Calculus

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

August 21, 2017

Contents

Chapter 1

Limit

1.1 Limit of Sequence

A *sequence* is an infinite list

$$x_1, x_2, \dots, x_n, \dots .$$

The n-th term of the sequence is x_n, and n is the *index* of the term. In this course, we will always assume that all the terms are real numbers. Here are some examples

$$\begin{aligned}
x_n = n\colon &\quad 1,\ 2,\ 3,\ \dots,\ n,\ \dots;\\
y_n = 2\colon &\quad 2,\ 2,\ 2,\ \dots,\ 2,\ \dots;\\
z_n = \frac{1}{n}\colon &\quad 1,\ \frac{1}{2},\ \dots,\ \frac{1}{n},\ \dots;\\
u_n = (-1)^n\colon &\quad 1,\ -1,\ 1,\ \dots,\ (-1)^n,\ \dots;\\
v_n = \sin n\colon &\quad \sin 1,\ \sin 2,\ \sin 3,\ \dots,\ \sin n,\ \dots .
\end{aligned}$$

Note that the index does not have to start from 1. For example, the sequence v_n actually starts from $n = 0$ (or any even integer). Moreover, a sequence does not have to be given by a formula. For example, the decimal expansions of π give a sequence

$$w_n\colon 3,\ 3.1,\ 3.14,\ 3.141,\ 3.1415,\ 3.14159,\ 3.141592,\ \dots .$$

If n is the number of digits after the decimal point, then the sequence w_n starts at $n = 0$.

Now we look at the trend of the examples above as n gets bigger. We find that x_n gets bigger and can become as big as we want. On the other hand, y_n remains constant, z_n gets smaller and can become as small as we want. This means that y_n approaches 2 and z_n approaches 0. Moreover, u_n and v_n jump around and do not approach anything. Finally, w_n is equal to π up to the n-th decimal place, and therefore approaches π.

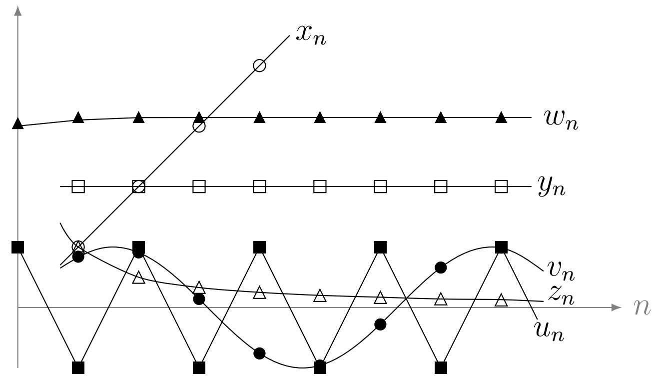


Figure 1.1.1: Sequences.

Definition 1.1.1 (Intuitive)**.** If x_n approaches a finite number l when n gets bigger and bigger, then we say that the sequence x_n *converges* to the *limit* l and write

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

A sequence *diverges* if it does not approach a specific finite number when n gets bigger.

The sequences y_n, z_n, w_n converge respectively to 2, 0 and π. The sequences x_n, u_n, v_n diverge. Since the limit describes the behavior when n gets very big, we have the following property.

Proposition 1.1.2. *If y_n is obtained from x_n by adding, deleting, or changing finitely many terms, then* $\lim_{n\to\infty} x_n = \lim_{n\to\infty} y_n$.

The equality in the proposition means that x_n converges if and only if y_n converges. Moreover, the two limits have equal value when both converge.

Example 1.1.1. The sequence $\dfrac{1}{\sqrt{n+2}}$ is obtained from $\dfrac{1}{\sqrt{n}}$ by deleting the first two terms. By $\lim_{n\to\infty} \dfrac{1}{\sqrt{n}} = 0$ and Proposition 1.1.2, we get $\lim_{n\to\infty} \dfrac{1}{\sqrt{n}} = \lim_{n\to\infty} \dfrac{1}{\sqrt{n+2}} = 0$.

In general, we have $\lim_{n\to\infty} x_{n+k} = \lim_{n\to\infty} x_n$ for any integer k.

The example assumes $\lim_{n\to\infty} \dfrac{1}{\sqrt{n}} = 0$, which is supposed to be intuitively obvious. Although mathematics is inspired by intuition, a critical feature of mathematics is rigorous logic. This means that we need to be clear what basic facts are assumed in any argument. For the moment, we will always assume that we already know

$\lim_{n\to\infty} c = c$ and $\lim_{n\to\infty} \dfrac{1}{n^p} = 0$ for $p > 0$. After the two limits are rigorously established in Examples 1.2.2 and 1.2.3, the conclusions based on the two limits become solid.

1.1.1 Arithmetic Rule

Intuitively, if x is close to 3 and y is close to 5, then the arithmetic combinations $x+y$ and xy are close to $3+5=8$ and $3\cdot 5=15$. The intuition leads to the following property of limit.

Proposition 1.1.3 (Arithmetic Rule). *Suppose* $\lim_{n\to\infty} x_n = l$ *and* $\lim_{n\to\infty} y_n = k$. *Then*

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

where c is a constant and $k \neq 0$ in the last equality.

The proposition says $\lim_{n\to\infty}(x_n+y_n) = \lim_{n\to\infty} x_n + \lim_{n\to\infty} y_n$. However, the equality is of different nature from the equality in Proposition 1.1.2, because the convergence of the limits on two sides are *not* equivalent: If the two limits on the right converge, then the limit on the left also converges and the two sides are equal. However, for $x_n = (-1)^n$ and $y_n = (-1)^{n+1}$, the limit $\lim_{n\to\infty}(x_n+y_n) = 0$ on the left converges, but both limits on the right diverge.

Exercise 1.1.1. Explain that $\lim_{n\to\infty} x_n = l$ if and only if $\lim_{n\to\infty}(x_n - l) = 0$.

Exercise 1.1.2. Suppose x_n and y_n converge. Explain that $\lim_{n\to\infty} x_n y_n = 0$ implies either $\lim_{n\to\infty} x_n = 0$ or $\lim_{n\to\infty} y_n = 0$. Moreover, explain that the conclusion fails if x_n and y_n are not assumed to converge.

Example 1.1.2. We have

$$\begin{aligned}
\lim_{n\to\infty} \frac{2n^2+n}{n^2-n+1} &= \lim_{n\to\infty} \frac{2+\dfrac{1}{n}}{1-\dfrac{1}{n}+\dfrac{1}{n^2}} = \frac{\lim_{n\to\infty}\left(2+\dfrac{1}{n}\right)}{\lim_{n\to\infty}\left(1-\dfrac{1}{n}+\dfrac{1}{n^2}\right)} \\
&= \frac{\lim_{n\to\infty} 2 + \lim_{n\to\infty} \dfrac{1}{n}}{\lim_{n\to\infty} 1 - \lim_{n\to\infty} \dfrac{1}{n} + \lim_{n\to\infty} \dfrac{1}{n} \cdot \lim_{n\to\infty} \dfrac{1}{n}} \\
&= \frac{2+0}{1-0+0\cdot 0} = 2.
\end{aligned}$$

The arithmetic rule is used in the second and third equalities. The limits $\lim_{n\to\infty} c = c$ and $\lim_{n\to\infty} \frac{1}{n} = 0$ are used in the fourth equality.

Exercise 1.1.3. Find the limits.

1. $\dfrac{n+2}{n-3}$.
2. $\dfrac{n+2}{n^2-3}$.
3. $\dfrac{2n^2-3n+2}{3n^2-4n+1}$.
4. $\dfrac{n^3+4n^2-2}{2n^3-n+3}$.
5. $\dfrac{(n+1)(n+2)}{2n^2-1}$.
6. $\dfrac{2n^2-1}{(n+1)(n+2)}$.
7. $\dfrac{(n^2+1)(n+2)}{(n+1)(n^2+2)}$.
8. $\dfrac{(2-n)^3}{2n^3+3n-1}$.
9. $\dfrac{(n^2+3)^3}{(n^3-2)^2}$.

Exercise 1.1.4. Find the limits.

1. $\dfrac{\sqrt{n}+2}{\sqrt{n}-3}$.
2. $\dfrac{\sqrt{n}+2}{n-3}$.
3. $\dfrac{2\sqrt{n}-3n+2}{3\sqrt{n}-4n+1}$.
4. $\dfrac{\sqrt[3]{n}+4\sqrt{n}-2}{2\sqrt[3]{n}-n+3}$.
5. $\dfrac{(\sqrt{n}+1)(\sqrt{n}+2)}{2n-1}$.
6. $\dfrac{2n-1}{(\sqrt{n}+1)(\sqrt{n}+2)}$.
7. $\dfrac{(\sqrt{n}+1)(n+2)}{(n+1)(\sqrt{n}+2)}$.
8. $\dfrac{(2-\sqrt[3]{n})^3}{2\sqrt[3]{n}+3n-1}$.
9. $\dfrac{(\sqrt[3]{n}+3)^3}{(\sqrt{n}-2)^2}$.

Exercise 1.1.5. Find the limits.

1. $\dfrac{n+a}{n+b}$.
2. $\dfrac{\sqrt{n}+a}{n+b}$.
3. $\dfrac{n+a}{n^2+bn+c}$.
4. $\dfrac{\sqrt{n}+a}{n+b\sqrt{n}+c}$.
5. $\dfrac{(\sqrt{n}+a)(\sqrt{n}+b)}{cn+d}$.
6. $\dfrac{cn+d}{(\sqrt{n}+a)(\sqrt{n}+b)}$.
7. $\dfrac{an^3+b}{(c\sqrt{n}+d)^6}$.
8. $\dfrac{(a\sqrt[3]{n}+b)^2}{(c\sqrt{n}+d)^3}$.
9. $\dfrac{(a\sqrt{n}+b)^2}{(c\sqrt[3]{n}+d)^3}$.

Exercise 1.1.6. Show that

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

Exercise 1.1.7. Find the limits.

1. $\dfrac{10^{10}n}{n^2-10}$.
2. $\dfrac{5^5(2n+1)^2-10^{10}}{10n^2-5}$.
3. $\dfrac{5^5(2\sqrt{n}+1)^2-10^{10}}{10n-5}$.

