, there is $\delta > 0$, such that F is defined on the ball $B(\vec{x}_0, \delta)$ and

$$\|\vec{x} - \vec{x}_0\| < \delta \implies \|R'(\vec{x})\| = \|F'(\vec{x}) - F'(\vec{x}_0)\| < \epsilon. \qquad (6.2.5)$$

If we know $\vec{x}_k, \vec{x}_{k-1} \in B(\vec{x}_0, \delta)$, then by (6.2.4), (6.2.5) and Proposition 6.1.4,

$$\begin{aligned}
\|\vec{x}_{k+1} - \vec{x}_k\| &= \|F'(\vec{x}_0)^{-1}(R(\vec{x}_k) - R(\vec{x}_{k-1}))\| \\
&\leq \|F'(\vec{x}_0)^{-1}\| \|R(\vec{x}_k) - R(\vec{x}_{k-1})\| \\
&\leq \epsilon \|F'(\vec{x}_0)^{-1}\| \|\vec{x}_k - \vec{x}_{k-1}\|.
\end{aligned} \qquad (6.2.6)$$

Note that the estimation in Proposition 6.1.4 holds for the Euclidean norm only. Therefore this proof will be restricted to the Euclidean norm for vectors. Now fix some $0 < \alpha < 1$. By choosing $\epsilon < \dfrac{\alpha}{\|F'(\vec{x}_0)^{-1}\|}$ at the beginning, we get $\|\vec{x}_{k+1} - \vec{x}_k\| \leq \alpha \|\vec{x}_k - \vec{x}_{k-1}\|$. This further implies

$$\|\vec{x}_{k+1} - \vec{x}_k\| \leq \alpha^k \|\vec{x}_1 - \vec{x}_0\|.$$

This further tells us that, if we know $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k \in B(\vec{x}_0, \delta)$, then for any $0 \leq j < i \leq k + 1$, we have

$$\|\vec{x}_i - \vec{x}_j\| \leq (\alpha^j + \alpha^{j+1} + \cdots + \alpha^{i-1})\|\vec{x}_1 - \vec{x}_0\| < \frac{\alpha^j}{1 - \alpha}\|\vec{x}_1 - \vec{x}_0\|. \quad (6.2.7)$$

This will imply that $\{\vec{x}_k\}$ is a Cauchy sequence and therefore converges. Taking the limit of (6.2.1), we further see that the limit $\vec{x} = \lim_{k \to \infty} \vec{x}_k$ satisfies $\vec{y} = F(\vec{x})$.

How do we know $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k \in B(\vec{x}_0, \delta)$, so that the estimations (6.2.6) and (6.2.7) hold? The estimation (6.2.7) tells us that if $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k \in B(\vec{x}_0, \delta)$, then

$$\|\vec{x}_{k+1} - \vec{x}_0\| < \frac{1}{1 - \alpha}\|\vec{x}_1 - \vec{x}_0\| \leq \frac{1}{1 - \alpha}\|F'(\vec{x}_0)^{-1}\| \|\vec{y} - F(\vec{x}_0)\|.$$

The right side will be $< \delta$ if we assume $\vec{y}$ satisfies

$$\|\vec{y} - \vec{y}_0\| = \|\vec{y} - F(\vec{x}_0)\| < \frac{(1-\alpha)\delta}{\|F'(\vec{x}_0)^{-1}\|}. \qquad (6.2.8)$$

Under the assumption, we have $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k, \vec{x}_{k+1} \in B(\vec{x}_0, \delta)$ and the inductive estimation can continue.

In summary, we should make the following rigorous setup at the very beginning of the proof: Assume $0 < \alpha < 1$ is fixed, $0 < \epsilon < \dfrac{\alpha}{\|F'(\vec{x}_0)^{-1}\|}$ is given, and $\delta > 0$ is found so that (6.2.5) holds. Then for any $\vec{y}$ satisfying (6.2.8), we may inductively construct $\vec{x}_k$ by the formula (6.2.1). Moreover, an inductive argument shows that the sequence lies in $B(\vec{x}_0, \delta)$. Therefore the limit $\vec{x} = \lim \vec{x}_k$, which solves the equation $F(\vec{x}) = \vec{y}$, still satisfies

$$\|\vec{x} - \vec{x}_0\| = \lim \|\vec{x}_{k+1} - \vec{x}_0\| \leq \frac{1}{1-\alpha}\|F'(\vec{x}_0)^{-1}\|\|\vec{y} - F(\vec{x}_0)\| < \delta.$$

Geometrically, this means

$$F(B(\vec{x}_0, \delta)) \supset B\left(F(\vec{x}_0), \frac{(1-\alpha)\delta}{\|F'(\vec{x}_0)^{-1}\|}\right). \qquad (6.2.9)$$

Note that so far we have only used the fact that F' is continuous at $\vec{x}_0$ and $F'(\vec{x}_0)$ is invertible.

Since F' is continuous and $F'(\vec{x}_0)$ is invertible, by choosing δ even smaller, we may further assume that F' is actually invertible on the ball $U = B(\vec{x}_0, \delta)$. We show that the image $F(U)$ is open. A point in $F(U)$ is of the form $F(\vec{z})$ for some $\vec{z} \in U$. For $F(\vec{x})$ near $\vec{z}$, we find α_z, ϵ_z and sufficiently small δ_z as above, with $B(\vec{z}, \delta_z) \subset U$. Then for $\delta'_z = \dfrac{(1-\alpha_z)\delta_z}{\|F'(\vec{z})^{-1}\|}$, we have by (6.2.9) that

$$B(F(\vec{z}), \delta'_z) \subset F(B(\vec{z}, \delta_z)) \subset F(U).$$

This proves that $F(U)$ is open.

Next we prove that F is one-to-one on the ball U, so that U can be used as the open subset in the statement of the theorem. By the approximation (6.2.2) and the estimation (6.2.5), for any $\vec{x}, \vec{x}' \in U = B(\vec{x}_0, \delta)$,

$$\begin{aligned}
\|F(\vec{x}) - F(\vec{x}')\| &= \|F'(\vec{x}_0)(\vec{x} - \vec{x}') + R(\vec{x}) - R(\vec{x}')\| \\
&\geq \|F'(\vec{x}_0)(\vec{x} - \vec{x}')\| - \epsilon\|\vec{x} - \vec{x}'\| \\
&\geq \left(\frac{1}{\|F'(\vec{x}_0)^{-1}\|} - \epsilon\right)\|\vec{x} - \vec{x}'\| \\
&\geq \frac{1-\alpha}{\|F'(\vec{x}_0)^{-1}\|}\|\vec{x} - \vec{x}'\|. \qquad (6.2.10)
\end{aligned}$$

Since $\alpha < 1$, $\vec{x} \neq \vec{x}'$ implies $F(\vec{x}) \neq F(\vec{x}')$.

It remains to prove the differentiability of the inverse map. By (6.2.2), for $\vec{y} \in F(U)$ and $\vec{x} = F^{-1}(\vec{y})$, we have

$$\vec{x} = \vec{x}_0 + F'(\vec{x}_0)^{-1}(\vec{y} - \vec{y}_0) - F'(\vec{x}_0)^{-1}R(\vec{x}).$$

Then by $R(\vec{x}_0) = \vec{0}$, $\|R'(\vec{x})\| < \epsilon$ along the line connecting $\vec{x}_0$ and $\vec{x}$, and Proposition 6.1.4, we get

$$
\begin{aligned}
\|F'(\vec{x}_0)^{-1}R(\vec{x})\| &\leq \|F'(\vec{x}_0)^{-1}\|\|R(\vec{x}) - R(\vec{x}_0)\| \\
&\leq \epsilon\|F'(\vec{x}_0)^{-1}\|\|\vec{x} - \vec{x}_0\| \\
&\leq \frac{\epsilon\|F'(\vec{x}_0)^{-1}\|^2}{1 - \alpha}\|\vec{y} - \vec{y}_0\|,
\end{aligned}
$$

where the last inequality makes use of (6.2.10). The estimation shows that $\vec{x}_0 + F'(\vec{x}_0)^{-1}(\vec{y} - \vec{y}_0)$ linearly approximates $F^{-1}(\vec{y})$. Therefore the inverse map is differentiable at $\vec{x}_0$, with $(F^{-1})'(\vec{x}_0) = F'(\vec{x}_0)^{-1}$.

$\square$

Note that most of the proof only makes use of the assumption that the derivative F' is continuous at $\vec{x}_0$. The assumption implies that the image of a ball around $\vec{x}_0$ contains a ball around $F(\vec{x}_0)$. Based on this fact, the continuity of the derivative everywhere is (only) further used to conclude that the image of open subsets must be open.
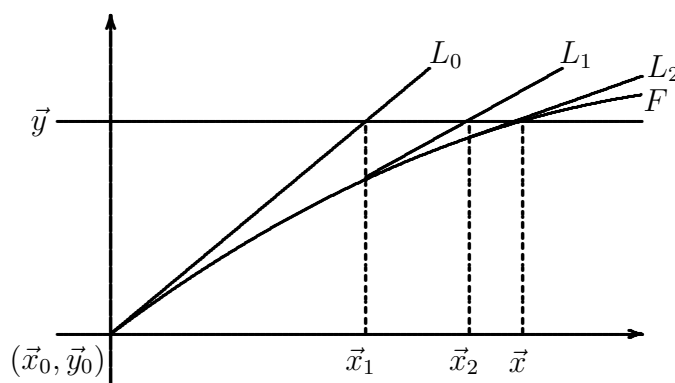
Figure 6.3: Newton's method

The proof actually includes a method for computing the inverse function. In fact, it appears to be more suitable to use

$$
L_n(\vec{x}) = F(\vec{x}_k) + F'(\vec{x}_k)(\vec{x} - \vec{x}_k)
$$

to approximate F at $\vec{x}_k$ (with F' at $\vec{x}_k$ instead of at $\vec{x}_0$). The method is quite effective in case the dimension is small (so that $F'(\vec{x}_k)^{-1}$ is easier to compute). In particular, for a single variable function $f(x)$, the solution to $f(x) = 0$ can be found by starting from x_0 and successively constructing

$$
x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.
$$

This is Newton's method.

Example 6.2.1. In Example 6.1.8, we computed the differentials of the cartesian coordinate in terms of the polar coordinate. The Jacobian matrix is invertible away from the origin, so that the polar coordinate can also be expressed locally

in terms of the cartesian coordinate. In fact, the map $(r, \theta) \to (x, y)$ is invertible for $(r, \theta) \in (0, \infty) \times (a, a + 2\pi)$. By solving the system $dx = \cos \theta dr - r \sin \theta d\theta$, $dy = \sin \theta dr + r \cos \theta d\theta$, we get

$$dr = \cos \theta dx + \sin \theta dy, \quad d\theta = -r^{-1} \sin \theta dx + r^{-1} \cos \theta dy.$$

By the inverse function theorem, the coefficients form the Jacobian matrix

$$\frac{\partial(r, \theta)}{\partial(x, y)} = \begin{pmatrix} \cos \theta & \sin \theta \\ -r^{-1} \sin \theta & r^{-1} \cos \theta \end{pmatrix}.$$

Exercise 6.2.1. Use the differential in Exercise 6.1.21 to find the differential of the change from the cartesian coordinate (x, y, z) to the spherical coordinate (r, ϕ, θ).

Exercise 6.2.2. Suppose $x = e^u + u \cos v$, $y = e^u + u \sin v$. Find the places where u and v can be locally expressed as differentiable functions of x and y and then compute $\dfrac{\partial(u, v)}{\partial(x, y)}$.

Exercise 6.2.3. Find the places where z can be locally expressed as a function of x and y and then compute z_x and z_y.

1. $x = s + t$, $y = s^2 + t^2$, $z = s^3 + t^3$.

2. $x = e^{s+t}$, $y = e^{s-t}$, $z = st$.

Exercise 6.2.4. Change the partial differential equation $(x + y)u_x - (x - y)u_y = 0$ to an equation with respect to the polar coordinate (r, θ).

6.2.2 Implicit Differentiation

Suppose a continuous function $f(x, y)$ satisfies $f(x_0, y_0) = 0$ and has continuous partial derivative f_y near (x_0, y_0). Assume $f_y(x_0, y_0) > 0$. Then $f_y(x, y) > 0$ for (x, y) near (x_0, y_0). Thus for any fixed x, $f(x, y)$ is strictly increasing in y. In particular, for some small $\epsilon > 0$, we have $f(x_0, y_0 + \epsilon) > f(x_0, y_0) = 0$ and $f(x_0, y_0 - \epsilon) < f(x_0, y_0) = 0$. By the continuity in x, there is $\delta > 0$, such that $f(x, y_0 + \epsilon) > 0$ and $f(x, y_0 - \epsilon) < 0$ for any $x \in (x_0 - \delta, x_0 + \delta)$. Now for any fixed $x \in (x_0 - \delta, x_0 + \delta)$, $f(x, y)$ is strictly increasing in y and has different signs at $y_0 + \epsilon$ and $y_0 - \epsilon$. Therefore there is a unique $y \in (y_0 - \epsilon, y_0 + \epsilon)$ satisfying $f(x, y) = 0$.

We say y is an *implicit function* of x because the function is only implicitly given by the equation $f(x, y) = 0$.

The argument shows that if $f(x_0, y_0) = 0$ and $f_y(x_0, y_0) \neq 0$, plus some continuity condition, then the equation $f(x, y) = 0$ can be solved to define a function $y = y(x)$ near (x_0, y_0). If f is differentiable at (x_0, y_0), then the equation $f(x, y) = 0$ is approximated by the linear equation

$$f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) = 0.$$

The assumption $f_y(x_0, y_0) \neq 0$ makes it possible to solve the linear equation and get $y = y_0 - \dfrac{f_x(x_0, y_0)}{f_y(x_0, y_0)}(x - x_0)$. So the conclusion is that if the linear
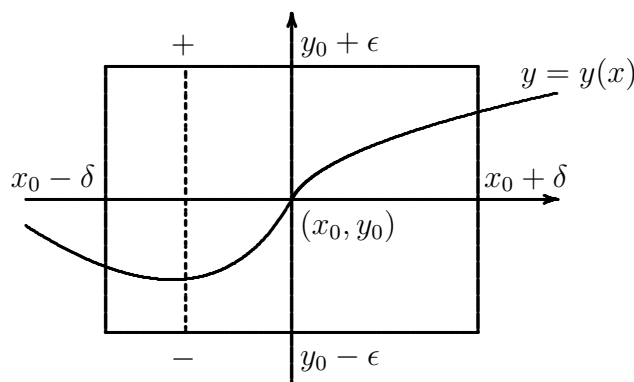
Figure 6.4: implicit function

approximation implicitly defines a function, then the original equation also implicitly defines a function.

In general, consider a differentiable map $F \colon \mathbb{R}^n \times \mathbb{R}^m = \mathbb{R}^{m+n} \to \mathbb{R}^m$. We have

$$
\begin{aligned}
F'(\vec{x}_0, \vec{y}_0)(\vec{u}, \vec{v}) &= F'(\vec{x}_0, \vec{y}_0)(\vec{u}, \vec{0}) + F'(\vec{x}_0, \vec{y}_0)(\vec{0}, \vec{v}) \\
&= F_{\vec{x}}(\vec{x}_0, \vec{y}_0)(\vec{u}) + F_{\vec{y}}(\vec{x}_0, \vec{y}_0)(\vec{v}),
\end{aligned}
$$

where $F_{\vec{x}}(\vec{x}_0, \vec{y}_0) \colon \mathbb{R}^n \to \mathbb{R}^m$ and $F_{\vec{y}}(\vec{x}_0, \vec{y}_0) \colon \mathbb{R}^m \to \mathbb{R}^m$ are linear transforms and generalize the partial derivatives. In terms of the Jacobian matrix, F' can be written as the block matrix $(F_{\vec{x}}(\vec{x}_0, \vec{y}_0) \ \ F_{\vec{y}}(\vec{x}_0, \vec{y}_0))$.

The equation $F(\vec{x}, \vec{y}) = \vec{0}$ is approximated by the linear equation

$$
F_{\vec{x}}(\vec{x}_0, \vec{y}_0)(\vec{x} - \vec{x}_0) + F_{\vec{y}}(\vec{x}_0, \vec{y}_0)(\vec{y} - \vec{y}_0) = \vec{0}. \tag{6.2.11}
$$

The linear equation defines $\vec{y}$ as a linear map of $\vec{x}$ (in other words, the linear equation has a unique solution $\vec{x}$ for each $\vec{y}$) if and only if the linear transform $F_{\vec{y}}(\vec{x}_0, \vec{y}_0)$ is invertible. We expect the condition is also the condition for the original equation $F(\vec{x}, \vec{y}) = \vec{0}$ to locally define $\vec{y}$ as a map of $\vec{x}$. The expectation is summarized in the next result.

Theorem 6.2.2 (Implicit Function Theorem). *Suppose $F \colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ is continuously differentiable near $(\vec{x}_0, \vec{y}_0)$ and satisfies $F(\vec{x}_0, \vec{y}_0) = \vec{0}$. Suppose the derivative $F_{\vec{y}}(\vec{x}_0, \vec{y}_0) \colon \mathbb{R}^m \to \mathbb{R}^m$ in $\vec{y}$ is an invertible linear map. Then there is an open subset U containing $\vec{x}_0$ and a map $G \colon U \subset \mathbb{R}^n \to \mathbb{R}^m$, such that G is continuously differentiable near $\vec{x}_0$, $G(\vec{x}_0) = \vec{y}_0$ and $F(\vec{x}, G(\vec{x})) = \vec{0}$. Moreover, the derivative of G is given by $G' = -F_{\vec{y}}^{-1} F_{\vec{x}}$.*

The solution $\Delta \vec{y} = -F_{\vec{y}}^{-1} F_{\vec{x}} \Delta \vec{x}$ to the linear approximation equation (6.2.11) should be the linear approximation of the solution $\vec{y} = G(\vec{x})$ of the original equation $F(\vec{x}, \vec{y}) = \vec{0}$. In other words, the linear transform $-F_{\vec{y}}^{-1} F_{\vec{x}}$ should be the derivative of G, or the differential $d\vec{y} = G' d\vec{x}$ is the solution of the equation

$$
dF = F_{\vec{x}} d\vec{x} + F_{\vec{y}} d\vec{y} = 0.
$$

This is the last claim of the theorem.

Another way of understanding the formula $G' = -F_{\vec{y}}^{-1} F_{\vec{x}}$ is that vectors of the form
$$L(\Delta \vec{x}) = (\Delta \vec{x}, -F_{\vec{y}}^{-1} F_{\vec{x}} \Delta \vec{x})$$
are solutions of the linear approximation equation
$$F'(\vec{x}_0, \vec{y}_0)(\Delta \vec{x}, \Delta \vec{y}) = F_{\vec{x}}(\vec{x}_0, \vec{y}_0)\Delta \vec{x} + F_{\vec{y}}(\vec{x}_0, \vec{y}_0)\Delta \vec{y} = \vec{0}.$$

Proof. The map $H(\vec{x}, \vec{y}) = (\vec{x}, F(\vec{x}, \vec{y})) \colon \mathbb{R}^{m+n} \to \mathbb{R}^{m+n}$ has continuous derivative
$$H'(\vec{u}, \vec{v}) = (\vec{u}, F_{\vec{x}}(\vec{u}) + F_{\vec{y}}(\vec{v}))$$
near $\vec{x}_0$. The invertibility of $F_{\vec{y}}(\vec{x}_0, \vec{y}_0)$ implies the invertibility of $H'(\vec{x}_0, \vec{y}_0)$. Then by the inverse function theorem, H has inverse $H^{-1} = (S, T)$ that is continuously differentiable near $H(\vec{x}_0, \vec{y}_0) = (\vec{x}_0, \vec{0})$, where $S \colon \mathbb{R}^{m+n} \to \mathbb{R}^n$ and $T \colon \mathbb{R}^{m+n} \to \mathbb{R}^m$. Since $HH^{-1} = (S, F(S, T))$ is the identity, we have $S(\vec{x}, \vec{z}) = \vec{x}$ and $F(\vec{x}, T(\vec{x}, \vec{z})) = \vec{z}$. Then
$$F(\vec{x}, \vec{y})) = \vec{0} \iff H(\vec{x}, \vec{y}) = (\vec{x}, \vec{0})$$
$$\iff (\vec{x}, \vec{y}) = (S(\vec{x}, \vec{0}), T(\vec{x}, \vec{0})) = (\vec{x}, T(\vec{x}, \vec{0})).$$

Therefore for $\vec{y} = G(\vec{x}) = T(\vec{x}, \vec{0})$ is exactly the solution of $F(\vec{x}, \vec{y}) = \vec{0}$.

Finally, for the derivative of G, we differentiate the equality $F(\vec{x}, G(\vec{x})) = \vec{0}$ and get
$$F_{\vec{x}} + F_{\vec{y}}G' = 0.$$
Since $F_{\vec{y}}$ is invertible, we get $G' = -F_{\vec{y}}^{-1} F_{\vec{x}}$.
$$\qquad \square$$

Example 6.2.2. The unit sphere $S^2 \subset \mathbb{R}^3$ is given by the equation $f(x, y, z) = x^2 + y^2 + z^2 = 1$. By $f_z = 2z$ and the implicit function theorem, z can be expressed as a function of (x, y) near any place where $z \neq 0$. In fact, the expression is $z = \pm\sqrt{1 - x^2 - y^2}$, where the sign is the same as the sign of z. By solving the equation
$$df = 2xdx + 2ydy + 2zdz = 0,$$
we get $dz = -\dfrac{x}{z}dx - \dfrac{y}{z}dy$. Therefore $z_x = -\dfrac{x}{z}$ and $z_y = -\dfrac{y}{z}$.

Exercise 6.2.5. Find the places where the map is implicitly defined and compute the derivatives of the map.

1. $x^3 + y^3 - 3axy = 0$, find $\dfrac{dy}{dx}$.

2. $x^2 + y^2 + z^2 - 4x + 6y - 2z = 0$, find $\dfrac{\partial z}{\partial(x, y)}$ and $\dfrac{\partial x}{\partial(y, z)}$.

3. $x^2 + y^2 = z^2 + w^2$, $x + y + z + w = 1$, find $\dfrac{\partial(z, w)}{\partial(x, y)}$ and $\dfrac{\partial(x, w)}{\partial(y, z)}$.

4. $z = f(x + y + z, xyz)$, find $\dfrac{\partial z}{\partial(x, y)}$ and $\dfrac{\partial x}{\partial(y, z)}$.

Exercise 6.2.6. Verify that implicitly defined function satisfy the partial differential equation.

1. $f(x - az, y - bz) = 0$, $z = z(x, y)$ satisfies $az_x + bz_y = 1$.

2. $x^2 + y^2 + z^2 = yf\left(\dfrac{z}{y}\right)$, $z = z(x, y)$ satisfies $(x^2 - y^2 - z^2)z_x + 2xyz_y = 4xz$.

Exercise 6.2.7. In solving the equation $f(x, y) = 0$ for two variable functions, we did not assume anything about the partial derivative in x. Extend the discussion to the general multivariable case and point out what conclusion in the implicit function theorem may not hold.

6.2.3 Hypersurface

For a continuously differentiable map $F \colon \mathbb{R}^n \to \mathbb{R}^m$, the graph $(\vec{x}, F(\vec{x})) \colon \mathbb{R}^n \to \mathbb{R}^{m+n}$ is an *n-dimensional hypersurface* in $\mathbb{R}^{m+n}$. However, we should not insist that it is always the last m coordinates that can be written as a map of the first n coordinates. For example, the unit circle $S^1 \subset \mathbb{R}^2$ is the graph of a function of y in x near $(x_0, y_0) \in S^1$ if $y_0 \neq 0$, and the circle is the graph of a function of x in y if $x_0 \neq 0$. Thus an n-dimensional hypersurface in $\mathbb{R}^{m+n}$ is a subset such that near any point, the subset is the graph of a continuously differentiable map of some choice of m variables in terms of the other n variables.

There are generally two ways of specifying a hypersurface. For $n \leq m$, the image of a continuously differentiable map $F \colon \mathbb{R}^n \to \mathbb{R}^m$ may give an n-dimensional *parametrized hypersurface* in $\mathbb{R}^m$. For $n \geq m$, the preimage of a continuously differentiable map $F \colon \mathbb{R}^n \to \mathbb{R}^m$ may give an $(n - m)$-dimensional *level hypersurface* in $\mathbb{R}^n$.

A continuously differentiable map $F \colon \mathbb{R}^n \to \mathbb{R}^m$ is *regular* if the derivative $F'(\vec{x}) \colon \mathbb{R}^n \to \mathbb{R}^m$ is always an injective linear map (this necessarily implies $n \leq m$). For example, a parametrized curve $\phi(t)$ is regular if its tangent vector $\phi'(t)$ is not zero, and a parametrized surface $\sigma(u, v)$ is regular if the tangent vectors σ_u and σ_v are not parallel. From linear algebra, the injective linear map $F'(\vec{x})$ is invertible when projected to certain n coordinates in $\mathbb{R}^m$. Assume these n coordinates are the first n coordinates. Then $F(\vec{x}) = (G(\vec{x}), H(\vec{x}))$, where $G \colon \mathbb{R}^n \to \mathbb{R}^n$ and $H \colon \mathbb{R}^n \to \mathbb{R}^{m-n}$ are continuously differentiable and $G'(\vec{x})$ is invertible. By the inverse function theorem, G is locally invertible, and we can locally change the variable from $\vec{x}$ to $\vec{y} = G(\vec{x})$ and get $F(G^{-1}(\vec{y})) = (\vec{y}, H(G^{-1}(\vec{y})))$. In other words, by a local change of variable in $\mathbb{R}^n$, F becomes the graph of a continuously differentiable map $H \circ G^{-1} \colon \mathbb{R}^n \to \mathbb{R}^{m-n}$. Thus F specifies an n-dimensional hypersurface S in $\mathbb{R}^m$.

The *tangent space* $T_{F(\vec{x})}S$ of the hypersurface is the collection of tangent vectors of the curves $F(\phi(t))$ in the surface. By $F(\phi(t))' = F'(\phi(t))(\phi'(t))$ (linear transform $F'(\phi(t))$ applied to vector $\phi'(t)$), the tangent space is the image of the linear transform $F'(\vec{x})$

$$T_{F(\vec{x})}S = \{F'(\vec{x})(\vec{v}) \colon \vec{v} \in \mathbb{R}^n\}. \tag{6.2.12}$$

Example 6.2.3. For the parametrized sphere (5.1.19), we have

$$\frac{\partial(x, y, z)}{\partial(\phi, \theta)} = \begin{pmatrix} \cos\phi\cos\theta & -\sin\phi\sin\theta \\ \cos\phi\sin\theta & \sin\phi\cos\theta \\ -\sin\phi & 0 \end{pmatrix}.$$

The matrix has rank 2 as long as $\sin\phi \neq 0$ (i.e., $z = \cos\phi \neq \pm 1$). Therefore the parametrization is a regular surface away from the north and south poles. Moreover, if $\sin\phi \neq 0$ and $\cos\phi \neq 0$ (i.e., $z \neq 0$ or ± 1), then $\dfrac{\partial(x, y)}{\partial(\phi, \theta)}$ has full rank 2. This suggests that away from the two poles and the equator, the map $(\phi, \theta) \mapsto (x, y)$ is locally invertible and the surface can be reparametrized by (x, y). As a matter of fact, on the whole northern hemisphere (which means $0 < z < 1$), we have the explicit formula for the reparametrization

$$(\phi, \theta) = \left(\arcsin\sqrt{x^2 + y^2}, \arctan\frac{y}{x}\right), \quad z = \cos\arcsin\sqrt{x^2 + y^2} = \sqrt{1 - x^2 - y^2}.$$

The southern hemisphere can be similarly reparametrized.

Exercise 6.2.8. Study the reparametrization of the sphere (5.1.19) by two coordinates along the equator.

Exercise 6.2.9. For the parametrized torus (5.1.20), find the places where the surfaces are regular and then compute the partial derivatives of some coordinate in terms of the other two.

A vector $\vec{y}_0 \in \mathbb{R}^m$ is a *regular value* of a continuously differentiable map $F \colon \mathbb{R}^n \to \mathbb{R}^m$ if $F'(\vec{x}) \colon \mathbb{R}^n \to \mathbb{R}^m$ is a surjective linear map (this necessarily implies $n \geq m$) for all $\vec{x}$ satisfying $F(\vec{x}) = \vec{y}_0$ ($\vec{x}$ is a preimage of $\vec{y}_0$). From linear algebra, the surjective linear map $F'(\vec{x})$ is invertible when restricted to certain choice of m coordinates in $\mathbb{R}^n$ (by taking all the other $n - m$ coordinates to be 0). By taking these m coordinates to be $\mathbb{R}^m$ and the remaining $n - m$ coordinates to be $\mathbb{R}^n$ in the implicit function theorem, we find the equation $F(\vec{x}) = \vec{y}_0$ defines the graph of a continuously differentiable map that expresses the m coordinates in terms of the remaining $n - m$ coordinates. In particular, this shows that $F(\vec{x}) = \vec{y}_0$ is an $(n - m)$-dimensional hypersurface S in $\mathbb{R}^n$.

The *tangent space* $T_{\vec{x}}S$ is given by the derivative of the map produced by the implicit function theorem. The derivative of the map is computed from the linear approximation of the defining equation $F(\vec{x}) = \vec{y}_0$. Thus the tangent space

$$T_{\vec{x}}S = \{\vec{v} \in \mathbb{R}^n \colon F'(\vec{x})(\vec{v}) = \vec{0}\} \tag{6.2.13}$$

is the kernel of the linear transform $F'(\vec{x}) \colon \mathbb{R}^n \to \mathbb{R}^m$.

For a continuously differentiable function $f(\vec{x})$, a number y_0 is a regular value if $f(\vec{x}) = y_0$ implies the gradient $\nabla f \neq \vec{0}$ at $\vec{x}$ (or at least one partial derivative is nonzero). The equation $f(\vec{x}) = y_0$ specifies an $(n - 1)$-dimensional hypersurface, with the tangent hyperplane given by $\nabla f \cdot \vec{v} = 0$. In other words, the gradient ∇f is the *normal vector* of the tangent hyperplane.

In general, for a continuously differentiable map $F = (f_1, f_2, \ldots, f_m) \colon \mathbb{R}^n \to \mathbb{R}^m$, the derivative $F'(\vec{x}) \colon \mathbb{R}^n \to \mathbb{R}^m$ is surjective if and only if the gradients

$\nabla f_1, \nabla f_2, \ldots, \nabla f_m$ are *linearly independent*

$$c_1 \nabla f_1 + c_2 \nabla f_2 + \cdots + c_m \nabla f_m = 0 \implies c_1 = c_2 = \cdots = c_m = 0.$$

For a regular value $\vec{y}_0$, the hypersurface S defined by the equation $F(\vec{x}) = \vec{y}_0$ is the intersection of the $(n-1)$-dimensional hypersurfaces $f_i(\vec{x}) = y_{i0}$. The tangent space $T_{\vec{x}} S$ is the intersection of the tangent spaces $\nabla f_i \cdot \vec{v} = 0$ for the coordinate functions. The normal vectors $\nabla f_1, \nabla f_2, \ldots, \nabla f_m$ span the *normal space* of the hypersurface S at $\vec{x}$.

Example 6.2.4. The unit sphere $S^{n-1} = f^{-1}(1)$, where $f(\vec{x}) = \vec{x} \cdot \vec{x} = x_1^2 + x_2^2 + \cdots + x_n^2$. The number 1 is a regular value because $f(\vec{x}) = 1 \implies \vec{x} \neq \vec{0} \implies \nabla f = 2\vec{x} \neq \vec{0}$. As a linear map, the restriction of the derivative $\vec{v} \mapsto 2\vec{x} \cdot \vec{v}$ to the i-th coordinate is invertible if and only if $x_i \neq 0$. By the implicit function theorem, if $x_i \neq 0$, then x_i can be expressed as a function of $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ near $\vec{x}$. Indeed, the part of the sphere with $x_i \neq 0$ consists of the graphs of the functions $x_i = \pm\sqrt{1 - x_1^2 - \cdots - x_{i-1}^2 - x_{i+1}^2 - \cdots - x_n^2}$ (which graph $\vec{x}$ belongs to depends on the sign of its i-th coordinate). The whole sphere is then covered by these pieces of hypersurfaces, called *coordinate charts*.

Exercise 6.2.10. Find the regular values of the map and determine the tangent space of the preimage of regular value.

1. $f(x, y) = x^2 + y^2 + z^2 + xy + yz + zx + x + y + z$.

2. $F(x, y, z) = (x + y + z, xy + yz + zx)$.

Exercise 6.2.11. Find a regular value λ for xyz, so that the surface $xyz = \lambda$ is tangential to the ellipse $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c^2} = 1$ at some point.

Exercise 6.2.12. Prove that any sphere $x^2 + y^2 + z^2 = a^2$ and any cone $x^2 + y^2 = b^2 z^2$ are orthogonal at their intersections. Can you extend this to higher dimension?

Exercise 6.2.13. The space of $n \times n$ symmetric matrices can be identified with $R^{\frac{n(n+1)}{2}}$. By considering the map $F(X) = X^T X \colon \mathbb{R}^{n^2} \to \mathbb{R}^{\frac{n(n+1)}{2}}$ in Exercise 6.1.6, prove that the orthogonal matrices $O(n) = \{X \colon X^T X = I\}$ is an $\dfrac{n(n-1)}{2}$ dimensional hypersurface in $\mathbb{R}^{n^2}$.

6.2.4 Additional Exercise

Orthogonal Change of Variable

A change of variable $\vec{x} = F(\vec{y}) \colon \mathbb{R}^n \to \mathbb{R}^n$ is *orthogonal* if the vectors

$$\vec{x}_{y_1} = \frac{\partial \vec{x}}{\partial y_1}, \ \ \vec{x}_{y_2} = \frac{\partial \vec{x}}{\partial y_2}, \ \ \ldots, \ \ \vec{x}_{y_n} = \frac{\partial \vec{x}}{\partial y_n}$$

are orthogonal.

Exercise 6.2.14. Prove that an orthogonal change of variable satisfies $\dfrac{\partial y_i}{\partial x_j} = \dfrac{1}{\|\vec{x}_{y_i}\|_2^2} \dfrac{\partial x_j}{\partial y_i}$.

Exercise 6.2.15. Is the inverse $\vec{y} = F^{-1}(\vec{x})$ of an orthogonal change of variable also an orthogonal change of variable?

Exercise 6.2.16. Prove that under an orthogonal change of variable, the gradient in $\vec{x}$ can be written in terms of the new variable by

$$\nabla f = \frac{\partial f}{\partial y_1} \frac{\vec{x}_{y_1}}{\|\vec{x}_{y_1}\|_2^2} + \frac{\partial f}{\partial y_2} \frac{\vec{x}_{y_2}}{\|\vec{x}_{y_2}\|_2^2} + \cdots + \frac{\partial f}{\partial y_n} \frac{\vec{x}_{y_n}}{\|\vec{x}_{y_n}\|_2^2}.$$

Elementary Symmetric Polynomial

The *elementary symmetric polynomials* for n variables $x_1, x_2, \ldots, x_n$ are

$$\sigma_k = \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} x_{i_1} x_{i_2} \cdots x_{i_k}, \quad k = 1, 2, \ldots, n.$$

Vieta's formulae says that they appear as the coefficients of the polynomial

$$\begin{aligned} p(x) &= (x - x_1)(x - x_2) \cdots (x - x_n) \\ &= x^n - \sigma_1 x^{n-1} + \sigma_2 x^{n-2} - \cdots + (-1)^{n-1} \sigma_{n-1} x + (-1)^n \sigma_n. \quad (6.2.14) \end{aligned}$$

Therefore $\vec{x} \mapsto \vec{\sigma} \colon \mathbb{R}^n \to \mathbb{R}^n$ is the map that takes the roots of polynomials to polynomials.

Exercise 6.2.17. Prove that the derivative $\dfrac{\partial \vec{\sigma}}{\partial \vec{x}}$ of the polynomial with respect to the roots satisfy

$$\begin{pmatrix} x_1^{n-1} & x_1^{n-2} & \cdots & 1 \\ x_2^{n-1} & x_2^{n-2} & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ x_n^{n-1} & x_n^{n-2} & \cdots & 1 \end{pmatrix} \frac{\partial \vec{\sigma}}{\partial \vec{x}} + \begin{pmatrix} p'(x_1) & 0 & \cdots & 0 \\ 0 & p'(x_2) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & p'(x_n) \end{pmatrix} = O.$$

Then prove that when the roots are distinct, the roots can be locally written as continuously differentiable functions of the polynomial.

Power Sum and Newton's Identity

The *power sums* for n variables $x_1, x_2, \ldots, x_n$ are

$$s_k = \sum_{1 \leq i \leq n} x_i^k = x_1^k + x_2^k + \cdots + x_n^k, \quad k = 1, 2, \ldots, n.$$

For the polynomial $p(x)$ in (6.2.14), by adding $p(x_i) = 0$ together for $i = 1, 2, \ldots, n$, we get

$$s_n - \sigma_1 s_{n-1} + \sigma_2 s_{n-2} - \cdots + (-1)^{n-1} \sigma_{n-1} s_1 + (-1)^n n \sigma_n = 0. \quad (6.2.15)$$

Exercise 6.2.18. For

$$u_{l,k} = \sum_{i_1 < i_2 < \cdots < i_l, j \neq i_p} x_{i_1} x_{i_2} \cdots x_{i_l} x_j^k, \quad l \geq 0, k \geq 1, l + k \leq n,$$

prove that

$$s_k = u_{0,k},$$
$$\sigma_1 s_{k-1} = u_{0,k} + u_{1,k-1},$$
$$\sigma_2 s_{k-2} = u_{1,k-1} + u_{2,k-2},$$
$$\vdots$$
$$\sigma_{k-2} s_2 = u_{k-3,3} + u_{k-2,2},$$
$$\sigma_{k-1} s_1 = u_{k-2,2} + k\sigma_k.$$

and derive Newton's identities

$$s_k - \sigma_1 s_{k-1} + \sigma_2 s_{k-2} - \cdots + (-1)^{k-1}\sigma_{k-1}s_1 + (-1)^k k\sigma_k = 0, \quad k = 1, 2, \ldots, n. \tag{6.2.16}$$

Exercise 6.2.19. Prove that there is a polynomial invertible map that relates $\vec{s} = (s_1, s_2, \ldots, s_n)$ and $\vec{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_n)$. Then discuss the local invertibility of the map $\vec{x} \mapsto \vec{\sigma} \colon \mathbb{R}^n \to \mathbb{R}^n$ when there are multiple roots (see Exercise 6.2.17).

Functional Dependence

A collection of functions are *functionally dependent* if some can be written as functions of the others. For example, the functions $f = x + y$, $g = x^2 + y^2$, $h = x^3 + y^3$ are functionally dependent because $h = \dfrac{1}{2}f^3 - \dfrac{1}{2}fg$. In the following exercises, all functions are continuously differentiable.

Exercise 6.2.20. Prove that if $f_1(\vec{x})$, $f_2(\vec{x})$, ..., $f_n(\vec{x})$ are functionally dependent, then the gradients ∇f_1, ∇f_2, ..., ∇f_n are linearly dependent.

Exercise 6.2.21. Prove that $f_1(\vec{x})$, $f_2(\vec{x})$, ..., $f_n(\vec{x})$ are functionally dependent near $\vec{x}_0$ if and only if there is a function $h(\vec{y})$ defined near $\vec{y}_0 = (f_1(\vec{x}_0), f_2(\vec{x}_0), \ldots, f_n(\vec{x}_0))$, such that $\nabla h(\vec{y}_0) \neq \vec{0}$ and $h(f_1(\vec{x}), f_2(\vec{x}), \ldots, f_n(\vec{x})) = 0$.

Exercise 6.2.22. Suppose the gradients ∇f, ∇g are linearly dependent everywhere, and $\nabla g(\vec{x}_0) \neq \vec{0}$. Prove that there is a function $h(y)$ defined for y near $g(\vec{x}_0)$, such that $f(\vec{x}) = h(g(\vec{x}))$ for $\vec{x}$ near $\vec{x}_0$.

Hint: If $\dfrac{\partial g}{\partial x_1} \neq 0$, then $(x_1, x_2, \ldots, x_n) \mapsto (g, x_2, \ldots, x_n)$ is invertible near $\vec{x}_0$. After changing the variables from $(x_1, x_2, \ldots, x_n)$ to $(g, x_2, \ldots, x_n)$, verify that $\dfrac{\partial f}{\partial x_2} = \cdots = \dfrac{\partial f}{\partial x_n} = 0$.

Exercise 6.2.23. Suppose the gradient vectors ∇f, ∇g_1, ..., ∇g_k are linearly dependent near $\vec{x}_0$. Suppose ∇g_1, ..., ∇g_k are linearly independent at $\vec{x}_0$. Prove that there is a function $h(\vec{y})$ defined for $\vec{y}$ near $(f_1(\vec{x}_0), \ldots, f_k(\vec{x}_0))$, such that $f(\vec{x}) = h(g(\vec{x}))$ for $\vec{x}$ near $\vec{x}_0$.

Exercise 6.2.24. Suppose the rank of the gradient vectors ∇f_1, ∇f_2, ..., ∇f_m is always k near $\vec{x}_0$. Prove that there are k functions from f_1, f_2, ..., f_m, such that the other $m - k$ functions are functionally dependent on these k functions.

Exercise 6.2.25. Determine functional dependence.

1. $x + y + z$, $x^2 + y^2 + z^2$, $x^3 + y^3 + z^3$.

2. $x + y - z$, $x - y + z$, $x^2 + y^2 + z^2 - 2yz$.

3. $\dfrac{x}{x^2+y^2+z^2}, \dfrac{y}{x^2+y^2+z^2}, \dfrac{z}{x^2+y^2+z^2}$.

4. $\dfrac{x}{\sqrt{x^2+y^2+z^2}}, \dfrac{y}{\sqrt{x^2+y^2+z^2}}, \dfrac{z}{\sqrt{x^2+y^2+z^2}}$.

6.3   High Order Differentiation

The high order differentiation and Taylor expansion can also be extended to multivariable. The concepts can be used to solve problems that require high order approximations, such as local maximum and minimum with or without constraint.

6.3.1   Quadratic Approximation

A function $f(\vec{x})$ defined near $\vec{x}_0 \in \mathbb{R}^n$ is approximated by a quadratic function

$$p(\vec{x}) = a + \sum_{1 \le i \le n} b_i(x_i - x_{i0}) + \sum_{1 \le i,j \le n} c_{ij}(x_i - x_{i0})(x_j - x_{j0})$$
$$= a + \vec{b} \cdot \Delta\vec{x} + C\Delta\vec{x} \cdot \Delta\vec{x} \tag{6.3.1}$$

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

$$\|\Delta\vec{x}\| < \delta \implies |f(\vec{x}) - p(\vec{x})| \le \epsilon\|\Delta\vec{x}\|^2. \tag{6.3.2}$$

In this case, we say f is second order differentiable at $\vec{x}_0$, with the quadratic form

$$f''(\vec{x}_0)(\vec{v}) = 2 \sum_{1 \le i,j \le n} c_{ij}v_iv_j = C\vec{v} \cdot \vec{v} \tag{6.3.3}$$

as the the second order derivative. The second order differential is

$$d^2f = 2 \sum_{1 \le i,j \le n} c_{ij}dx_idx_j = 2Cd\vec{x} \cdot d\vec{x}. \tag{6.3.4}$$

Similar to the single variable case, the second order differentiability implies the first order differentiability, so that $a = f(\vec{x}_0)$ and $b_i = \dfrac{\partial f(\vec{x}_0)}{\partial x_i}$. We expect the coefficients c_{ij} to be the second order partial derivatives

$$c_{ij} = \frac{1}{2}f_{x_jx_i}(\vec{x}_0) = \frac{1}{2}\frac{\partial^2 f(\vec{x}_0)}{\partial x_i\partial x_j} = \frac{1}{2}\frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right)_{\vec{x}=\vec{x}_0}. \tag{6.3.5}$$

However, a second order differentiable function can be very bad (discontinuous, for example) away from $\vec{x}_0$. Therefore the formula (6.3.5) may not make sense at all. On the other hand, Theorem 2.3.1 suggests that if the second order partial derivatives do exist and have good enough properties, then the quadratic function (6.3.1) with the coefficients given by suitable partial derivatives should approximate f.

Theorem 6.3.1. *Suppose a function f is differentiable near $\vec{x}_0$ and the partial derivatives $\dfrac{\partial f}{\partial x_i}$ are differentiable at $\vec{x}_0$. Then f is second order differentiable at $\vec{x}_0$, with the quadratic approximation*

$$T_2(\vec{x}) = f(\vec{x}_0) + \sum_{1 \leq i \leq n} \frac{\partial f(\vec{x}_0)}{\partial x_i} \Delta x_i + \frac{1}{2} \sum_{1 \leq i,j \leq n} \frac{\partial^2 f(\vec{x}_0)}{\partial x_i \partial x_j} \Delta x_i \Delta x_j.$$

Proof. We only prove the case $n = 2$ and $\vec{x}_0 = (0,0)$. The general case is similar. To show that

$$p(x,y) = f(0,0) + f_x(0,0)x + f_y(0,0)y$$
$$+ \frac{1}{2}(f_{xx}(0,0)x^2 + f_{xy}(0,0)xy + f_{yx}(0,0)yx + f_{yy}(0,0)y^2) \quad (6.3.6)$$

approximates f near $(0,0)$, we restrict the remainder $R_2(x,y) = f(x,y) - p(x,y)$ to the straight lines passing through $(0,0)$. Therefore for fixed (x,y) close to $(0,0)$, we introduce

$$r_2(t) = R_2(tx, ty) = f(tx, ty) - p(tx, ty).$$

Inspired by the proof of Theorem 2.3.1, we apply Cauchy's mean value theorem to get

$$R_2(x,y) = r_2(1) = \frac{r_2(1) - r_2(0)}{1^2 - 0^2} = \frac{r_2'(c)}{2c}, \quad 0 < c < 1.$$

To compute $r_2'(c)$, we note that the differentiability of f *near* $(0,0)$ allows us to use the chain rule to get

$$r_2'(t) = f_x(tx, ty)x + f_y(tx, ty)y - f_x(0,0)x - f_y(0,0)y$$
$$- (f_{xx}(0,0)x^2 + f_{xy}(0,0)xy + f_{yx}(0,0)yx + f_{yy}(0,0)y^2)t$$
$$= x[f_x(tx, ty) - f_x(0,0) - f_{xx}(0,0)tx - f_{xy}(0,0)ty]$$
$$+ y[f_y(tx, ty) - f_y(0,0) - f_{yx}(0,0)tx - f_{yy}(0,0)ty]. \quad (6.3.7)$$

Further by the differentiability of f_x and f_y, for any $\epsilon > 0$, there is $\delta > 0$, such that $\|(\xi, \eta)\| < \delta$ implies

$$|f_x(\xi, \eta) - f_x(0,0) - f_{xx}(0,0)\xi - f_{xy}(0,0)\eta| < \epsilon\|(\xi, \eta)\|,$$
$$|f_y(\xi, \eta) - f_y(0,0) - f_{yx}(0,0)\xi - f_{yy}(0,0)\eta| < \epsilon\|(\xi, \eta)\|.$$

Taking $(\xi, \eta) = (cx, cy)$, by (6.3.7) we find $\|(x,y)\| < \delta$ implies

$$|r_2'(c)| \leq |x|\epsilon\|(cx, cy)\| + |y|\epsilon\|(cx, cy)\| = c\epsilon(|x| + |y|)\|(x,y)\|.$$

Therefore

$$\|(x,y)\| < \delta \implies |R_2(x,y)| \leq \frac{1}{2}\epsilon(|x| + |y|)\|(x,y)\|.$$

By the equivalence of norms, the right side is comparable to $\epsilon\|(x,y)\|^2$. Therefore we conclude that $p(x,y)$ is a quadratic approximation of $f(x,y)$ near $(0,0)$.

$\square$

Under slightly stronger condition, the partial derivative coefficients are symmetric.

Proposition 6.3.2. *Suppose $f(x, y)$ has partial derivatives f_x, f_y, f_{xy} near (x_0, y_0) and f_{xy} is continuous at (x_0, y_0), then $f_{yx}(x_0, y_0)$ exists and $f_{xy}(x_0, y_0) = f_{yx}(x_0, y_0)$.*

Thus if a function has all the first and second order partial derivatives near $\vec{x}_0$, and the second order partial derivatives are continuous at $\vec{x}_0$, then the condition of Theorem 6.3.1 is satisfied. Moreover, the second order partial derivatives is independent of the order of the variables.