Exercise 1.1.8. Find the limits.

1. $\dfrac{n}{n+1}-\dfrac{n}{n-1}$.
2. $\dfrac{n^2}{n+1}-\dfrac{n^2}{n-1}$.
3. $\dfrac{n}{\sqrt{n}+1}-\dfrac{n}{\sqrt{n}-1}$.
4. $\dfrac{n+a}{n+b}-\dfrac{n+c}{n+d}$.
5. $\dfrac{n^2+a}{n+b}-\dfrac{n^2+c}{n+d}$.
6. $\dfrac{n+a}{\sqrt{n}+b}-\dfrac{n+c}{\sqrt{n}+d}$.
7. $\dfrac{n^3+a}{n^2+b}-\dfrac{n^3+c}{n^2+d}$.
8. $\dfrac{n^2+a}{n^3+b}-\dfrac{n^2+c}{n^3+d}$.
9. $\dfrac{\sqrt{n}+a}{\sqrt[3]{n}+b}-\dfrac{\sqrt{n}+c}{\sqrt[3]{n}+d}$.

Exercise 1.1.9. Find the limits.

1. $\dfrac{n^2+a_1n+a_0}{n+b}-\dfrac{n^2+c_1n+c_0}{n+d}$.
2. $\dfrac{n^2+a_1n+a_0}{n^2+b_1n+b_0}-\dfrac{n^2+c_1n+c_0}{n^2+d_1n+d_0}$.
3. $\left(\dfrac{n+a}{n+b}\right)^2-\left(\dfrac{n+c}{n+d}\right)^2$.
4. $\left(\dfrac{n^2+a}{n+b}\right)^2-\left(\dfrac{n^2+c}{n+d}\right)^2$.

Exercise 1.1.10. Find the limits, $p,q>0$.

1. $\dfrac{n^p+a}{n^q+b}$.
2. $\dfrac{an^p+bn^q+c}{an^q+bn^p+c}$.
3. $\dfrac{n^p+a}{n^q+b}-\dfrac{n^p+c}{n^q+d}$.
4. $\dfrac{n^{2p}+a_1n^p+a_2}{n^{2q}+b_1n^q+b_2}$.

1.1.2 Sandwich Rule

The following property reflects the intuition that if x and z are close to 3, then anything between x and z should also be close to 3.

Proposition 1.1.4 (Sandwich Rule). *Suppose $x_n \le y_n \le z_n$ for sufficiently big n. If $\lim_{n\to\infty} x_n = \lim_{n\to\infty} z_n = l$, then $\lim_{n\to\infty} y_n = l$.*

Note that something holds for sufficiently big n is the same as something fails for only finitely many n.

Example 1.1.3. By $2n-3>n$ for sufficiently big n (in fact, $n>3$ is enough), we have

$$0<\frac{1}{\sqrt{2n-3}}<\frac{1}{\sqrt{n}}.$$

Then by $\lim_{n\to\infty} 0 = \lim_{n\to\infty}\dfrac{1}{\sqrt{n}} = 0$ and the sandwich rule, we get $\lim_{n\to\infty}\dfrac{1}{\sqrt{2n-3}} = 0$.

On the other hand, for sufficiently big n, we have $n+1<2n$ and $n-1>\dfrac{n}{2}$, and therefore

$$0<\frac{\sqrt{n+1}}{n-1}<\frac{\sqrt{2n}}{\frac{n}{2}}=\frac{2\sqrt{2}}{\sqrt{n}}.$$

By $\lim_{n\to\infty} \dfrac{2\sqrt{2}}{\sqrt{n}} = 2\sqrt{2}\lim_{n\to\infty} \dfrac{1}{\sqrt{n}} = 0$ (arithmetic rule used) and the sandwich rule, we get $\lim_{n\to\infty} \dfrac{\sqrt{n+1}}{n-1} = 0$.

Example 1.1.4. By $-1 \le \sin n \le 1$, we have

$$-\frac{1}{n} \le \frac{\sin n}{n} \le \frac{1}{n}, \quad \frac{1}{\sqrt{n+1}} \le \frac{1}{\sqrt{n+\sin n}} \le \frac{1}{\sqrt{n-1}}.$$

By $\lim_{n\to\infty} \dfrac{1}{n} = 0$, $-\lim_{n\to\infty} \dfrac{1}{n} = 0$ and the sandwich rule, we get $\lim_{n\to\infty} \dfrac{\sin n}{n} = 0$. Moreover, by $\lim_{n\to\infty} \dfrac{1}{\sqrt{n+1}} = \lim_{n\to\infty} \dfrac{1}{\sqrt{n-1}} = 0$ (see argument in Example 1.1.1) and the sandwich rule, we get $\lim_{n\to\infty} \dfrac{1}{\sqrt{n+\sin n}} = 0$.

Exercise 1.1.11. Prove that $\lim_{n\to\infty} |x_n| = 0$ implies $\lim_{n\to\infty} x_n = 0$.

Exercise 1.1.12. Find the limits, $a > 0$.

1. $\dfrac{1}{\sqrt{3n-4}}$.
2. $\dfrac{\sqrt{2n+3}}{4n-1}$.
3. $\dfrac{1}{\sqrt{an+b}}$.
4. $\dfrac{\sqrt{an+b}}{cn+d}$.

Exercise 1.1.13. Find the limits.

1. $\dfrac{\cos n}{n}$.
2. $\dfrac{(-1)^n}{n}$.
3. $\dfrac{\sin\sqrt{n}}{n}$.
4. $\dfrac{\cos n}{\sqrt{n-2}}$.
5. $\dfrac{1}{n+(-1)^n}$.
6. $\dfrac{\cos n}{n+(-1)^n}$.
7. $\dfrac{\cos n}{\sqrt{n+(-1)^n 2}}$.
8. $\dfrac{\cos n}{\sqrt{n+\sin\sqrt{n}}}$.
9. $\dfrac{(-1)^n}{\sqrt{n+(-1)^n}}$.
10. $\dfrac{2+(-1)^n 3}{\sqrt[3]{n^2-2\cos n}}$.
11. $\dfrac{\sin n+(-1)^n\cos n}{\sqrt{n}+(-1)^n}$.
12. $\dfrac{|\sin n+\cos n|}{n}$.
13. $\dfrac{3\sqrt{n}+2}{2n+(-1)^n 3}$.
14. $\dfrac{\sqrt{n}\sin n+\cos n}{n-1}$.
15. $\dfrac{n+\sin\sqrt{n}}{n+\cos 2n}$.
16. $\dfrac{\sqrt{n}+\sin n}{\sqrt{n}-\cos n}$.
17. $\dfrac{(-1)^n(n+1)}{n^2+(-1)^{n+1}}$.
18. $\dfrac{(-1)^n(n+10)^2-10^{10}}{10(-1)^n n^2-5}$.

Exercise 1.1.14. Find the limits.

1. $\dfrac{\sqrt{n}+a}{n+(-1)^n b}$.
2. $\dfrac{1}{\sqrt[3]{n^2+an+b}}$.
3. $\dfrac{\sqrt{n+c}+d}{\sqrt[3]{n^2+an+b}}$.
4. $\dfrac{n+(-1)^n a}{n+(-1)^n b}$.
5. $\dfrac{(-1)^n(an+b)}{n^2+c(-1)^{n+1}n+d}$.
6. $\dfrac{(-1)^n(an+b)^2+c}{(-1)^n n^2+d}$.
7. $\dfrac{\cos\sqrt{n}+a}{n+b\sin n}$.
8. $\dfrac{\cos\sqrt{n}+a}{\sqrt{n+b\sin n}}$.
9. $\dfrac{an+b\sin n}{cn+d\sin n}$.

Exercise 1.1.15. Find the limits, $p>0$.

1. $\dfrac{\sin\sqrt{n}}{n^p}$.
2. $\dfrac{\sin(n+1)}{n^p+(-1)^n}$.
3. $\dfrac{a\sin n+b}{n^p+c}$.
4. $\dfrac{a\cos(\sin n)}{n^p-b\sin n}$.

Example 1.1.5. For $a>0$, the sequence $\sqrt{n+a}-\sqrt{n}$ satisfies

$$0<\sqrt{n+a}-\sqrt{n}=\frac{(\sqrt{n+a}-\sqrt{n})(\sqrt{n+a}+\sqrt{n})}{\sqrt{n+a}+\sqrt{n}}=\frac{a}{\sqrt{n+a}+\sqrt{n}}<\frac{a}{\sqrt{n}}.$$

By $\lim_{n\to\infty}\dfrac{a}{\sqrt{n}}=0$ and the sandwich rule, we get $\lim_{n\to\infty}(\sqrt{n+a}-\sqrt{n})=0$. Similar argument also shows the limit for $a<0$.

Example 1.1.6. The sequence $\sqrt{\dfrac{n+2}{n}}$ satisfies

$$1<\sqrt{\frac{n+2}{n}}<\frac{n+2}{n}=1+2\frac{1}{n}.$$

By $\lim_{n\to\infty}\left(1+2\dfrac{1}{n}\right)=1+2\cdot 0=1$ and the sandwich rule, we get $\lim_{n\to\infty}\sqrt{\dfrac{n+2}{n}}=1$.

Exercise 1.1.16. Show that $\lim_{n\to\infty}(\sqrt{n+a}-\sqrt{n})=0$ for $a<0$.

Exercise 1.1.17. Use the idea of Example 1.1.5 to estimate $\sqrt{\dfrac{n+2}{n}}-1$ and then find $\lim_{n\to\infty}\sqrt{\dfrac{n+2}{n}}$.

Exercise 1.1.18. Show that $\lim_{n\to\infty}\sqrt{\dfrac{n+a}{n+b}}=1$. You may need separate argument for $a>b$ and $a<b$.

Exercise 1.1.19. Find the limits.

1. $\sqrt{n+a}-\sqrt{n+b}$.
2. $\dfrac{\sqrt{n+a}}{\sqrt{n+c}+\sqrt{n+d}}$.

3. $\dfrac{\sqrt{n+a}+\sqrt{n+b}}{\sqrt{n+c}+\sqrt{n+d}}$.
4. $\dfrac{\sqrt{n+a}\sqrt{n+b}}{\sqrt{n+c}\sqrt{n+d}}$.
5. $\dfrac{\sqrt{n+a}+b}{\sqrt{n+c}+d}$.
6. $\sqrt{n}(\sqrt{n+a}-\sqrt{n+b})$.
7. $\sqrt{n+c}(\sqrt{n+a}-\sqrt{n+b})$.
8. $\sqrt{n+a}+\sqrt{n+b}-2\sqrt{n+c}$.
9. $\sqrt{\dfrac{n}{n^2+n+1}}$.
10. $\sqrt{\dfrac{n+a}{n^2+bn+c}}$.
11. $\sqrt{n^2+an+b}-\sqrt{n^2+cn+d}$.
12. $\sqrt{n+a}\sqrt{n+b}-\sqrt{n+c}\sqrt{n+d}$.
13. $\dfrac{n}{\sqrt{n^2+n+1}}$.
14. $\dfrac{n+a}{\sqrt{n^2+bn+c}}$.
15. $\sqrt{\dfrac{n^2+an+b}{n^2+cn+d}}$.

Exercise 1.1.20. Find the limits.

1. $\sqrt{n+a\sin n}-\sqrt{n+b\cos n}$.
2. $\sqrt{\dfrac{n+a\sin n}{n+b\cos n}}$.
3. $\sqrt{\dfrac{n+(-1)^n a}{n+(-1)^n b}}$.
4. $\dfrac{\sqrt{n+a}+\sin n}{\sqrt{n+c}+(-1)^n}$.
5. $\sqrt{n+(-1)^n}(\sqrt{n+a}-\sqrt{n+b})$.
6. $\sqrt{n^2+an+\sin n}-\sqrt{n^2+bn+\cos n}$.
7. $\sqrt{\dfrac{n^2+an+\sin n}{n^2+bn+\cos n}}$.
8. $\dfrac{\sqrt{n^2+an+b}}{n+(-1)^n c}$.

Exercise 1.1.21. Find the limits.

1. $\sqrt[3]{n+a}-\sqrt[3]{n+b}$.
2. $\sqrt[3]{\dfrac{n+a}{n+b}}$.
3. $\sqrt[3]{n^2}(\sqrt[3]{n+a}-\sqrt[3]{n+b})$.
4. $\sqrt[3]{n}(\sqrt[3]{\sqrt{n}+a}-\sqrt[3]{\sqrt{n}+b})$.

Exercise 1.1.22. Find the limits.

1. $\left(\dfrac{n-2}{n+1}\right)^5$.
2. $\left(\dfrac{n-2}{n+1}\right)^{5.4}$.
3. $\left(\dfrac{n-2}{n+1}\right)^{-\sqrt{2}}$.
4. $\left(\dfrac{n+a}{n+b}\right)^p$.

Exercise 1.1.23. Find the limits.

1. $\left(\dfrac{\sqrt{n}+a\sin n}{\sqrt{n}+b\cos 2n}\right)^p$.
2. $\left(\dfrac{n^2+an+b}{n^2+(-1)^n c}\right)^p$.
3. $\left(\dfrac{n+a}{n^2+bn+c}\right)^p$.

Exercise 1.1.24. Suppose $\lim_{n\to\infty} x_n = 1$. Use the arithmetic rule and the sandwich rule to prove that, if $x_n \le 1$, then $\lim_{n\to\infty} x_n^p = 1$. Of course we expect the condition $x_n \le 1$ to be unnecessary. See Example 1.1.21.

1.1.3 Some Basic Limits

Using $\lim_{n\to\infty}\dfrac{1}{n}=0$ and the sandwich rule, we may establish some basic limits.

Example 1.1.7. We show that

$$\lim_{n\to\infty}\sqrt[n]{a}=1,\ \text{for } a>0.$$

First assume $a\ge 1$. Then $x_n=\sqrt[n]{a}-1\ge 0$, and

$$a=(1+x_n)^n=1+nx_n+\frac{n(n-1)}{2}x_n^2+\cdots+x_n^n>nx_n.$$

This implies

$$0\le x_n<\frac{a}{n}.$$

By the sandwich rule and $\lim_{n\to\infty}\dfrac{a}{n}=0$, we get $\lim_{n\to\infty}x_n=0$. Then by the arithmetic rule, this further implies

$$\lim_{n\to\infty}\sqrt[n]{a}=\lim_{n\to\infty}(x_n+1)=\lim_{n\to\infty}x_n+1=1.$$

For the case $0<a\le 1$, let $b=\dfrac{1}{a}\ge 1$. Then by the arithmetic rule,

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

Example 1.1.8. Example 1.1.7 can be extended to

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

Let $x_n=\sqrt[n]{n}-1$. Then we have $x_n>0$ for sufficiently large n (in fact, $n\ge 2$ is enough), and

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

This implies

$$0\le x_n<\frac{\sqrt{2}}{\sqrt{n-1}}.$$

By $\lim_{n\to\infty}\dfrac{\sqrt{2}}{\sqrt{n-1}}=0$ (see Example 1.1.1 or 1.1.3) and the sandwich rule, we get $\lim_{n\to\infty}x_n=0$. This further implies

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

Example 1.1.9. The following "n-th root type" limits can be compared with the limits in Examples 1.1.7 and 1.1.8

$$\begin{aligned}
&1 < \sqrt[n]{n+1} < \sqrt[n]{2n} = \sqrt[n]{2}\,\sqrt[n]{n},\\
&1 < n^{\frac{1}{n+1}} < \sqrt[n]{n},\\
&1 < (n^2-n)^{\frac{n}{n^2-1}} < (n^2)^{\frac{n}{n^2/2}} = (\sqrt[n]{n})^4.
\end{aligned}$$

By Examples 1.1.7, 1.1.8 and the arithmetic rule, the sequences on the right converge to 1. Then by the sandwich rule, we get

$$\lim_{n\to\infty} \sqrt[n]{n+1} = \lim_{n\to\infty} n^{\frac{1}{n+1}} = \lim_{n\to\infty} (n^2-n)^{\frac{n}{n^2-1}} = 1.$$

Example 1.1.10. We have

$$3 = \sqrt[n]{3^n} < \sqrt[n]{2^n+3^n} < \sqrt[n]{3^n+3^n} = 3\sqrt[n]{2}.$$

By Example 1.1.7, we have $\lim_{n\to\infty} 3\sqrt[n]{2} = 3$. Then by the sandwich rule, we get $\lim_{n\to\infty} \sqrt[n]{2^n+3^n} = 3$.

For another example, we have

$$3^n > 3^n - 2^n = 3^{n-1} + 2\cdot 3^{n-1} - 2\cdot 2^{n-1} > 3^{n-1}.$$

Taking the n-th root, we get

$$3 > \sqrt[n]{3^n-2^n} > 3\frac{1}{\sqrt[n]{3}}.$$

By $\lim_{n\to\infty} 3\dfrac{1}{\sqrt[n]{3}} = 3$ and the sandwich rule, we get $\lim_{n\to\infty} \sqrt[n]{3^n-2^n} = 3$.

Exercise 1.1.25. Prove that if $a \le x_n \le b$ for some constants $a, b > 0$ and sufficiently big n, then $\lim_{n\to\infty} \sqrt[n]{x_n} = 1$.

Exercise 1.1.26. Find the limits, $a > 0$.

1. $n^{\frac{1}{2n}}$.
2. $n^{\frac{2}{n}}$.
3. $n^{\frac{c}{n}}$.
4. $(n+1)^{\frac{c}{n}}$.
5. $(an+b)^{\frac{c}{n}}$.
6. $(an^2+b)^{\frac{c}{n}}$.
7. $(\sqrt{n}+1)^{\frac{c}{n}}$.
8. $(n-2)^{\frac{1}{n+3}}$.
9. $(an+b)^{\frac{c}{n+d}}$.
10. $(an+b)^{\frac{cn}{n^2+dn+e}}$.
11. $(an^2+b)^{\frac{c}{n+d}}$.
12. $(an^2+b)^{\frac{cn+d}{n^2+en+f}}$.

Exercise 1.1.27. Find the limits, $a > 0$.

1. $\sqrt[n]{n+\sin n}$.
2. $\sqrt[n]{an+b\sin n}$.
3. $\sqrt[n]{n+(-1)^n\sin n}$.
4. $\sqrt[n]{an+(-1)^n b\sin n}$.
5. $(n-\cos n)^{\frac{1}{n+\sin n}}$.
6. $(an+b\sin n)^{\frac{n}{n^2+c\cos n}}$.

Exercise 1.1.28. Find the limits, $p, q > 0$.

1. $\sqrt[n]{n^p+\sin n}$.
2. $\sqrt[n]{n^p+n^q}$.
3. $\sqrt[n+2]{n^p+n^q}$.
4. $\sqrt[n-2]{n^p+n^q}$.

Exercise 1.1.29. Find the limits.

1. $\sqrt[n]{5^n-4^n}$.
2. $\sqrt[n]{5^n-3\cdot 4^n}$.
3. $\sqrt[n]{5^n-3\cdot 4^n+2^n}$.
4. $\sqrt[n]{5^n-3\cdot 4^n-2^n}$.
5. $\sqrt[n]{4^{2n-1}-5^n}$.
6. $\sqrt[n]{4^{2n-1}+(-1)^n 5^n}$.
7. $(5^n-4^n)^{\frac{1}{n+1}}$.
8. $(5^n-4^n)^{\frac{1}{n-2}}$.
9. $(5^n-4^n)^{\frac{n+1}{n^2+1}}$.

Exercise 1.1.30. Find the limits, $a > b > 0$.

1. $\sqrt[n]{a^n+b^n}$.
2. $\sqrt[n]{a^n-b^n}$.
3. $\sqrt[n]{a^n+(-1)^n b^n}$.
4. $\sqrt[n]{a^n b^{2n+1}}$.
5. $\sqrt[n+2]{a^n+b^n}$.
6. $\sqrt[n-2]{a^n-b^n}$.
7. $(a^n+b^n)^{\frac{n}{n^2-1}}$.
8. $(a^n-b^n)^{\frac{n}{n^2-1}}$.
9. $(a^n-(-1)^n b^n)^{\frac{n}{n^2-1}}$.