Proof. The second order partial derivative $f_{yx}(x_0, y_0)$ is defined by a repeated limit

$$f_{yx}(x_0, y_0) = \lim_{x \to x_0} \frac{f_y(x, y_0) - f_y(x_0, y_0)}{x - x_0}$$
$$= \lim_{x \to x_0} \lim_{y \to y_0} \frac{f(x, y) - f(x, y_0) - f(x_0, y) + f(x_0, y_0)}{(x - x_0)(y - y_0)}.$$

The other second order partial derivative $f_{xy}(x_0, y_0)$ is defined by switching the two limits in the repeated limit. We will show that the existence of f_y, f_{xy} near (x_0, y_0) and the continuity of f_{xy} at (x_0, y_0) actually imply the convergence of the whole limit

$$f_{xy}(x_0, y_0) = \lim_{x \to x_0, y \to y_0} \frac{f(x, y) - f(x, y_0) - f(x_0, y) + f(x_0, y_0)}{(x - x_0)(y - y_0)}.$$

Note that the existence of f_x implies the existence of the limit

$$\frac{f_y(x, y_0) - f_y(x_0, y_0)}{x - x_0} = \lim_{y \to y_0} \frac{f(x, y) - f(x, y_0) - f(x_0, y) + f(x_0, y_0)}{(x - x_0)(y - y_0)}.$$

Then by Proposition 5.1.11, the repeated limit $f_{yx}(x_0, y_0)$ exists and is equal to $f_{xy}(x_0, y_0)$.

For fixed y_0 and y, we apply the mean value theorem to the function $g(x) = f(x, y) - f(x, y_0)$. By the existence of f_x near (x_0, y_0), we get

$$f(x, y) - f(x, y_0) - f(x_0, y) + f(x_0, y_0)$$
$$= g(x) - g(x_0) = g'(c)(x - x_0) = (f_x(c, y) - f_y(c, y_0))(x - x_0)$$

for some c between x_0 and x. Then we fix c and apply the mean value theorem to the function $f_x(c, y)$ of y. By the existence of f_{xy} near (x_0, y_0), we get

$$f(x, y) - f(x_0, y) - f(x, y_0) + f(x_0, y_0) = f_{xy}(c, d)(x - x_0)(y - y_0)$$

for some d between y_0 and y. Then the continuity of f_{xy} at (x_0, y_0) further tells us

$$\lim_{x \to x_0, y \to y_0} \frac{f(x, y) - f(x_0, y) - f(x, y_0) + f(x_0, y_0)}{(x - x_0)(y - y_0)}$$
$$= \lim_{x \to x_0, y \to y_0} f_{xy}(c, d) = \lim_{c \to x_0, d \to y_0} f_{xy}(c, d) = f_{xy}(x_0, y_0).$$

$\square$

Example 6.3.1. The function $f = xy^2z^3$ has continuous partial derivatives

$$f_x = y^2z^3, \quad f_y = 2xyz^3, \quad f_z = 3xy^2z^2,$$
$$f_{xx} = 0, \quad f_{yy} = 2yz^3, \quad f_{zz} = 6xy^2z, \quad f_{xy} = 2yz^3, \quad f_{xz} = 3y^2z^2, \quad f_{yz} = 6xyz^2.$$

The values of the partial derivatives at $(3, 2, 1)$ are

$$f = 12,$$
$$f_x = 4, \quad f_y = 12, \quad f_z = 36,$$
$$f_{xx} = 0, \quad f_{yy} = 4, \quad f_{zz} = 72, \quad f_{xy} = 4, \quad f_{xz} = 12, \quad f_{yz} = 36.$$

The quadratic approximation of f at $(3, 2, 1)$ is

$$p = 12 + 4\Delta x + 12\Delta y + 36\Delta z + \frac{1}{2}(4\Delta y^2 + 72\Delta z^2 + 8\Delta x\Delta y + 24\Delta x\Delta z + 72\Delta y\Delta z),$$

where $\Delta x = x - 3$, $\Delta y = y - 2$, $\Delta z = z - 1$. The quadratic differential is

$$d^2_{(3,2,1)}f = 4dy^2 + 72dz^2 + 8dxdy + 24dxdz + 72dydz.$$

Exercise 6.3.1. Compute the partial derivatives up to second order.

1. $4xy^3 + 5x^3y - 3y^2$. 3. $e^{xyz}\sin(x+2y+3z)$. 5. x^{y^z}.

2. $\arctan\dfrac{y}{x}$. 4. $\log(x^2 + y^2)$. 6. $(xy)^z$.

Exercise 6.3.2. Construct a function that is second order differentiable at $(0, 0)$ but is not continuous anywhere away from $(0, 0)$.

Exercise 6.3.3. Show that the function

$$f(x, y) = \begin{cases} \dfrac{xy(x^2 - y^2)}{x^2 + y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0) \end{cases}$$

has all the second order partial derivatives near $(0, 0)$ but $f_{xy}(0, 0) \neq f_{yx}(0, 0)$.

Exercise 6.3.4. Consider the function $f(x, y)$ in Example 5.1.2. Show that $f(x, y)^2$ has all the second order derivatives but the function is still not continuous (so not differentiable) at $(0, 0)$. Can you find a function that has all the partial derivatives but is not continuous?

Exercise 6.3.5. Suppose f has continuous second order partial derivatives. Prove that if $ff_{xy} = f_xf_y$, then $f(x, y) = g(x)h(y)$.

Exercise 6.3.6. Derive the chain rule for the second order derivative by composing the quadratic approximations.

6.3.2 High Order Partial Derivative

The discussion on the quadratic approximation suggests the role to be played by high order partial derivatives in the high order approximation. In general, the *k-th order partial derivative*

$$\frac{\partial^k f}{\partial x_{i_1}\partial x_{i_2}\cdots\partial x_{i_k}} = D_{x_{i_1}x_{i_2}\cdots x_{i_k}}f = f_{x_{i_k}\cdots x_{i_2}x_{i_1}} \tag{6.3.8}$$

is obtained by successfully taking partial derivatives in x_{i_k}, $x_{i_{k-1}}$, $\ldots$, x_{i_2}, x_{i_1}. The partial derivative may depend on the order of variables in general. However, by Proposition 6.3.2, if all the partial derivatives up to the $(k-1)$-st order are continuous near $\vec{x}_0$, and the k-th order partial derivatives are continuous at $\vec{x}_0$, then the partial derivative is independent of the order. In this case, we can write

$$\frac{\partial^k f}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_k}} = \frac{\partial^k f}{\partial x_1^{k_1} \partial x_2^{k_2} \cdots \partial x_n^{k_n}},$$

where k_j is the number of x_j in the collection $\{x_{i_1}, x_{i_2}, \ldots, x_{i_k}\}$.

Many techniques for the computation of high order derivatives of single variable functions can be adapted to multivariable functions.

Example 6.3.2. For a function $f(x, y)$ and $x = x(u, v)$, $y = y(u, v)$, we have

$$\begin{aligned}
\frac{\partial^2 f}{\partial u \partial v} &= \frac{\partial}{\partial u}\left(\frac{\partial f}{\partial v}\right) = \frac{\partial}{\partial u}\left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v}\right) \\
&= \frac{\partial}{\partial u}\left(\frac{\partial f}{\partial x}\right)\frac{\partial x}{\partial v} + \frac{\partial f}{\partial x}\frac{\partial}{\partial u}\left(\frac{\partial x}{\partial v}\right) + \frac{\partial}{\partial u}\left(\frac{\partial f}{\partial y}\right)\frac{\partial y}{\partial v} + \frac{\partial f}{\partial y}\frac{\partial}{\partial u}\left(\frac{\partial y}{\partial v}\right) \\
&= \left(\frac{\partial^2 f}{\partial x^2}\frac{\partial x}{\partial u} + \frac{\partial^2 f}{\partial y \partial x}\frac{\partial y}{\partial u}\right)\frac{\partial x}{\partial v} + \frac{\partial f}{\partial x}\frac{\partial^2 x}{\partial u \partial v} \\
&\quad + \left(\frac{\partial^2 f}{\partial x \partial y}\frac{\partial x}{\partial u} + \frac{\partial^2 f}{\partial y^2}\frac{\partial y}{\partial u}\right)\frac{\partial y}{\partial v} + \frac{\partial f}{\partial y}\frac{\partial^2 y}{\partial u \partial v} \\
&= \frac{\partial x}{\partial u}\frac{\partial x}{\partial v}\frac{\partial^2 f}{\partial x^2} + \frac{\partial y}{\partial u}\frac{\partial y}{\partial v}\frac{\partial^2 f}{\partial y^2} + \left(\frac{\partial x}{\partial u}\frac{\partial y}{\partial v} + \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}\right)\frac{\partial^2 f}{\partial x \partial y} \\
&\quad + \frac{\partial^2 x}{\partial u \partial v}\frac{\partial f}{\partial x} + \frac{\partial^2 y}{\partial u \partial v}\frac{\partial f}{\partial y}.
\end{aligned}$$

Example 6.3.3. By Example 6.2.1, the partial derivatives of the polar coordinate in terms of the cartesian coordinate is

$$r_x = \cos\theta, \ \ r_y = \sin\theta, \ \ \theta_x = -r^{-1}\sin\theta, \ \ \theta_y = r^{-1}\cos\theta.$$

Thus

$$\begin{aligned}
r_{xx} &= -(\sin\theta)\theta_x = r^{-1}\sin^2\theta, \\
r_{yy} &= (\cos\theta)\theta_y = r^{-1}\cos^2\theta, \\
r_{xy} &= -(\sin\theta)\theta_y = -r^{-1}\sin\theta\cos\theta, \\
\theta_{xx} &= r^{-2}r_x\sin\theta - r^{-1}(\cos\theta)\theta_x = 2r^{-2}\sin\theta\cos\theta, \\
\theta_{yy} &= -r^{-2}r_y\cos\theta + r^{-1}(-\sin\theta)\theta_y = -2r^{-2}\sin\theta\cos\theta, \\
\theta_{xy} &= r^{-2}r_y\sin\theta - r^{-1}(\cos\theta)\theta_y = r^{-2}(\sin^2\theta - \cos^2\theta).
\end{aligned}$$

Example 6.3.4. By Example 6.2.2, the partial derivatives of one coordinate in terms of the other coordinates in the unit sphere $x^2 + y^2 + z^2 = 1$ is (when $z \neq 0$)

$$z_x = -\frac{x}{z}, \ \ z_y = -\frac{y}{z}.$$

Thus

$$z_{xx} = -\frac{z - xz_x}{z^2} = -\frac{x^2 + z^2}{z^3} = \frac{1 - y^2}{z^3},$$

$$z_{xy} = \frac{xz_y}{z^2} = -\frac{xy}{z^3},$$

$$z_{xxx} = -3\frac{(1 - y^2)z_x}{z^4} = 3\frac{(1 - y^2)x}{z^5},$$

$$z_{xxy} = \frac{-2yz^3 - 3(1 - y^2)z^2 z_y}{z^6} = \frac{(1 - 3y^2 - 2z^2)y}{z^5}.$$

Exercise 6.3.7. Compute the third order partial derivatives.

1. $4xy^3 + 5x^3y - 3y^2$.

2. $\arctan \dfrac{y}{x}$.

3. $e^{xyz} \sin(x + 2y + 3z)$.

4. $\log(x^2 + y^2)$.

5. x^{y^z}.

6. $(xy)^z$.

Exercise 6.3.8. Compute the partial derivatives.

1. $z = uv + \sin t$, $u = e^t$, $v = \cos t$, find z_{tt}, z_{ttt}.

2. $z = x^2 \log y$, $x = \dfrac{u}{v}$, $y = u - v$, find z_{uu}, z_{uv}, z_{vv}.

3. $u = f(x + y, x - y)$, find u_{xx}, u_{xy}, u_{xxx}, u_{xyy}.

4. $u = f(r \cos\theta, r \sin\theta)$, find u_{rr}, $u_{\theta\theta}$, $u_{r\theta}$.

5. $u = f\left(\dfrac{x}{y}, \dfrac{y}{z}, \dfrac{z}{x}\right)$, find u_{xx}, u_{xy}, u_{xyz}.

6. $u = xyzf(x, y, z)$, find $\dfrac{\partial^{i+j+k}u}{\partial x^i \partial y^j \partial z^k}$.

Exercise 6.3.9. For a function f on $\mathbb{R}^n$ and a linear transform $L\colon \mathbb{R}^m \to \mathbb{R}^n$. How are the high order partial derivatives of f and $f \circ L$ related?

Exercise 6.3.10. Verify the functions satisfy the partial differential equations.

1. $u = \dfrac{1}{2a\sqrt{\pi t}}e^{-\frac{(x-b)^2}{4a^2 t}}$, heat equation $u_t = a^2 u_{xx}$.

2. $u = \log((x - a)^2 + (y - b)^2)$, Laplace equation $u_{xx} + u_{yy} = 0$.

3. $u = A((x_1 - a_1)^2 + (x_2 - a_2)^2 + \cdots + (x_n - a_n)^2)^{\frac{2-n}{2}}$, Laplace equation $u_{x_1^2} + u_{x_2^2} + \cdots + u_{x_n^2} = 0$.

4. $u = \phi(x + at) + \psi(x - at)$, wave equation $u_{tt} = a^2 u_{xx}$.

Exercise 6.3.11. Derive the partial differential equations under the new variables.

1. $x = r \cos\theta$, $y = r \sin\theta$, Laplace equation $u_{xx} + u_{yy} = 0$.

2. $\xi = x + at$, $\eta = x - at$, wave equation $u_{tt} = a^2 u_{xx}$.

3. $s = xy$, $t = \dfrac{x}{y}$, equation $x^2 u_{xx} - y^2 u_{yy} = 0$.

Exercise 6.3.12. Compute the partial derivatives.

1. $x = s + t$, $y = s^2 + t^2$, $z = s^3 + t^3$, find z_{xx}, z_{xy}, z_{yy}.

2. $x = e^{s+t}$, $y = e^{s-t}$, $z = st$, find z_{xx}, z_{xxx}.

3. $x^3 + y^3 - 3axy = 0$, find y_{xx} and y_{xxx}.

4. $x^2 + y^2 + z^2 - 4x + 6y - 2z = 0$, find $\dfrac{\partial^2 z}{\partial(x,y)^2}$ and $\dfrac{\partial^2 x}{\partial(y,z)^2}$.

5. $x^2 + y^2 = z^2 + w^2$, $x + y + z + w = 1$, find $\dfrac{\partial^2(z,w)}{\partial(x,y)^2}$, $\dfrac{\partial^2(x,w)}{\partial(y,z)^2}$.

Exercise 6.3.13. Suppose $y = y(\vec{x})$ is given implicitly by $f(\vec{x}, y) = 0$. Compute the second order partial derivatives of y.

6.3.3 Taylor Expansion

A map $P \colon \mathbb{R}^n \to \mathbb{R}^m$ is a polynomial map of degree k if all its coordinate functions are polynomials of degree k. A map $F(\vec{x})$ defined near $\vec{x}_0$ is *k-th order differentiable* at $\vec{x}_0$ if it is approximated by a polynomial map P of degree k. In other words, for any $\epsilon > 0$, there is $\delta > 0$, such that

$$\|\Delta\vec{x}\| < \delta \implies \|F(\vec{x}) - P(\vec{x})\| \le \epsilon\|\Delta\vec{x}\|^k. \tag{6.3.9}$$

Express the approximate polynomial as

$$P(\vec{x}) = F(\vec{x}_0) + F'(\vec{x}_0)(\Delta x) + \frac{1}{2}F''(\vec{x}_0)(\Delta x) + \cdots + \frac{1}{k!}F^{(k)}(\vec{x}_0)(\Delta x), \tag{6.3.10}$$

where the coordinates of $F^{(i)}(\vec{x}_0)$ are i-th order forms. Then $F^{(i)}(\vec{x}_0)$ is called the *i-th order derivative* of F at $\vec{x}_0$.

It is easy to see that F is approximated by P if and only if each coordinate is approximated. A function $f(\vec{x})$ defined near $\vec{x}_0$ is k-th order differentiable at $\vec{x}_0$ if there is a polynomial

$$p(\vec{x}) = \sum_{k_1 + k_2 + \cdots + k_n \le k, k_i \ge 0} b^{k_1 k_2 \cdots k_n} \Delta x_1^{k_1} \Delta x_2^{k_2} \cdots \Delta x_n^{k_n}$$

of degree k, such that for any $\epsilon > 0$, there is $\delta > 0$, such that

$$\|\Delta\vec{x}\| < \delta \implies |f(\vec{x}) - p(\vec{x})| \le \epsilon\|\Delta\vec{x}\|^k.$$

The k-th order derivative

$$f^{(k)}(\vec{x}_0)(\vec{v}) = \sum_{k_1 + k_2 + \cdots + k_n = k, k_i \ge 0} k! b^{k_1 k_2 \cdots k_n} v_1^{k_1} v_2^{k_2} \cdots v_n^{k_n}$$

is a k-th order form, and the *k-th order differential* is

$$d^k f = \sum_{k_1 + k_2 + \cdots + k_n = k, k_i \ge 0} k! b^{k_1 k_2 \cdots k_n} dx_1^{k_1} dx_2^{k_2} \cdots dx_n^{k_n}.$$

Similar to single variable functions, a high order differentiable function may not be continuous away from the point. On the other hand, if f has

partial derivatives up to order k at $\vec{x}_0$, then we may construct the *k-th order Taylor expansion*

$$T_k(\vec{x}) = \sum_{1 \le i_1, i_2, \ldots, i_m \le n, 0 \le m \le k} \frac{1}{m!} \frac{\partial^m f(\vec{x}_0)}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_m}} \Delta x_{i_1} \Delta x_{i_2} \cdots \Delta x_{i_m}. \quad (6.3.11)$$

If all the partial derivatives exist near $\vec{x}_0$ and are continuous at $\vec{x}_0$, then the partial derivatives are independent of the order of the variables, and we have

$$T_k(\vec{x}) = \sum_{k_1 + k_2 + \cdots + k_n \le k, k_i \ge 0} \frac{1}{k_1! k_2! \cdots k_n!} \frac{\partial^{k_1 + k_2 + \cdots + k_n} f(\vec{x}_0)}{\partial x_1^{k_1} \partial x_2^{k_2} \cdots \partial x_n^{k_n}} \Delta x_1^{k_1} \Delta x_2^{k_2} \cdots \Delta x_n^{k_n}. \tag{6.3.12}$$

Theorem 6.3.3. *Suppose $f(\vec{x})$ has continuous partial derivatives up to order $k - 1$ near $\vec{x}_0$ and the $(k-1)$-st order partial derivatives are differentiable at $\vec{x}_0$. Then the k-th order Taylor expansion is the k-th order polynomial approximation at $\vec{x}_0$.*

Proof. Restrict the *remainder*

$$R_k(\vec{x}) = f(\vec{x}) - T_k(\vec{x})$$

to the straight line connecting $\vec{x}_0$ to $\vec{x}$

$$r_k(t) = R_k((1 - t)\vec{x}_0 + t\vec{x}) = R_k(\vec{x}_0 + t\Delta\vec{x}).$$

Since $f(\vec{x})$ has continuous partial derivatives up to order $k - 1$ near $\vec{x}_0$, the function $f(\vec{x})$ and its partial derivatives up to order $(k-2)$ are differentiable near $\vec{x}_0$. Then by the chain rule, r_k has derivatives up to order $(k - 1)$ for $t \in (-1, 1)$. Moreover, we have

$$r_k(0) = r_k'(0) = r_k''(0) = \cdots = r_k^{(k-1)}(0) = 0, \tag{6.3.13}$$

and

$$r_k^{(k-1)}(t) = \sum_{1 \le i_1, i_2, \ldots, i_{k-1} \le n} [\delta_{i_1, i_2, \ldots, i_{k-1}}(t\Delta\vec{x}) - \lambda_{i_1, i_2, \ldots, i_{k-1}}(t\Delta\vec{x})] \Delta x_{i_1} \Delta x_{i_2} \cdots \Delta x_{i_{k-1}},$$

where

$$\delta_{i_1, i_2, \ldots, i_{k-1}}(\Delta\vec{x}) = \frac{\partial^{k-1} f(\vec{x}_0 + \Delta\vec{x})}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_{k-1}}}$$

is the $(k - 1)$-st order partial derivative, and

$$\lambda_{i_1, i_2, \ldots, i_{k-1}}(\Delta\vec{x}) = \frac{\partial^{k-1} f(\vec{x}_0)}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_{k-1}}} + \sum_{1 \le i \le n} \frac{\partial^k f(\vec{x}_0)}{\partial x_i \partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_{k-1}}} \Delta x_i$$

is the linear approximation of the partial derivative.

By (6.3.13) and Cauchy's mean value theorem, we have

$$R_k(\vec{x}) = r_k(1) = \frac{r_k(1) - r_k(0)}{1^k - 0^k} = \frac{r_k'(c_1)}{kc_1^{k-1}} = \frac{r_k'(c_1) - r_k'(0)}{k(c_1^{k-1} - 0^{k-1})}$$

$$= \frac{r_k''(c_2)}{k(k-1)c_2^{k-2}} = \cdots = \frac{r_k^{(k-1)}(c_{k-1})}{k(k-1)\cdots 2c_{k-1}} = \frac{r_k^{(k-1)}(c_{k-1})}{k!c_{k-1}}$$

for some $1 > c_1 > c_2 > \cdots > c_{k-1} > 0$. By the assumption that the $(k-1)$-st order partial derivatives $\delta_{i_1,i_2,\dots,i_m}$ of f are differentiable, for any $\epsilon > 0$, there is $\delta > 0$, such that $\|\Delta\vec{x}\| < \delta$ implies

$$|\delta_{i_1,i_2,\dots,i_{k-1}}(\Delta\vec{x}) - \lambda_{i_1,i_2,\dots,i_{k-1}}(\Delta\vec{x})| \le \epsilon\|\Delta\vec{x}\|.$$

Then for $\|\Delta\vec{x}\| < \delta$ and $|t| < 1$, we have

$$|r_k^{(k-1)}(t)| \le \sum_{1 \le i_1,i_2,\dots,i_{k-1} \le n} \epsilon\|t\Delta\vec{x}\||\Delta x_{i_1}\Delta x_{i_2}\cdots\Delta x_{i_{k-1}}|$$

$$\le \epsilon n^{k-1}|t|\|\Delta\vec{x}\|\|\Delta\vec{x}\|_\infty^{k-1},$$

and

$$|R_k(\vec{x})| = \frac{|r_k^{(k-1)}(c_{k-1})|}{k!|c_{k-1}|} \le \epsilon\frac{n^{k-1}}{k!}\|\Delta\vec{x}\|\|\Delta\vec{x}\|_\infty^{k-1}.$$

By the equivalence of norms, this implies that $\lim_{\Delta\vec{x}\to\vec{0}} \dfrac{R_k(\vec{x})}{\|\Delta\vec{x}\|^k} = 0$.

$\square$

Exercise 6.3.14. Find Taylor expansions.

1. x^y at $(1,4)$, to third order.

2. $\sin(x^2 + y^2)$ at $(0,0)$, to fourth order.

3. $x^y y^z$ at $(1,1,1)$, to third order.

4. $\displaystyle\int_0^1 (1+x)^{t^2 y}dt$ at $(0,0)$, to third order.

Exercise 6.3.15. Find the second and third order derivatives of the map of taking the k-th power of matrices.

Exercise 6.3.16. Find the high order derivatives of the map of taking the inverse of matrices.

Exercise 6.3.17. Find the condition for a homogeneous function to be k-th order differentiable at $\vec{0}$. What about a multihomogeneous function?

Under stronger differentiability condition, the remainder formula in Proposition 2.3.3 can also be extended.

Proposition 6.3.4. *Suppose $f(\vec{x})$ has continuous partial derivatives up to order $k+1$ near $\vec{x}_0$. Then for any $\vec{x}$ near $\vec{x}_0$, there is $\vec{c}$ on the straight line connecting $\vec{x}_0$ to $\vec{x}$, such that*

$$|f(\vec{x}) - T_k(\vec{x})| \le \sum_{k_1+k_2+\cdots+k_n=k+1} \frac{1}{k_1!k_2!\cdots k_n!} \left|\frac{\partial^{k+1}f(\vec{c})}{\partial x_1^{k_1}\partial x_2^{k_2}\cdots\partial x_n^{k_n}}\right| |\Delta x_1|^{k_1}|\Delta x_2|^{k_2}\cdots|\Delta x_n|^{k_n}.$$

$$(6.3.14)$$

Proof. Under the assumption of the proposition, we have

$$R_k(\vec{x}) = \frac{r_k(1) - r_k(0)}{1^{k+1} - 0^{k+1}} = \cdots = \frac{r_k^{(k)}(c_k)}{(k+1)k \cdots 2c_k} = \frac{r_k^{(k+1)}(c_{k+1})}{(k+1)!}$$

for some $1 > c_1 > c_2 > \cdots > c_{k+1} > 0$. Since $T_k(\vec{x})$ is a polynomial of degree k, we have

$$r_k^{(k+1)}(t) = \frac{d^{k+1} f(\vec{x}_0 + t\Delta\vec{x})}{dt^{k+1}}$$

$$= \sum_{k_1 + k_2 + \cdots + k_n = k+1} \frac{(k+1)!}{k_1! k_2! \cdots k_n!} \frac{\partial^{k_1 + k_2 + \cdots + k_n} f(\vec{x}_0 + t\Delta\vec{x})}{\partial x_1^{k_1} \partial x_2^{k_2} \cdots \partial x_n^{k_n}} \Delta x_1^{k_1} \Delta x_2^{k_2} \cdots \Delta x_n^{k_n},$$

and the estimation for the remainder $R_k(\vec{x})$ follows.

$\square$

Example 6.3.5. For the function $f = xy^2 z^3$ in Example 6.3.1, we have

$$f_{xxx} = f_{xxy} = f_{xxz} = f_{xyy} = f_{yyy} = 0,$$
$$f_{xyz} = 6yz^2, \quad f_{xzz} = 6y^2 z, \quad f_{yyz} = 6yz^2, \quad f_{yzz} = 12xyz.$$

Thus at $(3, 2, 1)$, when $|\Delta x|, |\Delta y|, |\Delta z| < \delta$, the sum of the absolute values of the third order derivatives is bounded by $6(2 + \delta)(1 + \delta)^2 + 6(2 + \delta)^2(1 + \delta) + 6(2 + \delta)(1 + \delta)^2 + 12(3 + \delta)(2 + \delta)(1 + \delta) < 120(1 + \delta)^3$. The error of the quadratic approximation is $< \dfrac{120(1 + \delta)^3}{3!} \delta^3 = 20(1 + \delta)^3 \delta^3$.

Exercise 6.3.18. Estimate the errors of the approximations in Exercise 6.3.14.

6.3.4 Maximum and Minimum

A function $f(\vec{x})$ has a *local maximum* at $\vec{x}_0$ if there is $\delta > 0$, such that

$$\|\vec{x} - \vec{x}_0\| < \delta \implies f(\vec{x}) \le f(\vec{x}_0).$$

In other words, the value of f at $\vec{x}_0$ is biggest among the values of f at points near $\vec{x}_0$. Similarly, f has a *local minimum* at $\vec{x}_0$ if there is $\delta > 0$, such that

$$\|\vec{x} - \vec{x}_0\| < \delta \implies f(\vec{x}) \ge f(\vec{x}_0).$$

Suppose the function is defined near $\vec{x}_0$. Then $\vec{x}_0$ is a local extreme if and only if it is a local extreme in all directions. So a necessary condition is that if the derivative $D_{\vec{v}} f(\vec{x}_0)$ exists in the direction of $\vec{v}$, then $D_{\vec{v}} f(\vec{x}_0) = 0$. By taking $\vec{v}$ to be the coordinate directions, we get the following result.

Proposition 6.3.5. *Suppose $f(\vec{x})$ is defined near $\vec{x}_0$ and has a local extreme at $\vec{x}_0$. If a partial derivative $\dfrac{\partial f}{\partial x_i}(\vec{x}_0)$ exists, then the partial derivative must be zero.*

Thus if $f(\vec{x})$ is differentiable at a local extreme $\vec{x}_0$, then $\nabla f(\vec{x}_0) = \vec{0}$. The condition can also be expressed as $df = 0$.

Example 6.3.6. For $f(x,y) = x^2 + xy + y^2$, we have $f_x = 2x + y$ and $f_y = x + 2y$. The condition $f_x = f_y = 0$ tells us the only possible local extreme is $(0,0)$. Since

$$f(x,y) = \left(x + \frac{1}{2}y\right)^2 + \frac{3}{4}y^2 \geq 0 = f(0,0),$$

$(0,0)$ is a local minimum.

Similarly, by solving $\nabla g = 0$, the only possible local extreme for $g(x,y) = x^2 - y^2$ is again $(0,0)$. Since $g(x,y) > 0 = g(0,0)$ when $|x| > |y|$ and $g(x,y) < 0 = g(0,0)$ when $|x| < |y|$, $(0,0)$ is neither a local maximum nor a local minimum. Taking a clue from the graph of g, $(0,0)$ is called a *saddle point*.

Example 6.3.7. The function $f(x,y) = |x| + y$ satisfies $f_y = 1 \neq 0$. Therefore there is no local extreme, despite the fact that f_x may not exist.

Example 6.3.8. The function $f(x,y) = \sqrt{|xy|}$ has the partial derivatives

$$f_x = \begin{cases} \frac{1}{2}\sqrt{\left|\frac{y}{x}\right|} & \text{if } x > 0 \\ -\frac{1}{2}\sqrt{\left|\frac{y}{x}\right|} & \text{if } x < 0 \\ 0 & \text{if } y = 0 \\ \text{does not exist} & \text{if } x = 0, y \neq 0 \end{cases} \quad, \quad f_y = \begin{cases} \frac{1}{2}\sqrt{\left|\frac{x}{y}\right|} & \text{if } y > 0 \\ -\frac{1}{2}\sqrt{\left|\frac{x}{y}\right|} & \text{if } y < 0 \\ 0 & \text{if } x = 0 \\ \text{does not exist} & \text{if } x \neq 0, y = 0 \end{cases}.$$

Therefore the possible local extrema are $(0, y_0)$ for any y_0 and $(x_0, 0)$ for any x_0. Since

$$f(x,y) = \sqrt{|xy|} \geq 0 = f(0, y_0) = f(x_0, 0),$$

the points on the two axes are indeed local minima.

Example 6.3.9. Consider the differentiable function $z = z(x,y)$ given implicitly by $x^2 - 2xy + 4yz + z^3 + 2y - z = 1$. To find the possible local extrema of $z(x,y)$, we take the differential and get

$$(2x - 2y)dx + (-2x + 4z + 2)dy + (4y + 3z^2 - 1)dz = 0.$$

The possible local extrema are obtained by the condition $dz = z_x dx + z_y dy = 0$ and satisfy the implicit equation. Thus we try to solve

$$2x - 2y = 0, \quad -2x + 4z + 2 = 0, \quad x^2 - 2xy + 4yz + z^3 + 2y - z = 1.$$

From the first two equations, we have $x = y = 2z + 1$. Substituting into the third, we get $z^3 + 4z^2 + 3z + 1 = 1$, which has three solutions $z = 0, -1, -3$. Using $x = y = 2z + 1$, we find three possible local extrema $(1, 1, 0)$, $(-1, -1, -1)$, $(-5, -5, -3)$ for $z(x,y)$.

Example 6.3.10. The continuous function $f = x^2 - 2xy$ reaches its maximum and minimum on the compact subset $|x| + |y| \leq 1$. From $f_x = 2x - 2y = 0$ and $f_y = -2x = 0$, we find the point $(0,0)$ to be the only possible local extreme in the interior $|x| + |y| < 1$. Then we look for the local extrema of f restricted to the boundary $|x| + |y| = 1$, which may be devided into four (open) line segments and four points.

On the segment $x + y = 1$, $0 < x < 1$, we have $f = x^2 - 2x(1 - x) = 3x^2 - 2x$. From $f_x = 6x - 2 = 0$ and $x + y = 1$, we find the possible local extreme $\left(\frac{1}{3}, \frac{2}{3}\right)$ for

f on the segment. Similarly, we find a possible local extreme $\left(-\dfrac{1}{3}, -\dfrac{2}{3}\right)$ on the segment $x + y = -1$, $-1 < x < 0$, and no possible local extrema on the segments $x - y = 1$, $0 < x < 1$ and $-x + y = 1$, $-1 < x < 0$.

We also need to consider four points at the ends of the four segments, which are $(1, 0)$, $(-1, 0)$, $(0, 1)$, $(0, -1)$. Comparing the values of f at all the possible local extrema

$$f(0,0) = 0, \quad f(1/3, 2/3) = f(-1/3, -1/3) = -1/3,$$
$$f(1,0) = f(-1,0) = 1, \quad f(0,1) = f(0,-1) = 0,$$

we find $\pm(1/2, 2/3)$ are the absolute minima and $(\pm 1, 0)$ are the absolute maxima.

Exercise 6.3.19. Find possible local extrema.

1. $x^2 y^3 (a - x - y)$.

2. $x + y + 4 \sin x \sin y$.

3. $xyz \log(x^2 + y^2 + z^2)$.

4. $x_1 x_2 \cdots x_n + \dfrac{a_1}{x_1} + \dfrac{a_2}{x_2} + \cdots + \dfrac{a_n}{x_n}$.

5. $\dfrac{x_1 x_2 \cdots x_n}{(a + x_1)(x_1 + x_2) \cdots (x_n + b)}$.

Exercise 6.3.20. Find possible local extrema of implicitly defined function z.

1. $x^2 + y^2 + z^2 - 2x + 6z = 6$.

2. $(x^2 + y^2 + z^2)^2 = a^2(x^2 + y^2 - z^2)$.

3. $x^3 + y^3 + z^3 - axyz = 1$.

Exercise 6.3.21. Find the absolute extrema for the functions on the given domain.

1. $x + y + z$ on $\{(x, y, z) : x^2 + y^2 \le z \le 1\}$.

2. $\sin x + \sin y - \sin(x + y)$ on $\{(x, y) : x \ge 0, y \ge 0, x + y \le 2\pi\}$.

Exercise 6.3.22. What is the shortest distance between two straight lines?

Similar to the single variable case. The linear approximation may be used to find potential local extrema. The high order approximations may be used to determine whether the candidates are indeed local extrema.

If a function is continuously second order differentiable, then the second order derivative is a quadratic form, called the *Hessian* of the function

$$h_f(\vec{v}) = \sum_{1 \le i \le n} \frac{\partial^2 f}{\partial x_i^2} v_i^2 + 2 \sum_{1 \le i < j \le n} \frac{\partial^2 f}{\partial x_i x_j} v_i v_j$$

$$= \frac{\partial^2 f}{\partial x_1^2} v_1^2 + \frac{\partial^2 f}{\partial x_2^2} v_2^2 + \cdots + \frac{\partial^2 f}{\partial x_n^2} v_n^2$$

$$+ 2 \frac{\partial^2 f}{\partial x_1 \partial x_2} v_1 v_2 + 2 \frac{\partial^2 f}{\partial x_1 \partial x_3} v_1 v_3 + \cdots + 2 \frac{\partial^2 f}{\partial x_{n-1} \partial x_n} v_{n-1} v_n. \quad (6.3.15)$$

The following extends Proposition 2.3.4 in the second order case.

Proposition 6.3.6. *Suppose $f(\vec{x})$ has second order partial derivatives near $\vec{x}_0$, such that the first order derivative is continuous near $\vec{x}_0$ and the second order derivative is continuous at $\vec{x}_0$. Moreover, assume $\nabla f(\vec{x}_0) = 0$.*

1. *If the Hessian is positive definite, then $\vec{x}_0$ is a local minimum.*

2. *If the Hessian is negative definite, then $\vec{x}_0$ is a local maximum.*

3. *If the Hessian is indefinite, then $\vec{x}_0$ is not a local extreme.*

Proof. The Hessian is a continuous homogeneous function of order 2. Thus it reaches its maximum and minimum on the compact subset $\{\vec{v} \colon \|\vec{v}\| = 1\}$. Suppose the Hessian is positive definite. Then the minimum on the subset is reached at some point, with value $c > 0$. The homogeneity and the fact that $h_f(\vec{v}) \geq c$ for $\vec{v}$ satisfying $\|\vec{v}\| = 1$ implies that $h_f(\vec{v}) \geq c\|\vec{v}\|^2$ for any $\vec{v}$.

By Theorem 6.3.3 and $\nabla f(\vec{x}_0) = 0$, for $0 < \epsilon < \dfrac{c}{2}$, there is $\delta > 0$, such that $\|\Delta \vec{x}\| < \delta$ implies

$$\left| f(\vec{x}) - f(\vec{x}_0) - \frac{1}{2} h_f(\Delta \vec{x}) \right| \leq \epsilon \|\Delta \vec{x}\|^2.$$

By $h_f(\Delta \vec{x}) \geq c \|\Delta \vec{x}\|^2$, we get

$$f(\vec{x}) - f(\vec{x}_0) \geq \frac{1}{2} h_f(\Delta \vec{x}) - \epsilon \|\Delta \vec{x}\|^2 \geq 0.$$

The proof that the negative definiteness implies local maximum is similar.

The argument also shows that if $h_f(\vec{v}) > 0$ for some $\vec{v}$, then $f(\vec{x}_0 + t\vec{v}) > f(\vec{x}_0)$ for sufficiently small $t \neq 0$. Similarly, if $h_f(\vec{w}) < 0$ for some other $\vec{w}$, then $f(\vec{x}_0 + t\vec{w}) < f(\vec{x}_0)$ for sufficiently small $t \neq 0$. Thus if h_f is indefinite, then $\vec{x}_0$ is not a local extreme.

$\square$

Example 6.3.11. We try to find the local extrema of $f = x^3 + y^2 + z^2 + 12xy + 2z$. By solving

$$f_x = 3x^2 + 12y = 0, \quad f_y = 2y + 12x = 0, \quad f_z = 2z + 2 = 0,$$

we find two possible local extrema $\vec{a} = (0, 0, -1)$ and $\vec{b} = (24, -144, -1)$. The Hessian of f at the two points are

$$h_{f,\vec{a}}(u, v, w) = 2v^2 + 2w^2 + 24uv, \quad h_{f,\vec{b}}(u, v, w) = 144u^2 + 2v^2 + 2w^2 + 24uv.$$

Since $h_{f,\vec{a}}(1, 1, 0) = 26 > 0$, $h_{f,\vec{a}}(-1, 1, 0) = -22 < 0$, $\vec{a}$ is not a local extreme. Since $h_{f,\vec{b}} = (12u + v)^2 + v^2 + 2w^2 > 0$ for $(u, v, w) \neq \vec{0}$, $\vec{b}$ is a local minimum.

Example 6.3.12. We study whether the three possibilities in Example 6.3.9 are indeed local extrema. By $z_x = \dfrac{-2x + 2y}{4y + 3z^2 - 1}$, $z_y = \dfrac{2x - 4z - 2}{4y + 3z^2 - 1}$ and the fact that $z_x = z_y = 0$ at the three points, we have

$$z_{xx} = \frac{-2}{4y + 3z^2 - 1}, \quad z_{xy} = \frac{2}{4y + 3z^2 - 1}, \quad z_{yy} = \frac{-8(x - 2z - 1)}{(4y + 3z^2 - 1)^2},$$

at the three points, and get the Hessians

$$h_{(1,1,0)}(u,v) = -\frac{2}{3}u^2 + \frac{4}{3}uv,$$

$$h_{(-1,-1,-1)}(u,v) = u^2 - 2uv,$$

$$h_{(-5,-5,-3)}(u,v) = -\frac{1}{3}u^2 + \frac{2}{3}uv.$$

By taking $(u,v) = (1,1)$ and $(1,-1)$, we see the Hessians are indefinite and none are local extrema.

Example 6.3.13. The only possible local extreme for the function $f(x,y) = x^3 + y^2$ is $(0,0)$, where the Hessian $h_f(u,v) = 2v^2$. Although the Hessian is non-negative, it is not positive definite. In fact, the Hessian is semi-positive definite, and $(0,0)$ is not a local extreme.

The problem is that $h(1,0) = 0$, which corresponds to the fact that the local extreme problem cannot be solved for $f(x,0) = x^3$ by the quadratic approximation alone. What we have is a local extreme problem for which the quadratic approximation is needed in y-direction and the cubic approximation is needed in x-direction.

Exercise 6.3.23. Try your best to determine whether the possible local extrema in Exercises 6.3.19 and 6.3.20 are indeed local maxima or local minima.

Exercise 6.3.24. Suppose a two variable function $f(x,y)$ has continuous derivatives up to second order at (x_0, y_0). Suppose $f_x = f_y = 0$ at (x_0, y_0).

1. Prove that if $f_{xx} > 0$ and $f_{xx}f_{yy} - f_{xy}^2 > 0$, then (x_0, y_0) is a local minimum.

2. Prove that if $f_{xx} < 0$ and $f_{xx}f_{yy} - f_{xy}^2 > 0$, then (x_0, y_0) is a local maximum.

3. Prove that if $f_{xx} \neq 0$ and $f_{xx}f_{yy} - f_{xy}^2 < 0$, then (x_0, y_0) is not a local extreme.

Exercise 6.3.25. Show that the function

$$f(x,y) = \begin{cases} x^2 + y^2 + \dfrac{x^3 y^3}{(x^2+y^2)^3} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases}$$

has first and second order partial derivatives and satisfy $\nabla f(0,0) = 0$ and $h_f(u,v) > 0$ for $(u,v) \neq (0,0)$. However, the function does not have a local extreme at $(0,0)$.

The counterexample is not continuous at $(0,0)$. Can you find a counterexample that is differentiable at $(0,0)$?

Exercise 6.3.26. Suppose a function $f(\vec{x})$ has continuous derivatives up to third order at $\vec{x}_0$, such that all the first and second order partial derivatives vanish at $\vec{x}_0$. Prove that if some third order partial derivative is nonzero, then $\vec{x}_0$ is not a local extreme.

Example 6.3.13 shows that it is possible for the Hessian to be in none of the three cases, and the quadratic approximation is not enough for concluding the local extreme. Like the single variable case, we may consider the finer higher order approximation. However, two problems appear in the multivariable case. The first is that we may need approximations of different

orders in different directions, so we may not have a neat criterion like Proposition 2.3.4. The second is that even if we can use approximations of the same order in all directions, there is no general technique such as completing the squares to convert forms of higher order to some kind of standard form, from which we can see the positive or negative definiteness.

6.3.5 Constrained Extreme

Suppose $G \colon \mathbb{R}^n \to \mathbb{R}^m$ is a map and $G(\vec{x}_0) = \vec{c}$. A function $f(\vec{x})$ has a *local maximum at $\vec{x}_0$ under the constraint $G(\vec{x}) = \vec{c}$* if there is $\delta > 0$, such that

$$G(\vec{x}) = \vec{c}, \ \ \|\vec{x} - \vec{x}_0\| < \delta \implies f(\vec{x}) \le f(\vec{x}_0).$$

In other words, the value of f at $\vec{x}_0$ is biggest among the values of f at points near $\vec{x}_0$ and satisfying $G(\vec{x}) = \vec{c}$. Similarly, f has a *local minimum at $\vec{x}_0$ under the constraint $G(\vec{x}) = \vec{c}$* if there is $\delta > 0$, such that

$$G(\vec{x}) = \vec{c}, \ \ \|\vec{x} - \vec{x}_0\| < \delta \implies f(\vec{x}) \ge f(\vec{x}_0).$$

For example, the hottest and the coldest places in the world are the extrema of the temperature function $T(x, y, z)$ under the constraint $g(x, y, z) = x^2 + y^2 + z^2 = R^2$, where R is the radius of the earth.

Suppose $G \colon \mathbb{R}^n \to \mathbb{R}^m$ is continuously differentiable and $\vec{c}$ is a regular value, then $G(\vec{x}) = \vec{c}$ defines an $(n - m)$-dimensional hypersurface S in $\mathbb{R}^n$. Suppose f is a differentiable function defined on an open subset containing S. Then f has a local maximum at $\vec{x}_0$ under the constraint $G(\vec{x}) = \vec{c}$ if and only if the restriction of f on any curve in S passing through $\vec{x}_0$ has a local maximum at $\vec{x}_0$. In particular, this implies that the derivative $D_{\vec{v}} f$ in the direction of any tangent vector $\vec{v} \in T_{\vec{x}_0} S$ must vanish.

The tangent vectors $\vec{v} \in T_{\vec{x}_0} S$ are characterized by the linear equation $G'(\vec{x}_0)(\vec{v}) = \vec{0}$. Suppose $G = (g_1, g_2, \ldots, g_m)$. Then

$$G'(\vec{x}_0)(\vec{v}) = (\nabla g_1(\vec{x}_0) \cdot \vec{v}, \ \nabla g_2(\vec{x}_0) \cdot \vec{v}, \ \ldots, \ \nabla g_m(\vec{x}_0) \cdot \vec{v}).$$

Thus $D_{\vec{v}} f = \nabla f(\vec{x}_0) \cdot \vec{v}$ vanishes for all tangent vectors $\vec{v}$ means

$$\nabla g_1(\vec{x}_0) \cdot \vec{v} = \nabla g_2(\vec{x}_0) \cdot \vec{v} = \cdots = \nabla g_m(\vec{x}_0) \cdot \vec{v} = 0 \implies \nabla f(\vec{x}_0) \cdot \vec{v} = 0.$$

Linear algebra tells us that this is equivalent to $\nabla f(\vec{x}_0)$ being a linear combination of $\nabla g_1(\vec{x}_0)$, $\nabla g_2(\vec{x}_0)$, $\ldots$, $\nabla g_m(\vec{x}_0)$.

Proposition 6.3.7. *Suppose $G = (g_1, g_2, \ldots, g_m) \colon \mathbb{R}^n \to \mathbb{R}^m$ is continuously differentiable map and $G(\vec{x}_0) = \vec{c}$ is a regular value of G. Suppose a function f is defined near $\vec{x}_0$ and is differentiable at $\vec{x}_0$. If f has a local extreme at $\vec{x}_0$ under the constraint $G(\vec{x}) = \vec{c}$, then $\nabla f(\vec{x}_0)$ is a linear combination of $\nabla g_1(\vec{x}_0)$, $\nabla g_2(\vec{x}_0)$, $\ldots$, $\nabla g_m(\vec{x}_0)$:*

$$\nabla f(\vec{x}_0) = \lambda_1 \nabla g_1(\vec{x}_0) + \lambda_2 \nabla g_2(\vec{x}_0) + \cdots + \lambda_m \nabla g_m(\vec{x}_0).$$

The numbers λ_1, λ_2, ..., λ_m are called the *lagrange multipliers*. The condition can also be written as an equality of linear functionals

$$f'(\vec{x}_0) = \vec{\lambda} \cdot G'(\vec{x}_0).$$

Example 6.3.14. We try to find the possible local extrema of $f = xy^2$ on the circle $g = x^2 + y^2 = 3$. At local extrema, we have

$$\nabla f = (y^2, 2xy) = \lambda \nabla g = \lambda(2x, 2y).$$

Combined with the fact the point lies in the circle, we get

$$y^2 = 2\lambda x, \quad 2xy = 2\lambda y, \quad x^2 + y^2 = 3.$$

If $\lambda = 0$, then the first two equations become $xy = y^2 = 0$, which is the same as $y = 0$. From the third equation, we get two possible local extrema $(\pm\sqrt{3}, 0)$.

If $\lambda \neq 0$, then from the first equation, $y = 0$ would imply $x = 0$, which contradicts with the third equation. By $y \neq 0$ and the first equation, we get $\lambda = x$. Substituting into the first and the third equations, we get $y^2 = 2x^2$ and $x^2 = 1$. This gives us four possible local extrema $(\pm 1, \pm\sqrt{2})$.

Note that since the circle is compact, the function must reach its maximum and minimum. By comparing the values of the function at the six possible local extrema, we find f has absolute maximum 2 at $(1, \pm\sqrt{2})$, and has absolute minimum -2 at $(-1, \pm\sqrt{2})$.

Example 6.3.15. The local extrema problem in Example 6.3.9 can also be considered as the local extrema of the function $f(x, y, z) = z$ under the constraint $g(x, y, z) = x^2 - 2xy + 4yz + z^3 + 2y - z = 1$. At local extrema, we have

$$\nabla f = (0, 0, 1) = \lambda \nabla g = \lambda(2x - 2y, -2x + 4z + 2, 4y + 2z^2 - 1).$$

This is the same as

$$0 = \lambda(2x - 2y), \quad 0 = \lambda(-2x + 4z + 2), \quad 1 = \lambda(4y + 2z^2 - 1).$$

The third equality tells us $\lambda \neq 0$. Thus the first two equalities become $2x - 2y = 0$ and $-2x + 4z + 2 = 0$. These are the conditions we found in Example 6.3.9.

Example 6.3.16. We try to find the possible local extrema of $f = xy + yz + zx$ on the sphere $x^2 + y^2 + z^2 = 1$. By $\nabla f = (y + z, z + x, x + y)$ and $\nabla(x^2 + y^2 + z^2) = (2x, 2y, 2z)$, we have

$$y + z = 2\lambda x, \quad z + x = 2\lambda y, \quad x + y = 2\lambda z, \quad x^2 + y^2 + z^2 = 1$$

at local extrema. Adding the first three equalities together, we get $(\lambda - 1)(x + y + z) = 0$.

If $\lambda = 1$, then $x = y = z$ from the first three equations. Substituting into the fourth equation, we get $3x^2 = 1$ and two possible extrema $\pm\left(\dfrac{1}{\sqrt{3}}, \dfrac{1}{\sqrt{3}}, \dfrac{1}{\sqrt{3}}\right)$.

If $\lambda \neq 1$, then $x + y + z = 0$ and the first three equations become $(2\lambda + 1)x = (2\lambda + 1)y = (2\lambda + 1)z = 0$. Since $(0, 0, 0)$ does not satisfy $x^2 + y^2 + z^2 = 1$, we must have $2\lambda + 1 = 0$, and the four equations is equivalent to $x + y + z = 0$ and $x^2 + y^2 + z^2 = 1$, which is a circle in $\mathbb{R}^3$.

Thus the possible local extrema are $\pm \left(\dfrac{1}{\sqrt{3}}, \dfrac{1}{\sqrt{3}}, \dfrac{1}{\sqrt{3}} \right)$ and the points on the circle $x + y + z = 0$, $x^2 + y^2 + z^2 = 1$.

Note that the sphere is compact and the continuous function f reaches its maximum and minimum on the sphere. Since $f = 1$ at $\pm \left(\dfrac{1}{\sqrt{3}}, \dfrac{1}{\sqrt{3}}, \dfrac{1}{\sqrt{3}} \right)$ and $f = \dfrac{1}{2}[(x + y + z)^2 - (x^2 + y^2 + z^2)] = -\dfrac{1}{2}$ along the circle $x + y + z = 0$, $x^2 + y^2 + z^2 = 1$, we find the maximum is 1 and the minimum is $-\dfrac{1}{2}$.

We can also use $f_x = y + z = 0$, $f_y = z + x = 0$, $f_z = x + y = 0$ to find the possible local extreme $(0, 0, 0)$ of f in the interior $x^2 + y^2 + z^2 < 1$ of the ball $x^2 + y^2 + z^2 \le 1$. By comparing $f(0, 0, 0) = 0$ with the maximum and minimum of f on the sphere, we find f reaches its extrema on the ball at boundary points.

Example 6.3.17. We try to find the possible local extrema of $f = xyz$ on the circle given by $g_1 = x + y + z = 0$ and $g_2 = x^2 + y^2 + z^2 = 6$. The necessary condition is

$$\nabla f = (yz, zx, xy) = \lambda_1 \nabla g_1 + \lambda_2 \nabla g_2 = (\lambda_1 + 2\lambda_2 x, \lambda_1 + 2\lambda_2 y, \lambda_1 + 2\lambda_2 z).$$

Canceling λ_1 from the three equations, we get

$$(x - y)(z + 2\lambda_2) = 0, \quad (x - z)(y + 2\lambda_2) = 0.$$

If $x \ne y$ and $x \ne z$, then $y = -2\lambda_2 = z$. Therefore at least two of x, y, z are equal.

If $x = y$, then the two constraints $g_1 = 0$ and $g_2 = 6$ tell us $2x + z = 0$ and $2x^2 + z^2 = 6$. Canceling z, we get $6x^2 = 6$ and two possible local extrema $(1, 1, -2)$ and $(-1, -1, 2)$. By assuming $x = z$ or $y = z$, we get four more possible local extrema $(1, -2, 1)$, $(-1, 2, -1)$, $(-2, 1, 1)$, $(2, -1, -1)$.

By evaluating the function at the six possible local extrema, we find the absolute maximum 2 is reached at three points and the absolute minimum -2 is reached at the other three points.

Exercise 6.3.27. Find possible local extrema under the constraint.

1. $x^\alpha y^\beta z^\gamma$ under the constraint $x + y + z = a$, $x, y, z > 0$, where $\alpha, \beta, \gamma > 0$.

2. $\sin x \sin y \sin z$ under the constraint $x + y + z = \dfrac{\pi}{2}$.

3. $x_1 + x_2 + \cdots + x_n$ under the constraint $x_1 x_2 \cdots x_n = a$.

4. $x_1 x_2 \cdots x_n$ under the constraint $x_1 + x_2 + \cdots + x_n = a$.

5. $x_1 x_2 \cdots x_n$ under the constraint $x_1 + x_2 + \cdots + x_n = a$, $x_1^2 + x_2^2 + \cdots + x_n^2 = b$.

6. $x_1^p + x_2^p + \cdots + x_n^p$ under the constraint $a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = b$, where $p > 0$.

Exercise 6.3.28. Prove inequalities.

1. $\sqrt[n]{x_1 x_2 \cdots x_n} \le \dfrac{x_1 + x_2 + \cdots + x_n}{n}$ for $x_i \ge 0$.

2. $\dfrac{x_1^p + x_2^p + \cdots + x_n^p}{n} \ge \left(\dfrac{x_1 + x_2 + \cdots + x_n}{n} \right)^p$ for $p \ge 1$, $x_i \ge 0$. What if $0 < p < 1$?

Exercise 6.3.29. Derive Hölder inequality in Exercise 2.2.41 by considering the function $\sum b_i^q$ of $(b_1, b_2, \ldots, b_n)$ under the constraint $\sum a_i b_i = 1$.

Exercise 6.3.30. Suppose A is a symmetric matrix. Prove that if the quadratic form $q(\vec{x}) = A\vec{x} \cdot \vec{x}$ reaches maximum or minimum at $\vec{v}$ on the unit sphere $\{\vec{x}: \|\vec{x}\|_2 = 1\}$, then $\vec{v}$ is an eigenvector of A.

Exercise 6.3.31. Fix the base triangle and the height of a pyramid. When is the total area of the side faces smallest?

Exercise 6.3.32. The intersection of the plane $x + y - z = 0$ and the ellipse $x^2 + y^2 + z^2 - xy - yz - zx = 1$ is an ellipse centered at the origin. Find the lengths of the two axes of the ellipse.

After identifying the possible local extrema by using the linear approximation, we need to use the quadratic approximation to determine whether the candidates are indeed local extrema.

Assume we are in the situation described in Proposition 6.3.7. Further assume that f and $G = (g_1, g_2, \ldots, g_m)$ are continuously second order differentiable. Then we have quadratic approximations

$$p_f(\vec{x}) = f(\vec{x}_0) + \nabla f(\vec{x}_0) \cdot \Delta\vec{x} + \frac{1}{2}h_f(\Delta\vec{x}),$$

$$p_{g_i}(\vec{x}) = g_i(\vec{x}_0) + \nabla g_i(\vec{x}_0) \cdot \Delta\vec{x} + \frac{1}{2}h_{g_i}(\Delta\vec{x}).$$

of f and g_i at the possible local extreme $\vec{x}_0$. The local maximum problem

$$f(\vec{x}) < f(\vec{x}_0) \text{ for } \vec{x} \text{ near } \vec{x}_0, \ \vec{x} \neq \vec{x}_0, \text{ and } g_i(\vec{x}) = c_i$$

for the original function f is approximated by the similar problem

$$p_f(\vec{x}) < p_f(\vec{x}_0) = f(\vec{x}_0) \text{ for } \vec{x} \text{ near } \vec{x}_0, \ \vec{x} \neq \vec{x}_0, \text{ and } p_{g_i}(\vec{x}) = c_i$$

for the quadratic approximations (note that the constraint is also quadratically approximated). The local minimum problem can be similarly approximated.

The quadratic approximation problem is not easy to solve because of the mixture of linear and quadratic terms. To remove the mix, we introduce the function

$$\tilde{f}(\vec{x}) = f(\vec{x}) - \vec{\lambda} \cdot G(\vec{x}) = f(\vec{x}) - \lambda_1 g_1(\vec{x}) - \lambda_2 g_2(\vec{x}) - \cdots - \lambda_m g_m(\vec{x}),$$

where λ_i are the lagrange multipliers obtained from the linear approximation. Since $\tilde{f}(\vec{x}) = f(\vec{x}) - \vec{\lambda} \cdot \vec{c}$ for those $\vec{x}$ satisfying the constraint $G(\vec{x}) = \vec{c}$, and $\vec{\lambda} \cdot \vec{c}$ is a constant, f has a local maximum at $\vec{x}_0$ under the constraint $G(\vec{x}) = \vec{c}$ if and only if $\tilde{f}$ has a local maximum at $\vec{x}_0$ under the constraint $G(\vec{x}) = \vec{c}$. However, the later problem is simpler because the quadratic approximation

$$
\begin{aligned}
p_{\tilde{f}} &= p_f(\vec{x}) - \lambda_1 p_{g_1}(\vec{x}) - \lambda_2 p_{g_2}(\vec{x}) - \cdots - \lambda_m p_{g_m}(\vec{x}) \\
&= \tilde{f}(\vec{x}_0) + (\nabla f(\vec{x}_0) - \lambda_1 \nabla g_1(\vec{x}_0) - \lambda_2 \nabla g_2(\vec{x}_0) - \cdots - \lambda_m \nabla g_m(\vec{x}_0)) \cdot \Delta\vec{x} \\
&\quad + \frac{1}{2}(h_f(\Delta\vec{x}) - \lambda_1 h_{g_1}(\Delta\vec{x}) - \lambda_2 h_{g_2}(\Delta\vec{x}) - \cdots - \lambda_m h_{g_m}(\Delta\vec{x})) \\
&= \tilde{f}(\vec{x}_0) + \frac{1}{2}(h_f(\Delta\vec{x}) - \lambda_1 h_{g_1}(\Delta\vec{x}) - \lambda_2 h_{g_2}(\Delta\vec{x}) - \cdots - \lambda_m h_{g_m}(\Delta\vec{x}))
\end{aligned}
$$

has no first order term.

Since the quadratic approximation of $\tilde{f}$ at $\vec{x}_0$ contains no first order term, the local maximum problem for $\tilde{f}$ at $\vec{x}_0$ is not changed if the constraint is modified by terms of second or higher order (this remains to be rigorously proved). Thus the local maximum problem becomes

$$p_{\tilde{f}}(\vec{x}) < p_{\tilde{f}}(\vec{x}_0) = \tilde{f}(\vec{x}_0) \text{ for } \vec{x} \text{ near } \vec{x}_0, \ \Delta\vec{x} \neq \vec{0}, \text{ and } \nabla g_i(\vec{x}_0) \cdot \Delta\vec{x} = 0.$$

The problem is easy to solve: First we solve the linear problem $g_i(\vec{x}_0) \cdot \Delta\vec{x} = 0$. Then we substitute the solution $\Delta\vec{x}$ into $p_{\tilde{f}}(\vec{x}) < \tilde{f}(\vec{x}_0)$, which is the same as the quadratic problem

$$h_f(\Delta\vec{x}) - \lambda_1 h_{g_1}(\Delta\vec{x}) - \lambda_2 h_{g_2}(\Delta\vec{x}) - \cdots - \lambda_m h_{g_m}(\Delta\vec{x}) < 0.$$

Proposition 6.3.8. *Suppose $G = (g_1, g_2, \ldots, g_m) \colon \mathbb{R}^n \to \mathbb{R}^m$ is a continuously second order differentiable map and $G(\vec{x}_0) = \vec{c}$ is a regular value of G. Suppose f is defined near $\vec{x}_0$ and is continuously second order differentiable at $\vec{x}_0$. Suppose $f'(\vec{x}_0) = \vec{\lambda} \cdot G'(\vec{x}_0)$ for a vector $\vec{\lambda} \in \mathbb{R}^m$. Denote the quadratic form*

$$q = h_f - \vec{\lambda} \cdot h_G = h_f - \lambda_1 h_{g_1} - \lambda_2 h_{g_2} - \cdots - \lambda_m h_{g_m}.$$

1. *If $q(\vec{v})$ is positive definite for vectors $\vec{v}$ satisfying $G'(\vec{x}_0)(\vec{v}) = \vec{0}$, then $\vec{x}_0$ is a local minimum of f under the constraint $G(\vec{x}) = \vec{c}$.*

2. *If $q(\vec{v})$ is negative definite for vectors $\vec{v}$ satisfying $G'(\vec{x}_0)(\vec{v}) = \vec{0}$, then $\vec{x}_0$ is a local maximum of f under the constraint $G(\vec{x}) = \vec{c}$.*

3. *If $q(\vec{v})$ is indefinite for vectors $\vec{v}$ satisfying $G'(\vec{x}_0)(\vec{v}) = \vec{0}$, then $\vec{x}_0$ is not a local extreme of f under the constraint $G(\vec{x}) = \vec{c}$.*

Proof. Since $G(\vec{x}) = \vec{c}$ implies $\tilde{f}(\vec{x}) = f(\vec{x}) - \vec{\lambda} \cdot G(\vec{x}) = f(\vec{x}) - \vec{\lambda} \cdot \vec{c}$, the extreme problem for $f(\vec{x})$ under the constraint $G(\vec{x}) = \vec{c}$ is the same as the extreme problem for $\tilde{f}(\vec{x})$ under the constraint $G(\vec{x}) = \vec{c}$.

Since $\vec{c}$ is a regular value, the constraint $G(\vec{x}) = \vec{c}$ means that some choice of m coordinates $\vec{z}$ of $\vec{x}$ can be written as a continuously differentiable map $H(\vec{y})$ of the other $n - m$ coordinates $\vec{y}$. In other words, after rearranging the coordinates if necessary, we may write $\vec{x} = (\vec{y}, \vec{z})$, and $\vec{x} = (\vec{y}, H(\vec{y}))$ is exactly the solution of the constraint $G(\vec{x}) = \vec{c}$ near $\vec{x}_0 = (\vec{y}_0, \vec{z}_0)$. Then the problem is to determine whether $\vec{y}_0$ is a (unconstrained) local extreme of $\tilde{f}(\vec{y}, H(\vec{y}))$. We will study the problem by substituting the linear approximation of $H(\vec{y})$ near $\vec{y}_0$ into the quadratic approximation of $\tilde{f}(\vec{y}, \vec{z})$ near $(\vec{y}_0, \vec{z}_0)$. In the process, we need to keep track of the remainder terms and show that they do not affect the local extreme problem.

The linear approximation of $\vec{z} = H(\vec{y})$ near $\vec{y}_0$ is

$$\Delta\vec{z} = \vec{z} - \vec{z}_0 = H'(\vec{y}_0)(\Delta\vec{y}) + R_1(\vec{y}), \ R_1(\vec{y}) = o(\|\Delta\vec{y}\|).$$

The quadratic approximation of $\tilde{f}(\vec{x})$ near $\vec{x}_0$ is

$$\tilde{f}(\vec{x}) = p_{\tilde{f}}(\vec{x}) + R_2(\vec{x}) = \tilde{f}(\vec{x}_0) + \frac{1}{2}q(\Delta\vec{x}) + R_2(\vec{x}), \ R_2(\vec{x}) = o(\|\Delta\vec{x}\|^2),$$

where $p_{\tilde{f}}$ has been computed before the statement of the theorem. Denote the linear transform $L(\vec{u}) = (\vec{u}, H'(\vec{y}_0)(\vec{u}))$. Then

$$\Delta\vec{x} = (\Delta\vec{y}, \Delta\vec{z}) = (\Delta\vec{y}, H'(\vec{y}_0)(\Delta\vec{y}) + R_1(\vec{y})) = L(\Delta\vec{y}) + (\vec{0}, R_1(\vec{y}))$$

is the solution of $G(\vec{x}) = \vec{c}$. Substituting into $\tilde{f}(\vec{x})$, we get

$$\tilde{f}(\vec{x}) = \tilde{f}(\vec{x}_0) + \frac{1}{2}q(L(\Delta\vec{y}) + (\vec{0}, R_1(\vec{y}))) + R_2(\vec{x})$$
$$= \tilde{f}(\vec{x}_0) + \frac{1}{2}q(L(\Delta\vec{y})) + \frac{1}{2}q(\vec{0}, R_1(\vec{y})) + b(L(\Delta\vec{y}), (\vec{0}, R_1(\vec{y}))) + R_2(\vec{x}),$$

where b be the symmetric bilinear form that induces q. Since q is quadratic and b is bilinear, the remainder terms may be estimated as follows

$$q(\vec{0}, R_1(\vec{y})) = O(\|(0, R_1(\vec{y}))\|^2) = O(\|R_1(\vec{y})\|^2) = o(\|\Delta\vec{y}\|^2),$$
$$b(L(\Delta\vec{y}), (\vec{0}, R_1(\vec{y}))) = O(\|L(\Delta\vec{y})\|\|(0, R_1(\vec{y}))\|) = O(\|\Delta\vec{y}\|\|R_1(\vec{y})\|) = o(\|\Delta\vec{y}\|^2),$$
$$R_2(\vec{x}) = o(\|\Delta\vec{x}\|^2) = o(\|L(\Delta\vec{y})\|^2 + \|(\vec{0}, R_1(\vec{y}))\|^2) = o(\|\Delta\vec{y}\|^2).$$

Therefore under the constraint $G(\vec{x}) = \vec{c}$, we get

$$\tilde{f}(\vec{x}) = \tilde{f}(\vec{y}, H(\vec{y})) = \tilde{f}(\vec{x}_0) + \frac{1}{2}q(L(\Delta\vec{y})) + R(\vec{y}), \quad R(\vec{y}) = o(\|\Delta\vec{y}\|^2).$$

Then we may apply the proof of Proposition 6.3.6 for the unconstrained extrema to conclude that the nature of the quadratic form $q(L(\vec{u}))$ determines the nature of the local extreme.

What are vectors of the form $L(\vec{u})$? Note that $\Delta\vec{x} = L(\Delta\vec{y})$ is the linear approximation of the solution of the equation $G(\vec{x}) = \vec{c}$ at $\vec{x}_0$. Therefore $\Delta\vec{x} = L(\Delta\vec{y})$ should be the solution of the linear approximation $G'(\vec{x}_0)(\Delta x) = \vec{0}$ of the constraint. See the discussion after the implicit function theorem. Thus vectors of the form $L(\vec{u})$ are exactly the solutions of the linear equation $G'(\vec{x}_0)(\vec{v}) = 0$, and we get the three statements in the proposition.

$\square$

Example 6.3.18. In Example 6.3.14 we found six possible local extrema of $f = xy^2$ on the circle $g = x^2 + y^2 = 3$. To determine whether they are indeed local extrema, we compute the Hessians $h_f(u, v) = 4yuv + 2xv^2$ and $h_g(u, v) = 2u^2 + 2v^2$. Together with $\nabla g = (2x, 2y)$, we have the following table showing what happens at the six points.

(x_0, y_0)	λ	$h_f - \lambda h_g$	$\nabla g \cdot (u, v) = 0$	$h_f - \lambda h_g$ restricted
$(\sqrt{3}, 0)$	0	$2\sqrt{3}v^2$	$2\sqrt{3}v = 0$	0
$(-\sqrt{3}, 0)$	0	$-2\sqrt{3}v^2$	$-2\sqrt{3}v = 0$	0
$(1, \sqrt{2})$	1	$-2u^2 + 4\sqrt{2}uv$	$2u + 2\sqrt{2}v = 0$	$-6u^2$
$(1, -\sqrt{2})$	1	$-2u^2 - 4\sqrt{2}uv$	$2u - 2\sqrt{2}v = 0$	$-6u^2$
$(-1, \sqrt{2})$	-1	$2u^2 + 4\sqrt{2}uv$	$-2u + 2\sqrt{2}v = 0$	$6u^2$
$(-1, -\sqrt{2})$	-1	$2u^2 - 4\sqrt{2}uv$	$-2u - 2\sqrt{2}v = 0$	$6u^2$

Therefore $(1, \pm\sqrt{2})$ are local maxima and $(-1, \pm\sqrt{2})$ are local minima. However, we still do not know whether $(0, \pm\sqrt{3})$ are local extrema (an approximation more refined than the quadratic one is needed).

Example 6.3.19. In Example 6.3.16, we found that $\pm \left(\dfrac{1}{\sqrt{3}}, \dfrac{1}{\sqrt{3}}, \dfrac{1}{\sqrt{3}} \right)$ are the possible extrema of $f = xy + yz + zx$ on the sphere $x^2 + y^2 + z^2 = 1$. To determine whether they are indeed local extrema, we compute the Hessians $h_f(u, v, w) = 2uv + 2vw + 2wu$ and $h_{x^2+y^2+z^2}(u, v, w) = 2u^2 + 2v^2 + 2w^2$. At the two points, we have $\lambda = 1$ and

$$h_f - \lambda h_{x^2+y^2+z^2} = -2u^2 - 2v^2 - 2w^2 + 2uv + 2vw + 2wu.$$

By $\nabla(x^2 + y^2 + z^2) = (2x, 2y, 2z)$, we need to consider the sign of $h_f - \lambda h_{x^2+y^2+z^2}$ for those (u, v, w) satisfying

$$\pm \left(\frac{1}{\sqrt{3}}u + \frac{1}{\sqrt{3}}v + \frac{1}{\sqrt{3}}w \right) = 0.$$

Substituting the solution $w = -u - v$, we get

$$h_f - \lambda h_{x^2+y^2+z^2} = -6u^2 - 6v^2 - 6uv = -6 \left(u + \frac{1}{2}v \right)^2 - \frac{9}{2}v^2,$$

which is negative definite. Thus $\pm \left(\dfrac{1}{\sqrt{3}}, \dfrac{1}{\sqrt{3}}, \dfrac{1}{\sqrt{3}} \right)$ are local maxima.

Exercise 6.3.33. Determine whether the possible local extrema in Exercise 6.3.27 are indeed local maxima or local minima.

6.3.6 Additional Exercise

Convex Subset and Convex Function

A subset $A \subset \mathbb{R}^n$ is convex if $\vec{x}, \vec{y} \in A$ implies the straight line connecting $\vec{x}$ and $\vec{y}$ still lies in A. In other words, $(1 - \lambda)\vec{x} + \lambda\vec{y} \in A$ for any $0 < \lambda < 1$.

A function $f(\vec{x})$ defined on a convex subset A is convex if

$$0 < \lambda < 1 \implies (1 - \lambda)f(\vec{x}) + \lambda f(\vec{y}) \geq f((1 - \lambda)\vec{x} + \lambda\vec{y}). \qquad (6.3.16)$$

Exercise 6.3.34. For any $\lambda_1, \lambda_2, \ldots, \lambda_k$ satisfying $\lambda_1 + \lambda_2 + \cdots + \lambda_k = 1$ and $0 \leq \lambda_i \leq 1$, extend *Jensen inequality* in Exercise 2.3.40 to multivariable convex functions

$$f(\lambda_1\vec{x}_1 + \lambda_2\vec{x}_2 + \cdots + \lambda_k\vec{x}_k) \leq \lambda_1 f(\vec{x}_1) + \lambda_2 f(\vec{x}_2) + \cdots + \lambda_k f(\vec{x}_k). \qquad (6.3.17)$$

Exercise 6.3.35. Prove that a function $f(\vec{x})$ is convex if and only if its restriction on any straight line is convex.

Exercise 6.3.36. Prove that a function $f(\vec{x})$ is convex if and only if for any linear function $L(\vec{x}) = a + b_1 x_1 + b_2 x_2 + \cdots + b_n x_n$, the subset $\{\vec{x} \colon f(\vec{x}) \leq L(\vec{x})\}$ is convex.

Exercise 6.3.37. Extend Proposition 2.3.5 to multivariable: A function $f(\vec{x})$ defined on an open convex subset is convex if and only if for any $\vec{z}$ in the subset, there is a linear function $K(\vec{x})$, such that $K(\vec{z}) = f(\vec{z})$ and $K(\vec{x}) \leq f(\vec{x})$.

Exercise 6.3.38. Prove that any convex function on an open convex subset is continuous.

Exercise 6.3.39. Prove that a second order continuously differentiable function $f(\vec{x})$ on an open convex subset is convex if and only if the Hessian is semipositive definite: $h_f(\vec{v}) \geq 0$ for any $\vec{v}$.

Snell's Law

Fermat's principle says that light travels along the path of least traveling time. Because of the principle, the direction of light changes as it enter from air to water, as shown in the picture. The phenomenon is called *refraction*.
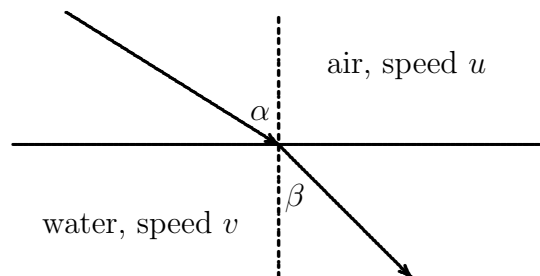
Figure 6.5: Snell's Law

Exercise 6.3.40. Suppose a light travels at the respective speed of u and v in the air and the water. Use Fermat's principle to prove that the angle α of incidence and the angle β of refraction are related by *Snell's law*: $\dfrac{u}{\sin \alpha} = \dfrac{v}{\sin \beta}$.

Laplacian

The *Laplacian* of a function $f(\vec{x})$ is

$$\Delta f = \frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} + \cdots + \frac{\partial^2 f}{\partial x_n^2}. \tag{6.3.18}$$

Functions satisfying the *Laplace equation* $\Delta f = 0$ are called *harmonic*.

Exercise 6.3.41. Prove that $\Delta(f+g) = \Delta f + \Delta g$ and $\Delta(fg) = g\Delta f + f\Delta g + 2\nabla f \cdot \nabla g$.

Exercise 6.3.42. Suppose a function $f(\vec{x}) = u(r)$ depends only on the Euclidean norm $r = \|\vec{x}\|_2$ of the vector. Prove that $\Delta f = u''(r) + (n-1)r^{-1}u'(r)$ and find the condition for the function to be harmonic.

Exercise 6.3.43. Derive the Laplacian in $\mathbb{R}^2$

$$\Delta f = \frac{\partial^2 f}{\partial r^2} + \frac{1}{r}\frac{\partial f}{\partial r} + \frac{1}{r^2}\frac{\partial^2 f}{\partial \theta^2}. \tag{6.3.19}$$

in the polar coordinates. Also derive the Laplacian in $\mathbb{R}^3$

$$\Delta f = \frac{\partial^2 f}{\partial r^2} + \frac{2}{r}\frac{\partial f}{\partial r} + \frac{1}{r^2}\frac{\partial^2 f}{\partial \theta^2} + \frac{\cos \phi}{r^2 \sin \phi}\frac{\partial f}{\partial \theta} + \frac{1}{r^2 \sin^2 \phi}\frac{\partial^2 f}{\partial \theta^2}. \tag{6.3.20}$$

in the spherical coordinates.

Exercise 6.3.44. Let $\vec{x} = F(\vec{y})\colon \mathbb{R}^n \to \mathbb{R}^n$ be an orthogonal change of variable (see Exercise (6.2.15)). Let $|J| = \|\vec{x}_{y_1}\|_2 \|\vec{x}_{y_2}\|_2 \cdots \|\vec{x}_{y_n}\|_2$ be the absolute value of the determinant of the Jacobian matrix. Prove the Laplacian is

$$\Delta f = \frac{1}{|J|} \left(\frac{\partial}{\partial y_1} \left(\frac{|J|}{\|\vec{x}_{y_1}\|_2^2} \frac{\partial f}{\partial y_1} \right) + \frac{\partial}{\partial y_2} \left(\frac{|J|}{\|\vec{x}_{y_2}\|_2^2} \frac{\partial f}{\partial y_2} \right) + \cdots + \frac{\partial}{\partial y_n} \left(\frac{|J|}{\|\vec{x}_{y_n}\|_2^2} \frac{\partial f}{\partial y_n} \right) \right).$$
(6.3.21)

In particular, if the change of variable is orthonormal (i.e., $\|\vec{x}_{y_i}\|_2 = 1$), then the Laplacian is not changed.

Euler Equation

Exercise 6.3.45. Suppose a function f is differentiable away from $\vec{0}$. Prove that f is homogeneous of degree α if and only if it satisfies the *Euler equation*

$$x_1 f_{x_1} + x_2 f_{x_2} + \cdots + x_n f_{x_n} = \alpha f. \tag{6.3.22}$$

What about a multihomogeneous function?

Exercise 6.3.46. Extend the characterization of homogeneous function (6.3.22) to high order derivatives

$$\sum_{1 \leq i_1, i_2, \ldots, i_k} x_{i_1} x_{i_2} \cdots x_{i_k} \frac{\partial^k f}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_k}} = \alpha(\alpha - 1) \cdots (\alpha - k + 1) f. \tag{6.3.23}$$

Exercise 6.3.47. Prove that a change of variable $\vec{x} = H(\vec{z})\colon \mathbb{R}^n \to \mathbb{R}^n$ preserves $\sum x_i f_{x_i}$

$$z_1 (f \circ H)_{z_1} + z_2 (f \circ H)_{z_2} + \cdots + z_n (f \circ H)_{z_n} = x_1 f_{x_1} + x_2 f_{x_2} + \cdots + x_n f_{x_n}.$$

for any differentiable function $f(\vec{x})$ if and only if it preserves the scaling

$$H(c\vec{z}) = cH(\vec{z}) \text{ for any } c > 0.$$

Chapter 7

Multivariable Integration

7.1 Riemann Integration

The integration of multivariable functions is closely tied to the concept of volume in two ways. First we can only integrate functions defined on subsets with volumes. Second, the integral of an n-variable function is the volume of an $(n + 1)$-dimensional subset (the region enclosed by the graph of the function). The connection makes the theory of integration equivalent to the theory of volume.

The theory of volume can be developed from the obvious positivity, additivity, and translation invariant properties. The finite additivity leads to the Jordan volume theory and the Riemann integral. In the future, the countable additivity will lead to the Lebesgue measure theory and the Lebesgue integral.

7.1.1 Volume in Euclidean Space

It is easy to imagine how to extend the Riemann integration to multivariable functions on rectangles

$$I = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n] \subset \mathbb{R}^n. \tag{7.1.1}$$

However, integration on rectangles alone is too restrictive. The subsets on which the Riemann integration can be defined should be subsets with volumes. After all, the integral of the constant function 1 is supposed to be the volume of the underlying subset. Thus a theory of volume needs to be established.

Let $A \subset \mathbb{R}^n$ be a bounded subset. Then $A \subset [-M, M]^n$ for some big M. By taking a partition of $[-M, M]$ for each coordinate, we get a partition P of $[-M, M]^n$ consisting of rectangles that intersect only along boundaries. The size of the partition can be measured by

$$\|P\| = \max\{d(I) \colon I \in P\}, \ \ d(I) = \max\{b_1 - a_1, b_2 - a_2, \ldots, b_n - a_n\}.$$

For any partition P, the unions of rectangles

$$A_P^+ = \cup\{I \colon I \in P, I \cap A \neq \emptyset\}, \ \ A_P^- = \cup\{I \colon I \in P, I \subset A\} \tag{7.1.2}$$

form the outer and the inner approximations of A, satisfying

$$A_P^+ \supset A \supset A_P^-.$$

Define the volume of the rectangle (7.1.1) by

$$\mu([a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]) = (b_1 - a_1)(b_2 - a_2) \cdots (b_n - a_n).$$

Then define

$$\mu(A_P^+) = \sum_{I \in P, I \cap A \neq \emptyset} \mu(I), \ \ \mu(A_P^-) = \sum_{I \in P, I \subset A} \mu(I). \tag{7.1.3}$$

We have

$$\mu(A_P^+) - \mu(A_P^-) = \sum_{I \in P, I \cap A \neq \emptyset, I - A \neq \emptyset} \mu(I),$$

and the volume of A is supposed to be between $\mu(A_P^+)$ and $\mu(A_P^-)$.

Definition 7.1.1. A subset $A \subset \mathbb{R}^n$ *has volume* (or is *Jordan measurable*) if for any $\epsilon > 0$, there is $\delta > 0$, such that

$$\|P\| < \delta \implies \mu(A_P^+) - \mu(A_P^-) < \epsilon.$$

The *volume* (or the *Jordan measure*) of A is the unique number $\mu(A)$ satisfying $\mu(A_P^+) \geq \mu(A) \geq \mu(A_P^-)$ for all partitions.

Example 7.1.1. Consider $A = (a, b) \subset \mathbb{R}$. For any $\epsilon > 0$, take P to be $a < a + \epsilon < b - \epsilon < b$. Then $\mu(A_P^+) = \mu([a, b]) = b - a$ and $\mu(A_P^-) = \mu([a + \epsilon, b - \epsilon]) = b - a - 2\epsilon$. This suggests that A has volume (or length) $b - a$.

Note that we cannot yet conclude that A has volume according to the definition, because we have not verified that $\mu(A_P^+) - \mu(A_P^-)$ is small for *all* P with small $\|P\|$. The rigorous reason is given by Proposition 7.1.2.

Consider $B = \{n^{-1} \colon n \in \mathbb{N}\}$. For any N and $\epsilon < \dfrac{1}{2N^2}$, take P to be $0 < N^{-1} + \epsilon < (N - 1)^{-1} - \epsilon < (N - 1)^{-1} + \epsilon < \cdots < 2^{-1} - \epsilon < 2^{-1} + \epsilon < 1 - \epsilon < 1$. Then $\mu(A_P^+) = N^{-1} + 2N\epsilon < 2N^{-1}$ and $\mu(A_P^-) = 0$. Thus A should have volume 0.

Example 7.1.2. Let A be the set of rational numbers in $[0, 1]$. Then any interval $[a, b]$ with $a < b$ contains points inside and outside A. Therefore for any partition P of $[0, 1]$, we have $\mu(A_P^+) = 1$ and $\mu(A_P^-) = 0$. We conclude that A has no volume.

Example 7.1.3. Consider a non-negative bounded single variable function $f(x)$ on $[a, b]$. The subset

$$G(f) = \{(x, y) \colon a \leq x \leq b, 0 \leq y \leq f(x)\} \subset \mathbb{R}^2$$

is the subset under the graph of the function. For any partition $P \colon a = x_0 < x_1 < \cdots < x_k = b$ of $[a, b]$ and any partition $Q \colon 0 = y_0 < y_1 < \cdots < y_l$ of the y-axis, we have

$$G(f)_{P \times Q}^+ = \cup [x_{i-1}, x_i] \times [0, y_q], \quad G(f)_{P \times Q}^- = \cup [x_{i-1}, x_i] \times [0, y_s].$$

where y_q is the smallest $y_j > f$ on $[x_{i-1}, x_i]$ and y_s is the biggest $y_j \leq f$ on $[x_{i-1}, x_i]$. The description of y_q and y_s tells us

$$y_q \geq \sup_{[x_{i-1}, x_i]} f(x) \geq y_{q-1} \geq y_q - \|Q\|, \quad y_s \leq \inf_{[x_{i-1}, x_i]} f(x) \leq y_{s+1} \leq y_s + \|Q\|,$$

Therefore $\mu\left(G(f)_{P \times Q}^+\right) = \sum y_q \Delta x_i$ satisfies

$$\mu\left(G(f)_{P \times Q}^+\right) \geq \sum \sup_{[x_{i-1}, x_i]} f(x) \Delta x_i \geq \mu\left(G(f)_{P \times Q}^+\right) - \|Q\|(b - a).$$

Similarly, we have

$$\mu\left(G(f)_{P \times Q}^-\right) \leq \sum \inf_{[x_{i-1}, x_i]} f(x) \Delta x_i \leq \mu\left(G(f)_{P \times Q}^-\right) + \|Q\|(b - a).$$

Thus

$$\left| \mu\left(G(f)_{P \times Q}^+\right) - \mu\left(G(f)_{P \times Q}^-\right) - \sum \omega_{[x_{i-1}, x_i]}(f) \Delta x_i \right| \leq 2\|Q\|(b - a).$$

This implies that $G(f)$ has volume (or area) if and only if f is Riemann integrable. Moreover, the volume of $G(f)$ is $\displaystyle\int_a^b f(x) dx$.

The following gives some criteria for subsets to have volume. Recall that $\vec{x}$ is a boundary point of A if and only if for any $\epsilon > 0$, there are $\vec{a} \in A$ and $\vec{b} \notin A$, such that $\|\vec{x} - \vec{a}\| < \epsilon$ and $\|\vec{x} - \vec{b}\| < \epsilon$.

Proposition 7.1.2. *The following are equivalent for a bounded subset A.*

1. *A has volume.*

2. *For any $\epsilon > 0$, there is a partition P, such that $\mu(A_P^+) - \mu(A_P^-) < \epsilon$.*

3. *For any $\epsilon > 0$, there are rectangles I_1, I_2, $\dots$, I_k, such that*

$$\partial A \subset I_1 \cup I_2 \cup \cdots \cup I_k, \quad \mu(I_1) + \mu(I_2) + \cdots + \mu(I_k) < \epsilon. \qquad (7.1.4)$$

Proof. The first property clearly implies the second.

Now assume the second property holds. A boundary point $\vec{x}$ of A lies either on the boundary of some $I \in P$, or inside the interior of some $I \in P$. If $\vec{x}$ is in the interior of I, then I contains points inside as well as outside A. In other words, I appears in the outer approximation A_P^+ but not in the inner approximation A_P^-. We conclude that

$$\partial A \subset I_1 \cup I_2 \cup \cdots \cup I_k,$$

where each I_i is either part of the boundary of some $I \in P$ or $I_i \in P$ appears in A_P^+ but not in A_P^-. Since the boundary parts of rectangles have volume 0, we have

$$\mu(I_1) + \mu(I_2) + \cdots + \mu(I_k) \leq \mu(A_P^+) - \mu(A_P^-) < \epsilon.$$

This proves the third property.

Finally we assume the third property. For I in (7.1.1) and $\delta > 0$, denote

$$I^\delta = \{\vec{y} \colon \|\vec{y} - \vec{x}\|_\infty \leq \delta \text{ for some } \vec{x} \in I\} = \prod [a_i - \delta, b_i + \delta].$$

Then for sufficiently small δ, we still have

$$\mu(I_1^\delta) + \mu(I_2^\delta) + \cdots + \mu(I_k^\delta) < \epsilon.$$

Suppose P is a partition with $\|P\| < \delta$. For any rectangle $I \in P$ that appears in A_P^+ but not in A_P^-, we can find $\vec{a} \in I \cap A$ and $\vec{b} \in I - A$. Then some point on the straight line connecting $\vec{a}$ and $\vec{b}$ is on the boundary ∂A of A (see Exercise 5.1.23). Denote the point by $\vec{c} = (1 - \tau)\vec{a} + \tau\vec{b}$, $0 \leq \tau \leq 1$. Then for any $\vec{x} \in I$, we have

$$\|\vec{x} - \vec{c}\|_\infty = \|(1 - \tau)(\vec{x} - \vec{a}) + \tau(\vec{x} - \vec{b})\|_\infty$$
$$\leq (1 - \tau)\|\vec{x} - \vec{a}\|_\infty + \tau\|\vec{x} - \vec{b}\|_\infty \leq (1 - \tau)\|P\| + \tau\|P\| < \delta.$$

By $\vec{c} \in \partial A \subset I_1 \cup I_2 \cup \cdots \cup I_k$, the estimation above implies $\vec{x} \in I_1^\delta \cup I_2^\delta \cup \cdots \cup I_k^\delta$. Therefore $I \subset I_1^\delta \cup I_2^\delta \cup \cdots \cup I_k^\delta$, and

$$\mu(A_P^+) - \mu(A_P^-) = \sum_{I \cap A \neq \emptyset, I - A \neq \emptyset} \mu(I) \leq \mu(I_1^\delta) + \mu(I_2^\delta) + \cdots + \mu(I_k^\delta) < \epsilon.$$

This proves the first property.

We remark that, in the proof above, we did not use the fact that the subsets I in P are rectangles. The observation will be used in the proof of Proposition 7.1.4.

<div style="text-align: right">□</div>

Define the *outer volume* and the *inner volume*

$$\mu^+(A) = \inf_P \mu(A_P^+), \ \ \mu^-(A) = \sup_P \mu(A_P^-).$$

Since $\mu(A_P^+) \geq \mu(A_Q^+) \geq \mu(A_Q^-) \geq \mu(A_P^-)$ for a refinement Q of P, and any two partitions have a common refinement, we can easily get $\mu^+(A) \geq \mu^-(A)$. By Proposition 7.1.2, A has volume if and only if $\mu^+(A) = \mu^-(A)$.

The outer and inner volumes are clearly monotone

$$A \subset B \implies \mu^+(A) \leq \mu^+(B), \ \ \mu^-(A) \leq \mu^-(B).$$

Moreover, the outer volume is clearly *subadditive*

$$\mu^+(A_1 \cup A_2 \cup \cdots \cup A_k) \leq \mu^+(A_1) + \mu^+(A_2) + \cdots + \mu^+(A_k). \qquad (7.1.5)$$

Consequently, the volume is monotone and subadditive.

Proposition 7.1.3. *A subset has volume if and only if it boundary has volume 0. Moreover, if A and B have volumes, then $A \cup B$, $A \cap B$ and $A - B$ have volumes and satisfy the additive property*

$$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B). \qquad (7.1.6)$$

The additivity (7.1.6) means that the subadditivity property (7.1.5) becomes an equality when the subsets have volumes and are disjoint.

Proof. Suppose A has volume. By the criterion (7.1.4) in Proposition 7.1.2, for any $\epsilon > 0$, we have

$$\mu^+(\partial A) \leq \mu^+(I_1 \cup I_2 \cup \cdots \cup I_k) \leq \mu^+(I_1) + \mu^+(I_2) + \cdots + \mu^+(I_k)$$
$$= \mu(I_1) + \mu(I_2) + \cdots + \mu(I_k) < \epsilon.$$

Therefore we get $\mu^+(\partial A) = 0$. Then by $\mu^-(\partial A) \leq \mu^+(\partial A)$, we see that ∂A has volume 0. Conversely, if ∂A has volume 0, then we can derive the third property in Proposition 7.1.2 from the definition of $\mu^+(\partial A) = 0$. This completes the proof that A has volume if and only if ∂A has volume 0.

The boundaries of $A \cup B$, $A \cap B$ and $A - B$ are all contained in $\partial A \cup \partial B$. If A and B have volumes, then by $\mu^+(\partial A \cup \partial B) \leq \mu^+(\partial A) + \mu^+(\partial B) = 0$, we see that all the boundaries have volume 0, so that the three subsets have volumes.

Suppose A and B are disjoint. Then for any partition P, we have $A_P^- \cup B_P^- \subset (A \cup B)_P^-$, and A_P^-, B_P^- are disjoint. This leads to $\mu^-(A) + \mu^-(B) \leq \mu^-(A \cup B)$. On the other hand, we always have $\mu^+(A) + \mu^+(B) \geq \mu^+(A \cup B)$ by

the subadditivity. Therefore in case A and B have volumes, we get $\mu(A \cup B) = \mu(A) + \mu(B)$.

In general, suppose A and B have volumes. Then $A - B$, $B - A$, and $A \cap B$ have volumes and are disjoint. Then (7.1.6) follows from

$$\mu(A \cup B) = \mu(A - B) + \mu(B - A) + \mu(A \cap B),$$
$$\mu(A) = \mu(A - B) + \mu(A \cap B),$$
$$\mu(B) = \mu(B - A) + \mu(A \cap B).$$

$\square$

Exercise 7.1.1. Find subsets A, B of $\mathbb{R}$, such that A and B do not have volumes, but $A \cup B$, $A \cap B$ have volumes. Is it possible for $A - B$ and $B - A$ also not to have volumes?

Exercise 7.1.2. Suppose A has volume and $A - \partial A \subset B \subset \bar{A}$. Prove that B also has volume and $\mu(B) = \mu(A)$. In particular, the interior $A - \partial A$ and the closure $\bar{A}$ have the same volume as A. Conversely, construct a subset A such that both $A - \partial A$ and $\bar{A}$ have volumes, but $\mu(A - \partial A) \neq \mu(\bar{A})$, so that A does not have volume.

Exercise 7.1.3. Suppose $A \subset \mathbb{R}^m$ and $B \subset \mathbb{R}^n$ are bounded.

1. Prove that $\mu^+(A \times B) = \mu^+(A)\mu^+(B)$.

2. Prove that if A has volume 0, then $A \times B$ has volume 0.

3. Prove that if $\mu^+(A) \neq 0$ and $\mu^+(B) \neq 0$, then $A \times B$ has volume if and only if A and B have volumes.

4. Prove that if A and B have volumes, then $\mu(A \times B) = \mu(A)\mu(B)$.

Exercise 7.1.4. Prove that $\mu^+(\partial A) = \mu^+(A) - \mu^-(A)$. This gives another proof that A has volume if and only if ∂A has volume 0.

Exercise 7.1.5. Determine whether the subsets have volumes and compute the volumes.

1. $\{m^{-1} + n^{-1} \colon m, n \in \mathbb{N}\}$.

2. $[0, 1] - \mathbb{Q}$.

3. $\{(x, x^2) \colon 0 \leq x \leq 1\}$.

4. $\{(x, y) \colon 0 \leq x \leq 1, 0 \leq y \leq x^2\}$.

5. $\{(x, y) \colon 0 \leq x \leq y^2, 0 \leq y \leq 1\}$.

6. $\{(n^{-1}, y) \colon n \in \mathbb{N}, 0 < y < 1 + n^{-1}\}$.

7. $\{(x, y) \colon |x| < 1, |y| < 1, x \in \mathbb{Q} \text{ or } y \in \mathbb{Q}\}$.

8. $\{(x, y) \colon |x| < 1, |y| < 1, x \neq y\}$.

Exercise 7.1.6. Prove that any n-gon in $\mathbb{R}^2$ has volume.

So far the partitions are made of rectangles. To allow more flexibility, define a *general partition* of a subset B to be a finite collection P of subsets such that

1. Each $I \in P$ has volume.

2. $\mu(I \cap J) = 0$ for $I \neq J$ in P.

3. $B = \cup_{I \in P} I$.

The subset B must have volume and $\mu(B) = \sum_{I \in P} \mu(I)$. Extend the size of the partition by

$$\|P\| = \max\{d(I) \colon I \in P\}, \quad d(I) = \sup\{\|\vec{x} - \vec{y}\|_\infty \colon \vec{x}, \vec{y} \in I\}.$$

For a subset $A \subset B$, the outer and the inner approximations A_P^+, A_P^- may also be defined by (7.1.2).

Proposition 7.1.4. *Suppose B has volume and $A \subset B$ is a subset. The following are equivalent.*

1. *A has volume.*

2. *For any $\epsilon > 0$, there is $\delta > 0$, such that $\|P\| < \delta$ for a general partition P implies $\mu(A_P^+) - \mu(A_P^-) < \epsilon$.*

3. *For any $\epsilon > 0$, there is a general partition P of B, such that $\mu(A_P^+) - \mu(A_P^-) < \epsilon$.*

4. *For any $\epsilon > 0$, there are subsets A^+ and A^- with volumes, such that $A^- \subset A \subset A^+$ and $\mu(A^+) - \mu(A^-) < \epsilon$.*

Proof. To prove the first property implies the second, we use the last part of the proof of Proposition 7.1.2. As remarked at the end of the proof, that part of the proof does not make use of the assumption that the subsets in P are rectangles. Therefore the proof can be applied to any general partition P. The result is that if A has the third property in Proposition 7.1.2, then for any $\epsilon > 0$, there is $\delta > 0$, such that

$$\|P\| < \delta \implies \mu(A_P^+) - \mu(A_P^-) < \epsilon$$

for general partitions P. In other words, the third property in Proposition 7.1.2 implies the second property in the current theorem. Since A having volume implies the third property in Proposition 7.1.2, we conclude that the first property implies the second.

The second property also implies the first by restricting to rectangular partitions. Therefore the first two properties are equivalent.

The second property clearly implies the third, and the third implies the fourth by taking $A^+ = A_P^+$ and $A^- = A_P^-$.

Finally, assume the fourth property. Since A^+ and A^- have volumes, there is $\delta > 0$, such that if a partition P satisfies $\|P\| < \delta$, then

$$\mu((A^+)_P^+) - \mu((A^+)_P^-) < \epsilon, \quad \mu((A^-)_P^+) - \mu((A^-)_P^-) < \epsilon.$$

By $A_P^+ \subset (A^+)_P^+$ and $(A^+)_P^- \subset A^+$, we get

$$\mu(A_P^+) - \mu(A^+) \leq \mu((A^+)_P^+) - \mu((A^+)_P^-) < \epsilon.$$

Similarly, we get $\mu(A^-) - \mu(A_P^-) < \epsilon$. Then

$$\mu(A_P^+) - \mu(A_P^-) = (\mu(A_P^+) - \mu(A^+)) + (\mu(A^+) - \mu(A^-)) + (\mu(A^-) - \mu(A_P^-)) < 3\epsilon.$$

This shows that A has volume.

$\square$

Exercise 7.1.7. Prove that the disks in $\mathbb{R}^2$ have volumes by comparing the inscribed and circumscribed regular n-gons.