Exercise 1.1.31. For $a, b, c > 0$, find $\lim_{n\to\infty}\sqrt[n]{a^n+b^n+c^n}$.

Exercise 1.1.32. For $a \ge 1$, prove $\lim_{n\to\infty}\sqrt[n]{a}=1$ by using

$$a-1=(\sqrt[n]{a}-1)\left((\sqrt[n]{a})^{n-1}+(\sqrt[n]{a})^{n-2}+\cdots+\sqrt[n]{a}+1\right).$$

Example 1.1.11. We show that

$$\lim_{n\to\infty} a^n = 0, \text{ for } |a| < 1.$$

First assume $0 < a < 1$ and write $a = \dfrac{1}{1+b}$. Then $b > 0$ and

$$0 < a^n = \frac{1}{(1+b)^n} = \frac{1}{1+nb+\dfrac{n(n-1)}{2}b^2+\cdots+b^n} < \frac{1}{nb}.$$

By $\lim_{n\to\infty}\dfrac{1}{nb}=0$ and the sandwich rule, we get $\lim_{n\to\infty} a^n = 0$.

If $-1 < a < 0$, then $0 < |a| < 1$ and $\lim_{n\to\infty}|a^n| = \lim_{n\to\infty}|a|^n = 0$. By Exercise 1.1.11, we get $\lim_{n\to\infty} a^n = 0$.

Example 1.1.12. Example 1.1.11 can be extended to

$$\lim_{n\to\infty} na^n = 0, \text{ for } |a| < 1.$$

This follows from

$$0 < na^n = \frac{n}{(1+b)^n} = \frac{n}{1 + nb + \frac{n(n-1)}{2}b^2 + \cdots + b^n} < \frac{n}{\frac{n(n-1)}{2}b^2} = \frac{2}{(n-1)b^2},$$

the limit $\lim_{n\to\infty} \dfrac{2}{(n-1)b^2} = 0$ and the sandwich rule.

Exercise 1.1.58 gives further extension of the limit.

Example 1.1.13. We show that

$$\lim_{n\to\infty} \frac{a^n}{n!} = 0, \text{ for any } a$$

for the special case $a = 4$. For $n > 4$, we have

$$0 < \frac{4^n}{n!} = \frac{4\cdot 4\cdot 4\cdot 4}{1\cdot 2\cdot 3\cdot 4} \cdot \frac{4}{5} \cdot \frac{4}{6} \cdots \frac{4}{n} \le \frac{4\cdot 4\cdot 4\cdot 4}{1\cdot 2\cdot 3\cdot 4} \cdot \frac{4}{n} = \frac{4^5}{4!}\frac{1}{n}.$$

By $\lim_{n\to\infty} \dfrac{4^5}{4!}\dfrac{1}{n} = 0$ and the sandwich rule, we get $\lim_{n\to\infty} \dfrac{4^n}{n!} = 0$.

Exercise 1.1.59 suggests how to show the limit in general.

Exercise 1.1.33. Show that $\lim_{n\to\infty} n^2 a^n = 0$ for $|a| < 1$ in two ways. The first is by using the ideas from Examples 1.1.11 and 1.1.12. The second is by using $\lim_{n\to\infty} na^n = 0$ for $|a| < 1$.

Exercise 1.1.34. Show that $\lim_{n\to\infty} n^{5.4} a^n = 0$ for $|a| < 1$. What about $\lim_{n\to\infty} n^{-5.4} a^n$? What about $\lim_{n\to\infty} n^p a^n$?

Exercise 1.1.35. Show that $\lim_{n\to\infty} \dfrac{a^n}{n!} = 0$ for $a = 5.4$ and $a = -5.4$.

Exercise 1.1.36. Show that $\lim_{n\to\infty} \dfrac{a^n}{\sqrt{n!}} = 0$ for $a = 5.4$ and $a = -5.4$.

Exercise 1.1.37. Show that $\lim_{n\to\infty} \dfrac{n!a^n}{(2n)!} = 0$ for any a.

Exercise 1.1.38. Find the limits.

1. $\dfrac{n+1}{2^n}$.
2. $\dfrac{n^2}{2^n}$.
3. $n^{99} 0.99^n$.
4. $\dfrac{(n^2+1)^{1001}}{1.001^{n-2}}$.
5. $\dfrac{n+2^n}{3^n}$.
6. $\dfrac{n2^n + (-3)^n}{4^n}$.
7. $\dfrac{n3^n}{(1+2^n)^2}$.
8. $\dfrac{5^n - n6^{n+1}}{3^{2n-1} - 2^{3n+1}}$.

Exercise 1.1.39. Find the limits. Some convergence depends on a and p. You may try some special values of a and p first.

1. $n^p a^n$.
2. $\dfrac{a^n}{n^p}$.
3. $\dfrac{n^p}{a^n}$.
4. $\dfrac{1}{n^p a^n}$.
5. $\dfrac{n^p}{n!}$.
6. $\dfrac{n^p a^n}{n!}$.
7. $\dfrac{n^p}{n!a^n}$.
8. $\dfrac{n^p}{\sqrt{n!}}$.
9. $\dfrac{n^p a^n}{\sqrt{n!}}$.
10. $\dfrac{n^p a^n}{\sqrt[3]{n!}}$.
11. $\dfrac{n!a^n}{(2n)!}$.
12. $\dfrac{n!n^p a^n}{(2n)!}$.

Exercise 1.1.40. Find the limits.

1. $\dfrac{n^2+3n+5^n}{n!}$.
2. $\dfrac{n^2+3n+5^n}{n!-n^2+2^n}$.
3. $\dfrac{n^2+n3^n+5!}{n!}$.
4. $\dfrac{n^2 3^{n+5}+5\cdot(n-1)!}{(n+1)!}$.
5. $\dfrac{n^2+n!+(n-1)!}{3^n-n!+(n-1)!}$.
6. $\dfrac{2^n n!+3^n(n-1)!}{4^n(2n-1)!-5^n n!}$.

Exercise 1.1.41. Prove $\sqrt[n]{n!} > \sqrt{\dfrac{n}{2}}$. Then use this to prove $\lim_{n\to\infty} \dfrac{1}{\sqrt[n]{n!}} = 0$.

Exercise 1.1.42. Prove $\dfrac{n!}{n^n} < \dfrac{1}{n}$ and $\dfrac{(n!)^2}{(2n)!} < \dfrac{1}{n+1}$ for $n > 2$. Then use this to prove $\lim_{n\to\infty} \dfrac{n!}{n^n} = \lim_{n\to\infty} \dfrac{(n!)^2}{(2n)!} = 0$. What about $\lim_{n\to\infty} \dfrac{(n!)^k}{(kn)!}$ where $k \geq 2$ is an integer?

1.1.4 Order Rule

The following property reflects the intuition that bigger sequence should have bigger limit.

Proposition 1.1.5 (Order Rule). *Suppose* $\lim_{n\to\infty} x_n = l$ *and* $\lim_{n\to\infty} y_n = k$.

1. *If $x_n \leq y_n$ for sufficiently big n, then $l \leq k$.*
2. *If $l < k$, then $x_n < y_n$ for sufficiently big n.*

By taking $y_n = l$, we get the following special cases of the property for a converging sequence x_n.

1. If $x_n \leq l$ for sufficiently big n, then $\lim_{n\to\infty} x_n \leq l$.
2. If $\lim_{n\to\infty} x_n < l$, then $x_n < l$ for sufficiently big n.

Similar statements with reversed inequalities also hold (see Exercise 1.1.43).

Note the non-strict inequality in the first statement of Proposition 1.1.5 and the strict inequality the second statement. For example, we have $x_n = \dfrac{1}{n^2} < y_n = \dfrac{1}{n}$, but $\lim_{n\to\infty} x_n \not< \lim_{n\to\infty} y_n$. The example also satisfies $\lim_{n\to\infty} x_n \geq \lim_{n\to\infty} y_n$ but $x_n \not\geq y_n$, even for sufficiently big n.

Exercise 1.1.43. Explain how to get the following special cases of the order rule.

1. If $x_n \geq l$ for sufficiently big n, then $\lim_{n\to\infty} x_n \geq l$.
2. If $\lim_{n\to\infty} x_n > l$, then $x_n > l$ for sufficiently big n.

Example 1.1.14. By $\lim_{n\to\infty} \dfrac{2n^2+n}{n^2-n+1} = 2$ and the order rule, we know $1 < \dfrac{2n^2+n}{n^2-n+1} < 3$ for sufficiently big n. This implies $1 < \sqrt[n]{\dfrac{2n^2+n}{n^2-n+1}} < \sqrt[n]{3}$ for sufficiently big n. By $\lim_{n\to\infty} \sqrt[n]{3} = 1$ and the sandwich rule, we get

$$\lim_{n\to\infty} \sqrt[n]{\frac{2n^2+n}{n^2-n+1}} = 1.$$

Example 1.1.15. We showed $\lim_{n\to\infty} \sqrt[n]{3^n - 2^n} = 3$ in Example 1.1.10. Here we use a different method, with the help of the order rule.

By $\sqrt[n]{3^n - 2^n} = 3\sqrt[n]{1 - \left(\dfrac{2}{3}\right)^n}$, we only need to find $\lim_{n\to\infty} \sqrt[n]{1 - \left(\dfrac{2}{3}\right)^n}$. By Example 1.1.11, we have $\lim_{n\to\infty} \left(1 - \left(\dfrac{2}{3}\right)^n\right) = 1$. By the order rule, therefore, we have

$$\frac{1}{2} < 1 - \left(\frac{2}{3}\right)^n < 2$$

for sufficiently big n. This implies that

$$\frac{1}{\sqrt[n]{2}} < \sqrt[n]{1 - \left(\frac{2}{3}\right)^n} < \sqrt[n]{2}$$

for sufficiently big n. Then by Example 1.1.7 and the sandwich rule, we get $\lim_{n\to\infty} \sqrt[n]{1 - \left(\dfrac{2}{3}\right)^n} = 1$, and we conclude that

$$\lim_{n\to\infty} \sqrt[n]{3^n - 2^n} = 3 \lim_{n\to\infty} \sqrt[n]{1 - \left(\frac{2}{3}\right)^n} = 3.$$

Exercise 1.1.44. Prove that if $\lim_{n\to\infty} x_n = l > 0$, then $\lim_{n\to\infty} \sqrt[n]{x_n} = 1$. Moreover, find a sequence satisfying $\lim_{n\to\infty} x_n = 0$ and $\lim_{n\to\infty} \sqrt[n]{x_n} = 1$. Can we have x_n converging to 0 and $\sqrt[n]{x_n}$ converging to 0.32?