7.1.2   Riemann Sum

Let $f(\vec{x})$ be a function defined on a subset $A \subset \mathbb{R}^n$ with volume. For any general partition P of A and choices $\vec{x}_I^* \in I$, define the *Riemann sum*

$$S(P, f) = \sum_{I \in P} f(\vec{x}_I^*)\mu(I). \tag{7.1.7}$$

Definition 7.1.5. A function $f(\vec{x})$ is *Riemann integrable* on a subset A with volume, with *integral* J, if for any $\epsilon > 0$, there is $\delta > 0$, such that

$$\|P\| < \delta \implies |S(P, f) - J| < \epsilon.$$

We denote

$$\int_A f(\vec{x})d\mu = J.$$

Sometimes we also use dV (dA for $n = 2$) or $dx_1 dx_2 \cdots dx_n$ in place of $d\mu$.

Riemann integrable functions might be unbounded if $\mu(A) = 0$ (or some "separated part" of A has volume 0). To avoid complicated statements, we will always insist Riemann integrable functions to be bounded. For a bounded function f, define the *oscillation*

$$\omega_A(f) = \sup_{\vec{x} \in A} f(\vec{x}) - \inf_{\vec{x} \in A} f(\vec{x}) \tag{7.1.8}$$

as in the single variable case. Then we have

$$\vec{c} \in A \implies |f(\vec{c})\mu(A) - S(P, f)| \leq \omega_A(f)\mu(A). \tag{7.1.9}$$

For a function f defined on $A \subset \mathbb{R}^n$, the subset ($y \in [f(\vec{x}), 0]$ if $f(\vec{x}) \leq 0$)

$$G_A(f) = \{(\vec{x}, y) \colon \vec{x} \in A, y \in [0, f(\vec{x})]\} \subset \mathbb{R}^{n+1} \tag{7.1.10}$$

is the region between the graph of the function and the $\vec{x}$-plane.

Proposition 7.1.6. *Suppose f is a bounded function on a subset A with volume. Then the following are equivalent.*

1. *f is Riemann integrable on A.*

2. *For any $\epsilon > 0$, there is $\delta > 0$, such that $\|P\| < \delta$ for a general partition P implies $\sum_{I \in P} \omega_I(f)\mu(I) < \epsilon$.*

3. *For any $\epsilon > 0$, there is a general partition P of A, such that $\sum_{I \in P} \omega_I(f)\mu(I) < \epsilon$.*

4. *The subset $G_A(f)$ has volume in $\mathbb{R}^{n+1}$.*

Proof. The proof that the first two properties are equivalent is the same as the single variable case. The second property clearly implies the third.

Let P be a general partition of A. Let $|f| < M$ and $Q \colon -M = y_0 < y_1 < y_2 < \cdots < y_k = M$ be a partition of the y-axis. Then

$$G_A(f)^+_{P \times Q} = \cup_{I \in P} I \times [y_p, y_q], \quad G_A(f)^-_{P \times Q} = \cup_{I \in P} I \times [y_r, y_s],$$

where $[y_p, y_q]$ and $[y_r, y_s]$ are determined as follows.

If $f \geq 0$ on I, then y_p is the biggest $y_j < 0$ and y_q is the smallest $y_j \geq f$ on I. Similar to the discussion in Example 7.1.3, this implies

$$y_q \geq \sup_I f \geq y_{q-1} \geq y_q - \|Q\|, \quad -\|Q\| \leq y_p < 0 \leq y_{p+1} \leq \|Q\|.$$

Moreover, $[y_r, y_s]$ is the biggest interval satisfying $0 \leq y_r < y_s \leq f$ on I. Note that $[y_r, y_s]$ exists if and only if $f \geq y_{p+2}$ on I. If $[y_r, y_s]$ exists, then $y_r = y_{p+1}$, and similar to the discussion in Example 7.1.3, we have

$$y_s \leq \inf_I f \leq y_{s+1} \leq y_s + \|Q\|.$$

Therefore

$$|\mu(I \times [y_p, y_q]) - \mu(I \times [y_r, y_s]) - \omega_I(f)\mu(I)|$$
$$=|(y_q - y_p) - (y_s - y_r) - \omega_I(f)|\mu(I)$$
$$\leq(|y_q - \sup_I f| + |y_s - \inf_I f| + |y_{p+1} - y_p|)\mu(I) \leq 3\|Q\|\mu(I).$$

If $[y_r, y_s]$ does not exist, then

$$0 \leq \inf_I f \leq y_{p+2} < y_{p+2} - y_p \leq 2\|Q\|,$$

and $I \times [y_r, y_s] = \emptyset$ in the formula for $G_A(f)^-_{P \times Q}$. Therefore

$$|\mu(I \times [y_p, y_q]) - \mu(I \times [y_r, y_s]) - \omega_I(f)\mu(I)|$$
$$=|y_q - y_p - \omega_I(f)|\mu(I)$$
$$\leq(|y_q - \sup_I f| + |y_p| + |\inf_I f|)\mu(I) \leq 4\|Q\|\mu(I).$$

If $f \leq 0$ on I, then similar argument can be carried out, and we also get

$$|\mu(I \times [y_p, y_q]) - \mu(I \times [y_r, y_s]) - \omega_I(f)\mu(I)| \leq 4\|Q\|\mu(I).$$

If f can be positive and negative on I, then y_p is the biggest $y_j < f$ on I, y_q is the smallest $y_j > f$ on I, and $I \times [y_r, y_s] = \emptyset$. We then have

$$y_p \leq \inf_I f \leq y_{p+1} \leq y_p + \|Q\|, \;\; y_q \geq \sup_I f \geq y_{q-1} \geq y_q - \|Q\|,$$

and

$$
\begin{aligned}
&|\mu(I \times [y_p, y_q]) - \mu(I \times [y_r, y_s]) - \omega_I(f)\mu(I)| \\
=&|y_q - y_p - \omega_I(f)|\mu(I) \\
\leq&(|y_q - \sup_I f| + |y_p - \inf_I f|)\mu(I) \leq 2\|Q\|\mu(I).
\end{aligned}
$$

Combining the estimations of the three cases together, we have

$$
\left| \mu\left(G_A(f)^+_{P \times Q}\right) - \mu\left(G_A(f)^-_{P \times Q}\right) - \sum_I \omega_I(f)\mu(I) \right|
$$
$$
\leq \sum_{I \in P} |(y_q - y_p) - (y_s - y_r) - \omega_I(f)|\mu(I) \leq 4\|Q\|\mu(A).
$$

By taking $\|Q\|$ sufficiently small, and with the help of Proposition 7.1.4, we find the second, the third, and the fourth statements are equivalent.

$\square$

Now assume f is integrable on A and we compare the Riemann sum $S(P, |f|) = \sum_{I \in P} |f(\vec{x}_I^*)|\mu(I)$ with the volume $\mu(G_A(f)^+_{P \times Q})$. In case $f \geq 0$ on I, we have

$$0 \leq (y_q - y_p) - |f(\vec{x}_I^*)| \leq \sup_I f + \|Q\| + \|Q\| - f(\vec{x}_I^*) \leq \omega_I(f) + 2\|Q\|.$$

In case $f \leq 0$ on I, we similarly have

$$0 \leq (y_q - y_p) - |f(\vec{x}_I^*)| \leq \omega_I(f) + 2\|Q\|.$$

In the remaining case, we have

$$0 \leq (y_q - y_p) - |f(\vec{x}_I^*)| \leq (\sup_I f + \|Q\|) - (\inf_I f - \|Q\|) \leq \omega_I(f) + 2\|Q\|.$$

Therefore

$$
\begin{aligned}
\left|\mu(G_A(f)^+_{P \times Q}) - S(P, |f|)\right| &\leq \sum_{I \in P} |(y_q - y_p) - |f(\vec{x}_I^*)||\mu(I) \\
&\leq \sum_I \omega_I(f)\mu(I) + 2\|Q\|\mu(A).
\end{aligned}
$$

Since the right side can be arbitrarily small, we conclude that

$$\mu(G_A(f)) = \int_A |f| dA. \tag{7.1.11}$$

Example 7.1.4. Consider the characteristic function of a bounded subset A

$$\chi_A(\vec{x}) = \begin{cases} 1 & \text{if } \vec{x} \in A \\ 0 & \text{if } \vec{x} \notin A \end{cases}.$$

Let $A \subset (-M, M)^n$ and let P be a partition of $[-M, M]^n$. Then the oscillation

$$\omega_I(\chi_A) = \begin{cases} 1 & \text{if } I \cap A \neq \emptyset \text{ and } I - A \neq \emptyset \\ 0 & \text{if } I \subset A \text{ or } I \cap A = \emptyset \end{cases},$$

and we get $\sum \omega_I(\chi_A)\mu(I) = \mu(A_P^+) - \mu(A_P^-)$. Thereofre χ_A is Riemann integrable if and only if A has volume.

Moreover, by taking $x_I^* \in I \cap A$ for all I satisfying $I \cap A \neq \emptyset$ and $I - A \neq \emptyset$, we get $S(P, f) = \mu(A_P^+)$. By taking $x_I^* \in I - A$ instead, we get $S(P, f) = \mu(A_P^-)$. Either formula implies $\mu(A) = \int \chi_A d\mu$ in case A has volume.

Exercise 7.1.8. Prove that if $\mu(A) = 0$, then any bounded function is Riemann integrable on A, with $\int_A f d\mu = 0$.

Exercise 7.1.9. The Thomae function in Examples 1.4.2 and 3.1.5 may be extended to two variables by $R(x, y) = \dfrac{1}{q_x} + \dfrac{1}{q_y}$ if x and y are rational numbers with irreducible denominators q_x and q_y and $R(x, y) = 0$ otherwise. Prove that $R(x, y)$ is Riemann integrable on any subset with volume, and the integral is 0.

Exercise 7.1.10. Suppose f is a bounded function on a subset A with volume. Prove that f is integrable if and only if $\mathring{G}_A(f) = \{(\vec{x}, y) : \vec{x} \in A, y \in [0, f(\vec{x}))\}$ has volume. Moreover, the integrability of f implies $H_A(f) = \{(\vec{x}, f(\vec{x})) : \vec{x} \in A\}$ has volume zero, but the converse is not true.

Exercise 7.1.11. Suppose $f \leq g \leq h$ on a subset A with volume. Suppose f and h are integrable, with $\int_A f d\mu = \int_A h d\mu$. Prove that g is also integrable.

Exercise 7.1.12. Prove that if f is integrable on A, and $B \subset A$ has volume, then f is integrable on B.

For a map $F \colon A \subset \mathbb{R}^n \to \mathbb{R}^m$ on a subset A with volume, we may also introduce the Riemann sum

$$S(P, F) = \sum_{I \in P} \mu(I) F(x_I^*).$$

The map is Riemann integrable on a subset A with volume, with integral $\vec{J} \in \mathbb{R}^m$, if for any $\epsilon > 0$, there is $\delta > 0$, such that

$$\|P\| < \delta \implies \|S(P, F) - \vec{J}\| < \epsilon.$$

Then we denote

$$\int_A F(\vec{x}) d\mu = \vec{J}.$$

It is easy to see that F is integrable if and only if each coordinate is integrable.

Exercise 7.1.13. Define the oscillation of a map $F\colon \mathbb{R}^n \to \mathbb{R}^m$ to be

$$\omega_A(F) = \sup_{\vec{x},\vec{y}\in A} \|F(\vec{x}) - F(\vec{y})\|.$$

Prove that a map F is Riemann integrable if and only if for any $\epsilon > 0$, there is $\delta > 0$, such that $\|P\| < \delta$ for a general partition P of A implies the Riemann sum of the oscillation $\sum_{I\in P} \omega_I(F)\mu(I) < \epsilon$. Moreover, the integrability is also equivalent to the existence of one partition P satisfying $\sum_{I\in P} \omega_I(F)\mu(I) < \epsilon$.

Exercise 7.1.14. For a Riemann integrable map, prove

$$\left\| \int_A F(\vec{x})d\mu \right\| \le \int_A \|F(\vec{x})\|d\mu \le \left(\sup_{\vec{x}\in A} \|F(\vec{x})\| \right) \mu(A).$$

7.1.3   Properties of Integration

The following property extends Proposition 3.1.4.

Proposition 7.1.7. *Bounded and continuous maps are Riemann integrable on subsets with volumes.*

Proof. Suppose f is continuous on a subset A with volume. For any $\epsilon > 0$, there is a partition P, such that $\mu(A - A_P^-) < \epsilon$. The subset $K = A_P^-$ is compact because it is a union of finitely many closed rectangles. Therefore f is uniformly continuous on K. This means that there is $\delta > 0$, such that $\vec{x}, \vec{y} \in K$ and $\|\vec{x} - \vec{y}\| < \delta$ implies $|f(\vec{x}) - f(\vec{y})| < \epsilon$. By further refine the rectangles in P such that $\|P\| < \delta$, we get $\omega_I(f) < \epsilon$ for any $I \in P$ satisfying $I \subset A$.

Now consider the general partition $Q = \{I \in P \colon I \subset A\} \cup \{A - K\}$ of A. If $|f| < b$ on A, then with respect to the partition, we have

$$\sum_{J\in Q} \omega_J(f)\mu(J) = \omega_{A-K}(f)\mu(A - K) + \sum_{I\in P, I\subset A} \omega_I(f)\mu(I)$$

$$\le 2b\epsilon + \epsilon \sum_{I\in P, I\subset A} \mu(I) \le 2b\epsilon + \mu(A)\epsilon.$$

This verifies the criterion for the integrability of f on A.

$\square$

Proposition 3.1.5 cannot be extended in general because there is no monotonicity concept for multivariable functions. With basically the same proof, Proposition 3.1.6 can be extended.

Proposition 7.1.8. *Suppose $F\colon \mathbb{R}^n \to \mathbb{R}^m$ is a Riemann integrable map on a subset A with volume. Suppose the values of F lie in a compact subset K, and Φ is a continuous map on K. Then the composition $\Phi \circ F$ is Riemann integrable on A.*

Propositions 3.1.7 and 3.1.8 can also be extended, by the same argument.

Proposition 7.1.9. *Suppose f and g are Riemann integrable on a subset A with volume.*

1. *The linear combination $af + bg$ and the product fg are Riemann integrable on A, and $\int_A (af + bg)d\mu = a \int_A fd\mu + b \int_A gd\mu$.*

2. *If $f \leq g$, then $\int_A fd\mu \leq \int_A gd\mu$. Moreover, $\left|\int_A fd\mu\right| \leq \int_A |f|d\mu$.*

Denote the "upper half" and the "lower half" of $\mathbb{R}^{n+1}$

$$H^+ = \{(\vec{x}, y)\colon y \geq 0\} = \mathbb{R}^n \times [0, +\infty),$$
$$H^- = \{(\vec{x}, y)\colon y \leq 0\} = \mathbb{R}^n \times (-\infty, 0].$$

For any subset $G \subset \mathbb{R}^{n+1}$ with volume, the intersections $G \cap H^+$ and $G \cap H^-$ have volumes. For the special case $G = G_A(f)$, we have

$$G_A(f) \cap H^+ = G_A(\max\{f, 0\}), \quad G_A(f) \cap H^- = G_A(\min\{f, 0\}).$$

If f is Riemann integrable on A, then both intersections have volumes, so that $\max\{f, 0\}$ and $\min\{f, 0\}$ are Riemann integrable on A. Moreover, by (7.1.11), we have

$$\int_A \max\{f, 0\}d\mu = \mu(G_A(f) \cap H^+), \quad \int_A \min\{f, 0\}d\mu = -\mu(G_A(f) \cap H^-).$$

Adding the two equalities, we get

$$\int_A fd\mu = \mu(G_A(f) \cap H^+) - \mu(G_A(f) \cap H^-). \tag{7.1.12}$$

The equality and Proposition 7.1.3 imply the following extension of Proposition 3.1.9.

Proposition 7.1.10. *Suppose f is Riemann integrable on subsets A and B with volume. Then f is Riemann integrable on $A \cup B$ and $A \cap B$, with*

$$\int_{A \cup B} fd\mu + \int_{A \cap B} fd\mu = \int_A fd\mu + \int_B fd\mu.$$

The next result will be needed in the theory of improper multivariable integrals.

Proposition 7.1.11. *Suppose f is Riemann integrable on a subset A with volume. Then*

$$\int_A \max\{f, 0\}d\mu = \sup_{C \subset A,\, C \text{ has volume}} \int_C fd\mu.$$

Intuitively, we wish C to be the largest subset of A on which $f \geq 0$. The natural choice of C would be $A^+ = \{\vec{x} \in A\colon f(\vec{x}) \geq 0\}$. Unfortunately, the subset A^+ may not have volume. A counterexample is the negative of the Thomae function in Example 3.1.5.

Proof. For any $\epsilon > 0$, there is $\delta > 0$, such that $\sum_{I \in P} \omega_I(f)\mu(I) < \epsilon$ for any general partition P of A satisfying $\|P\| < \delta$. Then $f^+ = \max\{f, 0\}$ satisfies (see Exercise 7.1.20)

$$\left| S(P, f^+) - \int_A f^+ d\mu \right| \leq \sum_{I \in P} \omega_I(f^+)\mu(I) \leq \sum_{I \in P} \omega_I(f)\mu(I) \leq \epsilon.$$

Moreover,
$$Q = \{I \in P \colon f(\vec{x}) \geq 0 \text{ for some } \vec{x} \in I\}$$

is a general partition of $C = \cup_{I \in Q} I$. For each $I \in P$, choose $\vec{x}_I^* \in I$. The choice can be made to satisfy $f(\vec{x}_I^*) \geq 0$ in case $I \in Q$. Then we have

$$\left| S(P, f^+) - \int_C f d\mu \right| = \left| S(Q, f) - \int_C f d\mu \right|$$
$$\leq \sum_{I \in Q} \omega_I(f)\mu(I) \leq \sum_{I \in P} \omega_I(f)\mu(I) \leq \epsilon.$$

Combined with the earlier estimation, we get

$$\left| \int_A f^+ d\mu - \int_C f d\mu \right| \leq 2\epsilon.$$

On the other hand, we always have

$$\int_A f^+ d\mu \geq \int_C f^+ d\mu \geq \int_C f d\mu.$$

Then the equality in the proposition follows.

$\square$

Example 7.1.5. For any $\vec{a} \neq \vec{0}$, consider a codimension 1 hyperplane

$$B = \{\vec{x} \colon \vec{a} \cdot \vec{x} = c\}.$$

If the last coordinate of $\vec{a}$ is nonzero, then we can solve the last coordinate from $\vec{a} \cdot \vec{x} = c$ and get
$$B = \{(\vec{y}, z) \colon z = \vec{b} \cdot \vec{y}\}.$$

For any $M > 0$, $B \cap [-M, M]^{n-1} \times \mathbb{R}$ is contained in the graph of the linear function $\vec{b} \cdot \vec{y}$ for $\vec{y} \in [-M, M]^{n-1}$. Since linear functions are Riemann integrable, we conclude that $\mu(B \cap [-M, M]^{n-1} \times \mathbb{R}) = 0$. In general, we always have $\mu(B \cap [-M, M]^{n-1} \times \mathbb{R}) = 0$ up to a permutation of coordinates.

A convex polytope A is the intersection of finitely many half spaces $\vec{a} \cdot \vec{x} \geq c$, $\vec{a} \neq \vec{0}$. The boundary of the convex polytope consists of those points such that some inequalities become equalities $\vec{a} \cdot \vec{x} = c$. Therefore the boundary is a finite union of subsets of the form B above. In particular, if A is bounded, then the boundary has volume zero. Therefore bounded convex polytopes have volume.

A bounded polyhedron is a finite union of bounded polytopes and therefore also has volume.

Exercise 7.1.15. Prove Proposition 7.1.8.

Exercise 7.1.16. Suppose f is Riemann integrable on $[a, b]$. Suppose $A \subset \mathbb{R}^2$ is a subset with area, such that $(x, y) \in A$ implies $a \leq x + y \leq b$.

1. Prove that there is b, such that $\mu\{(x, y) \in A \colon s \leq x + y \leq t\} < b(t - s)$ for any $s < t$.

2. Prove that $f(x + y)$ is Riemann integrable on A.

Moreover, compute the area of $\{(x, y) \colon |x| \leq M, |y| \leq M, s \leq xy \leq t\}$ and study the integrability of $f(xy)$.

Exercise 7.1.17. Prove Proposition 7.1.9. Explain that the Riemann integral is not changed if the function is arbitrarily modified on any subset of volume 0.

Exercise 7.1.18. Prove Proposition 7.1.10 by directly using the Riemann sum.

Exercise 7.1.19. Suppose a function $f(x, y)$ defined on a rectangle $[a, b] \times [c, d]$ is increasing in x for fixed y and increasing in y for fixed x. Prove that f is Riemann integrable. What if f in increasing in one variable and continuous in another?

Exercise 7.1.20. Suppose $f(\vec{x})$ is integrable on A. Prove that

$$\mu(A) \inf_A f \leq \int_A f d\mu \leq \mu(A) \sup_A f,$$

$$\left| S(P, f) - \int_A f d\mu \right| \leq \sum_{I \in P} \omega_I(f) \mu(I).$$

Exercise 7.1.21. Suppose A is a path connected and compact subset with volume. Suppose f is a continuous function on A. Prove that there is $\vec{a} \in A$, such that $\int_A f d\mu = f(\vec{a})\mu(A)$. Extend the conclusion to $\int_A f g d\mu$ for a non-negative integrable function g on A.

Exercise 7.1.22. Suppose A is an open subset with volume. Suppose f is a continuous function on A. Prove the following are equivalent.

1. $f = 0$.

2. $\int_A |f| d\mu = 0$.

3. $\int_B f d\mu = 0$ for any subset $B \subset A$ with volume.

4. $\int_I f d\mu = 0$ for any rectangle $I \subset A$.

5. $\int_A f g d\mu = 0$ for any continuous function g on A.

Does the conclusion still hold if A is not open?

Exercise 7.1.23. Extend Proposition 4.2.9 to multivariable Riemann integration. Suppose $f_n(\vec{x})$ is integrable on a subset A with volume and the sequence $\{f_n(\vec{x})\}$ uniformly converges on A. Then $\lim_{n \to \infty} f_n(\vec{x})$ is integrable on A, with

$$\int_A \lim_{n \to \infty} f_n(\vec{x}) d\mu = \lim_{n \to \infty} \int_A f_n(\vec{x}) d\mu.$$

7.1.4   Fubini Theorem

Theorem 7.1.12 (Fubini Theorem). *Suppose $A \subset \mathbb{R}^m$ and $B \subset \mathbb{R}^n$ have volumes. Suppose $f(\vec{x}, \vec{y})$ is Riemann integrable on $A \times B$. Suppose for each fixed $\vec{y}$, $f(\vec{x}, \vec{y})$ is Riemann integrable on A. Then $\displaystyle\int_A f(\vec{x}, \vec{y})d\mu_{\vec{x}}$ is Riemann integrable on B, and*

$$\int_{A \times B} f(\vec{x}, \vec{y})d\mu_{\vec{x},\vec{y}} = \int_B \left(\int_A f(\vec{x}, \vec{y})d\mu_{\vec{x}} \right) d\mu_{\vec{y}}.$$

Proof. Let P and Q be partitions of A and B. Let $\vec{x}_I^* \in I \in P$ and $\vec{y}_J^* \in J \in Q$. Let $g(\vec{y}) = \displaystyle\int_A f(\vec{x}, \vec{y})d\mu_{\vec{x}}$. Then we have the Riemann sums

$$S(P \times Q, f) = \sum_{I \in P, J \in Q} f(\vec{x}_I^*, \vec{y}_J^*)\mu_{\vec{x}}(I)\mu_{\vec{y}}(J) = \sum_{J \in Q} S(P, f(\vec{x}, \vec{y}_J^*))\mu_{\vec{y}}(J),$$

$$S(Q, g) = \sum_{J \in Q} g(\vec{y}_J^*)\mu_{\vec{y}}(J) = \sum_{J \in Q} \left(\int_A f(\vec{x}, \vec{y}_J^*)d\mu_{\vec{x}} \right) \mu_{\vec{y}}(J).$$

For any $\epsilon > 0$, by the integrability of f on $A \times B$, there is $\delta > 0$, such that $\|P\| < \delta$, $\|Q\| < \delta$ implies

$$\left| S(P \times Q, f) - \int_{A \times B} f(\vec{x}, \vec{y})d\mu_{\vec{x},\vec{y}} \right| < \epsilon.$$

Fix one partition Q satisfying $\|Q\| < \delta$ and fix one choice of $\vec{y}_J^*$. Then there is $\delta \geq \delta' > 0$, such that $\|P\| < \delta'$ implies

$$\left| S(P, f(\vec{x}, \vec{y}_J^*)) - \int_A f(\vec{x}, \vec{y}_J^*)d\mu_{\vec{x}} \right| < \epsilon$$

for all (finitely many) $J \in Q$. For any such P, we have

$$|S(P \times Q, f) - S(Q, g)| \leq \sum_{J \in Q} \left| S(P, f(\vec{x}, \vec{y}_J^*)) - \int_A f(\vec{x}, \vec{y}_J^*)d\mu_{\vec{x}} \right| \mu_{\vec{y}}(J)$$

$$\leq \sum_{J \in Q} \epsilon\mu_{\vec{y}}(J) = \epsilon\mu_{\vec{y}}(B),$$

and we get

$$\left| S(Q, g) - \int_{A \times B} f(\vec{x}, \vec{y})d\mu_{\vec{x},\vec{y}} \right|$$

$$\leq |S(P \times Q, f) - S(Q, g)| + \left| S(P \times Q, f) - \int_{A \times B} f(\vec{x}, \vec{y})d\mu_{\vec{x},\vec{y}} \right|$$

$$< \epsilon\mu_{\vec{y}}(B) + \epsilon.$$

In particular, the third property in Theorem 7.1.6 is satisfied. Therefore g is integrable on B, with $\displaystyle\int_B gd\mu_{\vec{y}} = \int_{A \times B} f(\vec{x}, \vec{y})d\mu_{\vec{x},\vec{y}}$.
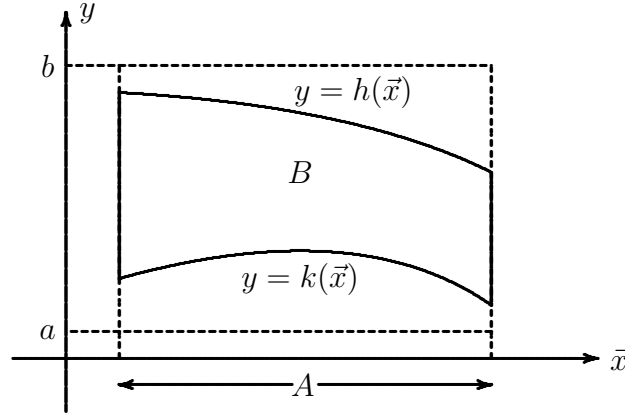
$\square$

Figure 7.1: integration between two graphs

Example 7.1.6. Suppose h and k are integrable functions on a subset A with volume. Suppose $h(\vec{x}) \geq k(\vec{x})$. Then the region

$$B = \{(\vec{x}, y) \colon \vec{x} \in A, h(\vec{x}) \geq y \geq k(\vec{x})\}$$

between the graphs of h and k has volume. For an integrable function $f(\vec{x}, y)$ on B, we may extend the function to $A \times [a, b]$ ($b > h(\vec{x}) \geq k(\vec{x}) > a$ for all $\vec{x} \in A$) by taking value 0 outside B. If for each fixed $\vec{x}$, $f(\vec{x}, y)$ is integrable for $y \in [k(\vec{x}), h(\vec{x})]$, then we have

$$\int_B f(\vec{x}, y) d\mu_{\vec{x}, y} = \int_{A \times [a,b]} f(\vec{x}, y) d\mu_{\vec{x}, y}$$

$$= \int_A \left(\int_a^b f(\vec{x}, y) dy \right) d\mu_{\vec{x}} = \int_A \left(\int_{k(\vec{x})}^{h(\vec{x})} f(\vec{x}, y) dy \right) d\mu_{\vec{x}}.$$

For a specific example, the integral of $(x + y)^2$ on the region bounded by $x = 0$, $y = 1$ and $x = y$ is

$$\int_{0 \leq x \leq 1, x \leq y \leq 1} (x+y)^2 dx dy = \int_0^1 \left(\int_x^1 (x + y)^2 dy \right) dx = \int_0^1 \frac{(x + 1)^3 - (2x)^3}{3} dx = \frac{37}{12}.$$

Example 7.1.7. Suppose $A \subset \mathbb{R}^m \times \mathbb{R}^n$ is a subset with volume. Suppose for each $\vec{x} \in \mathbb{R}^m$, the *section*

$$A_{\vec{x}} = \{\vec{y} \in \mathbb{R}^n \colon (\vec{x}, \vec{y}) \in A\}$$

also has volume. Then by applying Fubini theorem to the characteristic function (see Example 7.1.4), we find the volume of A to be

$$\mu_{m+n}(A) = \int_A \chi_A(\vec{x}, \vec{y}) d\mu_{\vec{x}, \vec{y}} = \int_{\mathbb{R}^m} \left(\int_{\mathbb{R}^n} \chi_A(\vec{x}, \vec{y}) d\mu_{\vec{y}} \right) d\mu_{\vec{x}} = \int_{\mathbb{R}^m} \mu_{\vec{y}}(A_{\vec{x}}) d\mu_{\vec{x}}.$$

The integral on the Euclidean spaces are actually the integral on some rectangles enclosing the subsets.

For a specific example, the area of the unit disk in $\mathbb{R}^2$ is

$$\int_{-1}^1 \mu \left[-\sqrt{1 - x^2}, \sqrt{1 - x^2} \right] dx = \int_{-1}^1 2\sqrt{1 - x^2} dx = \int_0^\pi 2\sqrt{1 - \cos^2 t} \sin t\, dt = \pi.$$

Example 7.1.8. The two variable Thomae function in Exercise 7.1.9 is Riemann integrable. However, for each rational y, we have $R(x, y) \geq \dfrac{1}{q_y} > 0$ for rational x and $R(x, y) = 0$ for irrational x. By the reason similar to the non-integrability of the Dirichlet function, $f(x, y)$ is not integrable in x. Thus the repeated integral $\int \left(\int R(x, y) dx \right) dy$ does not exist. The other repeated integral also does not exist for the same reason.

Note that since Riemann integrability is not changed if the function is modified on a subset of volume 0, Fubini theorem essentially holds even if those $\vec{y}$ for which $f(\vec{x}, \vec{y})$ is not integrable in $\vec{x}$ form a set of volume zero. Unfortunately, this is not the case for the two variable Thomae function, because the set of rational numbers does not have volume zero.

Example 7.1.9. If $f(\vec{x}, \vec{y})$ is integrable on $A \times B$, $f(\vec{x}, \vec{y})$ is integrable on A for each fixed $\vec{y}$, and $f(\vec{x}, \vec{y})$ is integrable on B for each fixed $\vec{x}$, then by Fubini theorem, the two repeated integrals are the same

$$\int_A \left(\int_B f(\vec{x}, \vec{y}) d\mu_{\vec{y}} \right) d\mu_{\vec{x}} = \int_B \left(\int_A f(\vec{x}, \vec{y}) d\mu_{\vec{x}} \right) d\mu_{\vec{y}}.$$

However, if two repeated integrals exist and are equal, it does not necessarily follow that the function is integrable.

Consider

$$S = \left\{ \left(\frac{k}{p}, \frac{l}{p} \right) : 0 \leq k, l \leq n, p \text{ is a prime number} \right\}.$$

The section $S_x = \{ y \colon (x, y) \in S \}$ is empty for any irrational x and is finite for any rational x. As a result, the section has volume (length) 0 for any x. The same holds for the sections S_y. On the other hand, for any 2 dimensional partition P of $[0, 1]^2$ by rectangles, we have $S_P^+ = [0, 1]^2$ (because S is dense in $[0, 1]^2$) and $S_P^- = \emptyset$ (because S contains no rectangles). Therefore $\mu(S_P^+) = 1$, $\mu(S_P^-) = 0$, and the set S has no volume.

By Example 7.1.4, a subset has volume if and only if its characteristic function is Riemann integrable. Therefore in terms of the characteristic function χ_S, the two repeated integrals $\int \left(\int \chi_S(x, y) dy \right) dx$ and $\int \left(\int \chi_S(x, y) dx \right) dy$ exist and are equal to 0, but the double integral $\int \chi_S(x, y) dx dy$ does not exist.

Exercise 7.1.24. Compute the integral.

1. $\displaystyle \int_{[0, \pi] \times [0, \pi]} |\cos(x + y)| dx dy$.

2. $\displaystyle \int_{[0, 1] \times [0, 1]} y^2 \cos(\pi x y) dx dy$.

3. $\displaystyle \int_{[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]} (x_1 + x_2 + \cdots + x_n)^p dx_1 dx_2 \cdots dx_n$.

4. $\displaystyle \int_A (x + y) dx dy$, A is the region bounded by $y = x^2$, $y = 4x^2$, $y = 1$ and with $x \geq 0$.

5. $\int_A \dfrac{\sin y}{y} dx dy$, A is the region bounded by $x = 0$, $y = x$ and $y = \pi$.

6. $\int_A xy dx dy$, A is the region bounded by $y = \sin x$ and $y = \cos x$ and between the intersection points $x = \dfrac{\pi}{4}$ and $x = \dfrac{5\pi}{4}$.

7. $\int_A \dfrac{dx dy dz}{x^2 + y^2}$, A is the region bounded by $x = 1$, $x = 2$, $z = 0$, $y = x$, $z = y$.

8. $\int_0^1 \left(\int_{-1}^1 \sqrt{|y - x^2|} dx \right) dy$.

9. $\int_0^1 \left(\int_x^1 \sin \pi(1 - y^2) dy \right) dx$.

10. $\int_A (1 - x_1 - x_2 - \cdots - x_n) dx_1 dx_2 \cdots dx_n$, A is the region bounded by $x_1 = 0$, $x_2 = 0$, ..., $x_n = 0$, $x_1 + x_2 + \cdots + x_n = 1$.

Exercise 7.1.25. Prove that

$$\int_{a \le x_1 \le \cdots \le x_n \le b} f(x_1) dx_1 dx_2 \cdots dx_n = \frac{1}{(n-1)!} \int_a^b (b-t)^{n-1} f(t) dt$$

$$\int_{a \le x_1 \le \cdots \le x_n \le b} f(x_1) f(x_2) \cdots f(x_n) dx_1 dx_2 \cdots dx_n = \frac{1}{n!} \left(\int_a^b f(t) dt \right)^n.$$

Exercise 7.1.26. Compute the volume.

1. The region bounded by $x = y$ and $x = y^2$.

2. The area between the graphs of $1 + 2x + 3x^3 + 4x^4$ and $2x + 3x^3 + 5x^4$, and $-1 \le x \le 1$.

3. The region bounded by the elliptic parabloid $z = x^2 + y^2$ and the plane $z = 1$.

4. The region bounded by the elliptic parabloid $z = 2x^2 + y^2$ and the cylinder $z = 2 - y^2$.

5. The region bounded by $x_1 = 0$, $x_2 = 0$, ..., $x_n = 0$, $x_1 + x_2 + \cdots + x_n = 1$.

Exercise 7.1.27. Suppose $f(x)$ is continuous and the improper integral $\displaystyle\int_0^{+\infty} f(x) dx$ converges. For $0 < a < b$, show that the improper integral $\displaystyle\int_0^{+\infty} \left(\int_a^b f(tx) dt \right) dx$ also converges and compute the improper integral.

Exercise 7.1.28. Study the existence and the equalities between the double integral and the repeated integrals.

1. $f(x, y) = \begin{cases} 1 & \text{if } x \text{ is rational} \\ 2y & \text{if } x \text{ is irrational} \end{cases}$ on $[0, 1] \times [0, 1]$.

2. $f(x, y) = \begin{cases} 1 & \text{if } x \text{ is rational} \\ 2y & \text{if } x \text{ is irrational} \end{cases}$ on $[0, 1] \times [0, 2]$.

3. $f(x, y) = \begin{cases} 1 & \text{if } x, y \text{ are rational} \\ 0 & \text{otherwise} \end{cases}$.

4. $f(x, y) = \begin{cases} R(x) & \text{if } x, y \text{ are rational} \\ 0 & \text{otherwise} \end{cases}$, where $R(x)$ is the Thomae function in Example 1.4.2.

5. $f(x, y) = \begin{cases} 1 & \text{if } x = \dfrac{1}{n}, n \in \mathbb{N} \text{ and } y \text{ is rational} \\ 0 & \text{otherwise} \end{cases}$.

6. $f(x, y) = \begin{cases} 1 & \text{if } x = \dfrac{1}{n}, n \in \mathbb{N} \text{ and } y \text{ is rational} \\ 1 & \text{if } x \text{ is rational and } y = \dfrac{1}{n}, n \in \mathbb{N} \\ 0 & \text{otherwise} \end{cases}$.

Exercise 7.1.29. Suppose $f(x)$ and $g(x)$ are increasing functions on $[0, 1]$. By considering the double integral of $(f(x) - f(y))(g(x) - g(y))$, prove that

$$\int_0^1 f(x)g(x)dx \geq \int_0^1 f(x)dx \int_0^1 g(x)dx.$$

More generally, for $p(x) \geq 0$, prove the *Chebychev inequality*

$$\int_a^b p(x)dx \int_a^b p(x)f(x)g(x)dx \geq \int_a^b p(x)f(x)dx \int_a^b p(x)g(x)dx.$$

Exercise 7.1.30. Suppose $f(x)$ is a continuous function on $[a, b]$. By considering the double integral of $(f(x) - f(y))^2$, prove that

$$\left(\int_a^b f(x)dx \right)^2 \leq (b - a) \int_a^b f(x)^2 dx.$$

Exercise 7.1.31. Suppose A and B are subsets with volumes. Suppose $f(\vec{x})$ and $g(\vec{y})$ are bounded functions on A and B.

1. Prove that if $f(\vec{x})$ and $g(\vec{y})$ are integrable, then $f(\vec{x})g(\vec{y})$ is integrable, and
$$\int_{A \times B} f(\vec{x})g(\vec{y})d\mu_{\vec{x},\vec{y}} = \int_A f(\vec{x})d\mu_{\vec{x}} \int_B g(\vec{y})d\mu_{\vec{y}}.$$

2. Prove that if $f(\vec{x})$ is integrable and $\int_A f d\mu \neq 0$, then $f(\vec{x})g(\vec{y})$ is integrable if and only if $g(\vec{y})$ is integrable.

7.1.5   Volume in Vector Space

A collection Σ of subsets is a *ring* if it satisfies

$$A, B \in \Sigma \implies A \cup B, A \cap B, A - B \in \Sigma.$$

A *volume theory* is a function ν defined on a ring Σ, such that the following properties are satisfied.

1. Non-negativity: $\nu(A) \geq 0$.

2. Additivity: If $A_1, A_2, \ldots, A_k$ are disjoint, then $\nu(A_1 \cup A_2 \cup \cdots \cup A_k) = \nu(A_1) + \nu(A_2) + \cdots + \nu(A_k)$.

The properties imply the monotonicity

$$A \subset B \text{ and } A, B \in \Sigma \implies \nu(A) \leq \mu(A) + \nu(B - A) = \nu(B).$$

We also say a subset A has ν-volume (or ν-*measurable*) if $A \in \Sigma$.

The notation μ will be reserved for the specific volume theory developed in Section 7.1.1. By Propositions 7.1.9 and 7.1.10, if f is a non-negative function on $\mathbb{R}^n$ that is Riemann integrable on any subset with μ-volume, then $\nu(A) = \int_A f d\mu$ is a volume theory on the collection of subsets with μ-volumes.

Exercise 7.1.32. Given a volume theory ν on a ring Σ, define the outer and inner volumes of any $A \subset X$ by

$$\nu^+(A) = \inf_{K \in \Sigma, K \supset A} \nu(K), \ \ \nu^-(A) = \sup_{L \in \Sigma, L \subset A} \nu(L).$$

We say A has $\bar{\nu}$-volume if $\nu^+(A) = \nu^-(A)$, and denote $\bar{\nu}(A) = \nu^+(A) = \nu^-(A)$. Prove that the collection $\bar{\Sigma}$ of subsets with $\bar{\nu}$-volume is still a ring. Moreover, prove that $\bar{\nu}$ is indeed a volume theory on $\bar{\Sigma}$.

Exercise 7.1.33. Prove that finite disjoint unions of all types of rectangles (i.e., products of all types of intervals, including single points) form a ring Σ of subsets of $\mathbb{R}^n$. Moreover, define the volume $\mu(I)$ of rectangles I in a usual way and define $\mu(A) = \sum \mu(I_j)$ for $A = \cup I_j \in \Sigma$, I_j disjoint.

1. Prove that μ is well-defined. In other words, $\mu(A)$ is independent of the way A is divided into disjoint rectangles.

2. Prove that μ is a volume theory on Σ.

The volume theory $\bar{\mu}$ produced by the process in Exercise 7.1.32 is then the Jordan measure theory.

A volume theory ν on a vector space V is *translation invariant* if $A \in \Sigma$ implies $A + \vec{v} \in \Sigma$, and

$$\nu(A + \vec{v}) = \nu(A) \text{ for any } \vec{v} \in V.$$

The following result shows that translation invariant volume theories are essentially unique.

Proposition 7.1.13. *Suppose ν is a translation invariant volume theory on $\mathbb{R}^n$, such that all subsets with μ-volumes also have ν-volumes. Then $\nu(A) = c\mu(A)$ for some constant c.*

Proof. For any x, N and distinct $x_1, x_2, \ldots, x_N \in [a_1, b_1]$, we have

$$N\nu(x \times [a_2, b_2] \times [a_3, b_3] \times \cdots \times [a_n, b_n])$$

$$= \sum_{i=1}^{N} \nu(x_i \times [a_2, b_2] \times [a_3, b_3] \times \cdots \times [a_n, b_n]) \qquad \text{(translation invariant)}$$

$$= \nu(\{x_1, x_2, \ldots, x_N\} \times [a_2, b_2] \times [a_3, b_3] \times \cdots \times [a_n, b_n]) \qquad \text{(additive)}$$

$$\leq \nu([a_1, b_1] \times [a_2, b_2] \times [a_3, b_3] \times \cdots \times [a_n, b_n]). \qquad \text{(monotone)}$$

Since N is arbitrary, we must have

$$\nu(x \times [a_2, b_2] \times [a_3, b_3] \times \cdots \times [a_n, b_n]) = 0.$$

Similar equality holds for other "reduced rectangles", including ones with open or partly open intervals. In particular, the additivity holds if the intersections of A_i are finite unions of reduced rectangles.

Let N be a natural number and $\epsilon = \dfrac{1}{N}$. The cube $[0, 1]^n$ is the union of translations of N^n copies of the small cube $[0, \epsilon]^n$, such that the intersections among the translations are reduced rectangles. Therefore

$$\nu([0, \epsilon]^n) = \epsilon^n N^n \nu([0, \epsilon]^n) = \epsilon^n \nu([0, 1]^n) = c\epsilon^n, \quad c = \nu([0, 1]^n).$$

Now for any rectangle $I = [a_1, b_1] \times [a_2, b_2] \times [a_3, b_3] \times \cdots \times [a_n, b_n]$, we find non-negative integers k_i satisfying $k_i \epsilon \leq b_i - a_i \leq (k_i + 1)\epsilon$. Then

$$I^+ = \prod [a_i, a_i + (k_i + 1)\epsilon] \supset I \supset I^- = \prod [a_i, a_i + k_i\epsilon].$$

By filling I^+ and I^- with translations of $[0, \epsilon]^n$, we get

$$\nu(I^+) = \left(\prod (k_i + 1)\right) c\epsilon^n \geq \nu(I) \geq \nu(I^-) = \left(\prod k_i\right) c\epsilon^n.$$

By $\lim_{N\to\infty} k_i \epsilon = \lim_{N\to\infty} (k_i + 1)\epsilon = b_i - a_i$, we conclude that

$$\nu(I) = c \prod (b_i - a_i) = c\mu(I).$$

Let A be any subset with μ-volume. For a partition P, we have the approximations A_P^+ and A_P^- of A. By the additivity and the fact that $\nu = c\mu$ on rectangles, we get $\nu(A_P^+) = c\mu(A_P^+) \geq \nu(A) \geq \nu(A_P^-) = c\mu(A_P^-)$. Since $\mu(A_P^+)$ and $\mu(A_P^-)$ can become arbitrarily close to $\mu(A)$ as P gets more refined, we conclude that $\nu(A) = c\mu(A)$.

$\square$

The discussion on translation invariant volume theory helps us to study the change of the volume under linear transforms.

Proposition 7.1.14. *Suppose $L \colon \mathbb{R}^n \to \mathbb{R}^n$ is a linear transform. Suppose a subset A has μ-volume. Then $L(A)$ has μ-volume, and*

$$\mu(L(A)) = |\det(L)|\mu(A).$$

Proof. For any rectangle I, $L(I)$ is a parallelepiped, which has μ-volume by Example 7.1.5. This implies that if A is a finite union of rectangles, then $L(A)$ has μ-volume.

Denote $c_L = \mu(L([0,1]^n))$. By the translation invariance property of $A \mapsto \mu(L(A))$ and the proof of Proposition 7.1.13, we have $\mu(L(A)) = c_L\mu(A)$ for any finite union A of rectangles. Repeating the last part of the proof, for any subset A with μ-volume, we get $\mu(L(A_P^+)) = c_L\mu(A_P^+) \geq \mu(L(A)) \geq \mu(L(A_P^-)) = c_L\mu(A_P^-)$. Then by the last criterion in Proposition 7.1.4, we conclude that $L(A)$ also has μ-volume. Moreover, the estimation tells us $\mu(L(A)) = c_L\mu(A)$.

It remains to show $c_L = |\det(L)|$. For linear transforms $L, K \colon \mathbb{R}^n \to \mathbb{R}^n$, we have $\mu(L \circ K(A)) = \mu(L(K(A)) = c_L\mu(K(A)) = c_Lc_K\mu(A)$. This tells us $c_{L \circ K} = c_Lc_K$. On the other hand, any linear transform is a composition of the following types (called elementary linear transforms).

1. Exchange: $L(x_1, x_2, x_3, \ldots, x_n) = (x_2, x_1, x_3, \ldots, x_n)$.

2. Scaling: $L(x_1, x_2, \ldots, x_n) = (cx_1, x_2, \ldots, x_n)$.

3. Shifting: $L(x_1, x_2, x_3, \ldots, x_n) = (x_1, x_2 + cx_1, x_3, \ldots, x_n)$.

By $c_{L \circ K} = c_Lc_K$ and $|\det(L \circ K)| = |\det(L)||\det(K)|$, it suffices to verify the formula $c_L = \mu(L([0,1]^n)) = |\det(L)|$ for the elementary linear transforms.

For the exchange, we have $L([0,1]^n) = [0,1]^n$, so that $c_L = \mu([0,1]^n) = 1 = |-1| = |\det L|$. For the scaling, we have $L([0,1]^n) = [0,c] \times [0,1]^{n-1}$, so that $c_L = \mu([0,c] \times [0,1]^{n-1}) = |c| = |\det L|$. For the shifting, we have $L([0,1]^n) = B \times [0,1]^{n-2}$, where $B \subset \mathbb{R}^2$ is the parallelogram with $(0,0)$, $(1,c)$, $(0,1)$, $(0,1+c)$ as vertices. It is easy to see that $\mu(B) = 1$, so that $c_L = \mu(B \times [0,1]^{n-2}) = 1 = |\det L|$.

$\square$

Proposition 7.1.14 suggests that the concept of having μ-volume is well-defined on any finite dimensional vector space V. Proposition 7.1.13 further tells us that translation invariant volumes on the collection of subsets of V with μ-volumes are unique up to multiplying a constant. In particular, such a volume theory is completely determined by the volume of a non-degenerate parallelepiped (rectangles are not well defined in vector spaces, parallelepipeds are well defined).

Exercise 7.1.34. Suppose $A \subset \mathbb{R}^n$ is a subset with volume. The cone on A with cone point $\vec{c} \in \mathbb{R}^{n+1}$ is

$$\vec{c}A = \{(1-t)\vec{c} + t(\vec{a}, 0) \colon \vec{a} \in A, t \in [0,1]\}.$$

Prove that the volume of $\vec{c}A$ is $\dfrac{1}{n+1}|c_{n+1}|\mu(A)$.

Exercise 7.1.35. Suppose α_n is the volume of the unit ball $B^n \subset \mathbb{R}^n$.

1. Prove that $\alpha_n = \int_{-1}^1 (1-r^2)^{\frac{n-1}{2}} \alpha_{n-1} dr$.

2. Find the general formula for α_n.

3. What is the volume of the ellipsoid $x^2 + 13y^2 + 14z^2 + 6xy + 2xz + 18yz \le 1$? The quadratic form appeared in Example 5.2.3.

Now we can justify the claim in the multilinear algebra that the volume of a parallelepiped spanned by $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k \in \mathbb{R}^n$ is the Euclidean norm $\|\vec{x}_1 \wedge \vec{x}_2 \wedge \cdots \wedge \vec{x}_k\|_2$ of the wedge of the spanning vectors.

Suppose $V \subset \mathbb{R}^n$ is the k-dimensional linear subspace spanned by the vectors $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k$. Then V has an inner product induced from the dot product of $\mathbb{R}^n$. For any orthonormal basis of V, there is a unique translation invariant volume theory on V such that the volume of the cube spanned by the orthonormal basis is 1. Since any two orthonormal basis are related by an orthogonal linear transform, which has determinant ± 1, the volume theory is actually independent of the choice of the orthonormal basis. Denote the volume theory by μ_V (which only depends on the inner product on V). Equivalently, we can take any isometry $L \colon \mathbb{R}^k \to V$ (linear transform preserving the inner product) and define $\mu_V = \mu \circ L^{-1}$.

Suppose $K \colon \mathbb{R}^k \to V$ is the linear transform that takes the standard basis $\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_k$ to $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k$. Then $A = K(I)$, where $I = [0,1]^k$ is the parallelepiped spanned by $\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_k$. For $\vec{v}_i = L^{-1}(\vec{x}_i) = L^{-1} \circ K(\vec{e}_i)$, we have

$$
\begin{aligned}
\mu_V(A) = \mu_V(K(I)) &= (\mu_V \circ L)(L^{-1} \circ K(I)) \\
&= \mu(L^{-1} \circ K(I)) && \text{(definition of } \mu_V) \\
&= |\det(L^{-1} \circ K)|\mu(I) && \text{(Proposition 7.1.14)} \\
&= \|\vec{v}_1 \wedge \vec{v}_2 \wedge \cdots \wedge \vec{v}_k\|_2 && \text{(definition of determinant)} \\
&= \|\vec{x}_1 \wedge \vec{x}_2 \wedge \cdots \wedge \vec{x}_k\|_2. && (L \text{ is an isometry})
\end{aligned}
$$

The computation also tells us that the equality

$$\mu_V(K(B)) = \|K(\vec{e}_1) \wedge K(\vec{e}_2) \wedge \cdots \wedge K(\vec{e}_k)\|_2 \mu(B)$$

holds for $B = I$. Since both sides are translation invariant volume theories on $\mathbb{R}^k$, the equality holds for any subset $B \subset \mathbb{R}^k$ with μ-volume. After changing the notations, we make a record of the fact.

Proposition 7.1.15. *Suppose $L \colon \mathbb{R}^k \to \mathbb{R}^n$ is an injective linear transform, with image subspace V. Then*

$$\mu_V(L(A)) = \|L(\vec{e}_1) \wedge L(\vec{e}_2) \wedge \cdots \wedge L(\vec{e}_k)\|_2 \mu(A).$$

7.1.6   Change of Variable

A change of variable on $\mathbb{R}^n$ is an invertible map $\Phi \colon U \subset \mathbb{R}^n \to \mathbb{R}^n$ defined on an open subset U, such that both Φ and Φ^{-1} are continuously differentiable. By the inverse function theorem, the condition is the same as Φ is invertible, continuously differentiable and the derivative Φ' is invertible everywhere. Moreover, we know $\Phi(U)$ is also open.

A change of variable on a subset A is a change of variable on an open subset U containing the closure $\bar{A}$ of A.

The formula for the Riemann integral under a change of variable was established for single variable functions in Theorems 3.2.4 and 3.2.5. For multivariable functions, we first study how the change of variable affects the volume.

Proposition 7.1.14 tells us how the volume changes under a linear change of variable. For a general change of variable on a subset A with volume, we consider a partition P of A and choices $\vec{x}_I^* \in I$ for $I \in P$. Then Φ is approximated by the linear map

$$L_I(\vec{x}) = \Phi(\vec{x}_I^*) + \Phi'(\vec{x}_I^*)(\vec{x} - \vec{x}_I^*)$$

on I, and we expect the volume of $\Phi(I)$ to be approximated by the volume of $L_I(I)$, which is $|\det \Phi'(\vec{x}_I^*)|\mu(I)$. Thus the volume of $\Phi(A)$ is approximated by $\sum_{I \in P} |\det \Phi'(\vec{x}_I^*)|\mu(I)$, which leads to the expectation that

$$\mu(\Phi(A)) = \int_A |\det \Phi'(\vec{x})|d\mu_{\vec{x}}.$$

This further leads to the formula for the Riemann integral under a change of variable.

Theorem 7.1.16 (Change of Variable). *Suppose $A \subset \mathbb{R}^n$ has volume. Suppose $\Phi \colon \mathbb{R}^n \to \mathbb{R}^n$ is a change of variable on A. Then $\Phi(A)$ has volume. Moreover, for any Riemann integrable function f on $\Phi(A)$, $f \circ \Phi$ is also Riemann integrable on A, and*

$$\int_A f(\Phi(\vec{x}))|\det \Phi'(\vec{x})|d\mu_{\vec{x}} = \int_{\Phi(A)} f(\vec{y})d\mu_{\vec{y}}. \qquad (7.1.13)$$

The sketchy argument leading to the formula (7.1.13) contains some technical difficulties. First, we do not yet know that $\Phi(I)$ has volume, so we can only talk about $\mu^+(\Phi(I))$ or $\mu^-(\Phi(I))$, at least in the beginning of a rigorous argument. Second, the estimation on the volume of $\Phi(I)$ can only be made by finding a nice subset I^Φ containing or contained in $\Phi(I)$. We expect I^Φ to be a "slightly inflated" version of the parallelepiped $L_I(I)$ (for containing $\Phi(I)$) or a "slightly shrunken" version (for being contained in $\Phi(I)$). However, the only information we have is the fact that $L_I(\vec{x})$ is a linear approximation of Φ, which is some inequality about norms. The subset containment that can be derived from the inequality must be the containment of balls. Thus we must consider partitions consisting of balls. This means that the partitions are made up of closed balls

$$\bar{B}_{L_\infty}(\vec{c}, r) = \{\vec{x} \colon \|\vec{x} - \vec{c}\|_\infty \leq r\} = \vec{c} + [-r, r]^n$$

in the L_∞-norm. Moreover, the containment argument between $\Phi(I)$ and I^Φ must be done after using a linear transform to convert the parallelepiped I^Φ to an L_∞-ball.

Proof. The change of variable Φ is defined on an open subset U containing the closure $\bar{A}$. We need to consider partitions P and take linear approximations

of Φ on $I \in P$ that have nonempty intersections with A. Therefore we need to consider Φ on a compact subset slightly bigger than A and still contained in U. By Proposition 5.1.9, there is $\delta_0 > 0$, such that

$$K = \{\vec{y} : \|\vec{y} - \vec{x}\|_\infty \leq \delta_0 \text{ for some } \vec{x} \in \bar{A}\} \subset U.$$

Since A is bounded, it is easy to see that K is a bounded and closed subset. By Proposition 5.1.3, K is compact.

With the help of Theorem 5.1.15, we can set up the bounds and the linear approximations for Φ on K. Specifically, the uniform continuity of Φ' and Proposition 6.1.4 tell us that for any $\epsilon > 0$, there is $\delta > 0$, such that if $\|\vec{y} - \vec{x}\|_\infty < \delta$ and the straight line connecting $\vec{x}$ to $\vec{y}$ is contained in K, then

$$\|\Phi(\vec{y}) - \Phi(\vec{x}) - \Phi'(\vec{x})(\vec{y} - \vec{x})\|_\infty \leq \epsilon \|\vec{y} - \vec{x}\|_\infty. \tag{7.1.14}$$

Moreover, the boundedness of Φ' and $(\Phi')^{-1}$ tells us that $|\det \Phi'| < M$ and $\|(\Phi')^{-1}\| < M$ on K for some constant M, where the norms of the linear transform is with respect to the L_∞-norm.
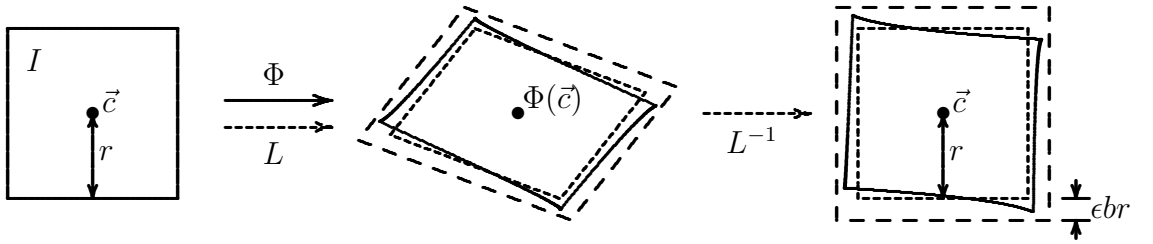
Figure 7.2: estimate the volume of $\Phi(I)$

Now consider a closed L_∞-ball $I = \bar{B}_{L_\infty}(\vec{c}, r) \subset K$ with radius $r < \delta$. We would like to estimate the volume of $\Phi(I)$. Denote the linear approximation of Φ at $\vec{c}$ by

$$L(\vec{x}) = \Phi(\vec{c}) + \Phi'(\vec{c})(\vec{x} - \vec{c}).$$

Then for any $\vec{x} \in I$, by (7.1.14) we have

$$\|\Phi(\vec{x}) - L(\vec{x})\|_\infty \leq \epsilon \|\vec{x} - \vec{c}\|_\infty \leq \epsilon r.$$

This implies

$$\|L^{-1}(\Phi(\vec{x})) - \vec{x}\|_\infty = \|\Phi'(\vec{c})^{-1}(\Phi(\vec{x}) - L(\vec{x}))\|_\infty \leq M\|\Phi(\vec{x}) - L(\vec{x})\|_\infty \leq \epsilon M r,$$

and

$$\|L^{-1}(\Phi(\vec{x})) - \vec{c}\|_\infty \leq \|L^{-1}(\Phi(\vec{x})) - \vec{x}\|_\infty + \|\vec{x} - \vec{c}\|_\infty \leq \epsilon M r + r = (1 + \epsilon M)r.$$

Therefore $L^{-1}(\Phi(I)) \subset \bar{B}_{L_\infty}(\vec{c}, (1 + \epsilon M)r)$, which is the same as $\Phi(I) \subset L(\bar{B}_{L_\infty}(\vec{c}, (1 + \epsilon M)r))$. Thus by Proposition 7.1.14,

$$\mu^+(\Phi(I)) \leq \mu(L(\bar{B}_{L_\infty}(\vec{c}, (1 + \epsilon M)r))) = |\det \Phi'(\vec{c})|\mu(\bar{B}_{L_\infty}(\vec{c}, (1 + \epsilon M)r))$$
$$= (1 + \epsilon M)^n |\det \Phi'(\vec{c})|\mu(I). \tag{7.1.15}$$

Next we need to put the estimations together. For the given $\epsilon > 0$, there is $\delta' > 0$, such that

$$\|P\| < \delta' \implies \mu(A_P^+) - \mu(A_P^-) < \epsilon.$$

Fix any $r < \min\left\{\dfrac{\delta_0}{2}, \delta, \dfrac{\delta'}{2}\right\}$ and choose P to consist of special rectangles of the form $I = \bar{B}_{L_\infty}(\vec{c}, r)$. By $2r < \delta_0$, we have

$$I \cap A \neq \emptyset \implies I \subset K.$$

Therefore the estimation (7.1.15) can be applied to $\Phi(I)$ and gives us

$$\mu^+(\Phi(A)) \leq \sum_{I \cap A \neq \emptyset} \mu^+(\Phi(I)) \leq (1 + \epsilon M)^n \sum_{I \cap A \neq \emptyset} |\det \Phi'(\vec{c})|\mu(I)$$

$$\leq (1 + \epsilon M)^n M \mu(A_P^+). \tag{7.1.16}$$

Applying the inequality (7.1.16) to ∂A in place of A, we get

$$\mu^+(\Phi(\partial A)) \leq (1 + \epsilon M)^n M \mu((\partial A)_P^+).$$

Since A has volume, $\mu((\partial A)_P^+)$ can become arbitrarily small. Therefore $\mu^+(\Phi(\partial A)) = 0$. By the continuity of Φ and Φ^{-1}, we have $\Phi(\partial A) = \partial \Phi(A)$. Then by the second criterion in Proposition 7.1.2, we conclude that $\Phi(A)$ has volume. Moreover, the inequality (7.1.16) (the first and the third terms, more precisely) further tells us

$$\mu(\Phi(A)) = \mu^+(\Phi(A))$$

$$\leq (1 + \epsilon M)^n \sum_{I \cap A \neq \emptyset} |\det \Phi'(\vec{c})|\mu(I)$$

$$\leq (1 + \epsilon M)^n \left(\sum_{I \subset A} + \sum_{I \cap A \neq \emptyset, I \cap (\mathbb{R}^n - A) \neq \emptyset}\right) |\det \Phi'(\vec{c})|\mu(I)$$

$$\leq (1 + \epsilon M)^n \left(\sum_{I \in P} |\det \Phi'(\vec{c})|\mu(I \cap A) + M \sum_{I \cap \partial A \neq \emptyset} \mu(I)\right)$$

$$= (1 + \epsilon M)^n \left(S(P \cap A, |\det \Phi'|) + M\mu((\partial A)_P^+)\right).$$

As $\epsilon \to 0$ and $\|P\| \to 0$, we get

$$\mu(\Phi(A)) \leq \int_A |\det \Phi'| d\mu.$$

We obtained an inequality for the change of volume. We will show that this implies a corresponding inequality for the change of Riemann integrals, at least for non-negative functions. Then by applying the inequality to the change of variable Φ^{-1}, we get an inequality for the change of Riemann integrals in the opposite direction. Two inequalities in both directions become an equality.

First we need to clarify the integrability. By Proposition 7.1.6, the integrability of f on $\Phi(A)$ means that the graph $G_{\Phi(A)}(f)$ has volume. By what we just proved, the change of variable (Φ^{-1}, id) takes $G_{\Phi(A)}(f)$ to another subset

$$(\Phi^{-1}, \mathrm{id})\left(G_{\Phi(A)}(f)\right) = G_A(f \circ \Phi)$$

with volume. Again by Proposition 7.1.6, this means that $f \circ \Phi$ is integrable on A.

Second we study the change of integral for an integrable function $f \geq 0$ on A. Suppose P is a general partition of A. Then $\Phi(P)$ is also a general partition. By choosing $\vec{x}_I^* \in I$ and the corresponding $\Phi(\vec{x}_I^*) \in \Phi(I)$, we have

$$S(f, \Phi(P)) = \sum_{I \in P} f(\Phi(x_I^*))\mu(\Phi(I)) \leq \sum_{I \in P} f(\Phi(x_I^*)) \int_I |\det \Phi'(\vec{x})| d\mu_{\vec{x}}.$$

$$(7.1.17)$$

On the other hand,

$$\left| S((f \circ \Phi)| \det \Phi'|, P) - \sum_{I \in P} f(\Phi(x_I^*)) \int_I |\det \Phi'(\vec{x})| d\mu_{\vec{x}} \right|$$

$$\leq \sum_{I \in P} f(\Phi(x_I^*)) \left| |\det \Phi'(x_I^*)|\mu(I) - \int_I |\det \Phi'(\vec{x})| d\mu_{\vec{x}} \right|$$

$$\leq \sum_{I \in P} f(\Phi(x_I^*))\omega_I(|\det \Phi'|)\mu(I). \qquad (7.1.18)$$

Since f is bounded and $|\det \Phi'|$ is integrable (by continuity), for any $\epsilon > 0$, there is $\delta > 0$, such that $\|P\| < \delta$ implies

$$\sum_{I \in P} f(\Phi(x_I^*))\omega_I(|\det \Phi'|)\mu(I) < \epsilon. \qquad (7.1.19)$$

Combining (7.1.17), (7.1.18), (7.1.19) together, we see $\|P\| < \delta$ implies

$$S(f, \Phi(P)) \leq S((f \circ \Phi)|\det \Phi'|, P) + \epsilon.$$

The boundedness of Φ' implies that $\|\Phi(P)\|$ is also small when $\|P\|$ is small. Thus the inequality above implies

$$\int_{\Phi(A)} f(\vec{y}) d\mu_{\vec{y}} \leq \int_A f(\Phi(\vec{x}))|\det \Phi'(\vec{x})| d\mu_{\vec{x}}. \qquad (7.1.20)$$

Applying the inequality (7.1.20) to Φ^{-1}, $(f \circ \Phi)|\det \Phi'|$, $\Phi(A)$ in place of Φ, f, A, we get the inequality in the opposite direction. Thus the formula for the change of variable is proved for non-negative functions. The general case can be proved by expressing any Riemann integrable function f as $f = \max\{0, f\} - \max\{0, -f\}$, where both $\max\{0, f\}$ and $\max\{0, -f\}$ are non-negative and Riemann integrable.

$\square$

Example 7.1.10. Suppose a region A in the cartesian coordinate corresponds to a region B in the polar coordinate. Then by taking the determinant of the Jacobian matrix in Example 6.1.8, we have

$$\int_A f(x,y)dxdy = \int_B f(r\cos\theta, r\sin\theta)rdrd\theta.$$

For example, the volume of the intersection of the ball $x^2 + y^2 + z^2 \le R^2$ and the cylinder $x^2 + y^2 \le Rx$ is the region between the graphs of $z = \sqrt{R^2 - x^2 - y^2}$ and $z = -\sqrt{R^2 - x^2 - y^2}$ over the disk $x^2 + y^2 \le Rx$. Since the boundary of the disk is given by $r = R\cos\theta$ for $-\dfrac{\pi}{2} \le \theta \le \dfrac{\pi}{2}$, the volume is

$$\int_{x^2+y^2\le Rx} 2\sqrt{R^2-x^2-y^2}dxdy = \int_0^R \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sqrt{R^2-r^2}rdrd\theta = \frac{\pi R^3}{3}.$$

Example 7.1.11. Suppose

$$F(\vec{\theta})\colon A \subset \mathbb{R}^{n-1} \to S^{n-1} = \{\vec{x}\colon \|\vec{x}\|_2 = 1\} \subset \mathbb{R}^n$$

is a parametrization of the unit sphere (see Example 6.1.10 for an example in case $n = 3$). Then

$$\Phi(r, \vec{\theta}) = rF(\vec{\theta})\colon (0, \infty) \times A \to \mathbb{R}^n$$

is a *spherical change of variable* (away from $\vec{0}$). We have

$$\Phi' = \begin{pmatrix} F & rF' \end{pmatrix}, \quad \det \Phi' = r^{n-1} \det \begin{pmatrix} F & F' \end{pmatrix}.$$

For a Riemann integrable function $f(t)$ on $[a, b]$, $b > a \ge 0$, we have

$$\int_{a\le\|\vec{x}\|_2\le b} f(\|\vec{x}\|_2)d\mu = \int_{a\le r\le b, \vec{\theta}\in A} f(r)r^{n-1}|\det\begin{pmatrix} F & F' \end{pmatrix}|drd\mu_{\vec{\theta}}$$

$$= \beta_{n-1}\int_a^b f(r)r^{n-1}dr.$$

The quantity

$$\beta_{n-1} = \int_A |\det\begin{pmatrix} F & F' \end{pmatrix}|d\mu_{\vec{\theta}}$$

is independent of the choice of parametrization F and is the $(n-1)$-dimensional volume of the unit sphere. For the special case $n = 2$ and $n = 3$, we have $\beta_1 = 2\pi$ and

$$\beta_2 = \int_{[0,\pi]\times[0,2\pi]} \left|\det\begin{pmatrix} \sin\phi\cos\theta & \cos\phi\cos\theta & -\sin\phi\sin\theta \\ \sin\phi\sin\theta & \cos\phi\sin\theta & \sin\phi\cos\theta \\ \cos\phi & -\sin\phi & 0 \end{pmatrix}\right| d\phi d\theta = 4\pi.$$

We also note that by taking $f = 1$, $a = 0$, $b = 1$, we get the volume

$$\alpha_n = \beta_{n-1}\int_0^1 r^{n-1}dr = \frac{\beta_{n-1}}{n}$$

of the unit ball B^n. Then β_{n-1} can be computed by using Exercise 7.1.35.

Example 7.1.12. Suppose $A \subset \mathbb{R}^2$ has area and all $(x, y) \in A$ satisfies $y \geq 0$. The rotation of A with respect to the x-axis is the subset

$$B = \{(x, y, z) \colon (x, r) \in A, x^2 + y^2 = r^2\} \subset \mathbb{R}^3.$$

By using the parametrization

$$\Phi(x, r, \theta) = (x, r \cos \theta, r \sin \theta) \colon A \times [0, 2\pi) \to B$$

and $\det \Phi' = r$, we find the volume

$$\mu(B) = \int_{A \times [0, 2\pi)} r \, dx dr d\theta = 2\pi \int_A y \, dx dy.$$

Introduce the *center of weight*

$$(x_A^*, y_A^*) = \frac{1}{\mu(A)} \int_A (x, y) dx dy$$

for the subset A. Then we get

$$\mu(B) = 2\pi y_A^* \mu(A).$$

The formula is called the Pappus-Guldinus theorem. We note that y_A^* is the average of the distance from points in A to the x-axis.

Next we consider the more general case of the rotation along a straight line $L \colon ax + bx = c$. Assume A lies in the "positive side" of the straight line, which means that all $(x, y) \in A$ satisfies $ax + by \geq c$. Denote by B the subset of $\mathbb{R}^3$ obtained by rotating A around L.

To compute the volume of B, we introduce a new coordinate system on $\mathbb{R}^2$ by

$$F(\tilde{x}, \tilde{y}) = (x_0 + \tilde{x} \cos \alpha - \tilde{y} \sin \alpha, y_0 + \tilde{x} \sin \alpha + \tilde{y} \cos \alpha) \colon \mathbb{R}^2 \to \mathbb{R}^2,$$

where (x_0, y_0) is a point on L and α is the angle of inclination of L. The transform takes the x-axis to L, and $F^{-1}(A)$ lies in the upper half plane. A corresponding rotation in $\mathbb{R}^3$ takes the rotation of $F(A)$ with respect to the x-axis to B. Since the rotation does not change distance and volume, we still have

$$\mu(B) = 2\pi \tilde{y}_A^* \mu(F^{-1}(A)) = 2\pi d\mu(A),$$

where d is the average distance from points in A to L. Since the distance from (x, y) to L is $\tilde{y} = \dfrac{(a, b)}{\sqrt{a^2 + b^2}} \cdot (x - x_0, y - y_0)$, we have

$$d = \frac{(a, b)}{\mu(A)\sqrt{a^2 + b^2}} \cdot \int_A (x - x_0, y - y_0) d\mu = \frac{a(x_A^* - x_0) + b(y_A^* - y_0)}{\sqrt{a^2 + b^2}} = \frac{ax_A^* + by_A^* - c}{\sqrt{a^2 + b^2}}.$$

Exercise 7.1.36. Find the formula for the integral under the changes of variable in Exercise 6.1.12.

Exercise 7.1.37. Use suitable change of variable to compute the integral.

1. $\displaystyle\int_{x^2 + y^2 \leq a^2} |xy| dx dy$.

2. $\displaystyle\int_A e^{\frac{x-y}{x+y}} dx dy$, A is the region bounded by $x = 0$, $y = 0$, $x + y = 1$.

3. $\displaystyle\int_A (ax^2 + 2bxy + cy^2)dxdy$, A is the part of the unit disk in the first quadrant.

4. $\displaystyle\int_A z^2 dxdydz$, A is the intersection of the ball $x^2 + y^2 + z^2 \le 1$ and the cylinder $x^2 + y^2 \le x$.

Exercise 7.1.38. Compute the area and volume.

1. The region on the plane bounded by the curve $r = \sin n\theta$ in polar coordinate.

2. The region on the plane bounded by the curve $(x^2 + y^2)^2 = 2ax^3$.

3. The intersection of the balls $x^2 + y^2 + z^2 \le 1$ and $x^2 + y^2 + z^2 \le 2az$.

Exercise 7.1.39. Extend the Pappus-Guldinus Theorem in Example 7.1.12 to high dimension. Specifically, we may rotate a subset $A \subset \mathbb{R}^n$ with volume and on the positive side of the hyperplane $\vec{a} \cdot \vec{x} = b$ around the hyperplane. Moreover, the rotation of a single point can be a sphere instead of a circle.

Exercise 7.1.40. Suppose $f(t)$ is a Riemann integrable function. Prove that

$$\int_{\|\vec{x}\|_2 \le R} f(\|\vec{x} - \vec{a}\|_2)d\mu = \frac{1}{2}\beta_{n-2} \int_{x^2 + y^2 \le R^2} f(\sqrt{(x - \|\vec{a}\|_2)^2 + y^2})|y|^{n-2}dxdy.$$

7.1.7 Improper Integration

Suppose $A \subset \mathbb{R}^n$ is an unbounded subset such that $A \cap B$ has volume for any bounded subset B with volume. Suppose f is a function on A that is Riemann integrable on $A \cap B$ for any bounded subset B with volume. The condition is the same as that for any R, the intersection $A \cap B_R^n$ with the ball of radius R has volume, and f is Riemann integrable on $A \cap B_R^n$. The *improper integral* $\displaystyle\int_A f d\mu$ converges to I if for any $\epsilon > 0$, there is R, such that

$$\left| \int_{A \cap B} f d\mu - I \right| < \epsilon$$

for any subset B with volume and containing B_R^n.

Proposition 7.1.17. *Suppose f is a function on an unbounded subset A, such that for any bounded subset B with volume, $A \cap B$ has volume and f is Riemann integrable on $A \cap B$. Then the following are equivalent.*

1. The improper integral $\displaystyle\int_A f d\mu$ converges.

2. The improper integral $\displaystyle\int_A |f| d\mu$ converges.

3. For any $\epsilon > 0$, there is $R > 0$, such that if a bounded subset C has volume and satisfies $C \cap B_R^n = \emptyset$, then $\left| \displaystyle\int_{A \cap C} f d\mu \right| < \epsilon$.

4. *There is $M > 0$, such that $\int_{A \cap B} |f| d\mu < M$ for all bounded subsets B with volume.*

5. *There is $M > 0$, such that $\left| \int_{A \cap B} f d\mu \right| < M$ for all bounded subsets B with volume.*

In the fourth statement, B can be replaced by B_R^n, or by any sequence B_k with the property that any B_R^n is contained in some B_k.

Proof. We have the Cauchy criterion for the convergence: For any $\epsilon > 0$, there is $R > 0$, such that if B_1 and B_2 have volumes and contain B_R^n, then

$$\left| \int_{A \cap B_1} f d\mu - \int_{A \cap B_2} f d\mu \right| < \epsilon.$$

By taking $B_1 = C \cup B_R^n$ and $B_2 = B_R^n$, the Cauchy criterion becomes the third statement. Conversely, suppose the third statement holds. Then for B_1 and B_2 in the Cauchy criterion, $B_1 - B_2$ and $B_2 - B_1$ satisfy the condition for C in the second statement. Therefore

$$\left| \int_{A \cap B_1} f d\mu - \int_{A \cap B_2} f d\mu \right| = \left| \int_{A \cap (B_1 - B_2)} f d\mu - \int_{A \cap (B_2 - B_1)} f d\mu \right| < 2\epsilon.$$

This proves that the first and the third statements are equivalent.

By an argument similar to the the convergence of monotone sequences, the second and the fourth statements are equivalent. The first statement implies the fifth. By Proposition 7.1.11, the fifth statement implies the fourth. Then the fourth statement, being equivalent to the second statement, implies the first by way of Cauchy criterion. This completes the proof that all statements are equivalent.

$\square$

In contrast to the single variable case, an improper integral converges if and only if it absolutely converges. The reason is that more freedom is allowed for the choice of B in the definition of the improper integral. If we restrict the choice of B to balls only, then we get a concept similar to the single variable case, and there will be a difference between absolute and conditional convergences.

The third condition in Proposition 7.1.17 is the Cauchy criterion for the convergence. This allows us to establish the comparison tests for the convergence similar to the single variable case. Moreover, the usual properties of the Riemann integrals, including the Fubini theorem and the change of variable formula, can be extended to improper integrals.

Example 7.1.13. Suppose $f(r)$ is a function on $[a, +\infty)$. By the computation in Example 7.1.11, we have

$$\int_{a \leq \|\vec{x}\|_2 \leq R} |f(\|\vec{x}\|_2)| d\mu = \beta_{n-1} \int_a^R |f(r)| r^{n-1} dr.$$

Since the convergence of $\displaystyle\int_{\|\vec{x}\|_2 \geq a} f(\|\vec{x}\|_2) d\mu$ and $\displaystyle\int_a^{+\infty} |f(r)| r^{n-1} dr$ are equivalent to the boundedness of $\displaystyle\int_{a \leq \|\vec{x}\|_2 \leq R} |f(\|\vec{x}\|_2)| d\mu$ and $\displaystyle\int_a^R |f(r)| r^{n-1} dr$, we conclude the convergence of the two improper integrals are equivalent. In particular, $\displaystyle\int_{\|\vec{x}\|_2 \geq a} \frac{d\mu}{\|\vec{x}\|_2^p}$ converges if and only if $p > n$, and we have

$$\int_{\|\vec{x}\|_2 \geq a} \frac{d\mu}{\|\vec{x}\|_2^p} = \beta_{n-1} \int_a^{+\infty} \frac{1}{r^p} r^{n-1} dr = \frac{\beta_{n-1}}{(p-n)a^{p-n}}.$$

Example 7.1.14. The improper integral $\displaystyle\int_{\mathbb{R}^2} e^{-(x^2+y^2)} dxdy$ converges because $e^{-(x^2+y^2)} \leq \dfrac{1}{(x^2+y^2)^2}$ for sufficiently big $\|(x,y)\|_2$ and $\displaystyle\int_{x^2+y^2 \geq 1} \frac{dxdy}{(x^2+y^2)^2}$ converges. By Fubini theorem, we have

$$\int_{\mathbb{R}^2} e^{-(x^2+y^2)} dxdy = \left(\int_{-\infty}^{+\infty} e^{-x^2} dx\right)\left(\int_{-\infty}^{+\infty} e^{-y^2} dy\right) = \left(\int_{-\infty}^{+\infty} e^{-x^2} dx\right)^2.$$

On the other hand, we have

$$\int_{\mathbb{R}^2} e^{-(x^2+y^2)} dxdy = \beta_1 \int_0^{+\infty} re^{-r^2} dr = \pi.$$

Thus we conclude that

$$\int_0^{+\infty} e^{-x^2} dx = \frac{1}{2} \int_{-\infty}^{+\infty} e^{-x^2} dx = \frac{\sqrt{\pi}}{2}.$$

Exercise 7.1.41. State and prove the comparison test for the convergence of improper integrals.

Exercise 7.1.42. Prove that if $f \geq 0$ and the improper integral $\displaystyle\int_A f d\mu$ converges, then

$$\int_A f d\mu = \sup_B \int_{A \cap B} f d\mu = \sup_{R>0} \int_{A \cap B_R^n} f d\mu,$$

where B are all bounded subsets with volume.

Exercise 7.1.43. Suppose $A \subset \mathbb{R}^n$ is an unbounded subset, such that $A \cap B$ has volume for any bounded subset B with volume. We say A has volume if $\mu(A \cap B)$ is bounded for all B and denote $\mu(A) = \sup_B \mu(A \cap B)$.

1. Prove that A has volume if and only if the characteristic function χ_A is integrable on $\mathbb{R}^n$, and $\mu(A) = \displaystyle\int_{\mathbb{R}^n} \chi_A d\mu$.

2. Prove that the extended volume satisfies $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$.

3. Prove that for unbounded A with volume, the improper integral $\displaystyle\int_A f d\mu$ converges if and only if $G_A(f)$ has volume.

Exercise 7.1.44. Suppose A and B have the property that $A \cap C$ and $B \cap C$ have volume for any bounded subset C with volume. Prove that $\int_{A \cup B} f d\mu$ converges if and only if $\int_A f d\mu$ and $\int_B f d\mu$ converge. Moreover, $\int_{A \cup B} f d\mu + \int_{A \cap B} f d\mu = \int_A f d\mu + \int_B f d\mu$.

Exercise 7.1.45. Suppose $\int_0^{+\infty} |f(r^2)| r^{n-1} dr$ converges and $q(\vec{x}) = A\vec{x} \cdot \vec{x}$ is a quadratic form given by a positive definite matrix A. Prove that

$$\int_{\mathbb{R}^n} f(q(\vec{x})) d\mu = \frac{\beta_{n-1}}{\sqrt{\det A}} \int_0^{+\infty} f(r^2) r^{n-1} dr.$$

Exercise 7.1.46. Suppose $\varphi(\vec{x})$ is a non-negative continuously differentiable homogeneous function of order $p > 0$ satisfying $\vec{x} \cdot \nabla \varphi(\vec{x}) \neq 0$. Prove that $\int_{\varphi(\vec{x}) \geq 1} f(\varphi(\vec{x})) d\mu$ converges if and only if $\int_1^{+\infty} |f(r^p)| r^{n-1} dr$ converges.

The result still holds even if restricted to a cone with origin as the vertex. A consequence is that for any $1 \leq p \leq \infty$, the improper integral $\int_{\|\vec{x}\|_p \geq 1} f(\|\vec{x}\|_p) d\mu$ converges if and only if $\int_1^{+\infty} |f(r)| r^{n-1} dr$ converges.

Exercise 7.1.47. Determine the convergence of the improper integrals and compute the convergent ones if possible.

1. $\int_{\mathbb{R}^2} \frac{e^{-|x-y|} dx dy}{1 + |x - y|}$.

2. $\int_{[1,+\infty)^2} \frac{dx dy}{(x + y)^p}$.

3. $\int_{\mathbb{R}^2} \frac{dx dy}{(1 + |x|)^p (1 + |y|)^q}$.

4. $\int_{[0,+\infty)^2} \frac{(x - y) dx dy}{(x + y + 1)^p}$.

5. $\int_{[1,+\infty)^2} \frac{\log(xy) dx dy}{x^2 + xy + y^2}$.

6. $\int_{\mathbb{R}^2} \frac{\sin x \sin y \, dx dy}{(x^2 + y^2)^p}$.

Exercise 7.1.48. Study the convergence of $\int_{x,y \geq 1} \frac{x^p y^q}{(x^m + y^n)^k} dx dy$, where all the parameters are positive. Extend the study to more variables.

Exercise 7.1.49. Study the convergence for any norm $\|\vec{x}\|$.

1. $\int_{\|\vec{x}\| \geq 1} \|\vec{x}\|^p d\mu$.

2. $\int_{\mathbb{R}^n} \|\vec{x}\|^{-\|\vec{x}\|} d\mu$.

3. $\int_{\mathbb{R}^n} \frac{\sin \|\vec{x}\|}{(1 + \|\vec{x}\|)^p} d\mu$.

Suppose f is an unbounded function on a bounded subset A with volume, such that f is Riemann integrable on $A - B(\vec{x}_0, \delta)$ for any $\delta > 0$. Then we say the improper integral $\int_A f d\mu$ converges to I if for any $\epsilon > 0$, there is $\delta > 0$, such that

$$\left| \int_{A-B} f d\mu - I \right| < \epsilon$$

for any subset B with volume satisfying $B(\vec{x}_0, \delta') \subset B \subset B(\vec{x}_0, \delta)$ for some $\delta' < \delta$.

The improperness of multivariable integration may also appear along more sophisticated subsets. See Example 7.1.16.

With the necessary modification, Proposition 7.1.17 remains largely true. In particular, the convergence may be determined by the uniform boundedness.

Example 7.1.15. Suppose $f(r)$ is a function on $(0, a]$. By the computation in Example 7.1.11, the improper integral $\int_{\|\vec{x}\|_2 \leq a} f(\|\vec{x}\|_2) d\mu$ converges if and only if $\int_0^a |f(r)| r^{n-1} dr$ converges. In particular, $\int_{\|\vec{x}\|_2 \leq a} \frac{d\mu}{\|\vec{x}\|_2^p}$ converges if and only if $p < n$, and we have

$$\int_{\|\vec{x}\|_2 \leq a} \frac{d\mu}{\|\vec{x}\|_2^p} = \beta_{n-1} \int_0^a \frac{1}{r^p} r^{n-1} dr = \frac{\beta_{n-1} a^{n-p}}{n - p}.$$

Example 7.1.16. The integral $\int_{(0,1)^2} \frac{dxdy}{\sqrt{(x+y)(1-x)(1-y)}}$ is improper along $1 \times [0,1] \cup [0,1] \times 1$ and $(0,0)$. Near $1 \times [0,1] \cup [0,1] \times 1$, we have

$$\frac{1}{\sqrt{2}\sqrt{(1-x)(1-y)}} \leq \frac{1}{\sqrt{(x+y)(1-x)(1-y)}} \leq \frac{1}{\sqrt{(1-x)(1-y)}}.$$

By the boundedness criterion and the convergence of $\int_{(0,1)^2} \frac{dxdy}{\sqrt{(1-x)(1-y)}} = \int_0^1 \frac{dx}{\sqrt{1-x}} \int_0^1 \frac{dy}{\sqrt{1-y}}$, we see that $\int_{(0,1)^2} \frac{dxdy}{\sqrt{(x+y)(1-x)(1-y)}}$ converges along $1 \times [0,1] \cup [0,1] \times 1$. Near $(0,0)$, we have

$$\frac{c_1}{\sqrt[4]{x^2+y^2}} \leq \frac{1}{\sqrt{(x+y)(1-x)(1-y)}} \leq \frac{c_2}{\sqrt[4]{x^2+y^2}}$$

for some $c_1, c_2 > 0$. By Example 7.1.15, the integral $\int_{x,y \geq 0, x^2+y^2 \leq 1} \frac{dxdy}{\sqrt[4]{x^2+y^2}}$ converges. Therefore $\int_{(0,1)^2} \frac{dxdy}{\sqrt{(x+y)(1-x)(1-y)}}$ converges at $(0,0)$. We conclude that the whole integral converges.

Exercise 7.1.50. Prove that the improper integral $\int_{x^2+y^2 \leq 1} \frac{(x^2-y^2)dxdy}{(x^2+y^2)^2}$ diverges but the *Cauchy principal value* $\lim_{\epsilon \to 0^+} \int_{\epsilon^2 \leq x^2+y^2 \leq 1} \frac{(x^2-y^2)dxdy}{(x^2+y^2)^2}$ converges.

Exercise 7.1.51. Study the existence of the repeated integrals and the improper integrals.

1. $\int_{[0,\pi] \times (0,1]} \frac{\cos x}{y} dxdy$.

2. $\int_{(0,+\infty)^2} e^{-xy} \sin x dxdy$.

Exercise 7.1.52. Determine the convergence of the improper integrals and compute the convergent ones if possible.

1. $\int_{(0,1)^2} \frac{|x-y|^p}{(x+y)^q} dxdy$.

2. $\int_{\|\vec{x}\|_2 < 1} \frac{dx_1 \cdots dx_n}{(1 - \|\vec{x}\|_2^2)^p}$.

3. $\displaystyle\int_{x+y+z\leq 1, x>0, y>0, z>0} \frac{z\,dxdydz}{(x+y+z)(y+z)^2}.$

4. $\displaystyle\int_{0\leq x<y<\pi} \log\sin(y-x)dxdy.$

5. $\displaystyle\int_{(0,1)^2} \frac{x^p y^q}{(x^m+y^n)^k}dxdy.$

7.1.8   Additional Exercise

Integrability and Continuity

Exercise 7.1.53. Prove that an integrable function f on A must be continuous at some interior points of A. In fact, for any ball $B(\vec{a}, \epsilon) \subset A$, f is continuous somewhere in the ball.

Exercise 7.1.54. Extend Exercise 3.1.35 to multivariable Riemann integral. Suppose f is integrable on a subset A with volume. Prove that for any $\epsilon > 0$, there is a union U of countably many rectangles, such that the sum of the volumes of the rectangles is $< \epsilon$, and all discontinuous points of f are inside U.

Exercise 7.1.55. Suppose f and g are integrable on A. Prove that if $f(\vec{x}) > g(\vec{x})$ in the interior of A, then $\displaystyle\int_A f d\mu > \int_A g(\vec{x})d\mu$.

Exercise 7.1.56. Suppose f is integrable on A. Prove the following are equivalent.

1. $\displaystyle\int_A |f|d\mu = 0.$

2. $\displaystyle\int_B f d\mu = 0$ for any subset $B \subset A$ with volume.

3. $\displaystyle\int_B f d\mu = 0$ for any rectangle $I \subset A$.

4. $\displaystyle\int_A fg d\mu = 0$ for any continuous function g on A.

5. $\displaystyle\int_A fg d\mu = 0$ for any integrable function g on A.

6. $f = 0$ at continuous points.

Darboux Sum and Darboux Integral

   The upper and the lower Darboux sums (3.1.18) and (3.1.19) and the upper and the lower Darboux integrals (3.1.20) may be extended to multivariable functions.

Exercise 7.1.57. For a single variable function on $[a, b]$, there are two versions of the Darboux integrals by choosing interval partitions or by choosing partitions with subsets with volumes (lengths). Prove that the two versions are the same. The result can be extended to functions defined on rectangles in $\mathbb{R}^n$.

Exercise 7.1.58. Prove that $\overline{\int_A} f(\vec{x})d\mu_{\vec{x}} = \lim_{\|P\|\to 0} U(P, f)$ and $\underline{\int_A} f(\vec{x})d\mu_{\vec{x}} = \lim_{\|P\|\to 0} L(P, f)$.

Exercise 7.1.59. Prove that if $f(\vec{x}) \geq 0$ on A, then $\overline{\int_A} f(\vec{x})d\mu_{\vec{x}} = \mu^+(G_A(f))$ and $\underline{\int_A} f(\vec{x})d\mu_{\vec{x}} = \mu^-(G_A(f))$.

Exercise 7.1.60. Prove that $\overline{\int_A} f(\vec{x})d\mu_{\vec{x}} \geq \underline{\int_A} f(\vec{x})d\mu_{\vec{x}}$. Moreover, the equality holds if and only if $f(\vec{x})$ is Riemann integrable on A, and $\int_A f(\vec{x})d\mu_{\vec{x}}$ is the common value.

Exercise 7.1.61. Prove that

$$\underline{\int_{A\times B}} f(\vec{x}, \vec{y})d\mu_{\vec{x},\vec{y}} \leq \underline{\int_B} \left(\overline{\int_A} f(\vec{x}, \vec{y})d\mu_{\vec{x}} \right) d\mu_{\vec{y}},$$

$$\overline{\int_{A\times B}} f(\vec{x}, \vec{y})d\mu_{\vec{x},\vec{y}} \geq \overline{\int_B} \left(\underline{\int_A} f(\vec{x}, \vec{y})d\mu_{\vec{x}} \right) d\mu_{\vec{y}}.$$

Also write down the similar inequalities with $\underline{\int_A}$ in place of $\overline{\int_A}$.

Exercise 7.1.62. Prove the extension of Fubini theorem: If f is Riemann integrable on $A \times B$, then any function $g(\vec{y})$ satisfying $\underline{\int_A} f(\vec{x}, \vec{y})d\mu_{\vec{x}} \leq g(\vec{y}) \leq \overline{\int_A} f(\vec{x}, \vec{y})d\mu_{\vec{x}}$ is Riemann integrable on B, and $\int_{A\times B} f(\vec{x}, \vec{y})d\mu_{\vec{x},\vec{y}} = \int_B g(\vec{y})d\mu_{\vec{y}}$.

Exercise 7.1.63. Prove that if $f(\vec{x}, \vec{y})$ is integrable on $A \times B$, then the collection of $\vec{y}$ for which f is not integrable in $\vec{x}$ is contained in a countable union of subsets of volume 0.

Exercise 7.1.64. Study the Darboux integrals and the repeated Darboux integrals for the functions in Exercise 7.1.28.

Integral Continuity

Exercise 7.1.65. Suppose A has volume. Suppose f is integrable on an open subset containing the closure $\bar{A}$. Prove that

$$\lim_{\vec{t}\to\vec{0}} \int_A |f(\vec{x} + \vec{t}) - f(\vec{x})|d\mu = 0.$$

Kepler's Second Law on Planet Motion

Kepler's Law says that the line from the sun to the planet will sweep out equal areas in equal intervals of time.

Exercise 7.1.66. Suppose $\phi(t) = (x(t), y(t)) \colon [a, b] \to \mathbb{R}^2$ is a curve, such that the movement from the vector ϕ to ϕ' is counterclockwise. Prove that the area swept by the line connecting the origin and $\phi(t)$ as t moves from a to b is $\frac{1}{2} \int_a^b (xy' - yx')dt$.

Exercise 7.1.67. Derive Kepler's law from Exercise 7.1.66 and Newton's second law of motion: $\phi'' = c\dfrac{\phi}{\|\phi\|_2^3}$, where c is a constant determined by the mass of the sun.

Dirichlet Transform

The *Dirichlet transform* is

$$x_1 + x_2 + \cdots + x_n = t_1,$$
$$x_2 + \cdots + x_n = t_1 t_2,$$
$$\vdots$$
$$x_{n-1} + x_n = t_1 t_2 \cdots t_{n-1},$$
$$x_n = t_1 t_2 \cdots t_{n-1} t_n.$$

Exercise 7.1.68. Use the Dirichlet transform to prove the relation between the Beta function and the Gamma function

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

Exercise 7.1.69. The standard simplex is

$$\Delta^n = \{(x_1, x_2, \ldots, x_n)\colon x_i \geq 0, x_1 + x_2 + \cdots + x_n \leq 1\}.$$

Use the Dirichlet transform to prove that that for $p_1, p_2, \ldots, p_n > 0$, we have

$$\int_{\Delta^n} x_1^{p_1-1} x_2^{p_2-1} \cdots x_n^{p_n-1} dx_1 dx_2 \cdots dx_n = \frac{\Gamma(p_1)\Gamma(p_2)\cdots\Gamma(p_n)}{(p_1 + p_2 + \cdots + p_n)\Gamma(p_1 + p_2 + \cdots + p_n)}.$$

7.2   Integration on Hypersurface

The concept of volume can be defined for good geometric objects such as curves and surfaces in Euclidean spaces. Although the definition makes use the parametrizations of the objects, the volume is independent of the parametrizations, thanks to the change of volume formula for the Riemann integral. Then the volume can be further used to define the Riemann integrals on such geometric objects.

When the Riemann integrals are used for the physical quantities such as the work or the flux, we get the Riemann integral of differential forms on curves and surfaces, etc. The formalism of differential forms is compatible with orientable change of variables.

7.2.1   Rectifiable Curve

The straight line connecting $\vec{x}$ to $\vec{y}$ has length $\|\vec{x} - \vec{y}\|$. To define the length of a continuous curve $\phi\colon [a, b] \to \mathbb{R}^n$, we take a partition P of $[a, b]$ and approximate the part of ϕ on $[t_{i-1}, t_i]$ by connecting a straight line between the end points $\phi(t_{i-1})$ and $\phi(t_i)$. The length of the approximate curve is

$$\mu_P(\phi) = \sum \|\phi(t_i) - \phi(t_{i-1})\|.$$

We say ϕ is *rectifiable* if $\mu_P(\phi)$ has an upper bound. The *length* of a rectifiable curve is

$$\mu(\phi) = \sup_P \mu_P(\phi).$$

Proposition 7.2.1. *A curve is rectifiable if and only if its coordinate functions have bounded variations.*

Proof. Because all norms are equivalent, the rectifiability is independent of the choice of the norm. If we take the L_1-norm, then

$$\|\phi(t_i) - \phi(t_{i-1})\|_1 = |x_1(t_i) - x_1(t_{i-1})| + |x_2(t_i) - x_2(t_{i-1})| + \cdots + |x_n(t_i) - x_n(t_{i-1})|,$$

and

$$\mu_P(\phi) = V_P(x_1) + V_P(x_2) + \cdots + V_P(x_n).$$

The proposition then follows.

$\square$

A change of variable (or *reparametrization*) for a parametrized curve is an invertible continuous map $u \colon [c,d] \to [a,b]$. The map induces a one-to-one correspondence between the partitions Q of $[c,d]$ and partitions $P = u(Q)$ of $[a,b]$. Since $\mu_P(\phi) = \mu_Q(\phi \circ u)$, and $\|P\|$ is small if and only if $\|Q\|$ is small, the rectifiability and the length are independent of the choice of parametrization.

Proposition 7.2.2. *Suppose $\Psi \colon [a,b] \to \mathbb{R}^n$ is Riemann integrable. Then $\phi(t) = \vec{x}_0 + \int_a^t \Psi(\tau) d\tau$ is a rectifiable curve, and its length is $\int_a^b \|\Psi(t)\| dt$.*

If ϕ is continuously differentiable, then $\Psi = \phi'$ and

$$\mu(\phi) = \int_a^b \|\phi'(t)\| dt. \tag{7.2.1}$$

Proof. For any partition P of $[a,b]$ and choices $t_i^* \in [t_{i-1}, t_i]$, we have

$$|\mu_P(\phi) - S(P, \|\Psi(t)\|)| \le \sum \left| \left\| \int_{t_{i-1}}^{t_i} \Psi(t) dt \right\| - \|\Psi(t_i^*)\| \Delta t_i \right|$$

$$\le \sum \left\| \int_{t_{i-1}}^{t_i} \Psi(t) dt - \Delta t_i \Psi(t_i^*) \right\|$$

$$\le \sum \sup_{t \in [t_{i-1}, t_i]} \|\Psi(t) - \Psi(t_i^*)\| \Delta t_i.$$

The integrability of Ψ shows the right side is small when $\|P\|$ is small.