Exercise 1.1.45. Find the limits.

1. $\sqrt[n]{5^n - n4^n}$.
2. $\sqrt[n+2]{5^n - n4^n}$.
3. $\sqrt[n-2]{5^n - n4^n}$.
4. $\sqrt[n]{5^n - (-1)^n n4^n}$.
5. $(5^n - n4^n)^{\frac{n-1}{n^2+1}}$.
6. $(5^n-(-1)^n n4^n)^{\frac{n+(-1)^n}{n^2+1}}$.

Exercise 1.1.46. Find the limits.

1. $\sqrt[n]{\dfrac{1}{n}5^n - n4^n}$.
2. $\sqrt[n]{\dfrac{1}{n}5^n - (-1)^n n4^n}$.
3. $\sqrt[n+2]{\dfrac{1}{n}5^n - n4^n \sin n}$.
4. $(n^2 4^{2n-1} - 5^n)^{\frac{n-1}{n^2+1}}$.
5. $\sqrt[n]{n2^n + 4^{n+1} + \dfrac{3^{n-1}}{n}}$.
6. $\sqrt[n-2]{n2^{3n} + \dfrac{3^{2n-1}}{n^2}}$.
7. $\sqrt[n-2]{2^{3n} + \dfrac{n-1}{n^2+1}3^{2n-1}}$.
8. $\left(n2^{3n} + \dfrac{3^{2n-1}}{n^2}\right)^{\frac{n-1}{n^2}}$.
9. $\left(n2^{3n} + \dfrac{3^{2n-1}}{n^2}\right)^{\frac{1}{n^2}}$.

Exercise 1.1.47. Find the limits, $a > b$.

1. $\sqrt[n]{a^{n+1} + b^n}$.
2. $\sqrt[n]{a^{n+1} + (-1)^n b^n}$.
3. $\sqrt[n-2]{a^{n+1} + (-1)^n b^n}$.
4. $\sqrt[n]{4a^n - 5b^n}$.
5. $\sqrt[n]{4a^n + 5b^{2n+1}}$.
6. $\sqrt[n]{a + b^n}$.
7. $\sqrt[n]{an + b^n}$.
8. $\sqrt[n-2]{na^n + (n^2+1)b^n}$.
9. $\sqrt[n+2]{\dfrac{1}{n}a^n + nb^{n+1}}$.
10. $(a^n + b^n)^{\frac{n+1}{n^2+1}}$.
11. $((n+1)a^n + b^n)^{\frac{n}{n^2-1}}$.
12. $(a^n + b^n)^{\frac{1}{n^2-1}}$.
13. $(a^n + (-1)^n b^n)^{\frac{(-1)^n}{n^2-1}}$.

Exercise 1.1.48. Find the limits, $a, b, c > 0$.

1. $\sqrt[n]{n^2a^n + nb^n + 2c^n}$.
2. $\sqrt[n]{a^n(b^n+1) + nc^n}$.
3. $\sqrt[n]{(n+\sin n)a^n + b^n + n^2c^n}$.
4. $\sqrt[n]{a^n(n + b^n(1+nc^n))}$.

Exercise 1.1.49. Suppose a polynomial $p(n) = a_pn^p + a_{n-1}n^{p-1} + \cdots + a_1n + a_0$ has leading coefficient $a_p > 0$. Prove that $p(n) > 0$ for sufficiently big n.

Exercise 1.1.50. Suppose $a, b, c > 0$, and p, q, r are polynomials with positive leading coefficients. Find the limit of $\sqrt[n]{p(n)a^n + q(n)b^n + r(n)c^n}$.

Exercise 1.1.51. Find the limits, $a, b, p, q > 0$.

1. $\sqrt[n]{an^p + b\sin n}$.
2. $\sqrt[n]{an^p + bn^q}$.
3. $\sqrt[n+2]{an^p + bn^q}$.
4. $\sqrt[n-2]{an^p + bn^q}$.
5. $\sqrt[n^2]{an^p + bn^q}$.
6. $(an^p + bn^q)^{\frac{1}{n^2-1}}$.

Example 1.1.16. The sequence $x_n = \dfrac{3^n(n!)^2}{(2n)!}$ satisfies

$$\lim_{n\to\infty} \frac{x_n}{x_{n-1}} = \lim_{n\to\infty} \frac{3n^2}{2n(2n-1)} = \frac{3}{4} = 0.75.$$

By the order rule, we have $\dfrac{x_n}{x_{n-1}} < 0.8$ for sufficiently big n, say for $n > N$ (in fact, $N = 8$ is enough). Then for $n > N$, we have

$$0 < x_n = \frac{x_n}{x_{n-1}} \frac{x_{n-1}}{x_{n-2}} \cdots \frac{x_{N+1}}{x_N} x_N < 0.8^{n-N} x_N = C \cdot 0.8^n, \quad C = 0.8^{-N} x_N.$$

By Example 1.1.11, we have $\lim_{n\to\infty} 0.8^n = 0$. Since C is a constant, by the sandwich rule, we get $\lim_{n\to\infty} x_n = 0$.

Exercises 1.1.52 and 1.1.53 summarise the idea of the example.

Exercise 1.1.52. Prove that if $\left|\dfrac{x_n}{x_{n-1}}\right| \le c$ for a constant $c < 1$, then x_n converges to 0.

Exercise 1.1.53. Prove that if $\lim_{n\to\infty} \dfrac{x_n}{x_{n-1}} = l$ and $|l| < 1$, then x_n converges to 0.

Exercise 1.1.54. Find a such that the sequence converges to 0, $p, q > 0$.

1. $\dfrac{(2n)!}{(n!)^2} a^n$.
2. $\dfrac{(n!)^2}{(3n)!} a^n$.
3. $\dfrac{(n!)^3}{(3n)!} a^n$.
4. $\dfrac{\sqrt{(2n)!}}{n!} a^n$.
5. $\sqrt{n!} a^{n^2}$.
6. $\dfrac{a^{n^2}}{n!}$.
7. $\dfrac{a^{n^2}}{\sqrt{n!}}$.
8. $(n!)^p a^n$.
9. $\dfrac{a^n}{(n!)^p}$.
10. $\dfrac{n^q a^n}{(n!)^p}$.
11. $\dfrac{(n!)^p}{((2n)!)^q} a^n$.
12. $\dfrac{n^5 (n!)^p}{((2n)!)^q} a^n$.

1.1.5 Subsequence

A *subsequence* is obtained by choosing infinitely many terms from a sequence. We denote a subsequence by

$$x_{n_k}\colon x_{n_1},\ x_{n_2},\ \dots,\ x_{n_k},\ \dots,$$

where the indices satisfy

$$n_1 < n_2 < \cdots < n_k < \cdots.$$

The following are some examples

$$\begin{aligned}
x_{2k}&\colon x_2,\ x_4,\ x_6,\ x_8,\ \dots, x_{2k},\ \dots,\\
x_{2k-1}&\colon x_1,\ x_3,\ x_5,\ x_7,\ \dots, x_{2k-1},\ \dots,\\
x_{2^k}&\colon x_2,\ x_4,\ x_8,\ x_{16},\ \dots, x_{2^k},\ \dots,\\
x_{k!}&\colon x_1,\ x_2,\ x_6,\ x_{24},\ \dots, x_{k!},\ \dots.
\end{aligned}$$

If x_n starts at $n = 1$, then $n_1 \geq 1$, which further implies $n_k \geq k$ for all k.

Proposition 1.1.6. *If a sequence converges to l, then any subsequence converges to l. Conversely, if a sequence is the union of finitely many subsequences that all converge to the same limit l, then the whole sequence converges to l.*

Example 1.1.17. Since $\dfrac{1}{n^2}$ is a subsequence of $\dfrac{1}{n}$, $\lim_{n\to\infty} \dfrac{1}{n} = 0$ implies $\lim_{n\to\infty} \dfrac{1}{n^2} = 0$. We also know $\lim_{n\to\infty} \dfrac{1}{\sqrt{n}} = 0$ implies $\lim_{n\to\infty} \dfrac{1}{n} = 0$ but not vice versa.

Example 1.1.18. The sequence $\dfrac{n + (-1)^n 3}{n - (-1)^n 2}$ is the union of the odd subsequence $\dfrac{(2k-1) - 3}{(2k-1) + 2} = \dfrac{2k-4}{2k+1}$ and the even subsequence $\dfrac{2k+3}{2k-2}$. Both subsequences converge to 1, either by direct computation, or by regarding them also as subsequences of $\dfrac{n-4}{n+1}$ and $\dfrac{n+3}{n-2}$, which converge to 1. Then we conclude $\lim_{n\to\infty} \dfrac{n + (-1)^n 3}{n - (-1)^n 2} = 1$.

Example 1.1.19. The sequence $(-1)^n$ has one subsequence $(-1)^{2k} = 1$ converging to 1 and another subsequence $(-1)^{2k-1} = -1$ converging to -1. Since the two limits are different, by Proposition 1.1.6, the sequence $(-1)^n$ diverges.

Example 1.1.20. The sequence $\sin na$ converges to 0 when a is an integer multiple of π. Now assume $0 < a < \pi$. For any natural number k, the interval $[k\pi, (k+1)\pi]$ of length π contains the following interval of length a (both intervals have the same middle point)

$$[a_k, b_k] = \left[k\pi + \frac{\pi - a}{2}, (k+1)\pi - \frac{\pi - a}{2}\right].$$

For even k, we have $\sin x \geq \sin\left(\dfrac{\pi - a}{2}\right) = \cos\dfrac{a}{2} > 0$ on $[a_k, b_k]$. For odd k, we have $\sin x \leq -\cos\dfrac{a}{2} < 0$ on $[a_k, b_k]$.