$\square$

Example 7.2.1. The graph of a function $f(x)$ on $[a,b]$ is parametrized by $\phi(x) = (x, f(x))$. If f is continuously differentiable, then the Euclidean length of the graph is $\int_a^b \sqrt{1 + f'^2} dt$ and the L_1-length is $\int_a^b (1 + |f'|) dt$.

Example 7.2.2. Consider the following "circular curves".

$$\phi_1(t) = (\cos t, \sin t), \ \ 0 \le t \le 2\pi$$
$$\phi_2(t) = (\cos t, \sin t), \ \ 0 \le t \le 4\pi$$
$$\phi_3(t) = (\cos 2t, \sin 2t), \ \ 0 \le t \le 2\pi$$
$$\phi_4(t) = (\cos t, -\sin t), \ \ 0 \le t \le 2\pi$$
$$\phi_5(t) = \begin{cases} (\cos t, \sin t) & \text{if } 0 \le t \le \pi \\ (\cos t, -\sin t) & \text{if } \pi < t \le 2\pi \end{cases}$$

Although the images of both ϕ_1 and ϕ_2 are the whole unit circle, the length of ϕ_1 and ϕ_2 are 2π and 4π. As a matter of fact, ϕ_1 wraps around the circle once, and ϕ_2 wraps around twice, and we cannot reparametrize ϕ_1 to get ϕ_2. Therefore the length of a curve depends on the "track" instead of the image only.

The curve ϕ_3 is a reparametrization of ϕ_2 by $t \to 2t$. Therefore the length of ϕ_3 is also 4π. The curve ϕ_4 is a reparametrization of ϕ_1 by $t \to 2\pi - t$. Therefore the length of ϕ_4 is also 2π.

The image of the curve ϕ_5 is the upper unit circle. The curve first moves from $(1,0)$ to $(-1,0)$ along the circle and then moves back to $(1,0)$ along the circle. The total distance 2π of the movement is the length of the curve.

Example 7.2.3. The astroid $x^{\frac{2}{3}} + y^{\frac{2}{3}} = a^{\frac{2}{3}}$ can be parametrized as $x = a\cos^3 t$, $y = a\sin^3 t$ for $0 \le t \le 2\pi$. The Euclidean length of the curve is

$$\int_0^{2\pi} \sqrt{(-3a\cos^2 t \sin t)^2 + (3a\sin^2 t \cos t)^2}\, dt = 6a.$$

The L_∞-length is

$$\int_0^{2\pi} \max\{|-3a\cos^2 t \sin t|, |3a\sin^2 t \cos t|\}\, dt = 8\left(1 - \frac{1}{2\sqrt{2}}\right)a.$$

Note that the equality $x^{\frac{2}{3}} + y^{\frac{2}{3}} = a^{\frac{2}{3}}$ only specifies a subset of $\mathbb{R}^2$, which is the image of the astroid. Strictly speaking, the image alone is not sufficient to determine the length, as indicated by Example 7.2.2. On the other hand, we do implicitly understand that the question is to compute the length of *one round* of the astroid, which is indicated by the range $[0, 2\pi]$ for the parametrization.

Exercise 7.2.1. Prove that the L_1-length of a rectifiable curve is the sum of the variations of the coordinates on the interval.

Exercise 7.2.2. Prove that if ϕ is a rectifiable curve on $[a, b]$, then $\mu(\phi) = \mu(\phi|_{[a,c]}) + \mu(\phi|_{[c,b]})$ for any $a < c < b$.

Exercise 7.2.3. Let $n \ge 2$. Prove that the image of any rectifiable curve in $\mathbb{R}^n$ has n-dimensional volume zero.

Exercise 7.2.4. Suppose ϕ is a rectifiable curve. Prove that for any $\epsilon > 0$, there is $\delta > 0$, such that $\|P\| < \delta$ implies $\mu_P(\phi) > \mu(\phi) - \epsilon$. This implies $\lim_{\|P\| \to 0} \mu_P(\phi) = \mu(\phi)$.

Exercise 7.2.5. Suppose $F \colon \mathbb{R}^n \to \mathbb{R}^n$ is a map satisfying $\|F(\vec{x}) - F(\vec{y})\| = \|\vec{x} - \vec{y}\|$. Prove that $\mu(F(\phi)) = \mu(\phi)$.

Exercise 7.2.6. Suppose $\phi(t)$ is a continuously differentiable curve on $[a.b]$. For any partition P of $[a.b]$ and choice t_i^*, the curve is approximated by the tangent lines $\phi(t_i^*) + \phi'(t_i^*)(t - t_i^*)$ on intervals $[t_{i-1}, t_i]$. Prove that the sum of the lengths of the tangent lines converges to the length of ϕ as $\|P\| \to 0$.

Given a parametrized curve $\phi(t)\colon [a,b] \to \mathbb{R}^n$, the *arc length* function

$$s(t) = \mu(\phi|_{[a,t]})$$

is increasing. By Exercise 3.3.36 and the relation between $\mu_P(\phi)$ and the variation, $s(t)$ is also continuous. If ϕ is not constant on any interval in $[a,b]$, then $s(t)$ is strictly increasing, and the curve can be reparametrized by the arc length. In general, we may pick out the intervals on which ϕ is constant and modify the parametrization by reducing such intervals to single points. This does not change the arc length and the "track" of ϕ, and therefore will not affect all the later developments. Thus without loss of generality, we will assume that all curves can be reparametrized by the arc length.

Suppose ϕ is continuously differentiable. Then $s(t) = \int_a^t \|\phi'(t)\| dt$ by (7.2.1), or $ds = \|\phi'(t)\| dt$. In the special case $t = s$ is the arc length, we get

$$\|\phi'(s)\| = 1.$$

In other words, $\phi'(s)$ is the tangent vector of unit length.

Geometrically, $s(t) = \mu(\phi|_{[a,t]})$ is the arc length counted from the beginning point $\phi(a)$ of the curve. It is also possible to count the length $\tilde{s}(t) = \mu(\phi|_{[t,b]})$ from the end point $\phi(b)$ of the curve. The two arc lengths satisfy $s + \tilde{s} = \mu(\phi)$ and represent different choices of the *directions* or *orientations* of the curve. The arc length s is compatible with the orientation of the parameter t and $\tilde{s}$ is opposite to the orientation of t. The choice of direction will not affect the length, but will affect certain integrals in the future.

Example 7.2.4. For the circle $\phi(\theta) = (a\cos\theta, a\sin\theta)$, the Euclidean arc length (counted from $\theta = 0$) is

$$s = \int_0^\theta \|\phi'(t)\|_2 dt = \int_0^\theta \sqrt{a^2 \sin^2 t + a^2 \cos^2 t}\, dt = a\theta.$$

Therefore the circle is parametrized as $\phi(s) = \left(a\cos\dfrac{s}{a}, a\sin\dfrac{s}{a}\right)$ by the arc length.

On the other hand, with respect to the L_1-norm, we have

$$s = \int_0^\theta \|\phi'(t)\|_1 dt = \int_0^\theta |a|(|\sin t| + |\cos t|) dt = |a|(1 - \cos\theta + \sin\theta), \quad \text{for } 0 \le \theta \le \frac{\pi}{2}.$$

Exercise 7.2.7. Find the formula for the Euclidean arc length of a curve in $\mathbb{R}^2$ from the parametrized polar coordinate $r = r(t)$, $\theta = \theta(t)$.

Exercise 7.2.8. Compute the Euclidean lengths of the curves.

1. Parabola $y = x^2$, $0 \le x \le 1$.

2. Spiral $r = a\theta$, $0 \le \theta \le \pi$.

3. Another spiral $r = e^{a\theta}$, $0 \le \theta \le \alpha$.

4. Cycloid $x = a(t - \sin t)$, $y = a(1 - \cos t)$, $0 \le t \le 2\pi$.

5. Cardioid $r = 2a(1 + \cos\theta)$.

6. Involute of the unit circle $x = \cos\theta + \theta\sin\theta$, $y = \sin\theta - \theta\cos\theta$, $0 \le \theta \le \alpha$.

7. Helix $x = a\cos\theta$, $y = a\sin\theta$, $z = b\theta$, $0 \le \theta \le \alpha$.

We will often use capital letters such as C to denote curves. By this we mean that C is presented by a parametrization but any reparametrization is equally good. Thus different parametrizations are considered to give the same curve C if they are related by invertible continuous maps. Moreover, C can be parametrized by arc length s, which means that only parametrizations that are not constant on any interval are allowed. Thus we have the length $\mu(C)$ without any ambiguity. Finally, sometimes we may wish to distinguish the two choice of the directions C may take. This is achieved by restricting the reparametrizations to strictly increasing functions only. This is also equivalent to the choice of one of two possible arc lengths. We will indicate one choice by C and the other choice by $-C$.

7.2.2   Integration of Function on Curve

Suppose $f(\vec{x})$ is a function defined along a rectifiable curve C. Let $\phi\colon [a,b] \to \mathbb{R}^n$ be a parametrization of C. For a partition Q of $[a,b]$, we get a partition P of C by the segments C_i between $\vec{x}_{i-1} = \phi(t_{i-1})$ and $\vec{x}_i = \phi(t_i)$. For choices $t_i^* \in [t_{i-1}, t_i]$, we get $\vec{x}_i^* = \phi(t_i^*) \in C_i$. Define the Riemann sum

$$S(P, f) = \sum f(\vec{x}_i^*)\mu(C_i) = \sum f(\phi(t_i^*))\mu(\phi|_{[t_{i-1}, t_i]}).$$

The function is Riemann integrable along the curve, with Riemann integral $I = \int_C f\, ds$, if for any $\epsilon > 0$, there is $\delta > 0$, such that

$$\|P\| = \max \mu(C_i) < \delta \implies |S(P, f) - I| < \epsilon.$$

The definition can be easily extended to $\int_C F\, ds$ for maps $F\colon \mathbb{R}^n \to \mathbb{R}^m$.

Since the Riemann sum $S(P, f)$ depends only on the partition points $\vec{x}_i$, C_i and $\vec{x}_i^* \in C_i$, the integral is independent of the choice of parameter t.

The arc length $s(t) = \mu(\phi|_{[a,t]})$ is an increasing function, and

$$S(P, f) = \sum f(\phi(t_i^*))(s(t_i) - s(t_{i-1})) = S(P, f \circ \phi, s)$$

is the Riemann-Stieltjes sum of $f(\phi(t))$ with respect to $s(t)$. Therefore we get

$$\int_C f\, ds = \int_a^b f(\phi(t))\, ds(t).$$

The connection to the Riemann-Stieltjes integral gives us many usual properties of the integral along curves. See Exercises 7.2.10 through 7.2.16. Moreover, if the parametrization is further given by $\phi(t) = \vec{x}_0 + \int_a^t \Psi(\tau)\, d\tau$ as in Proposition 7.2.2. Then by Theorem 3.3.5, we get

$$\int_C f\, ds = \int_a^b f(\phi(t))\|\Psi(t)\|\, dt.$$

Example 7.2.5. We try to compute the integral of a linear function $l(\vec{x}) = \vec{a} \cdot \vec{x}$ along the straight line C connecting $\vec{u}$ to $\vec{v}$. The straight line can be parametrized as $\phi(t) = \vec{u} + t(\vec{v} - \vec{u})$, with $ds = \|\vec{v} - \vec{u}\|dt$. Therefore

$$\int_C \vec{a} \cdot \vec{x}\,ds = \int_0^1 \vec{a} \cdot (\vec{u} + t(\vec{v} - \vec{u}))\|\vec{v} - \vec{u}\|dt = \frac{1}{2}(\vec{a} \cdot (\vec{u} + \vec{v}))\|\vec{v} - \vec{u}\|.$$

The computation holds to any norm.

Example 7.2.6. The integral of $|y|$ along the unit circle with respect to the Euclidean norm is

$$\int_{x^2+y^2=1} |y|\,ds = \int_0^{2\pi} |\sin\theta|d\theta = 4\int_0^{\frac{\pi}{2}} |\sin\theta|d\theta = 4.$$

The integral with respect to the L_1-norms is

$$\int_{x^2+y^2=1} |y|\,ds = \int_0^{2\pi} |\sin\theta|(|\sin\theta| + |\cos\theta|)d\theta = 4\int_0^{\frac{\pi}{2}} \sin\theta(\sin\theta + \cos\theta)d\theta = 4.$$

Example 7.2.7. The integrals of $|y|$ along the curves C_1, C_2, C_3, C_4 in Example 7.2.2 with respect to the Euclidean norm are

$$\int_{C_1} |y|\,ds = \int_{C_4} |y|\,ds = \int_0^{2\pi} |\sin t|\sqrt{\sin^2 t + \cos^2 t}\,dt = 4,$$

$$\int_{C_2} |y|\,ds = \int_{C_3} |y|\,ds = \int_{\phi_2|_{[0,2\pi]}} |y|\,ds + \int_{\phi_2|_{[2\pi,4\pi]}} |y|\,ds = 2\int_{C_1} |y|\,ds = 8,$$

$$\int_{C_5} |y|\,ds = \int_{\phi_5|_{[0,\pi]}} |y|\,ds + \int_{\phi_5|_{[\pi,2\pi]}} |y|\,ds = 2\int_{\phi_5|_{[0,\pi]}} |y|\,ds = 4.$$

Note that C_2 may be split into two parts for $0 \leq t \leq 2\pi$ and $2\pi \leq t \leq 4\pi$. The first part is C_1 and the second part is also C_1 but reparametrized by $t \to t + 2\pi$. Therefore the integral on C_2 is the double of the integral on C_1.

The curve C_4 may be split into two parts for $0 \leq t \leq \pi$ and $\pi \leq t \leq 2\pi$. The second part is the first part reparametrized by $t \to 2\pi - t$. Thus the integral on C_4, which is the sum of the two parts, is the double of the first part.

Exercise 7.2.9. Compute the integral along the curve with respect to the Euclidean norm.

1. $\int_C |y|\,ds$, C is the unit circle $x^2 + y^2 = 1$.

2. $\int_C xy\,ds$, C is the part of the ellipse $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$ in the first quadrant.

3. $\int_C (x^{\frac{4}{3}} + y^{\frac{4}{3}})\,ds$, C is the astroid $x^{\frac{2}{3}} + y^{\frac{2}{3}} = a^{\frac{2}{3}}$.

4. $\int_C (a_1 x + a_2 y + a_3 z)\,ds$, C is the circle $x^2 + y^2 + z^2 = 1$, $b_1 x + b_2 y + b_3 z = 0$.

Exercise 7.2.10. Prove that f is integrable along a curve C if and only if for any $\epsilon > 0$, there is $\delta > 0$, such that $\|P\| < \delta$ implies $\sum \omega_{C_i}(f)\mu(C_i) < \epsilon$.

Exercise 7.2.11. Prove that continuous functions and monotone functions (the concept is defined along curves) are integrable along rectifiable curves.

Exercise 7.2.12. Suppose f is integrable along a curve. Suppose g is a uniformly continuous function on values of f. Prove that $g \circ f$ is integrable along the curve.

Exercise 7.2.13. Prove that the sum and the product of integrable functions along a curve are still integrable along the curve, and

$$\int_C (f + g)ds = \int_C fds + \int_C gds, \quad \int_C cfds = c\int_C fds.$$

Moreover, prove that

$$f \leq g \implies \int_C fds \leq \int_C gds.$$

Exercise 7.2.14. Suppose f is a continuous function along a curve C. Prove that there is $\vec{c} \in C$, such that $\int_C fds = f(\vec{c})\mu(C)$.

Exercise 7.2.15. Suppose C is divided into two parts C_1 and C_2. Prove that a function is integrable on C if and only if it is integrable on C_1 and C_2. Moreover,

$$\int_C fds = \int_{C_1} fds + \int_{C_2} fds.$$

Exercise 7.2.16. Suppose $F\colon \mathbb{R}^n \to \mathbb{R}^n$ is a map satisfying $\|F(\vec{x}) - F(\vec{y})\| = \|\vec{x} - \vec{y}\|$. Prove that $\int_{F(C)} f(\vec{x})ds = \int_C f(F(\vec{x}))ds$.

Exercise 7.2.17. Suppose f is a Riemann integrable function along a rectifiable curve C connecting $\vec{a}$ to $\vec{b}$. For any $\vec{x} \in C$, denote by $C[\vec{a}, \vec{x}]$ the part of C between $\vec{a}$ and $\vec{x}$. Then $F(\vec{x}) = \int_{C[\vec{a},\vec{x}]} fds$ can be considered as the "antiderivative" of f along C. By using the concept, state and prove the integration by part formula for the integral along C.

Exercise 7.2.18. Consider a map $F\colon C \subset \mathbb{R}^n \to \mathbb{R}^m$ on a rectifiable curve C. Prove that F is integrable along C if and only if each coordinate of F is integrable. Moreover, discuss properties of the integral $\int_C Fds$.

7.2.3   Integration of 1-Form on Curve

In physics, the force can be represented as a vector. The work done by a constant force F along the straight line from $\vec{a}$ to $\vec{b}$ is $F \cdot (\vec{b} - \vec{a})$. Now consider the work done by a changing force along a curve. Specifically, consider a vector field $F\colon C \to \mathbb{R}^n$ along a rectifiable curve C. For a partition P of C and choices $\vec{x}_i^* \in C_i$, define the Riemann sum

$$S(P, F) = \sum F(\vec{x}_i^*) \cdot (\vec{x}_i - \vec{x}_{i-1}).$$

The integral $\int_C F \cdot d\vec{x}$ is defined as the limit of the Riemann sum as $\|P\| \to 0$.

If the direction of the curve is reversed, then the order of the partition points $\vec{x}_i$ is reversed, and the Riemann sum changes the sign. The observation leads to

$$\int_{-C} F \cdot d\vec{x} = -\int_C F \cdot d\vec{x}.$$

The equality can be compared with $\int_b^a f dx = -\int_a^b f dx$.

Denote $F = (f_1, f_2, \ldots, f_n)$. Let $\phi = (x_1, x_2, \ldots, x_n) \colon [a, b] \to \mathbb{R}^n$ be a parametrization of C. For a partition Q of $[a, b]$ and choices t_i^*, we have the corresponding partition $P = \phi(Q)$ of C and choices $\vec{x}_i^* = \phi(t_i^*)$, and

$$
\begin{aligned}
S(P, F) &= \sum F(\phi(t_i^*)) \cdot (\phi(t_i) - \phi(t_{i-1})) \\
&= \sum (f_1(\phi(t_i^*))\Delta x_{1i} + f_2(\phi(t_i^*))\Delta x_{2i} + \cdots + f_n(\phi(t_i^*))\Delta x_{ni}) \\
&= S(P, f_1 \circ \phi, x_1) + S(P, f_2 \circ \phi, x_2) + \cdots + S(P, f_n \circ \phi, x_n),
\end{aligned}
$$

which is the sum of the Riemann-Stieljes sums of $f_i(\phi(t))$ with respect to $x_i(t)$. Therefore if $f_i(\phi(t))$ are integrable with respect to $x_i(t)$, then

$$
\int_C F \cdot d\vec{x} = \int_a^b f_1(\phi(t))dx_1(t) + \int_a^b f_2(\phi(t))dx_2(t) + \cdots + \int_a^b f_n(\phi(t))dx_n(t).
$$

Because of the connection to the Riemann-Stieljes integral, we also denote

$$
\int_C F \cdot d\vec{x} = \int_C f_1 dx_1 + f_2 dx_2 + \cdots + f_n dx_n.
$$

The expression

$$
F \cdot d\vec{x} = f_1 dx_1 + f_2 dx_2 + \cdots + f_n dx_n
$$

is called a 1-*form* (a *differential form* of order 1).

Suppose $\phi(t) = \vec{x}_0 + \int_a^t \Psi(\tau)d\tau$ as in Proposition 7.2.2. Then

$$
\begin{aligned}
&|S(P, F) - S(Q, (F \circ \phi) \cdot \Psi)| \\
&\leq \sum \left| F(\phi(t_i)) \cdot \int_{t_{i-1}}^{t_i} \Psi(t)dt - F(\phi(t_i)) \cdot \Delta t_i \Psi(t_i^*) \right| \\
&\leq \sum \|F(\phi(t_i))\|_2 \left\| \int_{t_{i-1}}^{t_i} \Psi(t)dt - \Delta t_i \Psi(t_i^*) \right\|_2 \\
&\leq \sum \|F(\phi(t_i))\|_2 \sup_{t \in [t_{i-1}, t_i]} \|\Psi(t) - \Psi(t_i^*)\|_2 \Delta t_i
\end{aligned}
$$

If F is bounded, then the integrability of Ψ implies that the right side is very small when $\|Q\|$ is very small. Therefore the 1-form $F \cdot d\vec{x}$ is integrable if and only if the function $F(\phi(t)) \cdot \Psi(t)$ is Riemann integrable, and

$$
\int_C F \cdot d\vec{x} = \int_a^b F(\phi(t)) \cdot \Psi(t)dt.
$$

For more properties of the integral of a 1-form along a curve, see Exercises 7.2.22 through 7.2.29.

Example 7.2.8. Consider three curves connecting $(0,0)$ to $(1,1)$. The curve C_1 is the straight line $\phi(t) = (t,t)$. The curve C_2 is the parabola $\phi(t) = (t,t^2)$. The curve C_3 is the straight line from $(0,0)$ to $(1,0)$ followed by the straight line from $(1,0)$ to $(1,1)$. Then

$$\int_{C_1} ydx + xdy = \int_0^1 (tdt + tdt) = 1,$$

$$\int_{C_2} ydx + xdy = \int_0^1 (t^2dt + t \cdot 2tdt) = 1,$$

$$\int_{C_3} ydx + xdy = \int_0^1 0dx + \int_0^1 1dy = 1.$$

We note that the result is independent of the choice of the curve. In fact, for any continuously differentiable curve $\phi(t) = (x(t), y(t))$, $t \in [a,b]$, connecting $(0,0)$ to $(1,1)$, we have

$$\int_C ydx + xdy = \int_a^b (y(t)x'(t) + x(t)y'(t))dt = \int_a^b (x(t)y(t))'dt$$

$$= x(b)y(b) - x(a)y(a) = 1 - 0 = 1.$$

Example 7.2.9. Taking the three curves in Example 7.2.8 again, we have

$$\int_{C_1} xydx + (x+y)dy = \int_0^1 (t^2dt + 2tdt) = \frac{4}{3},$$

$$\int_{C_2} xydx + (x+y)dy = \int_0^1 (t^3dt + (t+t^2)2tdt) = \frac{17}{12},$$

$$\int_{C_3} xydx + (x+y)dy = \int_0^1 0dx + \int_0^1 (1+y)dy = \frac{3}{2}.$$

In contrast to the integral of $ydx + xdy$, the integral of $xydx + (x+y)dy$ depends on the curve connecting the two points.

Example 7.2.10. Consider the integral of $F = (-y,x)$ along the four parametrized curves in Example 7.2.2.

$$\int_{C_1} F \cdot d\vec{x} = \int_0^{2\pi} (-\sin t, \cos t) \cdot (-\sin t, \cos t)dt = \int_0^{2\pi} dt = 2\pi,$$

$$\int_{C_2} F \cdot d\vec{x} = \int_0^{4\pi} (-\sin t, \cos t) \cdot (-\sin t, \cos t)dt = \int_0^{4\pi} dt = 4\pi,$$

$$\int_{C_3} F \cdot d\vec{x} = \int_0^{2\pi} (-\sin 2t, \cos 2t) \cdot (-2\sin 2t, 2\cos 2t)dt = \int_0^{2\pi} 2dt = 4\pi,$$

$$\int_{C_4} F \cdot d\vec{x} = \int_0^{2\pi} (\sin t, \cos t) \cdot (-\sin t, -\cos t)dt = -\int_0^{2\pi} dt = -2\pi,$$

$$\int_{C_5} F \cdot d\vec{x} = \int_0^{\pi} (-\sin t, \cos t) \cdot (-\sin t, \cos t)dt + \int_{\pi}^{2\pi} (\sin t, \cos t) \cdot (-\sin t, -\cos t)dt$$

$$= \int_0^{\pi} dt - \int_{\pi}^{2\pi} dt = 0.$$

The discussion in Example 7.2.7 on the relation between the integrals on the four curves is still valid. Note that the reparametrizations $t \to t + 2\pi$ in C_2 and

$t \to \dfrac{t}{2}$ in C_3 preserve the direction, and the reparametrization $t \to 2\pi - t$ in C_4 and C_5 reverses the direction. In particular, the integral along C_5, which is the sum of the two parts, must be zero.

Exercise 7.2.19. Compute the integrals of the 1-forms on the three curves in Example 7.2.8. In case the three integrals are the same, can you provide a general reason?

1. $xdx + ydy$.

2. $ydx - xdy$.

3. $(2x + ay)ydx + (x + by)xdy$.

4. $e^x(ydx + ady)$.

Exercise 7.2.20. Compute the integral of 1-form.

1. $\displaystyle\int_C \dfrac{ydx - xdy}{x^2 + y^2}$, C is upper half circle in the counterclockwise direction.

2. $\displaystyle\int_C (2a - y)dx + dy$, C is the cycloid $x = at - b\sin t$, $y = a - b\cos t$, $0 \le t \le 2\pi$.

3. $\displaystyle\int_C xdy + ydz + zdx$, C is the straight line connecting $(0, 0, 0)$ to $(1, 1, 1)$.

4. $\displaystyle\int_C (x + z)dx + (y + z)dy + (x - y)dz$, C is the helix $x = \cos t$, $y = \sin t$, $z = t$, $0 \le t \le \alpha$.

5. $\displaystyle\int_C (y^2 - z^2)dx + (z^2 - x^2)dy + (x^2 - y^2)dz$, C is the boundary of the intersection of the sphere $x^2 + y^2 + z^2 = 1$ with the first quadrant, and the direction is (x, y)-plane part, followed by (y, z)-plane part, followed by (z, x)-plane part.

Exercise 7.2.21. Suppose A is a symmetric matrix. Compute the integral of $A\vec{x} \cdot d\vec{x}$ on any continuously differentiable parametrized curve.

Exercise 7.2.22. Prove that if F is continuous, then the 1-form $F \cdot d\vec{x}$ is integrable along a rectifiable curve.

Exercise 7.2.23. Prove that if the 1-form $F \cdot d\vec{x}$ is integrable along a rectifiable curve and G is uniformly continuous on values of F, then $(G \circ F) \cdot d\vec{x}$ is integrable.

Exercise 7.2.24. Prove that the sum and scalar multiplication of integrable 1-forms are integrable, and

$$\int_C (F + G) \cdot d\vec{x} = \int_C F \cdot d\vec{x} + \int_C G \cdot d\vec{x}, \quad \int_C cF \cdot d\vec{x} = c\int_C F \cdot d\vec{x}.$$

Exercise 7.2.25. Prove the inequality

$$\left| \int_C F \cdot d\vec{x} \right| \le \mu(C) \sup_C \|F\|_2,$$

where $\mu(C)$ is the Euclidean length of the curve C.

Exercise 7.2.26. Suppose C is divided into two parts C_1 and C_2. Prove that a 1-form is integrable on C if and only if it is integrable on C_1 and C_2. Moreover,

$$\int_C F \cdot d\vec{x} = \int_{C_1} F \cdot d\vec{x} + \int_{C_2} F \cdot d\vec{x}.$$

Exercise 7.2.27. Suppose $\vec{a}$ is a constant vector and U is an orthogonal linear transform. Prove that $\int_{\vec{a}+U(C)} U(F(\vec{x})) \cdot d\vec{x} = \int_C F(\vec{a} + U(\vec{x})) \cdot d\vec{x}$.

Exercise 7.2.28. Suppose c is a constant. Prove that $\int_{cC} F(\vec{x}) \cdot d\vec{x} = c \int_C F(c\vec{x}) \cdot d\vec{x}$.

Exercise 7.2.29. Define $\omega_C(F) = \sup_{\vec{x}, \vec{y} \in C} \|F(\vec{x}) - F(\vec{y})\|$. Prove that if for any $\epsilon > 0$, there is $\delta > 0$, such that $\|P\| < \delta$ implies $\sum \omega_{C_i}(F)\mu(C_i) < \epsilon$, where C_i are the segments of C cut by the partition P, then $F \cdot d\vec{x}$ is integrable along C. Show that the converse is not true.

7.2.4 Surface Area

Suppose $\sigma(\vec{u}) = \sigma(u, v) \colon A \subset \mathbb{R}^2 \to \mathbb{R}^n$ is a parametrized surface on a subset A with area. For a partition P of $\mathbb{R}^2$ by triangles, we consider the triangles $I \in P$ that are contained in A. Each triangle is determined by its three vertices $\vec{u}_0$, $\vec{u}_1$, $\vec{u}_2$. Let $I^\sigma \subset \mathbb{R}^n$ be the triangle with $\sigma(\vec{u}_0)$, $\sigma(\vec{u}_1)$, $\sigma(\vec{u}_2)$ as the vertices. Then it is natural to think of the union of such I^σ as an approximation of the parametrized surface. By (5.2.13), the area of I^σ is

$$\frac{1}{2} \sum_I \|(\sigma(\vec{u}_1) - \sigma(\vec{u}_0)) \times (\sigma(\vec{u}_2) - \sigma(\vec{u}_0))\|_2.$$

We expect the sum of the areas approaches the area of the parametrized surface as $\|P\| \to 0$.

It turns out that the expectation is right for curves but wrong for surfaces. For curves, the direction of the straight line connecting two nearby points $\phi(t_{i-1})$ and $\phi(t_i)$ is very close to the direction $\phi'(t_i^*)$. For surfaces, the direction of the plane passing through three nearby points $\sigma(\vec{u}_0)$, $\sigma(\vec{u}_1)$, $\sigma(\vec{u}_2)$ may be far away from the direction of the tangent plane $\sigma(\vec{u}_I^*) + \sigma'(\vec{u}_I^*)(\vec{u} - \vec{u}_I^*)$.

Example 7.2.11. Consider the cylinder $\sigma(\theta, z) = (\cos\theta, \sin\theta, z)$ for $0 \le z \le 1$. Let $\alpha = \dfrac{\pi}{m}$ and $d = \dfrac{1}{2n}$. We cut the cylinder at heights $0, d, 2d, \dots, 2nd$ and get $2n + 1$ unit circles. On the unit circles at heights $0, 2d, 4d, \dots, 2nd$, we plot m points at angles $0, 2\alpha, 4\alpha, \dots, (2m - 2)\alpha$. On the unit circles at the heights $d, 3d, 5d, \dots, (2n - 1)d$, we plot m points at angles $\alpha, 3\alpha, \dots, (2m - 1)\alpha$. By connecting nearby triple points, we get $2mn$ identical isosceles triangles with base $2\sin\alpha$ and height $\sqrt{(1 - \cos\alpha)^2 + d^2}$. The total area is

$$4mn\frac{1}{2}2\sin\alpha\sqrt{(1 - \cos\alpha)^2 + d^2} = 2\pi\frac{\sin\alpha}{\alpha}\sqrt{\left(\frac{1 - \cos\alpha}{d}\right)^2 + 1}.$$

This has no limit as $\alpha, d \to 0$.

So it is necessary to make sure that the direction of the approximate planes to be really close to the direction of the tangent plane. Assume the parametrization σ is continuously differentiable. Take any general partition P of A, choose $\vec{u}_I^* \in I$. Then σ is approximated on I by the linear map

$$L_I(\vec{u}) = \sigma(\vec{u}_I^*) + \sigma'(\vec{u}_I^*)(\vec{u} - \vec{u}_I^*).$$
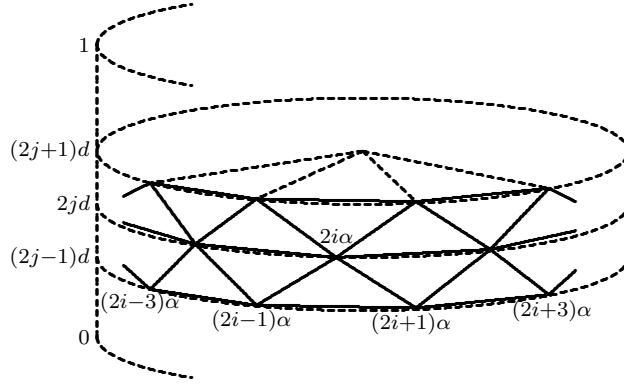
Figure 7.3: a bad approximation of the cylinder

The area of $\sigma(I)$ is approximated by the area of $L_I(I)$. By Proposition 7.1.15, the area of $L_I(I)$ is

$$\|\sigma'(\vec{u}_I^*)(\vec{e}_1) \times \sigma'(\vec{u}_I^*)(\vec{e}_2)\|_2 \mu(I) = \|\sigma_u(\vec{u}_I^*) \times \sigma_v(\vec{u}_I^*)\|_2 \mu(I).$$

Thus the area of the surface is approximated by $\sum_{I \in P} \|\sigma_u(\vec{u}_I^*) \times \sigma_v(\vec{u}_I^*)\|_2 \mu(I)$, which is the Riemann sum for the integral of the function $\|\sigma_u \times \sigma_v\|_2$ on A. Therefore the area of a surface parametrized by a continuously differentiable map $\sigma(u, v) \colon A \subset \mathbb{R}^2 \to \mathbb{R}^n$ is

$$\mu(S) = \int_A \|\sigma_u \times \sigma_v\|_2 dudv = \int_A \sqrt{\|\sigma_u\|_2^2 \|\sigma_v\|_2^2 - (\sigma_u \cdot \sigma_v)^2} dudv.$$

We also denote

$$dA = \|\sigma_u \times \sigma_v\|_2 dudv,$$

where A indicates the area.

Intuitively, the area of a surface should be independent of the parametrization. To verify the intuition, we consider a change of variable $(u, v) = \Phi(s, t) \colon B \to A$. We have $\sigma_s = u_s \sigma_u + v_s \sigma_v$, $\sigma_t = u_t \sigma_u + v_t \sigma_v$, and by (5.2.12),

$$\sigma_s \times \sigma_t = \det \frac{\partial(u, v)}{\partial(s, t)} \sigma_u \times \sigma_v. \tag{7.2.2}$$

Then by the change of variable formula, we have

$$\int_B \|\sigma_s \times \sigma_t\|_2 dsdt = \int_B \left|\det \frac{\partial(u, v)}{\partial(s, t)}\right| \|\sigma_u \times \sigma_v\|_2 dsdt = \int_A \|\sigma_u \times \sigma_v\|_2 dudv.$$

Example 7.2.12. The graph of a continuously differentiable function $f(x, y)$ defined for $(x, y) \in A$ is naturally parametrized as $\sigma(x, y) = (x, y, f(x, y))$. By

$$\sigma_x \times \sigma_y = (1, 0, f_x) \times (0, 1, f_y) = (-f_x, -f_y, 1),$$

the area of the surface is

$$\int_A \|(1, 0, f_x) \times (0, 1, f_y)\|_2 dxdy = \int_A \sqrt{1 + f_x^2 + f_y^2} dxdy.$$

In case $z = f(x, y)$ is implicitly defined by $g(x, y, z) = c$, the formula becomes

$$\int_A \frac{\sqrt{g_x^2 + g_y^2 + g_z^2}}{|g_z|} dxdy = \int_A \frac{dxdy}{\cos\theta},$$

where $\nabla g = (g_x, g_y, g_z)$ is normal to the tangent space, and θ is the angle between the tangent plane and the (x, y)-plane.

Example 7.2.13. The surface obtained by rotating a curve $x = x(t)$, $y = y(t)$, $t \in [a, b]$, around the x-axis is $\sigma(t, \theta) = (x(t), y(t)\cos\theta, y(t)\sin\theta)$. The area is

$$\int_{[a,b]\times[0,2\pi]} \|(x', y'\cos\theta, y'\sin\theta) \times (0, -y\sin\theta, y\cos\theta)\|_2 dtd\theta$$

$$= 2\pi \int_a^b |y(t)|\sqrt{x'(t)^2 + y'(t)^2} dt.$$

In particular, the area of the unit sphere is (see Example 7.1.11)

$$\beta_2 = 2\pi \int_0^\pi |\sin t|\sqrt{(\cos t)'^2 + (\sin t)'^2} dt = 4\pi.$$

The torus in Example 6.1.10 is obtained by rotating the circle $x = a + b\cos\phi$, $z = b\sin\phi$ around the z-axis. So the torus has area

$$2\pi \int_0^{2\pi} (a + b\cos\phi)\sqrt{(a + b\cos\phi)'^2 + (b\sin\phi)'^2} d\phi = 2\pi ab.$$

Example 7.2.14. The parametrized surface $\sigma(u, v) = (\cos(u+v), \sin(u+v), \cos u, \sin u)$ for $0 \le u, v \le \dfrac{\pi}{2}$ has area

$$\int_{[0,\frac{\pi}{2}]} \sqrt{\|\sigma_u\|_2^2\|\sigma_v\|_2^2 - (\sigma_u \cdot \sigma_v)^2}\, dudv = \int_{[0,\frac{\pi}{2}]} \sqrt{2 \cdot 1 - 1^2}\, dudv = \frac{\pi^2}{4}.$$

Exercise 7.2.30. Find the area of the graph of $F\colon \mathbb{R}^2 \to \mathbb{R}^{n-2}$. The graph is a surface in $\mathbb{R}^n$.

Exercise 7.2.31. Study the effect of transforms of $\mathbb{R}^n$ on the areas of surfaces.

1. For the shifting $\vec{x} \to \vec{a} + \vec{x}$, prove $\mu(\vec{a} + S) = \mu(S)$.

2. For the scaling $\vec{x} \to c\vec{x}$, prove $\mu(cS) = c^2\mu(S)$.

3. For an orthogonal linear transform U, prove $\mu(U(S)) = \mu(S)$.

Exercise 7.2.32. Establish and prove the Pappus-Guldinus theorem (see Example 7.1.12) for the area of surfaces obtained by rotation.

Exercise 7.2.33. Find the area of surface.

1. The boundary surface of the common part of the ball $x^2 + y^2 + z^2 \le R^2$ and the cylinder $x^2 + y^2 \le Rx$.

2. The surface $(x^2 + y^2)^{\frac{1}{3}} + z^{\frac{2}{3}} = 1$.

3. The parametrized surface $\sigma(u, v) = (u+v, u-v, u^2+v^2, u^2-v^2)$ for $0 \le u \le a$, $0 \le v \le b$.

7.2.5   Integration of Function on Surface

Suppose $f(\vec{x})$ is a bounded function on a surface S parametrized by $\sigma(u, v)\colon A \subset \mathbb{R}^2 \to \mathbb{R}^n$. For a general partition P of A and choices $(u_I^*, v_I^*) \in I$ for $I \in P$, we have the Riemann sum

$$S(P, f, \sigma) = \sum f(\sigma(u_I^*, v_I^*))\mu(\sigma(I)). \qquad (7.2.3)$$

The integral $\displaystyle\int_S f d\mu = \int_S f dA$ of f on the surface is defined as the limit of the Riemann sum as $\|P\| \to 0$. The notation dA indicates the integration is against the area.

The Riemann sum satisfies

$$|S(P, f, \sigma) - S(P, (f \circ \sigma)\|\sigma_u \times \sigma_v\|_2)|$$

$$\leq \sum |f(\sigma(u_I^*, v_I^*))| \left| \int_I \|\sigma_u \times \sigma_v\|_2 du dv - \|\sigma_u(u_I^*, v_I^*) \times \sigma_v(u_I^*, v_I^*)\|_2 \mu(I) \right|$$

$$\leq \sum |f(\sigma(u_I^*, v_I^*))| \sup_{(u,v)\in I} \|\sigma_u(u, v) \times \sigma_v(u, v) - \sigma_u(u_I^*, v_I^*) \times \sigma_v(u_I^*, v_I^*)\|_2 \mu(I).$$

By the continuity of $\sigma_u \times \sigma_v$ and the boundedness of f, the right side is small when $\|P\|$ is small. Therefore f is Riemann integrable on the surface if and only if $(f \circ \sigma)\|\sigma_u \times \sigma_v\|_2$ is Riemann integrable on A. When the parametrization is regular, this is also equivalent to $f \circ \sigma$ is Riemann integrable on A. Moreover, we have

$$\int_S f dA = \int_A f(\sigma(u, v))\|\sigma_u(u, v) \times \sigma_v(u, v)\|_2 du dv.$$

The integral has properties similar to the integral of functions on rectifiable curves in Section 7.2.2. Moreover, the integral is independent of the choice of the parametrization, either by an argument similar to the area of the surface, or by the fact that the Riemann sum (7.2.3) is essentially given by a partition of the surface and a choice of points on the surface.

Example 7.2.15. Let S be the part of the cone $z = \sqrt{x^2 + y^2}$ cut by the cylinder $(x - 1)^2 + y^2 = 1$. Then $dA = \sqrt{1 + z_x^2 + z_y^2}\,dxdy = \sqrt{2}\,dxdy$, and

$$\int_S (xy + yz + zx)dA = \int_{(x-1)^2+y^2\leq 1} (xy + (x + y)\sqrt{x^2 + y^2})\sqrt{2}\,dxdy.$$

The region $(x - 1)^2 + y^2 \leq 1$ can be described as $0 \leq r \leq 2\cos\theta$, $-\dfrac{\pi}{2} \leq \theta \leq \dfrac{\pi}{2}$ in the polar coordinate. Therefore

$$\int_S (xy + yz + zx)dA = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \left(\int_0^{2\cos\theta} (r^2\cos\theta\sin\theta + r(\cos\theta + \sin\theta)r)\sqrt{2}\,rdr \right) d\theta$$

$$= 4\sqrt{2} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} (\cos\theta\sin\theta + \cos\theta + \sin\theta)\cos^4\theta\,d\theta = \frac{64}{15}\sqrt{2}.$$

Example 7.2.16. For a fixed vector $\vec{a} \in \mathbb{R}^3$, consider the integral $\int_{S^2} f(\vec{a} \cdot \vec{x}) dA$ on the unit sphere. There is a rotation that moves $\vec{a}$ to $(\|\vec{a}\|_2, 0, 0)$. Since the rotation preserves the areas of surfaces in $\mathbb{R}^3$, we have

$$\int_{S^2} f(\vec{a} \cdot \vec{x}) dA = \int_{S^2} f((\|\vec{a}\|_2, 0, 0) \cdot (x, y, z)) dA = \int_{S^2} f(\|\vec{a}\|_2 x) dA$$

Parametrize S^2 by $\sigma(x, \theta) = (x, \sqrt{1 - x^2} \cos\theta, \sqrt{1 - x^2} \sin\theta)$. Then $\|\sigma_x \times \sigma_\theta\|_2 = 1$ and

$$\int_{S^2} f(\|\vec{a}\|_2 x) dA = \int_{-1 \le x \le 1, 0 \le \theta \le 2\pi} f(\|\vec{a}\|_2 x) dx d\theta = 2\pi \int_{-1}^{1} f(\|\vec{a}\|_2 x) dx.$$

Exercise 7.2.34. Use the change of variable formula to prove that the Riemann integral on a surface is independent of the parametrization.

Exercise 7.2.35. Study the effect of transforms of $\mathbb{R}^n$ on the integrals on surfaces.

1. For the shifting $\vec{x} \to \vec{a} + \vec{x}$, prove $\int_{\vec{a}+S} f(\vec{x}) dA = \int_S f(\vec{a} + \vec{x}) dA$.

2. For the scaling $\vec{x} \to c\vec{x}$, prove $\int_{cS} f(\vec{x}) dA = c^2 \int_S f(c\vec{x}) dA$.

3. For an orthogonal linear transform U, prove $\int_{U(S)} f(\vec{x}) dA = \int_S f(U(\vec{x})) dA$.

Exercise 7.2.36. Compute the integral on surface.

1. $\int_{x^2+y^2+z^2=1, a \le z \le b} \dfrac{dA}{z}$, $1 \ge b > a > 0$.

2. $\int_S x^2 dA$ and $\int_S x^2 y^2 dA$, S is the sphere $x^2 + y^2 + z^2 = 1$.

3. $\int_S \dfrac{dA}{(1 + x + y)^2}$, S is the boundary of the tetrahedron $x + y + z \le 1$, $x \ge 0$, $y \ge 0$, $z \ge 0$.

4. $\int_{T^2} (x + y + z) dA$, T^2 is part of the torus in (5.1.20) in the quadrant $x, y \ge 0$.

5. $\int_S (x_1^2 + x_2^2) dA$, S is the surface $\sigma(u, v) = (u + v, u - v, u^2 + v^2, u^2 - v^2)$ for $0 \le u \le a$, $0 \le v \le b$.

Exercise 7.2.37. Compute the attracting force $\int_{S^2} \dfrac{\vec{x} - \vec{a}}{\|\vec{x} - \vec{a}\|_2^3} dA$ of the unit sphere on a point $\vec{a}$.

7.2.6   Integration of 2-Form on Surface

In physics, a flow can be represented as a vector. The flux of a constant flow F through a region A on a plane $\vec{a} \cdot \vec{x} = b$ in $\mathbb{R}^3$ is $(F \cdot \vec{n})\mu(A)$, where $\vec{n} = \dfrac{\vec{a}}{\|\vec{a}\|_2}$ has unit Euclidean length and is called a *normal vector* of the

plane. A plane in $\mathbb{R}^2$ has two normal vectors $\vec{n}$ and $-\vec{n}$ that indicate "two sides" of the plane. The sign of the flux indicates whether the flow is the same as or against the normal direction.

Suppose $S \subset \mathbb{R}^3$ is a surface parametrized by a regular continuously differentiable map $\sigma(u, v) \colon A \subset \mathbb{R}^2 \to \mathbb{R}^3$. Then the parametrization gives a normal vector

$$\vec{n} = \frac{\sigma_u \times \sigma_v}{\|\sigma_u \times \sigma_v\|_2}$$

of (the tangent plane of) the surface. Geometrically, the equality means that the order σ_u, σ_v, $\vec{n}$ follows the "right hand rule". Sometimes instead of one differentiable map that parametrizes the whole surface, we may need several regular parametrizations that combine to cover the whole surface. In this case, we require the normal vectors from different parametrizations coincide on their overlapping. Then the whole surface has a *continuous choice* of the normal vectors. Such a choice is called an *orientation* of the surface.

Example 7.2.17. The graph of a continuously differentiable function $f(x, y)$ is naturally parametrized as $\sigma(x, y) = (x, y, f(x, y))$. By the computation in Exercise 7.2.12, the normal vector induced by the parametrization is

$$\vec{n} = \frac{(-f_x, -f_y, 1)}{\|(-f_x, -f_y, 1)\|_2} = \frac{(-f_x, -f_y, 1)}{\sqrt{1 + f_x^2 + f_y^2}}.$$

Note that the vector points to the positive direction of z, which is upward in the usual way of depicting the (x, y, z) coordinates. To induce the normal vector in the downward direction, we may use the parametrization $\sigma(y, x) = (x, y, f(x, y))$ (or perhaps $\sigma(u, v) = (v, u, f(v, u))$ in the less confusing way).

Example 7.2.18. The sphere $S_R^2 = \{(x, y, z) \colon x^2 + y^2 + z^2 = R^2\}$ of radius R can be covered by the paramerizations

$$\sigma_1(x, y) = (x, y, \sqrt{R^2 - x^2 - y^2}), \ \ x^2 + y^2 < R^2,$$
$$\sigma_2(y, x) = (x, y, -\sqrt{R^2 - x^2 - y^2}), \ \ x^2 + y^2 < R^2,$$
$$\sigma_3(z, x) = (x, \sqrt{R^2 - x^2 - z^2}, z), \ \ x^2 + z^2 < R^2,$$
$$\sigma_4(x, z) = (x, -\sqrt{R^2 - x^2 - z^2}, z), \ \ x^2 + z^2 < R^2,$$
$$\sigma_5(y, z) = (\sqrt{R^2 - y^2 - z^2}, y, z), \ \ y^2 + z^2 < R^2,$$
$$\sigma_6(z, y) = (-\sqrt{R^2 - y^2 - z^2}, y, z), \ \ y^2 + z^2 < R^2.$$

By

$$(\sigma_1)_x \times (\sigma_1)_y = \left(1, 0, -\frac{x}{z}\right) \times \left(0, 1, -\frac{y}{z}\right) = \left(\frac{x}{z}, \frac{y}{z}, 1\right),$$

The normal vector from the first parametrization is

$$\frac{(\sigma_1)_x \times (\sigma_1)_y}{\|(\sigma_1)_x \times (\sigma_1)_y\|_2} = \frac{(x, y, z)}{\|(x, y, z)\|_2} = \frac{\vec{x}}{R}.$$

Similar computations also show that $\dfrac{\vec{x}}{R}$ is the normal vector from the other parametrizations. Therefore the choice $\vec{n} = \dfrac{\vec{x}}{R}$ gives an orientation of the sphere.