Since the arithmetic sequence $a, 2a, 3a, \dots$ has increment a, which is the length of $[a_k, b_k]$, we must have $n_k a \in [a_k, b_k]$ for some natural number n_k. Then $\sin n_{2k} a$ is a subsequence of $\sin na$ satisfying $\sin n_{2k} \geq \cos\dfrac{a}{2} > 0$, and $\sin n_{2k+1} a$ is a subsequence satisfying $\sin n_{2k} \leq -\cos\dfrac{a}{2}$. Therefore the two subsequences cannot converge to the same limit. As a result, the sequence $\sin na$ diverges.

Now for general a that is not an integer multiple of π, we have $a = 2N\pi \pm b$ for an integer N and b satisfying $0 < b < \pi$. Then we have $\sin na = \pm \sin nb$. We have shown that $\sin nb$ diverges, so that $\sin na$ diverges.

We conclude that $\sin na$ converges if and only if a is an integer multiple of π.

Exercise 1.1.55. Find the limit.

1. $\sqrt{n!+1}-\sqrt{n!-1}$.
2. $(n!)^{\frac{1}{n!}}$.
3. $((n+1)!)^{\frac{1}{n!}}$.
4. $((n+(-1)^n)!)^{\frac{1}{n!}}$.
5. $(n!)^{\frac{1}{(n+1)!}}$.
6. $(2^{n^2-1}+3^{n^2})^{\frac{1}{n^2}}$.

Exercise 1.1.56. Explain convergence or divergence.

1. $2^{(-1)^n}$.
2. $n^{\frac{(-1)^n}{n}}$.
3. $n^{(-1)^n}$.
4. $\dfrac{(-1)^n n+3}{n-(-1)^n 2}$.
5. $\dfrac{(-1)^n n^2}{n^3-1}$.
6. $\sqrt{n}\left(\sqrt{n+(-1)^n}-\sqrt{n}\right)$.
7. $(2^{(-1)^n n}+3^n)^{\frac{1}{n}}$.
8. $(2^n+3^{(-1)^n n})^{\frac{1}{n}}$.
9. $\tan\dfrac{n\pi}{3}$.
10. $(-1)^n\sin\dfrac{n\pi}{3}$.
11. $\sin\dfrac{n\pi}{2}\cos\dfrac{n\pi}{3}$.
12. $\dfrac{n\sin\frac{n\pi}{3}}{n\cos\frac{n\pi}{2}+2}$.
13. $\dfrac{n-\sin\frac{n\pi}{3}}{n+2\cos\frac{n\pi}{2}}$.

Exercise 1.1.57. Find all a such that the sequence $\cos na$ converges.

Example 1.1.21. We prove that $\lim_{n\to\infty} x_n = 1$ implies $\lim_{n\to\infty} x_n^p = 1$. Exercises 1.1.58 and 1.1.59 extend the result.

The sequence x_n is the union of two subsequences x'_k and x''_k (short for x_{m_k} and x_{n_k}) satisfying all $x'_k \ge 1$ and all $x''_k \le 1$. By Proposition 1.1.6, the assumption $\lim_{n\to\infty} x_n = 1$ implies that $\lim_{k\to\infty} x'_k = \lim_{k\to\infty} x''_k = 1$.

Pick integers M and N satisfying $M < p < N$. Then $x'_k \ge 1$ implies ${x'_k}^M \le {x'_k}^p \le {x'_k}^N$. By the arithmetic rule, we have $\lim_{k\to\infty} {x'_k}^M = (\lim_{k\to\infty} x'_k)^M = 1^M = 1$ and similarly $\lim_{k\to\infty} {x'_k}^N = 1$. Then by the sandwich rule, we get $\lim_{k\to\infty} {x'_k}^p = 1$.

Similar proof shows that $\lim_{k\to\infty} x''^p_k = 1$. Since the sequence x_n^p is the union of two subsequences ${x'_k}^p$ and x''^p_k, by Proposition 1.1.6 again, we get $\lim_{n\to\infty} x_n^p = 1$.

Exercise 1.1.58. Suppose $\lim_{n\to\infty} x_n = 1$ and y_n is bounded. Prove that $\lim_{n\to\infty} x_n^{y_n} = 1$.

Exercise 1.1.59. Suppose $\lim_{n\to\infty} x_n = l > 0$. By applying Example 1.1.21 to the sequence $\dfrac{x_n}{l}$, prove that $\lim_{n\to\infty} x_n^p = l^p$.

1.2 Rigorous Definition of Sequence Limit

The statement $\lim_{n\to\infty}\dfrac{1}{n} = 0$ means that $\dfrac{1}{n}$ gets smaller and smaller as n gets bigger and bigger. To make the statement rigorous, we need to be more specific about smaller and bigger.

Is 1000 big? The answer depends on the context. A village of 1000 people is big, and a city of 1000 people is small (even tiny). Similarly, a rope of diameter less than one millimeter is considered thin. But the hair is considered thin only if the diameter is less than 0.05 millimeter.

So big or small makes sense only when compared with some reference quantity. We say n is in the thousands if $n > 1000$ and in the millions if $n > 10000000$. The reference quantities 1000 and 1000000 give a sense of the scale of bigness. In this spirit, the statement $\lim_{n\to\infty} \dfrac{1}{n} = 0$ means the following list of infinitely many implications

$$\begin{aligned}
n > 1 &\implies \left|\frac{1}{n} - 0\right| < 1, \\
n > 10 &\implies \left|\frac{1}{n} - 0\right| < 0.1, \\
n > 100 &\implies \left|\frac{1}{n} - 0\right| < 0.01, \\
&\vdots \\
n > 1000000 &\implies \left|\frac{1}{n} - 0\right| < 0.000001, \\
&\vdots
\end{aligned}$$

For another example, $\lim_{n\to\infty} \dfrac{2^n}{n!} = 0$ means the following implications

$$\begin{aligned}
n > 10 &\implies \left|\frac{2^n}{n!} - 0\right| < 0.0003, \\
n > 20 &\implies \left|\frac{2^n}{n!} - 0\right| < 0.0000000000005, \\
&\vdots
\end{aligned}$$

So the general shape of the implications is

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

Note that the relation between N (measuring the bigness of n) and ϵ (measuring the smallness of $|x_n - l|$) may be different for different limits.

The problem with infinitely many implications is that our language is finite. In practice, we cannot verify all the implications one by one. Even if we have verified the truth of the first one million implications, there is no guarantee that the one million and the first implication is true. To mathematically establish the truth of all implications, we have to formulate one finite statement that includes the consideration for *all* N and *all* ϵ.

1.2.1 Rigorous Definition

Definition 1.2.1 (Rigorous)**.** A sequence x_n converges to a finite number l, and denoted $\lim_{n\to\infty} x_n = l$, if for any $\epsilon > 0$, there is N, such that $n > N$ implies $|x_n - l| < \epsilon$.

In case N is a natural number (which can always be arranged if needed), the definition means that, for any given horizontal ϵ-band around l, we can find N, such that all the terms $x_{N+1}, x_{N+2}, x_{N+3}, \dots$ after N lie in the shaded area in Figure 1.2.1.

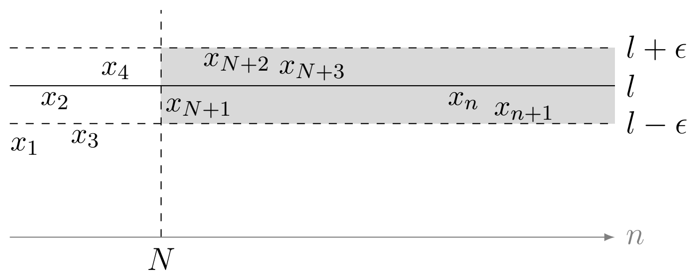


Figure 1.2.1: $n > N$ implies $|x_n - l| < \epsilon$.

Example 1.2.1. For any $\epsilon > 0$, choose $N = \dfrac{1}{\epsilon}$. Then

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

This verifies the rigorous definition of $\lim_{n\to\infty} \dfrac{1}{n} = 0$.

By applying the rigorous definition to $\epsilon = 0.1, 0.01, \dots$, we recover the infinitely many implications we wish to achieve. This justifies the rigorous definition of limit.

Example 1.2.2. For the constant sequence $x_n = c$, we rigorously prove

$$\lim_{n\to\infty} c = c.$$

For any $\epsilon > 0$, choose $N = 0$. Then

$$n > 0 \implies |x_n - c| = |c - c| = 0 < \epsilon.$$

In fact, the right side is always true, regardless of the left side.

Example 1.2.3. We rigorously prove

$$\lim_{n\to\infty} \frac{1}{n^p} = 0, \text{ for } p > 0.$$

For any $\epsilon > 0$, choose $N = \dfrac{1}{\epsilon^{\frac{1}{p}}}$. Then

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

We need to be more specific on the logical foundation for the arguments. We will assume the basic knowledge of real numbers, which are the four arithmetic operations $x+y, x-y, xy, \dfrac{x}{y}$, the exponential operation x^y (for $x > 0$), the order $x < y$ (or $y > x$, and $x \le y$ means $x < y$ or $x = y$), and the properties for these operations. For example, we assume that we already know $x > y > 0$ implies $\dfrac{1}{x} < \dfrac{1}{y}$ and $x^p > y^p$ for $p > 0$. These properties are used in the example above.

More important about the knowledge assumed above is the knowledge that are not assumed and therefore cannot be used until after the knowledge is established. In particular, we do not assume any knowledge about the logarithm. The logarithm and its properties will be rigorously established in Example 1.7.15 as the inverse of exponential.

Example 1.2.4. To rigorously prove $\lim_{n\to\infty} \dfrac{n^2-1}{n^2+1} = 1$, for any $\epsilon > 0$, we have

$$n > N = \sqrt{\frac{2}{\epsilon} - 1} \implies \left|\frac{n^2-1}{n^2+1} - 1\right| = \frac{2}{n^2+1} < \frac{2}{N^2+1} = \frac{2}{\left(\frac{2}{\epsilon} - 1\right) + 1} = \epsilon.$$

Therefore the sequence converges to 1.

How did we choose $N = \sqrt{\dfrac{2}{\epsilon} - 1}$? We want to achieve $\left|\dfrac{n^2-1}{n^2+1} - 1\right| < \epsilon$. Since this is equivalent to $\dfrac{2}{n^2+1} < \epsilon$, which we can solve to get $n > \sqrt{\dfrac{2}{\epsilon} - 1}$, choosing $N = \sqrt{\dfrac{2}{\epsilon} - 1}$ should work.

Example 1.2.5. To rigorously prove the limit in Example 1.1.5, we estimate the difference between the sequence and the expected limit

$$|\sqrt{n+2} - \sqrt{n} - 0| = \frac{(n+2) - n}{\sqrt{n+2} + \sqrt{n}} < \frac{2}{\sqrt{n}}.$$

This shows that for any $\epsilon > 0$, it is sufficient to have $\dfrac{2}{\sqrt{n}} < \epsilon$, or $n > \dfrac{4}{\epsilon^2}$. In other words, we should choose $N = \dfrac{4}{\epsilon^2}$.

The discussion above is the analysis of the problem, which you may write on your scratch paper. The formal rigorous argument you are supposed to present is the following: For any $\epsilon > 0$, choose $N = \dfrac{4}{\epsilon^2}$. Then

$$n > N \implies |\sqrt{n+2} - \sqrt{n} - 0| = \frac{2}{\sqrt{n+2}+\sqrt{n}} < \frac{2}{\sqrt{n}} < \frac{2}{\sqrt{N}} = \epsilon.$$

1.2.2 The Art of Estimation

In Examples 1.2.4, the formula for N is obtained by solving $\left|\dfrac{n^2-1}{n^2+1} - 1\right| < \epsilon$ in exact way. However, this may not be so easy in general. For example, for the limit in Example 1.1.2, we need to solve

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

While the exact solution can be found, the formula for N is rather complicated. For more complicated example, it may not even be possible to find the formula for the exact solution.

We note that finding the exact solution of $|x_n - l| < \epsilon$ is the same as finding $N = N(\epsilon)$, such that

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

However, in order to rigorously prove the limit, only $\implies$ direction is needed. The weaker goal can often be achieved in much simpler way.

Example 1.2.6. Consider the limit in Example 1.1.2. For $n > 1$, we have

$$\left|\frac{2n^2+n}{n^2-n+1} - 2\right| = \frac{3n-2}{n^2-n+1} < \frac{3n}{n^2-n} = \frac{3}{n-1}.$$

Since $\dfrac{3}{n-1} < \epsilon$ implies $\left|\dfrac{2n^2+n}{n^2-n+1} - 2\right| < \epsilon$, and $\dfrac{3}{n-1} < \epsilon$ is equivalent to $n > \dfrac{3}{\epsilon} + 1$, we find that choosing $N = \dfrac{3}{\epsilon} + 1$ is sufficient

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

Exercise 1.2.1. Show that $\left|\dfrac{n^2-1}{n^2+1} - 1\right| < \dfrac{2}{n}$ and then rigorously prove $\lim_{n\to\infty} \dfrac{n^2-1}{n^2+1} = 1$.

The key for the rigorous proof of limits is to find a simple and good enough estimation. We emphasize that there is no need to find the best estimation. Any estimation that can fulfill the rigorous definition of limit is good enough.

Everyday life is full of good enough estimations. Mastering the art of such estimations is very useful for not just learning calculus, but also for making smart judgement in real life.

Example 1.2.7. If a bottle is 20% bigger in size than another bottle, how much bigger is in volume?

The exact formula is the cube of the comparison in size

$$(1+0.2)^3 = 1 + 3\cdot 0.2 + 3\cdot 0.2^2 + 0.2^3.$$

Since $3\cdot 0.2 = 0.6$, $3\cdot 0.2^2 = 0.12$, and 0.2^3 is much smaller than 0.1, the bottle is a little more than 72% bigger in volume.

Example 1.2.8. The 2013 GDP per capita is 9,800USD for China and 53,100USD for the United States, in terms of PPP (purchasing power parity). The percentage of the annual GDP growth for the three years up to 2013 are 9.3, 7.7, 7.7 for China and 1.8, 2.8, 1.9 for the United States. What do we expect the number of years for China to catch up to the United States?

First we need to estimate how much faster is the Chinese GDP growing compared to the United States. The comparison for 2013 is

$$\frac{1+0.077}{1+0.019} \approx 1 + (0.077 - 0.019) = 1 + 0.058.$$

Similarly, we get the (approximate) comparisons $1 + 0.075$ and $1 + 0.049$ for the other two years. Among the three comparisons, we may choose a more conservative $1+0.05$. This means that we assume Chinese GDP per capita grows 5% faster than the United States for the next many years.

Based on the assumption of 5%, the number of years n for China to catch up to the United States is obtained exactly by solving

$$(1+0.05)^n = 1 + n\,0.05 + \frac{n(n-1)}{2}0.05^2 + \cdots + 0.05^n = \frac{53,100}{9,800} \approx 5.5.$$

If we use $1 + n\,0.05$ to approximate $(1+0.05)^n$, then we get $n \approx \dfrac{5.5-1}{0.05} = 90$. However, 90 years is too pessimistic because for $n = 90$, the third term $\dfrac{n(n-1)}{2}0.05^2$ is quite sizable, so that $1 + n\,0.05$ is not a good approximation of $(1+0.05)^n$.

An an exercise for the art of estimation, we try to avoid using calculator in getting better estimation. By $\dfrac{53,100}{9,800} \approx 2.3^2$, we may solve

$$n = 2m, \quad (1+0.05)^m = 1 + m\,0.05 + \frac{m(m-1)}{2}0.05^2 + \cdots + 0.05^m \approx 2.3.$$

We get $n \approx 2 \cdot \dfrac{2.3-1}{0.05} = 52$. Since $\dfrac{m(m-1)}{2}0.05^2$ is still sizable for $m = 26$ (but giving much better approximation than $n = 90$), the actual n should be somewhat smaller than 52. We try $n = 40$ and estimate $(1+0.05)^n$ by the first three terms

$$(1+0.05)^{40} \approx 1 + 40 \cdot 0.05 + \frac{40 \cdot 39}{2}0.05^2 \approx 5.$$

So it looks like somewhere between 40 and 45 is a good estimation.

We conclude that, if Chinese GDP per capita growth is 5% (a very optimistic assumption) faster than the United States in the next 50 years, then China will catch up to the United States in 40 some years.

Exercise 1.2.2. I wish to paint a wall measuring 3 meters tall and 6 meters wide, give or take 10% in each direction. If the cost of paint is \$13.5 per square meters, how much should I pay for the paint?

Exercise 1.2.3. In a supermarket, I bought four items at \$5.95, \$6.35, \$15.50, \$7.20. The sales tax is 8%. The final bill is around \$38. Is the bill correct?

Exercise 1.2.4. In 1900, Argentina and Canada had the same GDP per capita. In 2000, the GDP per capita is 9,300USD for Argentina and 24,000USD for Canada. On average, how much faster is Canadian GDP growing annually compared with Argentina in the 20th century?

Next we leave real life estimations and try some examples in calculus.

Example 1.2.9. If x is close to 3 and y is close to 5, then $2x-3y$ is close to $2\cdot 3-3\cdot 5 = -9$. We wish to be more precise about the statement, say, we want to find a tolerance for x and y, such that $2x-3y$ is within ± 0.2 of -9.

We have

$$|(2x-3y)-(-9)| = |2(x-3)-3(y-5)| \le 2|x-3| + 3|y-5|.$$

For the difference to be within ± 0.2, we only need to make sure $2|x-3|+3|y-5| < 0.2$. This can be easily achieved by $|x-3| < \frac{0.2}{2+3} = 0.04$ and $|y-5| < 0.04$.

Example 1.2.10. Again we assume x and y are close to 3 and 5. Now we want to find the percentage of tolerance, such that $2x-3y$ is within ± 0.2 of -9.

We can certainly use the answer in Example 1.2.9 and find the percentage $\frac{0.04}{3} \approx 1.33\%$ for x and $\frac{0.04}{5} \approx 0.8\%$ for y. This implies that, if both x and y are within 0.8% of 3 and 5, then $2x-3y$ is within ± 0.2 of -9.

The better (or more honest) way is to directly solve the problem. Let δ_1 and δ_2 be the percentage of tolerance for x and y. Then $x = 3(1+\delta_1)$ and $y = 5(1+\delta_2)$, and

$$|(2x-3y)-(-9)| = |2(x-3)-3(y-5)| \le |2\cdot 3\delta_1 - 3\cdot 5\delta_2| \le 21\delta, \quad \delta = \max\{|\delta_1|, |\delta_2|\}.$$

To get $21|\delta|$ to be within our target of 0.2, we may take our tolerance $\delta = 0.9\% < \frac{0.2}{21} \approx 0.0095$.

Example 1.2.11. Assume x and y are close to 3 and 5. We want to find the tolerance for x and y, such that xy is within ± 0.2 of $3 \cdot 5 = 15$. This means finding $\delta > 0$, such that

$$|x-3| < \delta,\ |y-5| < \delta \implies |xy-15| < 0.2.$$

Under the assumptions $|x-3| < \delta$ and $|y-5| < \delta$, we have

$$|xy-15| \le |xy-3y| + |3y-15| \le |x-3||y| + 3|y-5| \le (|y|+3)\delta.$$

We also note that, if we postulate $\delta \le 1$, then $|y-5| < \delta$ implies $4 < y < 6$, so that

$$|xy-15| \le (|y|+3)\delta \le (6+3)\delta = 9\delta.$$

To get $9|\delta|$ to be within our target of 0.2, we may take our tolerance $\delta = 0.02 < \frac{0.2}{9}$. Since this indeed satisfies $\delta \le 1$, we conclude that we can take $\delta = 0.02$.

If the targeted error ± 0.2 is changed to some other amount $\pm\epsilon$, then the same argument shows that we can take the tolerance to be $\delta = \frac{\epsilon}{10}$. Strictly speaking, since we also use $\delta \le 1$ in the argument above, we should take $\delta = \min\{\frac{\epsilon}{10}, 1\}$.