Note that the order of the variables in σ_2 are deliberately arranged to have y as the first and x as the second. Thus the normal vector should be computed from $(\sigma_2)_y \times (\sigma_2)_x$ instead of the other way around.

Under a change of variable $u = u(s,t)$, $v = v(s,t)$, by (7.2.2), we find the normal vectors computed from the two parametrizations are the same if and only if $\det \dfrac{\partial(u,v)}{\partial(s,t)} > 0$. Now on the overlapping of two parametrizations $\sigma_1(u,v)$, $\sigma_2(s,t)$ of a surface, the parameters (u,v) and (s,t) are related as a change of variables. We say the two parametrizations are *compatibly oriented* if the Jacobian matrix of the change of variable on the overlapping has positive determinant. Then we conclude that an orientation of the surface is the same as a collection of compatibly oriented parametrizations.

It is possible for a surface to be *nonorientable* in the sense that there is no continuous choice of the normal vectors. This is equivalent to the nonexistence of a collection of compatibly oriented parametrizations. A typical example is the Möbius band.
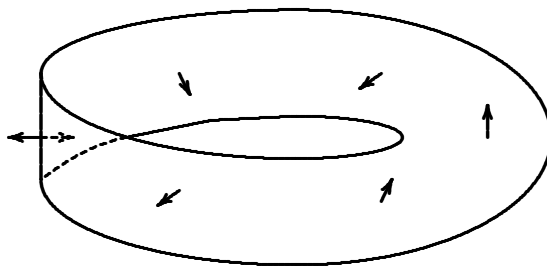

Figure 7.4: Möbius band

Back to flux, suppose $S \subset \mathbb{R}^3$ is a surface with orientation $\vec{n}$. The flow is a vector field $F \colon S \to \mathbb{R}^n$ on the surface. For a regular continuously differentiable parametrization $\sigma \colon A \to S$ compatible with the orientation, a general partition P of A and choices $(u_I^*, v_I^*) \in I$ for $I \in P$, the total flux of the flow through the surface is approximated by the Riemann sum

$$S(P, F, \vec{n}, \sigma) = \sum F(\sigma(u_I^*, v_I^*)) \cdot \vec{n}(\sigma(u_I^*, v_I^*)) \mu(\sigma(I)). \qquad (7.2.4)$$

The integral of the flow along the surface is defined as the limit of the Riemann sum as $\|P\| \to 0$. Since the Riemann sum $S(P, F, \vec{n}, \sigma)$ is the same as the Riemann sum $S(P, F \cdot \vec{n}, \sigma)$ for the integral of the function $F \cdot \vec{n}$ on the surface, the integrability of F along the surface is the same as the integrability of $F \cdot \vec{n}$, and

$$\int_S F \cdot \vec{n} \, dA = \int_A F(\sigma(u,v)) \cdot (\sigma_u(u,v) \times \sigma_v(u,v)) du \, dv. \qquad (7.2.5)$$

The integral has properties similar to the integral of 1-forms along rectifiable curves in Section 7.2.3. Moreover, the integral is independent of the choice of the orientation compatible parametrization, either by an argument similar to the area of the surface, or by the fact that the Riemann sum (7.2.4) is essentially given by a partition of the surface, a choice of points on the surface, and the orientation (indicated by the normal vector $\vec{n}$).

Example 7.2.19. By the computation of the normal vector in Example 7.2.18, the (outgoing) flux of the flow $F = \vec{x} = (x, y, z)$ through the sphere of radius R is

$$\int_{S_R^2} \vec{x} \cdot \frac{\vec{x}}{R} dA = \int_{S_R^2} R dA = R\mu(S_R^2) = R^3 \beta_2 = 4\pi R^3.$$

Example 7.2.20. The outgoing flux of the flow $F = (x^2, y^2, z^2)$ through the sphere $\|\vec{x} - \vec{x}_0\|_2^2 = (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = R^2$ is

$$\int_{\|\vec{x}-\vec{x}_0\|_2=R} (x^2, y^2, z^2) \cdot \frac{(x - x_0, y - y_0, z - z_0)}{R} dA.$$

By the shifting $\vec{x} \to \vec{x}_0 + \vec{x}$, the integral is equal to

$$\frac{1}{R} \int_{\|\vec{x}\|_2=R} ((x + x_0)^2 x + (y + y_0)^2 y + (z + z_0)^2 z) dA.$$

By suitable rotations that exchange the axis, we have $\int_{\|\vec{x}\|_2=R} x^n dA = \int_{\|\vec{x}\|_2=R} y^n dA = \int_{\|\vec{x}\|_2=R} z^n dA$. By the transform $(x, y, z) \to (-x, y, z)$, we also have $\int_{\|\vec{x}\|_2=R} x^n dA = 0$ for odd integers n. Thus the integral above becomes

$$\frac{1}{R} \int_{\|\vec{x}\|_2=R} (2x_0 x^2 + 2y_0 y^2 + 2z_0 z^2) dA = \frac{2}{3R} (x_0 + y_0 + z_0) \int_{\|\vec{x}\|_2=R} (x^2 + y^2 + z^2) dA$$

$$= \frac{8}{3} \pi (x_0 + y_0 + z_0) R^3.$$

Exercise 7.2.38. Use the change of variable formula to prove that the Riemann integral (7.2.5) on a surface is independent of the orientation compatible parametrization.

Exercise 7.2.39. Study the effect of transforms of $\mathbb{R}^n$ on the flux, similar to Exercise 7.2.35.

Exercise 7.2.40. Compute the flux.

1. $F = (x, y, z)$, S is a triangle with vertices $\vec{a}$, $\vec{b}$, $\vec{c}$ and normal vector $\vec{n}$ (which is necessarily orthogonal to $\vec{b} - \vec{a}$ and $\vec{c} - \vec{a}$).

2. $F = (f(x), g(y), h(z))$, S is the boundary of the rectangle $[a_1, a_2] \times [b_1, b_2] \times [c_1, c_2]$ with $\vec{n}$ pointing outward.

3. $F = (xy^2, yz^2, zx^2)$, S is the sphere $(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = 1$ with $\vec{n}$ pointing outward.

The right side of (7.2.5) requires the compatibility between the normal vector $\vec{n}$ and the parametrization $\sigma(u, v)$. The compatibility condition can be included in the integral if a more suitable formal notation is used.

We formally extend the cross product to the differential 1-forms. Then for the parametrization $\sigma(u, v) = (x(u, v), y(u, v), z(u, v))$, we have

$$dx \times dy = (x_u du + x_v dv) \times (y_u du + y_v dv)$$

$$= (x_u y_v - x_v y_u) du \times dv = \det \frac{\partial(x, y)}{\partial(u, v)} du \times dv$$

and the similar formulae for $dy \times dz$ and $dz \times dx$. Furthermore, denote $F = (f, g, h)$. Then

$$
\begin{aligned}
&[F \cdot (\sigma_u \times \sigma_v)] du \times dv \\
={}&[f(y_u z_v - y_v z_u) + g(z_u x_v - z_v x_u) + h(x_u y_v - x_v y_u)] du \times dv \\
={}&f \, dy \times dz + g \, dz \times dx + h \, dx \times dy.
\end{aligned}
$$

We also note the formal property $dv \times du = -du \times dv$, which is compatible with the fact that exchanging u and v would reverse the orientation (changing $\vec{n}$ to $-\vec{n}$) and change the sign of $\int_S F \cdot \vec{n} \, dA$.

The discussion suggests that the integral $\int_S F \cdot \vec{n} \, dA$ should really be denoted $\int_S f \, dy \times dz + g \, dz \times dx + h \, dx \times dy$. Since the cross product is a special case of the exterior product, we call

$$
\omega = f \, dy \wedge dz + g \, dz \wedge dx + h \, dx \wedge dy
$$

a *differential 2-form* on $\mathbb{R}^3$ and, inspired by the computation of $[F \cdot (\sigma_u \times \sigma_v)] du \times dv$ above, define the integral of the differential 2-form on the oriented surface by

$$
\int_S f \, dy \wedge dz + g \, dz \wedge dx + h \, dx \wedge dy = \int_A \left(f \det \frac{\partial(y, z)}{\partial(u, v)} + g \det \frac{\partial(z, x)}{\partial(u, v)} + h \det \frac{\partial(x, y)}{\partial(u, v)} \right) du \, dv.
$$

The use of exterior product allows us to extend the integral of 2-forms to surfaces in the other dimensions. Define the integral of a differential 2-form

$$
\omega = \sum_{i < j} f_{ij}(\vec{x}) dx_i \wedge dx_j
$$

on a surface $S \subset \mathbb{R}^n$ parametrized by a regular continuously differentiable map $\sigma(u, v) \colon A \subset \mathbb{R}^2 \to \mathbb{R}^n$ to be

$$
\int_S \omega = \sum_{i < j} \int_A f_{ij}(\sigma(u, v)) \det \frac{\partial(x_i, x_j)}{\partial(u, v)} du \, dv. \tag{7.2.6}
$$

Similar to the case $n = 3$, we are concerned about whether the integral depends on the choice of the parametrization. Under a change of variable $u = u(s, t)$, $v = v(s, t)$, we have

$$
\begin{aligned}
&\int_{A_{uv}} f(\sigma(u, v)) \det \frac{\partial(x_i, x_j)}{\partial(u, v)} du \, dv \\
={}&\int_{A_{st}} f(\sigma(s, t)) \det \frac{\partial(x_i, x_j)}{\partial(u, v)} \left| \det \frac{\partial(u, v)}{\partial(s, t)} \right| ds \, dt.
\end{aligned}
$$

By $\dfrac{\partial(x_i, x_j)}{\partial(s, t)} = \dfrac{\partial(x_i, x_j)}{\partial(u, v)} \dfrac{\partial(u, v)}{\partial(s, t)}$, the right side fits into the definition (7.2.6) in terms of the new variable (s, t) if and only if $\det \dfrac{\partial(u, v)}{\partial(s, t)} > 0$. Therefore

the definition of the integral of a 2-form on a surface is not changed if the parametrizations are compatibly oriented.

In general, a surface $S \subset \mathbb{R}^n$ may be covered by finitely many parametrized pieces $\sigma_i \colon A_i \subset \mathbb{R}^2 \to S_i \subset \mathbb{R}^n$, $S = \cup S_i$. On the overlap $S_i \cap S_j$ of the pieces, we get the change of variable among different parametrizations (called *transition map*)

$$\varphi_{ij} = \sigma_j^{-1}\sigma_i \colon A_{ij} = \sigma_i^{-1}(S_i \cap S_j) \to A_{ji} = \sigma_j^{-1}(S_i \cap S_j).$$

The parametrizations are compatibly oriented if $\det \varphi'_{ij} > 0$ for any i, j. The surface is orientable if there is a compatibly oriented collection of parametrizations that covers the surface. Such a collection gives an orientation of the surface. Any other parametrization $\sigma_* \colon A \subset \mathbb{R}^2 \to S_* \subset S$ is compatible with the orientation if $\det(\sigma_i^{-1} \circ \sigma_*)' > 0$ on $\sigma^{-1}(S_* \cap S_i)$.

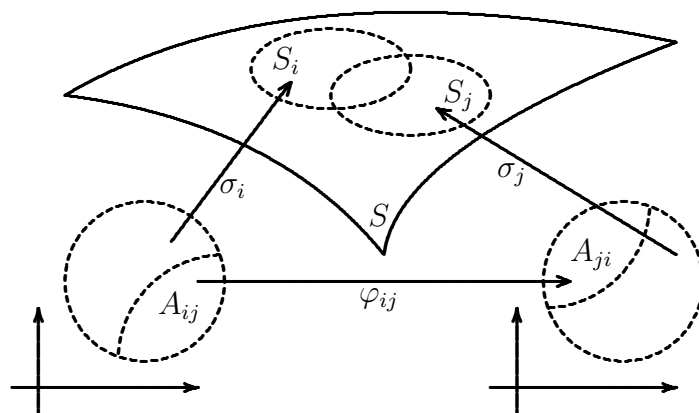Figure 7.5: transition between overlapping parametrizations

Choose orientation compatible parametrizations σ_i that cover the surface, such that the intersections $S_i \cap S_j$ have area zero. Then the integral of a 2-form ω on the surface is defined as

$$\int_S \omega = \sum \int_{S_i} \omega, \tag{7.2.7}$$

where the integrals inside the sum on the right side are defined by (7.2.6). The definition is independent of the choice of orientation compatible parametrizations.

Example 7.2.21. The flux in Example 7.2.19 is the integral of the 2-form $xdy \wedge dz + ydz \wedge dx + zdx \wedge dy$ on the sphere S_R^2 with orientation given by compatible parametrizations in Example 7.2.18. The integral of $zdx \wedge dy$ may be computed by using σ_1 and σ_2

$$\int_{S_R^2} zdx \wedge dy = \int_{x^2+y^2 \le R^2} \sqrt{1 - x^2 - y^2} \det \frac{\partial(x,y)}{\partial(x,y)} dxdy$$

$$+ \int_{x^2+y^2 \le R^2} -\sqrt{1 - x^2 - y^2} \det \frac{\partial(x,y)}{\partial(y,x)} dxdy$$

$$= 2\int_0^{2\pi} \left(\int_0^R \sqrt{1 - r^2}\, rdr \right) d\theta = \frac{4}{3}\pi R^3.$$

Similarly, we get $\displaystyle\int_{S_R^2} x\,dy \wedge dz = \int_{S_R^2} y\,dz \wedge dx = \frac{4}{3}\pi R^3$.

Example 7.2.22. We compute the outgoing flux of $F = (x^2, y^2, z^2)$ in Example 7.2.20 through the ellipse S given by $\dfrac{(x - x_0)^2}{a^2} + \dfrac{(y - y_0)^2}{b^2} + \dfrac{(z - z_0)^2}{c^2} = 1$. By the shifting $\vec{x} \to \vec{x}_0 + \vec{x}$, the flux is the integral

$$\int_{S + \vec{x}_0} (x + x_0)^2 dy \wedge dz + (y + y_0)^2 dz \wedge dx + (z + z_0)^2 dx \wedge dy.$$

The integral of $(z + z_0)^2 dx \wedge dy$ may be computed by using parametrizations similar to σ_1 and σ_2 in Example 7.2.18

$$\int_{S + \vec{x}_0} (z + z_0)^2 dx \wedge dy = \int_{\frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1} \left(c\sqrt{1 - \frac{x^2}{a^2} - \frac{y^2}{b^2}} + z_0 \right)^2 \det \frac{\partial(x, y)}{\partial(x, y)} dx\,dy$$

$$+ \int_{\frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1} \left(-c\sqrt{1 - \frac{x^2}{a^2} - \frac{y^2}{b^2}} + z_0 \right)^2 \det \frac{\partial(x, y)}{\partial(y, x)} dx\,dy$$

$$= 4z_0 c \int_{\frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1} \sqrt{1 - \frac{x^2}{a^2} - \frac{y^2}{b^2}} dx\,dy$$

$$= 4abc z_0 \int_{u^2 + v^2 \leq 1} \sqrt{1 - u^2 - v^2} du\,dv = \frac{16}{3}\pi abc z_0.$$

The total flux is $\dfrac{16\pi}{3} abc(x_0 + y_0 + z_0)$.

Exercise 7.2.41. Compute the integral of 2-form.

1. $\displaystyle\int_S xy^2 dy \wedge dz + yz^2 dz \wedge dx + zx^2 dx \wedge dy$, S is the ellipse $\dfrac{(x - x_0)^2}{a^2} + \dfrac{(y - y_0)^2}{b^2} + \dfrac{(z - z_0)^2}{c^2} = 1$ with orientation given by the inward normal vector.

2. $\displaystyle\int_S xyz\,dx \wedge dy$, S is the part of the ellipse $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c^2} = 1$, $x \geq 0$, $y \geq 0$, with orientation given by outward normal vector.

3. $\displaystyle\int_S xy\,dy \wedge dz$ and $\displaystyle\int_S x^2 y\,dy \wedge dz$, S is the boundary of the solid enclosed by $z = x^2 + y^2$ and $z = 1$, with orientation given by outward normal vector.

4. $\displaystyle\int_S dx_1 \wedge dx_2 + dx_2 \wedge dx_3 + \cdots + dx_{n-1} \wedge dx_n$, S is the surface $\sigma(u, v) = (u + v, u^2 + v^2, \ldots, u^n + v^n)$ for $0 \leq u \leq a$, $0 \leq v \leq b$.

Exercise 7.2.42. Prove that the area of a surface $S \subset \mathbb{R}^3$ given by an equation $g(x, y, z) = c$ is $\displaystyle\int_S \frac{g_x dy \wedge dz + g_y dz \wedge dx + g_z dx \wedge dy}{\sqrt{g_x^2 + g_y^2 + g_z^2}}$.

7.2.7   Volume and Integration on Hypersurface

Suppose $S \subset \mathbb{R}^k$ is a k-dimensional hypersurface parametrized by a continuously differentiable map $\sigma(\vec{u}) = \sigma(u_1, u_2, \ldots, u_k) \colon \mathbb{R}^k \to \mathbb{R}^n$ on a subset

$A \subset \mathbb{R}^k$ with volume. The (k-dimensional) volume of the hypersurface is

$$\mu(S) = \int_A \|\sigma_{u_1} \wedge \sigma_{u_2} \wedge \cdots \wedge \sigma_{u_k}\|_2 du_1 du_2 \cdots du_k$$
$$= \int_A \sqrt{\det(\sigma_{u_i} \cdot \sigma_{u_j})_{1 \leq i,j \leq k}} du_1 du_2 \cdots du_k. \qquad (7.2.8)$$

We also denote

$$dV = \|\sigma_{u_1} \wedge \sigma_{u_2} \wedge \cdots \wedge \sigma_{u_k}\|_2 du_1 du_2 \cdots du_k,$$

where V indicates the volume. The definition is independent of the choice of the parametrization.

Example 7.2.23. In Example 7.1.11, for a parametrization $F(\vec{\theta}) \colon \mathbb{R}^{n-1} \to \mathbb{R}^n$ of the unit sphere S^{n-1}, we claimed that the integral $\beta_{n-1} = \int_A |\det \begin{pmatrix} F & F' \end{pmatrix}| d\mu_{\vec{\theta}}$ is the volume of the sphere. Now the claim can be justified.

Since $F \cdot F = \|F\|_2^2 = 1$, by taking partial derivatives, we get $F_{\theta_i} \cdot F = 0$. Therefore F is a unit length vector orthogonal to F_{θ_i}. This implies that

$$\|F_{\theta_1} \wedge F_{\theta_2} \wedge \cdots \wedge F_{\theta_{n-1}}\|_2 = \|F \wedge F_{\theta_1} \wedge F_{\theta_2} \wedge \cdots \wedge F_{\theta_{n-1}}\|_2 = |\det \begin{pmatrix} F & F' \end{pmatrix}|,$$

where the second equality follows from the interpretation of the determinant as the wedge product of n vectors in $\mathbb{R}^n$. Thus the volume of the sphere is

$$\int_A \|F_{\theta_1} \wedge F_{\theta_2} \wedge \cdots \wedge F_{\theta_{n-1}}\|_2 d\mu_{\vec{\theta}} = \int_A |\det \begin{pmatrix} F & F' \end{pmatrix}| d\mu_{\vec{\theta}}.$$

Continuing with the computation of the volume of the unit sphere, we note that the parametrization of S^{n-1} induces a parametrization of S^n

$$G(\vec{\theta}, \phi) = (F(\vec{\theta}) \cos \phi, \sin \phi) \colon A \times \left[-\frac{\pi}{2}, \frac{\pi}{2} \right] \to \mathbb{R}^{n+1}.$$

We have

$$\|G_{\theta_1} \wedge G_{\theta_2} \wedge \cdots \wedge G_{\theta_{n-1}} \wedge G_\phi\|_2$$
$$= \|(F_{\theta_1} \cos \phi, 0) \wedge (F_{\theta_2} \cos \phi, 0) \wedge \cdots \wedge (F_{\theta_{n-1}} \cos \phi, 0) \wedge (-F \sin \phi, \cos \phi)\|_2$$
$$= \|(F_{\theta_1} \cos \phi, 0) \wedge (F_{\theta_2} \cos \phi, 0) \wedge \cdots \wedge (F_{\theta_{n-1}} \cos \phi, 0)\|_2 \|(-F \sin \phi, \cos \phi)\|_2$$
$$= \|F_{\theta_1} \wedge F_{\theta_2} \wedge \cdots \wedge F_{\theta_{n-1}}\|_2 \cos^{n-1} \phi,$$

where the second equality is due to the fact that $(-F \sin \phi, \cos \phi)$ is orthogonal to $(F_{\theta_i} \cos \phi, 0)$, and we use $\|(-F \sin \phi, \cos \phi)\|_2 = \sqrt{\|F\|_2^2 \sin^2 \phi + \cos^2 \phi} = \sqrt{\sin^2 \phi + \cos^2 \phi} = 1$ in the last equality. Therefore

$$\beta_n = \int_{\vec{\theta} \in A, -\frac{\pi}{2} \leq \phi \leq \frac{\pi}{2}} \|F_{\theta_1} \wedge F_{\theta_2} \wedge \cdots \wedge F_{\theta_{n-1}}\|_2 \cos^{n-1} \phi d\mu_{\vec{\theta}} d\phi = \beta_{n-1} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^{n-1} \phi d\phi.$$

Using integration by parts, we get (see Exercise 3.2.19)

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^n \phi d\phi = \frac{n-1}{n} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^{n-2} \phi d\phi = \cdots = \begin{cases} \dfrac{(n-1)(n-3)\cdots 1}{n(n-2)\cdots 2} \pi & \text{if } n \text{ is even} \\[2mm] \dfrac{(n-1)(n-3)\cdots 2}{n(n-2)\cdots 3} 2 & \text{if } n \text{ is odd} \end{cases}.$$

Therefore $\beta_n = \dfrac{2\pi}{n-1}\beta_{n-2}$. Combined with $\beta_1 = 2\pi$, $\beta_2 = 4\pi$, we get

$$\beta_{n-1} = \begin{cases} \dfrac{2\pi^k}{(k-1)!} & \text{if } n = 2k \\ \dfrac{2^{k+1}\pi^k}{(2k-1)(2k-3)\cdots 1} & \text{if } n = 2k+1 \end{cases}.$$

Finally, by Example 7.1.11, the volume of the unit ball is

$$\alpha_n = \frac{\beta_{n-1}}{n} = \begin{cases} \dfrac{\pi^k}{k!} & \text{if } n = 2k \\ \dfrac{2^{k+1}\pi^k}{(2k+1)(2k-1)(2k-3)\cdots 1} & \text{if } n = 2k+1 \end{cases}.$$

The result may be compared with Exercise 7.1.35.

A function $f(\vec{x})$ is Riemann integrable on the hypersurface S if $f(\sigma(\vec{u}))$ is Riemann integrable on A. The integral is

$$\int_S f dV = \int_A f(\sigma(u_1, u_2, \dots, u_k))\|\sigma_{u_1} \wedge \sigma_{u_2} \wedge \cdots \wedge \sigma_{u_k}\|_2 du_1 du_2 \cdots du_k$$

The integral is independent of the choice of regular parametrizations.

Example 7.2.24. Consider a k-dimensional hypersurface S parametrized by

$$\sigma(\vec{u}) = (\xi(\vec{u}), r(\vec{u})) \colon A \subset \mathbb{R}^k \to \mathbb{R}^n \times \mathbb{R}.$$

Assume $r(\vec{u}) \geq 0$, so that the hypersurface is inside the upper half Euclidean space. The l-dimensional rotation of S around the axis $\mathbb{R}^n \times 0$ is a $(k+l)$-dimensional hypersurface

$$\rho_l(S) = \{(\vec{x}, \vec{y}) \in \mathbb{R}^n \times \mathbb{R}^{l+1} \colon (\vec{x}, \|\vec{y}\|_2) \in S\}.$$

Let $F(\vec{\theta}) \colon B \subset \mathbb{R}^l \to \mathbb{R}^{l+1}$ be a parametrization of the unit sphere S^l. Then the rotation hypersurface $\rho_l(S)$ is parametrized by

$$\rho(\vec{u}, \vec{\theta}) = (\xi(\vec{u}), r(\vec{u})F(\vec{\theta})) \colon A \times B \to \mathbb{R}^n \times \mathbb{R}^{l+1}.$$

By $\rho_{u_i} = (\xi_{u_i}, r_{u_i}F)$, $\rho_{\theta_j} = (\vec{0}, rF_{\theta_j})$, and $\rho_{u_i} \cdot \rho_{\theta_j} = r_{u_i}F \cdot rF_{\theta_j} = 0$, we get

$$\|\rho_{u_1} \wedge \rho_{u_2} \wedge \cdots \wedge \rho_{u_k} \wedge \rho_{\theta_1} \wedge \rho_{\theta_2} \wedge \cdots \wedge \rho_{\theta_l}\|_2 = \|\rho_{u_1} \wedge \rho_{u_2} \wedge \cdots \wedge \rho_{u_k}\|_2 \|\rho_{\theta_1} \wedge \rho_{\theta_2} \wedge \cdots \wedge \rho_{\theta_l}\|_2,$$

and

$$\|\rho_{\theta_1} \wedge \rho_{\theta_2} \wedge \cdots \wedge \rho_{\theta_l}\|_2 = r^l \|F_{\theta_1} \wedge F_{\theta_2} \wedge \cdots \wedge F_{\theta_l}\|_2.$$

Moreover, there is an orthogonal transform U_F on $\mathbb{R}^{l+1}$ such that $U_F(F) = (1, 0, \dots, 0)$. Then (id, U_F) is an orthogonal transform on $\mathbb{R}^n \times \mathbb{R}^{l+1}$ such that

$$(\mathrm{id}, U_F)\rho_{u_i} = (\xi_{u_i}, r_{u_i}U_F(F)) = (\xi_{u_i}, r_{u_i}, 0, \dots, 0) = (\sigma_{u_i}, 0, \dots, 0).$$

Applying the orthogonal transform to $\|\rho_{u_1} \wedge \rho_{u_2} \wedge \cdots \wedge \rho_{u_k}\|_2$, we get

$$\|\rho_{u_1} \wedge \rho_{u_2} \wedge \cdots \wedge \rho_{u_k}\|_2 = \|\sigma_{u_1} \wedge \sigma_{u_2} \wedge \cdots \wedge \sigma_{u_k}\|_2$$

Thus the volume of the rotation hypersurface is

$$\mu(\rho_l(S)) = \int_{A \times B} \|\sigma_{u_1} \wedge \sigma_{u_2} \wedge \cdots \wedge \sigma_{u_k}\|_2 r^l \|F_{\theta_1} \wedge F_{\theta_2} \wedge \cdots \wedge F_{\theta_l}\|_2 d\mu_{\vec{u}} d\mu_{\vec{\theta}}$$

$$= \beta_l \int_A \|\sigma_{u_1} \wedge \sigma_{u_2} \wedge \cdots \wedge \sigma_{u_k}\|_2 r^l d\mu_{\vec{u}} = \beta_l \int_S r^l dV.$$

Exercise 7.2.43. Use the rotation to prove that $\beta_{k+l+1} = \beta_k \beta_l \int_0^{\frac{\pi}{2}} \cos^k \theta \sin^l \theta d\theta$.

Exercise 7.2.44. Extend the result of Example 7.2.24 to the rotation around a hyperplane $\vec{a} \cdot \vec{x} = b$ (the surface is on the positive side of the hyperplane).

A *differential k-form* on $\mathbb{R}^n$ is an expression of the form

$$\omega = \sum_{i_1 < i_2 < \cdots < i_k} f_{i_1 i_2 \cdots i_k}(\vec{x}) dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_k}.$$

The integral of the differential k-form on the hypersurface S with respect to a regular parametrization σ is

$$\int_S \omega = \int_A f_{i_1 i_2 \cdots i_k}(\sigma(\vec{u})) \det \frac{\partial(x_{i_1}, x_{i_2}, \ldots, x_{i_k})}{\partial(u_1, u_2, \ldots, u_k)} du_1 du_2 \cdots du_k.$$

Similar to the integral of differential 2-forms on surfaces, the integral is independent of the parametrizations as long as they are compatibly oriented. The compatibility means that the determinant of the Jacobian for the change of variables (so called transition map) is positive. In general, the hypersurface may be divided into several compatibly oriented pieces and the integral of the differential form is the sum of the integrals on the pieces.

Now we consider the special case of the integral of an $(n-1)$-form on an oriented $(n-1)$-dimensional hypersurface $S \subset \mathbb{R}^n$. At any $\vec{x} \in S$, there are two unit length vectors orthogonal to the tangent hyperplane. Specifically, if $\sigma(\vec{u})$ is a regular parametrization, then the tangent hyperplane is spanned by $\sigma_{u_1}, \sigma_{u_2}, \ldots, \sigma_{u_{n-1}}$. Define the normal vector compatible with the orientation of the parametrization to be

$$\vec{n} = (-1)^{n-1} \frac{(\sigma_{u_1} \wedge \sigma_{u_2} \wedge \cdots \wedge \sigma_{u_{n-1}})^\star}{\|\sigma_{u_1} \wedge \sigma_{u_2} \wedge \cdots \wedge \sigma_{u_{n-1}}\|_2}.$$

The compatible normal vector is characterized by the property that $\vec{n}$ has length 1, is orthogonal to σ_{u_i}, and the basis $\vec{n}, \sigma_{u_1}, \sigma_{u_2}, \ldots, \sigma_{u_{n-1}}$ has positive orientation (meaning $\det \begin{pmatrix} \vec{n} & \sigma_{u_1} & \sigma_{u_2} & \cdots & \sigma_{u_{n-1}} \end{pmatrix} > 0$).

Let ($\widehat{x}_i$ means the term x_i is missing from the list)

$$\vec{a} = (\sigma_{u_1} \wedge \sigma_{u_2} \wedge \cdots \wedge \sigma_{u_{n-1}})^\star = \sum \det \frac{\partial(x_1, x_2, \ldots, \widehat{x}_i, \ldots, x_n)}{\partial(u_1, u_2, \ldots, u_{n-1})} \vec{e}^{\,\star}_{\wedge([n]-i)}$$

$$= \sum (-1)^{n-i} \det \frac{\partial(x_1, x_2, \ldots, \widehat{x}_i, \ldots, x_n)}{\partial(u_1, u_2, \ldots, u_{n-1})} \vec{e}_i.$$

Then define the flux of a vector field $F = (f_1, f_2, \ldots, f_n) \colon S \to \mathbb{R}^n$ to be

$$\int_S F \cdot \vec{n} dV = (-1)^{n-1} \int_A F \cdot \vec{a} \, du_1 du_2 \cdots du_k$$

$$= \int_A \sum (-1)^{i-1} f_i \det \frac{\partial(x_1, x_2, \ldots, \widehat{x}_i, \ldots, x_n)}{\partial(u_1, u_2, \ldots, u_{n-1})} du_1 du_2 \cdots du_k$$

$$= \int_S \sum (-1)^{i-1} f_i dx_1 \wedge dx_2 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_n. \qquad (7.2.9)$$

This extends the relation between the flux in $\mathbb{R}^3$ and the integration of 2-forms in Section 7.2.6.

For the special case $n = 2$, we have an oriented curve $C \subset \mathbb{R}^2$ parametrized by $\phi(t)$, and the compatible normal vector $\vec{n}$ is the unit length vector with direction obtained by rotating the tangent vector ϕ' in clockwise direction by 90 degrees. Moreover,

$$\int_C (f, g) \cdot \vec{n} ds = \int_C f dy - g dx = \int_C (-g, f) \cdot d\vec{x}.$$

Exercise 7.2.45. Denote

$$\pi(v_0, v_1, \dots, v_n) = (v_1, v_0, v_2, \dots, v_n),$$
$$\rho_i(v_0, v_1, \dots, v_n) = (v_{i+1}, v_{i+2}, \dots, v_n, v_0, v_1, \dots, v_{i-1}, v_i).$$

Prove that the unit sphere S^n is covered by compatibly oriented regular parametrizations

$$\sigma_i^+(\vec{u}) = \rho_i \left(\sqrt{1 - \|\vec{u}\|_2^2}, \vec{u} \right), \quad \sigma_i^-(\vec{u}) = \rho_i \pi \left(-\sqrt{1 - \|\vec{u}\|_2^2}, \vec{u} \right)$$

defined for $1 \le i \le n$ and $\vec{u} \in \mathbb{R}^n$ satisfying $\|\vec{u}\|_2 < 1$.

Exercise 7.2.46. Use (7.2.9) to show that the length of a curve C in $\mathbb{R}^2$ given by an equation $g(x, y) = c$ is $\int_C \dfrac{-g_y dx + g_x dy}{\sqrt{g_x^2 + g_y^2}}$, similar to the formula in Exercise 7.2.42. Extend the formula to the volume of an $(n-1)$-dimensional hypersurface in $\mathbb{R}^n$ given by $g(x_1, x_2, \dots, x_n) = c$. What about the $(n-k)$-dimensional hypersurface defined by a map $G \colon \mathbb{R}^n \to \mathbb{R}^k$?

Exercise 7.2.47. Compute the integral.

1. $\int_{S_R^{n-1}} (\vec{x} - \vec{a}) \cdot \vec{n} dV$, the normal vector $\vec{n}$ points to the outside of the sphere.

2. $\int_{S_R^{n-1}} (a_1 x_1 + a_2 x_2 + \cdots + a_n x_n) dx_2 \wedge dx_3 \wedge \cdots \wedge dx_n$, the orientation of the sphere is given by the outward normal vector.

3. $\int_S dx_1 \wedge dx_2 \wedge dx_3 + dx_2 \wedge dx_3 \wedge dx_4 + \cdots + dx_{n-2} \wedge dx_{n-1} \wedge dx_n$, S is the surface $\sigma(u, v, w) = \rho(u) + \rho(v) + \rho(w)$, $\rho(u) = (u, u^2, \dots, u^n)$, $0 \le u, v, w \le a$.

We end the section with the discussion of the direction of the normal vector for the hypersurfaces given by graphs of functions.

Consider the graph $\sigma(\vec{u}) = (\vec{u}, h(\vec{u}))$ of a continuously differentiable function h on $\mathbb{R}^{n-1}$. By

$$(\sigma_{u_1} \wedge \sigma_{u_2} \wedge \cdots \wedge \sigma_{u_{n-1}})^\star$$
$$= ((\vec{e}_1 + h_{u_1}\vec{e}_n) \wedge (\vec{e}_2 + h_{u_2}\vec{e}_n) \wedge \cdots \wedge (\vec{e}_{n-1} + h_{u_{n-1}}\vec{e}_n))^\star$$
$$= (\vec{e}_1 \wedge \vec{e}_2 \wedge \cdots \wedge \vec{e}_{n-1})^\star + \sum (-1)^{n-1-i} h_{u_i}(\vec{e}_1 \wedge \vec{e}_2 \wedge \cdots \wedge \widehat{\vec{e}_i} \wedge \cdots \wedge \vec{e}_n)^\star$$
$$= \vec{e}_n - \sum h_{u_i}\vec{e}_i = (-h_{u_1}, -h_{u_2}, \dots, -h_{u_{n-1}}, 1),$$

the normal vector

$$\vec{n} = \frac{(-1)^{n-1}(-h_{u_1}, -h_{u_2}, \ldots, -h_{u_{n-1}}, 1)}{\sqrt{1 + h_{u_1}^2 + h_{u_2}^2 + \ldots + h_{u_{n-1}}^2}}$$

points in the direction of x_n when n is odd and opposite to the direction of x_n when n is even.

In general, consider the graph

$$\sigma_i(\vec{u}) = \sigma_i(u_1, u_2, \ldots, u_{n-1}) = (u_1, u_2, \ldots, u_{i-1}, h(\vec{u}), u_i, \cdots, u_{n-1})$$

of a continuously differentiable function h. We have $\sigma_i(\vec{u}) = U_i(\sigma(\vec{u}))$ for the orthogonal transform

$$U(x_1, x_2, \ldots, x_n) = (x_1, x_2, \ldots, x_{i-1}, x_n, x_{i+1}, \ldots, x_{n-1}).$$

By the characterization of the compatible normal vector, the normal vector compatible with the parametrization σ_i is

$$(\det U)U(\vec{n}) = \frac{(-1)^{i-1}(-h_{u_1}, -h_{u_2}, \ldots, -h_{u_{i-1}}, 1, -h_{u_i}, \ldots, -h_{u_{n-1}})}{\sqrt{1 + h_{u_1}^2 + h_{u_2}^2 + \ldots + h_{u_{n-1}}^2}}.$$

The normal direction points in the direction of x_i when i is odd and opposite to the direction of x_i when i is even.

7.3 Stokes Theorem

The fundamental theorem of calculus relates the integration of a function on an interval to the values of the antiderivative at the end points of the integral. The Stokes theorem extends the result to high dimensional integrations.

7.3.1 Green Theorem

A curve $\phi\colon [a, b] \to \mathbb{R}^n$ is *simple* if it has no self intersection. It is *closed* if $\phi(a) = \phi(b)$. The concept of simple closed curve is independent of the choice of parametrization.

A simple closed curve in $\mathbb{R}^2$ divides the plane into two path connected pieces, one bounded and the other unbounded. Therefore we have *two sides* of a simple closed curve in $\mathbb{R}^2$. Suppose $A \subset \mathbb{R}^2$ is a bounded subset with finitely many rectifiable simple closed curves $C_1, C_2, \ldots, C_k$ as the boundary. Since rectifiable curves have area zero (see Exercise 7.2.3), the subset A has area. We further assume that A is contained in only one side of each C_i. Then C_i has a *compatible orientation* such that A is "on the left" of C. This means that C_i has *counterclockwise* orientation if A is in the bounded side of C_i, and has *clockwise* orientation if A is in the unbounded side. This also means that if C_i is oriented and differentiable, then rotating the tangent vector $\vec{t}$ by 90 degrees in clockwise direction will produce a normal vector $\vec{n}$ pointing away from A.
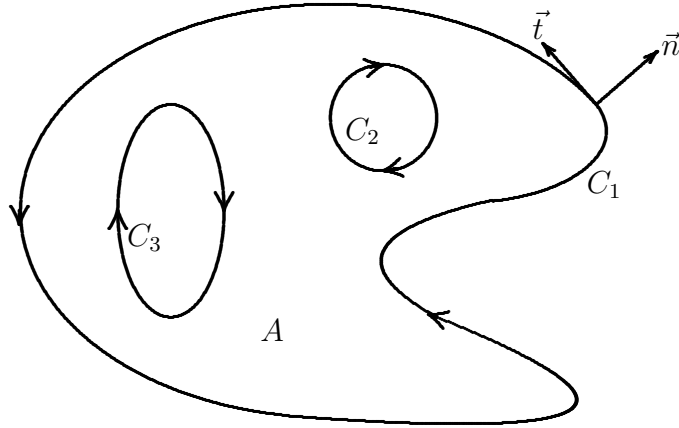
Figure 7.6: orientation of boundary

Theorem 7.3.1 (Green Theorem). *Suppose $A \subset \mathbb{R}^2$ is a region with compatibly oriented rectifiable simple closed boundary curves C. Then for any continuously differentiable functions f and g, we have*

$$\int_C f\,dx + g\,dy = \int_A (-f_y + g_x)\,dxdy. \tag{7.3.1}$$

In Green theorem, C is understood to be the union of simple closed curves $C_1, C_2, \ldots, C_k$, each oriented in the compatible way. The integral in (7.3.1) along C is understood to be

$$\int_C = \int_{C_1} + \int_{C_2} + \cdots + \int_{C_k}.$$

Proof. We first consider the special case that

$$A = \{(x, y) \colon x \in [a.b], h(x) \geq y \geq k(x)\}$$

is a region between the graphs of functions $h(x)$ and $k(x)$ with bounded variations and $g = 0$. We have

$$\int_A f_y\,dxdy = \int_a^b \left(\int_{k(x)}^{h(x)} f_y(x, y)\,dy \right) dx \qquad \text{(Fubini Theorem)}$$

$$= \int_a^b (f(x, h(x)) - f(x, k(x)))\,dx. \quad \text{(fundamental theorem)}$$

On the other hand, the boundary C consists of four segments with parametrizations compatible with the orientation

$$\begin{aligned}
C_b &\colon \phi(t) = (t, k(t)), & t &\in [a, b] \\
C_r &\colon \phi(t) = (b, t), & t &\in [k(b), h(b)] \\
C_t &\colon \phi(t) = (-t, h(-t)), & t &\in [-b, -a] \\
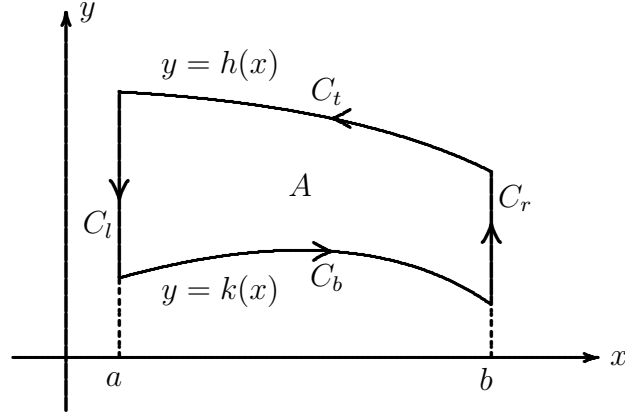C_l &\colon \phi(t) = (a, -t), & t &\in [-h(a), -k(a)]
\end{aligned}$$

Figure 7.7: Green theorem for a special case

Then

$$\int_C f dx = \left(\int_{C_b} + \int_{C_r} + \int_{C_t} + \int_{C_l} \right) f dx$$

$$= \int_a^b f(t, k(t)) dt + \int_{k(b)}^{h(b)} f(b, t) db$$

$$+ \int_{-b}^{-a} f(-t, h(-t)) d(-t) + \int_{-h(a)}^{-k(a)} f(a, -t) da$$

$$= \int_a^b f(x, k(x)) dx + 0 + \int_b^a f(x, h(x)) dx + 0.$$

Therefore we have $\int_A f_y dx dy = - \int_C f dx$.

Many regions can be divided by vertical lines into several special regions A_j studied above. Let C_j be the compatibly oriented boundary of A_j. Then the sum of $\int_{C_j} f dx$ is $\int_C f dx$, because the integrations along the vertical lines are all zero. Moreover, the sum of $\int_{A_j} f_y dx dy$ is $\int_A f_y dx dy$. This proves the equality $\int_C f dx = - \int_A f_y dx dy$ for the more general regions.

Now consider the most general case that the boundary curves are assumed to be only rectifiable. To simplify the presentation, we will assume that A has only one boundary curve C. Parametrize the curve C by the Euclidean arc length $\phi(s) \colon [a, b] \to \mathbb{R}^2$. For any natural number n, divide $[a, b]$ evenly into n intervals of length $\delta = \dfrac{b - a}{n}$. Denote by C_j the segment of C between $\phi(s_{j-1})$ and $\phi(s_j)$. Let L_j be the straight line connecting $\phi(s_{j-1})$ and $\phi(s_j)$. Let L be the curve constructed by taking the union of L_j. Let A' be the region enclosed by the curve L. Then A' can be divided by vertical lines into special regions. As explained above, we have

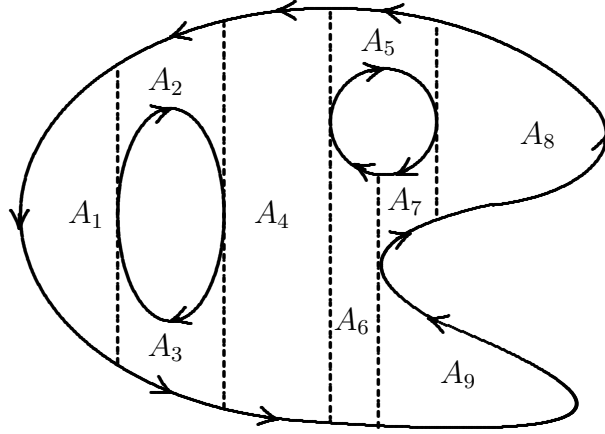$$\int_L f dx = - \int_{A'} f_y dx dy.$$

Figure 7.8: divide a general case to special cases

Let M be the upper bound of $\|\nabla f\|_2 = \sqrt{f_x^2 + f_y^2}$. Since s is the Euclidean arc length, we have $\|\phi(s) - \phi(t)\|_2 \le |s - t|$. Therefore for $\vec{x} \in C_j$ and $\vec{y} \in L_j$, we have

$$|f(\vec{x}) - f(\vec{y})| \le M\|\vec{x} - \vec{y}\|_2 \le M\|\vec{x} - \phi(s_j)\|_2 + M\|\phi(s_j) - \vec{y}\|_2 \le 2M\delta.$$

This implies

$$\left| S(P, f dx, \phi) - \int_L f dx \right| = \left| \sum \left(f(\phi(s_j^*))\Delta x_j - \int_{L_j} f dx \right) \right|$$

$$\le \sum \left| \int_{L_j} (f(\phi(s_j^*)) - f(\vec{x})) dx \right|$$

$$\le \sum 2M\delta|\Delta x_j| \le 2M\delta \sum \Delta s_j = 2M(b - a)\delta.$$

The estimation shows that $\lim_{n\to\infty} \int_L f dx = \int_C f dx$.

On the other hand, let B_j be the region bounded by C_j and L_j. Then $B_j \subset B(\phi(s_j), \delta)$. Therefore $\mu(B_j) \le \pi\delta^2$ and

$$\left| \int_A f dx dy - \int_{A'} f dx dy \right| \le \sum \int_{B_j} |f| dx dy$$

$$\le \sum M\mu(B_j) \le nM\pi\delta^2 = \frac{\pi M(b - a)^2}{n}.$$

This shows that $\lim_{n\to\infty} \int_{A'} f dx dy = \int_A f dx dy$.

This completes the proof of $\int_C f dx = -\int_A f_y dx dy$ for regions with rectifiable boundaries. The formula $\int_C g dy = \int_A g_x dx dy$ can be similarly proved.

$\square$

Suppose a curve C is closed but not simple. Then C may enclose some parts of the region more than once, and the orientation of C may not be
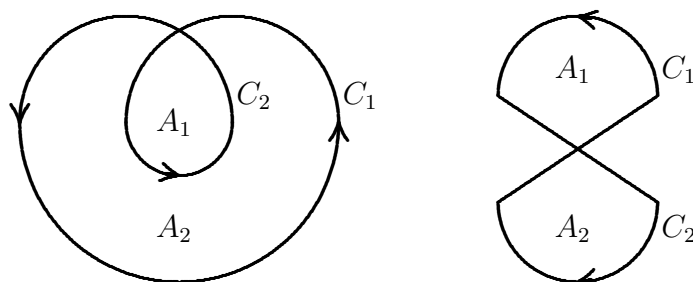
Figure 7.9: a closed but not simple curve

compatible with some parts. For example, in Figure 7.9, the left curve intersects itself once, which divides the curve into the outside part C_1 and the inside part C_2. Then

$$\int_C fdx + gdy = \left(\int_{C_1} + \int_{C_2}\right) fdx + gdy = \left(\int_{A_1 \cup A_2} + \int_{A_1}\right)(-f_x + g_y)dxdy$$
$$= \left(2\int_{A_1} + \int_{A_2}\right)(-f_x + g_y)dxdy.$$

For the curve on the right, we have

$$\int_C fdx + gdy = \left(\int_{A_1} - \int_{A_2}\right)(-f_x + g_y)dxdy$$

In general, we still have Green theorem for regions enclosed by (not necessarily simple) closed curves, as long as the regions are counted in the corresponding way. In fact, this remark is already used in the proof of Green theorem above, since the curve L in the proof may not be simple.

Example 7.3.1. The area of the region A enclosed by a simple closed curve C is

$$\int_A dxdy = \int_C xdy = -\int_C ydx.$$

For example, suppose C is the counterclockwise unit circle arc from $(1,0)$ to $(0,1)$. To compute $\int_C xdy$, we add $C_1 = [0,1] \times 0$ in the rightward direction and $C_2 = 0 \times [0,1]$ in the downward direction. Then the integral $\left(\int_C + \int_{C_1} + \int_{C_2}\right) xdy$ is the area $\dfrac{\pi}{4}$ of the quarter unit disk. Since $\int_{C_1} xdy = \int_0^1 xd0 = 0$ and $\int_{C_2} xdy = \int_1^0 0dy = 0$, we conclude that $\int_C xdy = \dfrac{\pi}{4}$.

Exercise 7.3.1. Compute the integral.

1. $\int_C (e^x + 2y - 3x)dx + (8x + \cos y^2)dy$, C is the triangle with vertices $(0,0)$, $(1,2)$, $(2,3)$, in the clockwise direction.

2. $\int_C (x^2 + y^2)dx - 2xydy$, C is the boundary of the half disk $x^2 + y^2 \le 4$, $y \ge 0$, in the counterclockwise direction.

Exercise 7.3.2. Compute the area.

1. Region enclosed by the astroid $x^{\frac{3}{2}} + y^{\frac{3}{2}} = 1$.

2. Region enclosed by the cardioid $r = 2a(1 + \cos\theta)$.

3. Region enclosed by the parabola $(x + y)^2 = x$ and the x-axis.

Exercise 7.3.3. Do Exercise 7.1.66 again by using Green theorem.

Exercise 7.3.4. Extend integration by parts by finding a relation between $\int_A (uf_x + vg_y)dxdy$ and $\int_A (u_x f + v_y g)dxdy$.

Exercise 7.3.5. The *divergence* of a vector field $F = (f, g)$ on $\mathbb{R}^2$ is

$$\mathrm{div}F = f_x + g_y.$$

Prove that

$$\int_C F \cdot \vec{n}ds = \int_A \mathrm{div}F dA,$$

where $\vec{n}$ is the normal vector pointing away from A.

Exercise 7.3.6. The Laplacian of a two variable function f is $\Delta f = f_{xx} + f_{yy} = \mathrm{div}\nabla f$. Prove Green identities

$$\int_A f\Delta g dA = \int_C f\nabla g \cdot \vec{n}ds - \int_A \nabla f \cdot \nabla g dA,$$

and

$$\int_A (f\Delta g - g\Delta f)dA = \int_C (f\nabla g - g\nabla f) \cdot \vec{n}ds.$$

7.3.2   Independence of Integral on Path

If $f_y = g_x$ on a region $A \subset \mathbb{R}^2$, then Green theorem tells us $\int_C f dx + g dy = 0$. The conclusion can be interpreted and utilized in different ways.

Suppose C_1 and C_2 are two curves connecting $\vec{a}$ to $\vec{b}$ in $\mathbb{R}^2$. Then $C_1 \cup (-C_2)$ is an oriented closed curve. If $f_y = g_x$ on the region enclosed by $C_1 \cup (-C_2)$, then we have $\int_{C_1 \cup (-C_2)} f dx + g dy = 0$, which means

$$\int_{C_1} f dx + g dy = \int_{C_2} f dx + g dy.$$

In other words, if $f_y = g_x$ on the region between two curves with the same beginning and end points, then the integral of the 1-form $f dx + g dy$ along the two curves are the same.

Another explanation of the independence of the integral of a 1-form on the choice of paths is given in Examples 7.2.8. The general argument is the following. Suppose $f = \varphi_x$ and $g = \varphi_y$ for a differentiable function
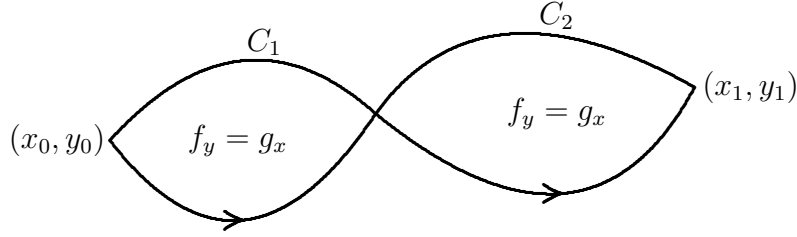
Figure 7.10: when $\displaystyle\int_{C_1} f\,dx + g\,dy = \int_{C_2} f\,dx + g\,dy$

$\varphi(x, y)$. Suppose a curve C connecting (x_0, y_0) to (x_1, y_1) is parametrized by differentiable $\phi(t) = (x(t), y(t))$. Then

$$\int_C f\,dx + g\,dy = \int_a^b (\varphi_x x' + \varphi_y y')dt = \int_a^b \frac{d\varphi(\phi(t))}{dt}dt$$
$$= \varphi(\phi(b)) - \varphi(\phi(a)) = \varphi(x_1, y_1) - \varphi(x_0, y_0).$$

The condition $f = \varphi_x$ and $g = \varphi_y$ means the 1-form $f\,dx + g\,dy$ is the differential $d\varphi$. It also means the vector field (f, g) is the gradient $\nabla\varphi$. The function φ, called a *potential* for the 1-form $f\,dx + g\,dy$ or for the vector field (f, g), is simply the antiderivative for the pair of two variable functions!

Any continuous single variable function $f(x)$ has essentially unique antiderivative

$$\varphi(x) = \varphi(x_0) + \int_{x_0}^x f(t)dt,$$

by integrating along the line (interval) $[x_0, x]$ connecting x_0 to x. However, if $\varphi_x = f$, $\varphi_y = g$, and f and g have continuous partial derivatives, then f and g must satisfy $f_y = \varphi_{yx} = \varphi_{xy} = g_x$. Therefore $f_y = g_x$ is a necessary condition for the vector field (f, g) to have antiderivative (potential). Moreover, similar to the single variable case, we expect the potential to be

$$\varphi(x, y) = \varphi(x_0, y_0) + \int_{(x_0,y_0)}^{(x,y)} f\,dx + g\,dy = \varphi(x_0, y_0) + \int_C f\,dx + g\,dy, \quad (7.3.2)$$

except there are many curves C connecting (x_0, y_0) to (x, y). Therefore the two variable antiderivative problem is closely related to the independence of the integral on the choice of paths.

Theorem 7.3.2. *Suppose f and g are continuous functions on an open subset $U \subset \mathbb{R}^2$. Then the following are equivalent.*

1. *The integral $\displaystyle\int_C f\,dx + g\,dy$ along an oriented rectifiable curve C in U depends only on the beginning and end points of C.*

2. *There is a differentiable function φ on U, such that $\varphi_x = f$ and $\varphi_y = g$.*

Moreover, if U is simply connected and f and g are continuously differentiable, then the above is also equivalent to

$$f_y = g_x.$$

The simply connected condition means that the subset has no holes. The rigorous definition is that any continuous map $S^1 \to U$ from the unit circle (i.e., a closed curve in U) extends to a continuous map $B^2 \to U$ from the unit disk. Any two curves in U with the same beginning and end points form a closed curve in U. The extension to the unit disk means that the region between the two curves still lies in U, so that the integral along the two curves are the same.

Proof. Suppose the integral depends only on the beginning and end points. Then we consider $\varphi(x, y)$ given by the formula (7.3.2), in which (x_0, y_0) is a fixed point in U and $\varphi(x_0, y_0)$ is the arbitrary constant allowed in the antiderivative. To study the differentiability of φ at $(x_1, y_1) \in U$, for any $\epsilon > 0$, we find $\delta > 0$, such that $\|\Delta(x, y)\|_2 = \|(x, y) - (x_1, y_1)\|_2 < \delta$ implies $(x, y) \in U$ and $\|(f(x, y) - f(x_1, y_1), g(x, y) - g(x_1, y_1))\|_2 < \epsilon$. Then for such (x, y), we have

$$\varphi(x, y) = \varphi(x_1, y_1) + \int_I f dx + g dy,$$

where I is the straight line connecting (x_1, y_1) to (x, y) and I still lies in U. For the linear function

$$l(x, y) = \varphi(x_1, y_1) + f(x_1, y_1)\Delta x + g(x_1, y_1)\Delta y,$$

we have

$$
\begin{aligned}
|\varphi(x, y) - l(x, y)| &= \left| \int_I (f(x, y) - f(x_1, y_1))dx + (g(x, y) - g(x_1, y_1))dy \right| \\
&\leq \sup_{(x,y) \in I} \|(f(x, y) - f(x_1, y_1), g(x, y) - g(x_1, y_1))\|_2 \mu(I) \\
&\leq \epsilon \|\Delta(x, y)\|_2.
\end{aligned}
$$

This shows that l is a linear approximation of φ. Moreover, we have $\varphi_x(x_1, y_1) = f(x_1, y_1)$ and $\varphi_y(x_1, y_1) = g(x_1, y_1)$.

Conversely, suppose φ is differentiable and $\varphi_x = f$, $\varphi_y = g$. By the earlier discussion, we have

$$\int_C f dx + g dy = \varphi(x_1, y_1) - \varphi(x_0, y_0) \tag{7.3.3}$$

for the special case that C is a differentiably parametrized curve connecting (x_0, y_0) to (x_1, y_1). Now consider a general rectifiable curve C parametrized by $\phi\colon [a, b] \to \mathbb{R}^2$ connecting (x_0, y_0) to (x_1, y_1). Let P be a partition of $[a, b]$. Let L_i be the straight line connecting $\phi(t_{i-1})$ to $\phi(t_i)$. Let L be obtained by combining L_i together. Since L_i can be differentiably parametrized, we have

$$
\begin{aligned}
\int_L f dx + g dy &= \sum \int_{L_i} f dx + g dy \\
&= \sum (\varphi(\phi(t_i)) - \varphi(\phi(t_{i-1}))) = \varphi(x_1, y_1) - \varphi(x_0, y_0).
\end{aligned}
$$

On the other hand, in the proof of Theorem 7.3.1, we have shown that $\int_C f\,dx + g\,dy = \lim_{\|P\|\to 0} \int_L f\,dx + g\,dy$. Therefore we conclude that (7.3.3) also holds for the general C.

Finally, we know that in case f and g are continuously differentiable, the second statement implies $f_y = g_x$. Conversely, we also explained that by Green theorem, the condition $f_y = g_x$ on a simply connected open subset implies the first statement.

$\square$

Example 7.3.2. The integral of $y\,dx + x\,dy$ in Example 7.2.8 is independent of the choice of the paths because the equality $\dfrac{\partial y}{\partial y} = \dfrac{\partial x}{\partial x}$ holds on the whole plane, which is simply connected. The potential of the 1-form is xy, up to adding constants.

In Example 7.2.9, the integrals of $xy\,dx + (x+y)\,dy$ along the three curves are different. Indeed, we have $(xy)_y = x \neq (x+y)_x = 1$, and the 1-form has no potential.

Example 7.3.3. The vector field $\dfrac{1}{y^2}(xy^2 + y, 2y - x)$ is defined on $y > 0$ and $y < 0$, both simply connected. It satisfies $\dfrac{d}{dy}\left(\dfrac{xy^2 + y}{y^2}\right) = -\dfrac{1}{y^2} = \dfrac{d}{dx}\left(\dfrac{2y - x}{y^2}\right)$. Therefore it has a potential function φ. By $\varphi_x = \dfrac{y(xy+1)}{y^2}$, we get $\varphi = \displaystyle\int \dfrac{y(xy+1)}{y^2}\,dx + \vartheta(y) = \dfrac{x^2}{2} + \dfrac{x}{y} + \vartheta(y)$. Then by $\varphi_y = -\dfrac{x}{y^2} + \vartheta'(y) = \dfrac{2y - x}{y^2}$, we get $\vartheta'(y) = \dfrac{2}{y}$, so that $\vartheta(y) = 2\log|y| + c$. The potential function is

$$\varphi = \frac{x^2}{2} + \frac{x}{y} + 2\log|y| + c.$$

In particular, we have

$$\int_{(1,1)}^{(2,2)} \frac{(xy^2 + y)dx + (2y - x)dy}{y^2} = \left(\frac{x^2}{2} + \frac{x}{y} + 2\log|y|\right)\Big|_{(1,1)}^{(2,2)} = \frac{3}{2} + \log 2.$$

The example shows that the formula (7.3.2) may not be the most practical way of computing the potential.

Example 7.3.4. The 1-form $\dfrac{y\,dx - x\,dy}{x^2 + y^2}$ satisfies

$$f_y = \left(\frac{y}{x^2 + y^2}\right)_y = \frac{x^2 - y^2}{(x^2 + y^2)^2} = g_x = \left(\frac{-x}{x^2 + y^2}\right)_x$$

on $\mathbb{R}^2 - (0,0)$, which is unfortunately not simply connected. Let U be obtained by removing the non-positive x-axis $(-\infty, 0] \times 0$ from $\mathbb{R}^2$. Then U is simply connected, and Theorem 7.3.2 can be applied. The potential for $\dfrac{y\,dx - x\,dy}{x^2 + y^2}$ is $\varphi = -\theta$, where $-\pi < \theta < \pi$ is the angle in the polar coordinate. The potential can be used to compute

$$\int_{(1,0)}^{(0,1)} \frac{y\,dx - x\,dy}{x^2 + y^2} = -\theta(0,1) + \theta(1,0) = -\frac{\pi}{2} + 0 = -\frac{\pi}{2}$$

for any curve connecting $(1,0)$ to $(0,1)$ that does not intersect the non-positive x-axis. The unit circle arc $C_1\colon \phi_1(t) = (\cos t, \sin t)$, $0 \le t \le \dfrac{\pi}{2}$, and the straight line $C_2\colon \phi_2(t) = (1 - t, t)$, $0 \le t \le 1$, are such curves.

On the other hand, consider the unit circle arc $C_3\colon \phi_3(t) = (\cos t, -\sin t)$, $0 \le t \le \dfrac{3\pi}{2}$, connecting $(1,0)$ to $(0,1)$ in the clockwise direction. To compute the integral along the curve, let V be obtained by removing the non-negative diagonal $\{(x,x)\colon x \ge 0\}$ from $\mathbb{R}^2$. The 1-form still has the potential $\varphi = -\theta$ on V. The crucial difference is that the range for θ is changed to $\dfrac{\pi}{4} < \theta < \dfrac{9\pi}{4}$. Therefore

$$\int_{(1,0)}^{(0,1)} \frac{ydx - xdy}{x^2 + y^2} = -\theta(0,1) + \theta(1,0) = -\frac{\pi}{2} + 2\pi = \frac{3\pi}{2}$$

for any curve connecting $(1,0)$ to $(0,1)$ that does not intersect the non-negative diagonal.

In general, the integral $\displaystyle\int_{(1,0)}^{(0,1)} \frac{ydx - xdy}{x^2 + y^2}$ depends only on how the curve connecting $(1,0)$ to $(0,1)$ goes around the origin, the only place where $f_y = g_x$ fails.

Exercise 7.3.7. Explain the computations in Exercise 7.2.19 by Green theorem.

Exercise 7.3.8. Use potential function to compute the integral.

1. $\displaystyle\int_C 2xdx + ydy$, C is the straight line connecting $(2,0)$ to $(0,2)$.

2. $\displaystyle\int_C \frac{xdx + ydy}{x^2 + y^2}$, C is the circular arc connecting $(2,0)$ to $(0,2)$.

3. $\displaystyle\int_C \frac{ydx - xdy}{ax^2 + by^2}$, C is the unit circle in counterclockwise direction.

4. $\displaystyle\int_C \frac{(x - y)dx + (x + y)dy}{x^2 + y^2}$, C is the elliptic arc $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$ in the upper half plane connecting $(a,0)$ to $(-a,0)$.

5. $\displaystyle\int_C (e^x \sin 2y - y)dx + (2e^x \cos 2y - 1)dy$, C is the circular arc connecting $(0,1)$ to $(1,0)$ in clockwise direction.

6. $\displaystyle\int_C (2xy^3 - 3y^2 \cos x)dx + (-6y \sin x + 3x^2y^2)dy$, C is the curve $2x = \pi y^2$ connecting $(0,0)$ to $\left(\dfrac{\pi}{2}, 1\right)$.

Exercise 7.3.9. Find α so that the vector fields $\dfrac{(x,y)}{(x^2 + y^2)^\alpha}$ and $\dfrac{(-y,x)}{(x^2 + y^2)^\alpha}$ have potentials. Then find the potentials.

Exercise 7.3.10. Compute the integral $\displaystyle\int_C g(xy)(ydx + xdy)$, where C is the straight line connecting $(2,3)$ to $(1,6)$.

Exercise 7.3.11. Why is the potential function unique up to adding constants?
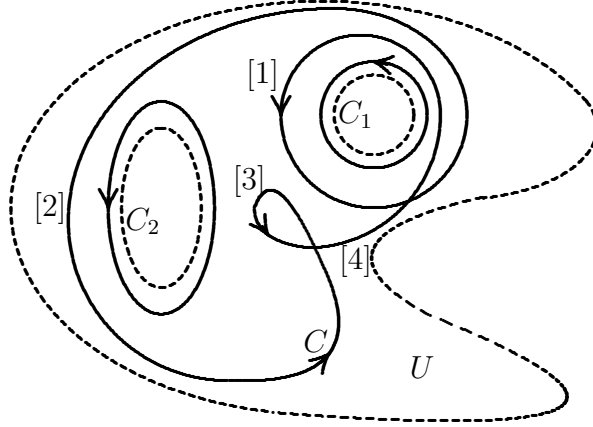
Figure 7.11: closed curve in open subset with holes

Now we discuss another application of Green theorem in case $f_y = g_x$. The open subset in Figure 7.11 has two holes. The two holes are enclosed by simple closed curves C_1 and C_2 in U oriented in counterclockwise direction. The closed curve C in U may be divided into four parts, denoted $C_{[1]}$, $C_{[2]}$, $C_{[3]}$, $C_{[4]}$. The union of oriented closed curves $C_{[1]} \cup (-C_1)$, $C_{[2]} \cup C_{[4]} \cup (-C_1) \cup (-C_2)$ and $C_{[3]}$ also enclose subsets contained in U. If $f_y = g_y$ on U, then by Green theorem,

$$\int_{C_{[1]} \cup (-C_1)} f dx + g dy = \int_{C_{[2]} \cup C_{[4]} \cup (-C_1) \cup (-C_2)} f dx + g dy = \int_{C_{[3]}} f dx + g dy = 0.$$

Thus

$$\int_C f dx + g dy = \left(\int_{C_1} + \int_{C_1 \cup C_2} \right) f dx + g dy = \left(2 \int_{C_1} + \int_{C_2} \right) f dx + g dy.$$

Note that the coefficients 2 for C_1 and 1 for C_2 mean that C wraps around C_1 twice and C_2 once.

In general, suppose U has finitely many holes. We enclose these holes with closed curves C_1, C_2, ..., C_k, all in counterclockwise orientation. Then any closed curve C in U wraps around the i-th hole n_i times. The sign of n_i is positive when the wrapping is counterclockwise and is negative if the wrapping is clockwise. Then we say C is *homologous* to $n_1 C_1 + n_2 C_2 + \cdots + n_k C_k$. If $f_y = g_x$ on U, then we have

$$\int_C f dx + g dy = \left(n_1 \int_{C_1} + n_2 \int_{C_2} + \cdots + n_k \int_{C_k} \right) f dx + g dy.$$

Finally, suppose C and D are two curves in U with the same end points. Because U is no longer simply connected, the equality no longer implies that the integrals $\int_C f dx + g dy$ and $\int_D f dx + g dy$ are equal. However, the difference between the two integrals can be computed by considering the homology of the closed curve $C \cup (-D)$.

Example 7.3.5. In Example 7.3.4, the 1-form $\dfrac{ydx - xdy}{x^2 + y^2}$ satisfies $f_y = g_x$ on $U = \mathbb{R}^2 - (0,0)$. The unit circle C in the counterclockwise direction encloses the only hole of U. We have

$$\int_C \frac{ydx - xdy}{x^2 + y^2} = \int_0^{2\pi} \frac{\sin t(-\sin t)dt - \cos t \cos t dt}{\cos^2 t + \sin^2 t} = -2\pi.$$

If C_1 is a curve on the first quadrangle connecting $(1,0)$ to $(0,1)$ and C_2 is a curve on the second, third and fourth quadrangles connecting the two points, then $C_1 \cup (-C_2)$ is homologous to C, and we have

$$\int_{C_1} \frac{ydx - xdy}{x^2 + y^2} = \int_{C_2} \frac{ydx - xdy}{x^2 + y^2} - 2\pi.$$

Exercise 7.3.12. Study how the integral of the 1-form $\dfrac{ydx - xdy}{x^2 + xy + y^2}$ depends on the curves.

Exercise 7.3.13. Study how the integral of the 1-form

$$\omega = \frac{(x^2 - y^2 - 1)dx + 2xydy}{((x-1)^2 + y^2)((x+1)^2 + y^2)}$$

depends on the curves. Note that if C_ϵ is the counterclockwise circle of radius ϵ around $(1,0)$, then $\displaystyle\int_{C_{\epsilon_0}} \omega = \lim_{\epsilon \to 0} \int_{C_\epsilon} \omega$.

7.3.3   Stokes Theorem

Stokes theorem is the extension of Green theorem to oriented surfaces $S \subset \mathbb{R}^3$ with compatibly oriented boundary curves C. Suppose the surface is given by an orientation compatible parametrization

$$\sigma(u,v) = (x(u,v), y(u,v), z(u,v)) \colon A \subset \mathbb{R}^2 \to S \subset \mathbb{R}^3,$$

and the boundary C of S corresponds to the boundary D of A. Recall that D is oriented in such a way that A is "on the left" of D. Correspondingly, C is oriented in such a way that S is "on the left" of C. Also recall the description in case D is differentiable, the clockwise rotation of the tangent vector of D by 90 degrees will be the normal direction pointing away from A. Correspondingly, in the tangent plane of S, the clockwise rotation of the tangent vector of C by 90 degrees will be the normal direction pointing away from S.

Suppose $F = (f, g, h)$ is a continuously differentiable vector field on $\mathbb{R}^3$. Then by Green theorem,

$$\begin{aligned}
\int_C F \cdot d\vec{x} &= \int_C fdx + gdy + hdz \\
&= \int_D (fx_u + gy_u + hz_u)du + (fx_v + gy_v + hz_v)dv \\
&= \int_A ((fx_v + gy_v + hz_v)_u - (fx_u + gy_u + hz_u)_v)dudv \\
&= \int_A (f_u x_v - f_v x_u + g_u y_v - g_v y_u + h_u z_v - h_v z_u)dudv.
\end{aligned}$$

By

$$f_u x_v - f_v x_u = (f_x x_u + f_y y_u + f_z z_u)x_v - (f_x x_v + f_y y_v + f_z z_v)x_u$$
$$= -f_y \det \frac{\partial(x, y)}{\partial(u, v)} + f_z \det \frac{\partial(z, x)}{\partial(u, v)},$$

we have

$$\int_A (f_u x_v - f_v x_u)dudv = \int_A -f_y \det \frac{\partial(x, y)}{\partial(u, v)}dudv + f_z \det \frac{\partial(z, x)}{\partial(u, v)}dudv$$
$$= \int_S -f_y dx \wedge dy + f_z dz \wedge dx.$$

Combined with the similar computations for g and h, we get the following result.

Theorem 7.3.3 (Stokes Theorem). *Suppose $S \subset \mathbb{R}^3$ is an oriented surface with compatibly oriented simple closed boundary curves C. Then for any continuously differentiable $F = (f, g, h)$, we have*

$$\int_C f dx + g dy + h dz = \int_S (g_x - f_y)dx \wedge dy + (h_y - g_z)dy \wedge dz + (f_z - h_x)dz \wedge dx.$$

Although the argument was made only for one parametrization, which may not cover the whole surface, the equality can be extended to oriented surfaces by adding the equalities on compatibly oriented parametrized pieces.

Introduce the symbol

$$\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right) \tag{7.3.4}$$

on $\mathbb{R}^3$. For any function f, we formally have the gradient

$$\text{grad} f = \nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right). \tag{7.3.5}$$

For any vector field $F = (f, g, h)$ on $\mathbb{R}^3$, we formally have the *curl*

$$\text{curl} F = \nabla \times F = \left(\frac{\partial h}{\partial y} - \frac{\partial g}{\partial z}, \frac{\partial f}{\partial z} - \frac{\partial h}{\partial z}, \frac{\partial g}{\partial x} - \frac{\partial f}{\partial y} \right).$$

Then Stokes theorem can be written as

$$\int_C F \cdot d\vec{x} = \int_S \text{curl} F \cdot \vec{n} dA.$$

Example 7.3.6. Suppose C is the circle given by $x^2 + y^2 + z^2 = 1$, $x + y + z = r$, with the counterclockwise orientation when viewed from the direction of the x-axis. We would like to compute the integral $\int_C (ax + by + cz + d)dx$. Let S be the disk $x^2 + y^2 + z^2 \leq 1$, $x + y + z = r$, with the normal direction given by $(1, 1, 1)$. Then by Stokes theorem and the fact that the radius of S is $\sqrt{1 - \frac{r^2}{3}}$, we have

$$\int_C (ax + by + cz + d)dx = \int_S -b dx \wedge dy + c dz \wedge dx = \int_S \frac{1}{\sqrt{3}}(-b + c)dA = \frac{c - b}{\sqrt{3}}\pi \left(1 - \frac{r^2}{3} \right).$$

Example 7.3.7. Faraday observed that a changing magnetic field B induces an electric field E. More precisely, *Faraday's induction law* says that the rate of change of the flux of the magnetic field through a surface S is the negative of the integral of the electric field along the boundary C of the surface S. The law is summarized in the formula

$$-\int_C E \cdot d\vec{x} = \frac{d}{dt}\int_S B \cdot \vec{n}dA. \tag{7.3.6}$$

By Stokes theorem, the left side is $-\int_S \text{curl}E \cdot \vec{n}dA$. Since the equality holds for any surface S, we conclude the differential version of Faraday's law

$$-\text{curl}E = \frac{\partial B}{\partial t}. \tag{7.3.7}$$

This is one of Maxwell's equations for electromagnetic fields.

Exercise 7.3.14. Compute $\int_C y^2 dx + (x+y)dy + yz dz$, where C is the ellipse $x^2 + y^2 = 2$, $x + y + z = 2$, with clockwise orientation as viewed from the origin.

Exercise 7.3.15. Suppose C is any closed curve on the sphere $x^2 + y^2 + z^2 = R^2$. Prove that $\int_C (y^2 + z^2)dx + (z^2 + x^2)dy + (x^2 + y^2)dz = 0$. In general, what is the condition for f, g, h so that $\int_C f dx + g dy + h dz = 0$ for any closed curve C on any sphere centered at the origin?

Exercise 7.3.16. Find the formulae for the curls of $F + G$, gF.

Exercise 7.3.17. Prove that $\text{curl}(\text{grad}f) = \vec{0}$. Moreover, compute $\text{curl}(\text{curl}F)$.

Exercise 7.3.18. The electric field E induced by a changing magnetic field is also changing and follows *Ampere's law*

$$\int_C B \cdot d\vec{x} = \mu_0 \int_S J \cdot \vec{n}dA + \epsilon_0\mu_0\frac{d}{dt}\int_S E \cdot \vec{n}dA, \tag{7.3.8}$$

where J is the current density, and μ_0, ϵ_0 are some physical constants. Derive the differential version of Ampere's law, which is the other Maxwell's equation for electromagnetic fields.

Green theorem can be further extended to surfaces in Euclidean spaces of dimension higher than 3. Suppose S is an oriented surface given by an orientation compatible regular parametrization $\sigma(u,v)\colon A \subset \mathbb{R}^2 \to \mathbb{R}^n$. The boundary C of S corresponds to the boundary D of A. The compatible orientation of C is described as follows. At any $\vec{x} = \sigma(u,v) \in C$, the tangent plane $T_{\vec{x}}S$ of the surface is spanned by the tangent vectors σ_u and σ_v. On the tangent plane is the normal vector $\vec{n} \in T_{\vec{x}}S$ pointing away from the surface S. Then the tangent vector of C is a unit length vector $\vec{t} \in T_{\vec{x}}S$ orthogonal to $\vec{n}$. Among the two choices of $\vec{t}$ (one is the negative of the other), the one compatible with the orientation of S is such that the rotation from σ_u to σ_v is in the same direction of the rotation from $\vec{n}$ to $\vec{t}$. The condition can be rephrased as follows. Since both $\{\sigma_u, \sigma_v\}$ and $\{\vec{n}, \vec{t}\}$ are bases of the 2-dimensional subspace $T_{\vec{x}}S$, we have

$$\sigma_u \wedge \sigma_v = \lambda\vec{n} \wedge \vec{t},$$

for some number $\lambda \neq 0$. The tangent vector $\vec{t}$ is the only unit length vector orthogonal to $\vec{n}$, such that $\lambda > 0$ in the equality above.

Now for a continuously differentiable f, by Green theorem, we have

$$
\begin{aligned}
\int_C f dx_j &= \int_D f \left(\frac{\partial x_j}{\partial u} du + \frac{\partial x_j}{\partial v} dv \right) \\
&= \int_A \left(\frac{\partial}{\partial v} \left(f \frac{\partial x_j}{\partial v} \right) - \frac{\partial}{\partial u} \left(f \frac{\partial x_j}{\partial u} \right) \right) du dv \\
&= \int_A \left(\frac{\partial f}{\partial v} \frac{\partial x_j}{\partial v} - \frac{\partial f}{\partial u} \frac{\partial x_j}{\partial u} \right) du dv \\
&= \int_A \sum_i \left(\frac{\partial f}{\partial x_i} \frac{\partial x_i}{\partial u} \frac{\partial x_j}{\partial v} - \frac{\partial f}{\partial x_i} \frac{\partial x_i}{\partial v} \frac{\partial x_j}{\partial u} \right) du dv \\
&= \int_A \sum_{i \neq j} \frac{\partial f}{\partial x_i} \det \frac{\partial(x_i, x_j)}{\partial(u, v)} du dv = \int_S \sum_{i \neq j} \frac{\partial f}{\partial x_i} dx_i \wedge dx_j.
\end{aligned}
$$

Thus for a continuously differentiable vector field $F = (f_1, f_2, \ldots, f_n)$ on $\mathbb{R}^n$, we have

$$
\begin{aligned}
\int_C F \cdot d\vec{x} = \int_C \sum_j f_j dx_j &= \int_S \sum_{i \neq j} \frac{\partial f_j}{\partial x_i} dx_i \wedge dx_j \\
&= \int_S \sum_{i < j} \left(\frac{\partial f_j}{\partial x_i} - \frac{\partial f_i}{\partial x_j} \right) dx_i \wedge dx_j.
\end{aligned}
\tag{7.3.9}
$$

This is Stokes theorem for oriented surfaces in $\mathbb{R}^n$.

Similar to Green theorem, Stokes theorem has implication on how the integral of a 1-form depends on the choice of the curve. Suppose a vector field F satisfies

$$
\frac{\partial f_j}{\partial x_i} = \frac{\partial f_i}{\partial x_j}
\tag{7.3.10}
$$

on an open subset $U \subset \mathbb{R}^n$. A collection C of oriented closed curves is *homologous to* 0 in U if it is the compatibly oriented boundary of an oriented surface S in U (in fact, a continuously differentiable map from S to U is sufficient). Stokes theorem tells us that if F satisfies (7.3.10) on U and C is homologous to 0 in U, then $\int_C F \cdot d\vec{x} = 0$. Moreover, two oriented curves C and D that have the same begining and end points are *homologous* in U if the closed curve $C \cap (-D)$ is homologous to 0. Then the equalities (7.3.10) imply that the integral $\int_C F \cdot d\vec{x}$ depends only on the homologous class of C.

A special case is when the subset U is *simply connected*, which means that any continuous map $S^1 \to U$ extends to a continuous map $B^2 \to U$. Then $C \cap (-D)$ is the boundary of a disk in U. With the same proof, Theorem 7.3.2 may be extended.

Theorem 7.3.4. *Suppose $F = (f_1, f_2, \ldots, f_n)$ is a continuous vector field on an open subset $U \subset \mathbb{R}^n$. Then the following are equivalent.*

1. *The integral $\int_C F \cdot d\vec{x}$ along an oriented rectifiable curve C in U depends only on the beginning and end points of C.*

2. *There is a differentiable function φ on U, such that $\nabla \varphi = F$.*

Moreover, if U is simply connected and F is continuously differentiable, then the above is also equivalent to

$$\frac{\partial f_j}{\partial x_i} = \frac{\partial f_i}{\partial x_j}$$

for any i, j.

In $\mathbb{R}^3$, the theorem says that $F = \text{grad}\varphi$ for some φ on a simply connected region if and only if $\text{curl}F = \vec{0}$.

The *potential function* φ is given by

$$\varphi(\vec{x}) = \varphi(\vec{x}_0) + \int_{\vec{x}_0}^{\vec{x}} F \cdot d\vec{x}$$

and also satisfies $d\varphi = F \cdot d\vec{x}$.

Example 7.3.8. The vector field $(f, g, h) = (yz(2x + y + z), zx(x + 2y + z), xy(x + y + 2z))$ satisfies

$$f_y = g_x = (2x + 2y + z)z, \quad h_y = g_z = (x + 2y + 2z)x, \quad f_z = h_x = (2x + y + 2z)y.$$

Therefore the vector field has potential. The potential function can be computed by integrating along successive straight lines connecting $(0, 0, 0)$, $(x, 0, 0)$, $(x, y, 0)$, (x, y, z). The integral is zero on the first two segments, so that

$$\varphi(x, y, z) = \int_0^z xy(x + y + 2z)dz = xyz(x + y + z).$$

Exercise 7.3.19. Determine whether the integral of vector field or the 1-form is independent of the choice of the curves. In the independent case, find the potential function.

1. $e^x(\cos yz, -z \sin yz, -y \sin yz)$.

2. $y^2 z^3 dx + 2xyz^3 dy + 2xyz^2 dz$.

3. $(y + z)dx + (z + x)dy + (x + y)dz$.

4. $(x_2, x_3, \ldots, x_n, x_1)$.

5. $(x_1^2, x_2^2, \ldots, x_n^2)$.

6. $x_1 x_2 \cdots x_n \left(\dfrac{dx_1}{x_1} + \dfrac{dx_2}{x_2} + \cdots + \dfrac{dx_n}{x_n} \right)$.

Exercise 7.3.20. Suppose $\vec{a}$ is a nonzero vector. Find condition on a function $f(\vec{x})$ so that the vector field $f(\vec{x})\vec{a}$ has a potential function.

Exercise 7.3.21. Find condition on a function $f(\vec{x})$ so that the vector field $f(\vec{x})\vec{x}$ has a potential function.

Exercise 7.3.22. Study the potential function of $\dfrac{(y-z)dx + (z-x)dy + (x-y)dz}{(x-y)^2 + (y-z)^2}$.

Exercise 7.3.23. Suppose a continuously second order differentiable function $g(\vec{x})$ satisfies $\nabla g(\vec{x}_0) \neq \vec{0}$. Suppose $f(\vec{x})$ is continuously differentiable near $\vec{x}_0$. Prove that the differential form

$$f \, dg = f g_{x_1} dx_1 + f g_{x_2} dx_2 + \cdots + f g_{x_n} dx_n$$

has a potential near $\vec{x}_0$ if and only if $f(\vec{x}) = h(g(\vec{x}))$ for a continuously differentiable $h(t)$.

7.3.4   Gauss Theorem

Green theorem can be extended to regions of higher dimension. For example, let

$$A = \{(x, y, z) \colon (x, y) \in B, k(x, y) \leq z \leq h(x, y)\}$$

be the 3-dimensional region between the graphs of functions $h(x, y)$ and $k(x, y)$ on $B \subset \mathbb{R}^2$ (see Figure 7.12). Then we have

$$\int_A f_z dx dy dz = \int_B \left(\int_{k(x,y)}^{h(x,y)} f_z(x, y, z) dz \right) dx dy$$

$$= \int_B (f(x, y, h(x, y)) - f(x, y, k(x, y))) dx dy.$$

On the other hand, the boundary S of A consists of three pieces. By taking the normal vector $\vec{n}$ of the boundary surface to point outward of A, the orientation compatible parametrizations of the three pieces are

$$
\begin{aligned}
&S_b \colon \sigma(y, x) = (x, y, k(x, y)), &&(x, y) \in B \\
&S_t \colon \sigma(x, y) = (x, y, h(x, y)), &&(x, y) \in B \\
&S_v \colon \sigma(t, z) = (x(t), y(t), z), &&k(x, y) \leq z \leq h(x, y)
\end{aligned}
$$

where $(x(t), y(t))$ is a parametrization of the boundary curve of B. We have

$$\int_S f dx \wedge dy = \left(\int_{S_b} + \int_{S_t} + \int_{S_v} \right) f dx \wedge dy$$

$$= \int_B f(x, y, k(x, y)) \det \frac{\partial(x, y)}{\partial(y, x)} dx dy$$

$$+ \int_B f(x, y, h(x, y)) \det \frac{\partial(x, y)}{\partial(x, y)} dx dy$$

$$+ \int_{a \leq t \leq b, k(x,y) \leq z \leq h(x,y)} f(x(t), y(t), z) \det \frac{\partial(x, y)}{\partial(t, z)} dt dz$$

$$= - \int_B f(x, y, k(x, y)) dx dy + \int_B f(x, y, h(x, y)) dx dy + 0.$$

Thus we proved $\displaystyle\int_S f dx \wedge dy = \int_A f_z dx dy dz$. The equality can be extended to regions that can be divided by vertical planes into regions between graphs
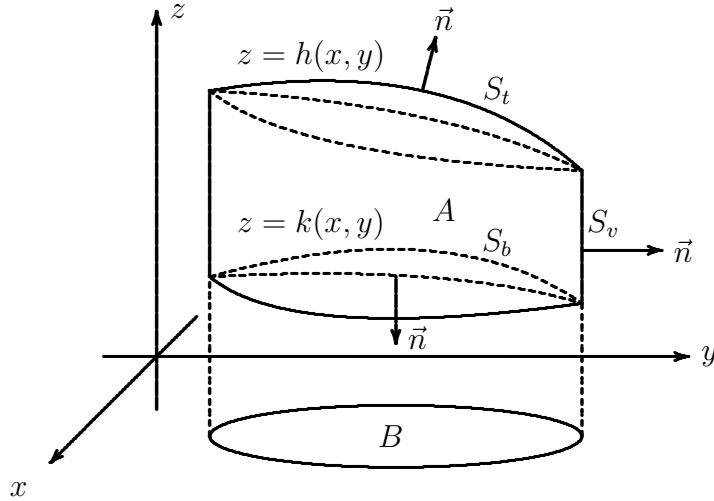
Figure 7.12: Gauss Theorem for a special case

of continuously differentiable functions. Similar equalities can be established by rotating x, y and z.

Similar to the 2-dimensional case, we may consider a subset $A \subset \mathbb{R}^3$ with surfaces S_1, S_2, ..., S_k as the boundary. We further assume that A is contained in only one side of each S_i. Let $\vec{n}$ be the normal vector of the surfaces pointing outward of A. This induces orientations on the surfaces. We denote by S the union of the oriented boundary surfaces and say A is a region with compatibly oriented boundary surfaces S.

Theorem 7.3.5 (Gauss Theorem). *Suppose $A \subset \mathbb{R}^3$ a region with boundary surfaces S compatibly oriented with respect to the outward normal vector $\vec{n}$. Then for any continuously differentiable $F = (f, g, h)$, we have*

$$\int_S f dy \wedge dz + g dz \wedge dx + h dx \wedge dy = \int_A (f_x + g_y + h_z) dx dy dz.$$

Define the *divergence*

$$\mathrm{div} F = \nabla \cdot F = \frac{\partial f}{\partial x} + \frac{\partial g}{\partial y} + \frac{\partial h}{\partial z}.$$

The Gauss theorem means that the outward flux of a flow $F = (f, g, h)$ is equal to the integral of the divergence of the flow on the solid

$$\int_S F \cdot \vec{n} dA = \int_A \mathrm{div} F dV.$$

Example 7.3.9. The volume of the region enclosed by a surface $S \subset \mathbb{R}^3$ without boundary is

$$\int_S x dy \wedge dz = \int_S y dz \wedge dx = \int_S z dx \wedge dy = \frac{1}{3} \int_S (x, y, z) \cdot \vec{n} dA.$$

Example 7.3.10. In Example 7.2.22, the outgoing flux of the flow $F = (x^2, y^2, z^2)$ through the ellipse $\dfrac{(x - x_0)^2}{a^2} + \dfrac{(y - y_0)^2}{b^2} + \dfrac{(z - z_0)^2}{c^2} = 1$ is computed by surface integral. Alternatively, by Gauss theorem, the flux is

$$\int F \cdot \vec{n} dA = \int_{\frac{(x-x_0)^2}{a^2} + \frac{(y-y_0)^2}{b^2} + \frac{(z-z_0)^2}{c^2} \leq 1} 2(x + y + z) dx dy dz$$

$$= \int_{\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} \leq 1} 2(x + x_0 + y + y_0 + z + z_0) dx dy dz.$$

By the transform $\vec{x} \to -\vec{x}$, we get

$$\int_{\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} \leq 1} (x + y + z) dx dy dz = 0.$$

Therefore the flux is

$$\int_{\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} \leq 1} 2(x_0 + y_0 + z_0) dx dy dz = \frac{8\pi R^3}{3} abc(x_0 + y_0 + z_0).$$

Example 7.3.11. To compute the upward flux of $F = (xz, -yz, (x^2 + y^2)z)$ through the surface S given by $0 \leq z = 4 - x^2 - y^2$, we introduce the disk $D = \{(x, y, 0) \colon x^2 + y^2 \leq 4\}$ on the (x, y)-plane. Taking the normal direction of D to be $(0, 0, 1)$, the surface $S \cup (-D)$ is the boundary of the region A given by $0 \leq z \leq 4 - x^2 - y^2$. By Gauss theorem, the flux through S is

$$\int_S F \cdot \vec{n} dA = \int_{S \cup (-D)} F \cdot \vec{n} dA + \int_D F \cdot \vec{n} dA$$

$$= \int_A (z - z + x^2 + y^2) dx dy dz + \int_{x^2 + y^2 \leq 4} F(x, y, 0) \cdot (0, 0, 1) dA$$

$$= \int_A (x^2 + y^2) dx dy dz = \int_{0 \leq r \leq 2, 0 \leq \theta \leq 2\pi} r^2 (4 - r^2) r dr d\theta = \frac{32\pi}{3}.$$

Example 7.3.12. The gravitational field created by a mass M at point $\vec{x}_0 \in \mathbb{R}$ is

$$G = -\frac{M}{\|\vec{x} - \vec{x}_0\|_2^3} (\vec{x} - \vec{x}_0).$$

A straightforward computation shows $\text{div} G = 0$. Suppose A is a region with compatibly oriented surfaces S as the boundary and $\vec{x}_0 \notin S$. If $\vec{x}_0 \notin A$, then by Gauss theorem, the outward flux $\int_S G \cdot \vec{n} dA = 0$. If $\vec{x}_0 \in A$, then let B_ϵ be the ball of radius ϵ centered at $\vec{x}_0$. The boundary of the ball is the sphere S_ϵ, which we given an orientation compatible with the outward normal vector $\vec{n} = \dfrac{\vec{x} - \vec{x}_0}{\|\vec{x} - \vec{x}_0\|_2}$. For sufficiently small ϵ, the ball is contained in A. Moreover, $A - B_\epsilon$ is a region not containing $\vec{x}_0$ and has compatibly oriented surfaces $S \cup (-S_\epsilon)$ as the boundary. Therefore

$$\int_S G \cdot \vec{n} dA = \int_{S_\epsilon} G \cdot \vec{n} dA = \int_{\|\vec{x} - \vec{x}_0\|_2 = \epsilon} -\frac{M}{\epsilon^3} (\vec{x} - \vec{x}_0) \cdot \frac{\vec{x} - \vec{x}_0}{\epsilon} dA = -\frac{M}{\epsilon^2} \int_{S_\epsilon} dA = -4\pi M.$$

More generally, the gravitational field created by several masses at various locations is the sum of the individual gravitational field. The outward flux of the field through S is then -4π multiplied to the total mass contained in A. In particular, the flux does not depend on the location of the mass, but whether the mass is contained in A or not. This is called *Gauss' Law.*

Exercise 7.3.24. Compute the flux.

1. Outward flux of (x^3, x^2y, x^2z) through boundary of the solid $x^2 + y^2 \leq a^2$, $0 \leq z \leq b$.

2. Inward flux of (xy^2, yz^2, zx^2) through the ellipse $\dfrac{(x - x_0)^2}{a^2} + \dfrac{(y - y_0)^2}{b^2} + \dfrac{(z - z_0)^2}{c^2} = 1$.

3. Upward flux of (x^3, y^3, z^3) through the surface $z = x^2 + y^2 \leq 1$.

4. Outward flux of $(x^2, y^2, -2(x + y)z)$ through the torus in Example 6.1.10.

Exercise 7.3.25. Suppose $A \subset \mathbb{R}^3$ is a convex region with boundary surface S. Suppose $\vec{a}$ is a vector in the interior of A, and p is the distance from $\vec{a}$ to the tangent plane of S. Compute $\displaystyle\int_S p \, dA$. Moreover, for the special case S is the ellipse $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c^2} = 1$ and $\vec{a} = (0, 0, 0)$, compute $\displaystyle\int_S \dfrac{1}{p} dA$.

Exercise 7.3.26. Find the formulae for the divergences of $F + G$, gF, $F \times G$.

Exercise 7.3.27. Prove that $\mathrm{div}(\mathrm{curl} F) = 0$. Moreover, compute $\mathrm{div}(\mathrm{grad} f)$ and $\mathrm{grad}(\mathrm{div} F)$.

Gauss theorem can be further extended to higher dimension. Suppose $A \subset \mathbb{R}^n$ is the n-dimensional region between the graphs $x_n = h(x_1, x_2, \ldots, x_{n-1})$ and $x_n = k(x_1, x_2, \ldots, x_{n-1})$ for $(x_1, x_2, \ldots, x_{n-1}) \in B$. Then we have

$$\int_A f_{x_n} dx_1 dx_2 \cdots dx_n = \int_B (f(\vec{u}, h(\vec{u})) - f(\vec{u}, k(\vec{u}))) du_1 du_2 \cdots du_{n-1}.$$

Suppose the boundary S of A is oriented to be compatible with the normal vector pointing away from A. Then the normal vector for the top boundary S_t points in the direction of x_n and the normal vector for the bottom boundary S_b points opposite to the direction of x_n. By the discussion at the end of Section 7.2.7, the parametrization $\sigma(\vec{u}) = (\vec{u}, h(\vec{u}))$ for S_t is compatibly oriented if n is odd and is oppositely oriented if n is even. Consequently, we get

$$\int_B f(\vec{u}, h(\vec{u})) du_1 du_2 \cdots du_{n-1}$$
$$= \int_B f(\sigma(\vec{u})) \frac{\partial(x_1, x_2, \ldots, x_{n-1})}{\partial(u_1, u_2, \ldots, u_{n-1})} du_1 du_2 \cdots du_{n-1}$$
$$= \int_{S_t} (-1)^{n-1} f dx_1 \wedge dx_2 \wedge \cdots \wedge dx_{n-1}.$$

Similar argument may be made for the bottom boundary S_b, for which we note that the parametrization $\sigma(\vec{u}) = (\vec{u}, k(\vec{u}))$ is compatibly oriented (the normal vector point opposite to the direction of x_n) if n is even and is oppositely oriented if n is odd. We then conclude that

$$\int_B (f(\vec{u}, h(\vec{u})) - f(\vec{u}, k(\vec{u}))) du_1 du_2 \cdots du_{n-1} = \int_S (-1)^{n-1} f dx_1 \wedge dx_2 \wedge \cdots \wedge dx_{n-1}.$$

The argument can be similarly applied to the case A is given by

$$h(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{n-1}) \geq x_i \geq k(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{n-1}).$$

In this case, we have

$$\int_A f_{x_i} dx_1 dx_2 \cdots dx_n = \int_B f(u_1, \ldots, u_{i-1}, h(\vec{u}), u_i, \ldots, u_n) du_1 du_2 \cdots du_{n-1}$$
$$- \int_B f(u_1, \ldots, u_{i-1}, k(\vec{u}), u_i, \ldots, u_n) du_1 du_2 \cdots du_{n-1}.$$

Moreover, S_t may be parametrized by

$$\sigma_i(\vec{u}) = \sigma_i(u_1, u_2, \ldots, u_{n-1}) = (u_1, u_2, \ldots, u_{i-1}, h(\vec{u}), u_i, \ldots, u_{n-1}).$$

The normal vector for S_t points in the direction of x_i. By the discussion at the end of Section 7.2.7, the orientation of the parametrization σ_i is compatible to the normal vector if i is odd and is not compatible if i is even. Consequently, we get

$$\int_B f(\sigma_i(\vec{u})) du_1 du_2 \cdots du_{n-1}$$
$$= \int_B f(\sigma_i(\vec{u})) \frac{\partial(x_1, x_2, \ldots, \widehat{x_i}, \ldots, x_n)}{\partial(u_1, u_2, \ldots, u_{n-1})} du_1 du_2 \cdots du_{n-1}$$
$$= \int_{S_t} (-1)^{i-1} f dx_1 \wedge dx_2 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_{n-1}.$$

Therefore we get the following extension of Gauss theorem. Suppose $A \subset \mathbb{R}^n$ is a region with boundary hypersurface S compatibly oriented with respect to the outward normal vector $\vec{n}$. Then for any continuously differentiable $F = (f_1, f_2, \ldots, f_n)$, we have

$$\int_S \sum (-1)^{i-1} f_i dx_1 \wedge dx_2 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_n$$
$$= \int_A \left(\frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2} + \cdots + \frac{\partial f_n}{\partial x_n} \right) dx_1 dx_2 \cdots dx_n. \tag{7.3.11}$$

Extend the definition of divergence to vector fields in $\mathbb{R}^n$

$$\text{div} F = \frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2} + \cdots + \frac{\partial f_n}{\partial x_n}.$$

The Gauss theorem can be rewritten as

$$\int_S F \cdot \vec{n} dV = \int_A \text{div} F d\mu_{\vec{x}}.$$