Exercise 1.2.5. Find a tolerance for x, y, z near $-2, 3, 5$, such that $5x - 3y + 4z$ is within $\pm\epsilon$ of 1.

Exercise 1.2.6. Find a tolerance for x and y near 2 and 2, such that xy is within $\pm\epsilon$ of 4.

Exercise 1.2.7. Find a tolerance for x near 2, such that x^2 is within $\pm\epsilon$ of 4.

Exercise 1.2.8. Find a percentage of tolerance for x near 2, such that $\dfrac{1}{x}$ is within ± 0.1 of 0.5.

1.2.3 Rigorous Proof of Limits

We revisit the limits derived before and make the argument rigorous.

Example 1.2.12. In Example 1.1.5, we argued that $\lim_{n\to\infty}(\sqrt{n+a} - \sqrt{n}) = 0$. To make the argument rigorous, we use the estimation in the earlier example. In fact, regardless of the sign of a, we always have

$$|\sqrt{n+a} - \sqrt{n} - 0| = \left|\frac{(\sqrt{n+a} - \sqrt{n})(\sqrt{n+a} + \sqrt{n})}{\sqrt{n+a} + \sqrt{n}}\right| = \frac{|a|}{\sqrt{n+a} + \sqrt{n}} < \frac{|a|}{\sqrt{n}}.$$

For the right side $\dfrac{|a|}{\sqrt{n}} < \epsilon$, it is sufficient to have $n > \dfrac{a^2}{\epsilon^2}$. Then we can easily get

$$n > \frac{a^2}{\epsilon^2} \implies |\sqrt{n+a} - \sqrt{n} - 0| < \epsilon.$$

This gives the rigorous proof of $\lim_{n\to\infty}(\sqrt{n+a}-\sqrt{n})=0$.

Example 1.2.13. In Example 1.1.6, we argued that $\lim_{n\to\infty}\sqrt{\dfrac{n+2}{n}}=1$. To make the argument rigorous, we use the estimation in the earlier example. The estimation suggested that it is sufficient to have $2\dfrac{1}{n}<\epsilon$. Thus we get the following rigorous argument for the limit

$$n<\frac{2}{\epsilon} \implies 0<\sqrt{\frac{n+2}{n}}-1<\frac{n+2}{n}-1=\frac{2}{n}<\epsilon \implies \left|\sqrt{\frac{n+2}{n}}-1\right|<\epsilon.$$

Exercise 1.2.9. Rigorously prove the limits.

1. $\dfrac{n+2}{n-3}$.
2. $\dfrac{n-2}{n+3}$.
3. $\dfrac{n+a}{n+b}$.
4. $\dfrac{2n^2-3n+2}{3n^2-4n+1}$.
5. $\dfrac{\sqrt{n}+2}{\sqrt{n}-3}$.
6. $\dfrac{\sqrt{n}+a}{n+b}$.
7. $\dfrac{n}{n+1}-\dfrac{n}{n-1}$.
8. $\dfrac{n+a}{n+b}-\dfrac{n+c}{n+d}$.
9. $\dfrac{1}{\sqrt{an+b}}$.
10. $\dfrac{\sin\sqrt{n}}{n}$.
11. $\dfrac{\cos\sqrt{n}+a}{n+b\sin n}$.
12. $\sqrt{n+a}-\sqrt{n+b}$.
13. $\sqrt{\dfrac{n+a}{n+b}}$.
14. $\sqrt[3]{n+1}-\sqrt[3]{n}$.

Exercise 1.2.10. Rigorously prove the limits, $p>0$.

1. $\dfrac{a}{\sqrt{n^p+b}}$.
2. $\dfrac{n^p+a}{n^p+b}$.
3. $\dfrac{a\sin n+b}{n^p+c}$.

Example 1.2.14. The estimation in Example 1.1.7 tells us that $|\sqrt[n]{a}-1|<\dfrac{a}{n}$ for $a>1$. This suggests that for any $\epsilon>0$, we may choose $N=\dfrac{a}{\epsilon}$. Then

$$n>N \implies |\sqrt[n]{a}-1|<\frac{a}{n}<\frac{a}{N}=\epsilon.$$

This rigorously proves that $\lim_{n\to\infty}\sqrt[n]{a}=1$ in case $a\ge 1$.

Example 1.2.15. We try to rigorously prove $\lim_{n\to\infty}n^2a^n=0$ for $|a|<1$.

Using the idea of Example 1.1.11, we write $|a|=\dfrac{1}{1+b}$. Then $|a|<1$ implies

$b > 0$, and for $n \geq 3$, we have

$$\begin{aligned}
|n^2a^n - 0| = n^2|a|^n &= \frac{n^2}{(1+b)^n} \\
&= \frac{n^2}{1 + nb + \dfrac{n(n-1)}{2!}b^2 + \dfrac{n(n-1)(n-2)}{3!}b^3 + \cdots + b^n} \\
&< \frac{n^2}{\dfrac{n(n-1)(n-2)}{3!}b^3} = \frac{3!n}{(n-1)(n-2)b^3} < \frac{3!n}{\dfrac{n}{2}\dfrac{n}{2}b^3} = \frac{3!2^2}{nb^3}.
\end{aligned}$$

Since $\dfrac{3!2^2}{nb^3} < \epsilon$ is the same as $n > \dfrac{3!2^2}{b^3\epsilon}$, we have

$$n > \frac{3!2^2}{b^3\epsilon} \text{ and } n \geq 3 \implies |n^2a^n - 0| < \frac{3!2^2}{nb^3} < \epsilon.$$

This shows that we may choose $N = \max\left\{\dfrac{3!2^2}{b^3\epsilon}, 3\right\}$.

It is clear from the proof that we generally have

$$\lim_{n\to\infty} n^pa^n = 0, \text{ for any } p \text{ and } |a| < 1.$$

Example 1.2.16. We rigorously prove $\lim_{n\to\infty} \dfrac{a^n}{n!} = 0$ in Example 1.1.13.

Choose a natural number M satisfying $|a| < M$. Then for $n > M$, we have

$$\left|\frac{a^n}{n!}\right| < \frac{M^n}{n!} = \frac{M \cdot M \cdots M}{1 \cdot 2 \cdots M} \cdot \frac{M}{M+1} \cdot \frac{M}{M+2} \cdots \frac{M}{n} \leq \frac{M^M}{M!} \cdot \frac{M}{n} = \frac{M^{M+1}}{M!} \cdot \frac{1}{n}.$$

Therefore for any $\epsilon > 0$, we have

$$n > \max\left\{\frac{M^{M+1}}{M!\epsilon}, M\right\} \implies \left|\frac{a^n}{n!} - 0\right| < \frac{M^{M+1}}{M!} \cdot \frac{1}{n} < \frac{M^{M+1}}{M!} \cdot \frac{1}{\dfrac{M^{M+1}}{M!\epsilon}} = \epsilon.$$

Exercise 1.2.11. Rigorously prove the limits.

1. $\sqrt[n]{n}$.
2. $\dfrac{n^{5.4}}{n!}$.
3. $\dfrac{n^p}{n!}$.
4. $\dfrac{n^{5.4}3^n}{n!}$.
5. $\dfrac{n^pa^n}{n!}$.
6. $\dfrac{n!}{n^n}$.
7. n^pa^n, $|a| < 1$.

1.2.4 Rigorous Proof of Limit Properties

The rigorous definition of limit allows us to rigorously prove some limit properties.

Example 1.2.17. Suppose $\lim_{n\to\infty} x_n = l > 0$. We prove that $\lim_{n\to\infty} \sqrt{x_n} = \sqrt{l}$.

First we clarify the problem. The limit $\lim_{n\to\infty} x_n = l$ means the implication

$$\text{For any } \epsilon > 0, \text{ there is } N, \text{ such that } n > N \implies |x_n - l| < \epsilon.$$

The limit $\lim_{n\to\infty} \sqrt{x_n} = \sqrt{l}$ means the implication

$$\text{For any } \epsilon > 0, \text{ there is } N, \text{ such that } n > N \implies |\sqrt{x_n} - \sqrt{l}| < \epsilon.$$

We need to argue is that the first implication implies the second implication.

We have

$$|\sqrt{x_n} - \sqrt{l}| = \frac{|(\sqrt{x_n} - \sqrt{l})(\sqrt{x_n} + \sqrt{l})|}{\sqrt{x_n} + \sqrt{l}} = \frac{|x_n - l|}{\sqrt{x_n} + \sqrt{l}} \le \frac{|x_n - l|}{\sqrt{l}}.$$

Therefore for any given $\epsilon > 0$, the second implication will hold as long as $\dfrac{|x_n - l|}{\sqrt{l}} < \epsilon$, or $|x_n - l| < \sqrt{l}\epsilon$. The inequality $|x_n - l| < \sqrt{l}\epsilon$ can be achieved from the first implication, *provided* we apply the first implication to $\sqrt{l}\epsilon$ in place of ϵ.

The analysis above leads to the following formal proof. Let $\epsilon > 0$. By applying the definition of $\lim_{n\to\infty} x_n = l$ to $\sqrt{l}\epsilon > 0$, there is N, such that

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

Then

$$\begin{aligned} n > N &\implies |x_n - l| < \sqrt{l}\epsilon \\ &\implies |\sqrt{x_n} - \sqrt{l}| = \frac{|(\sqrt{x_n} - \sqrt{l})(\sqrt{x_n} + \sqrt{l})|}{\sqrt{x_n} + \sqrt{l}} = \frac{|x_n - l|}{\sqrt{x_n} + \sqrt{l}} \le \frac{|x_n - l|}{\sqrt{l}} < \epsilon. \end{aligned}$$

In the argument, we take advantage of the fact that the definition of limit can be applied to *any* positive number, $\sqrt{l}\epsilon$ for example, instead of the given positive number ϵ.

Example 1.2.18. We prove the arithmetic rule $\lim_{n\to\infty}(x_n + y_n) = \lim_{n\to\infty} x_n + \lim_{n\to\infty} y_n$ in Proposition 1.1.3. The concrete Example 1.2.9 provides idea of the proof.

Let $\lim_{n\to\infty} x_n = l$ and $\lim_{n\to\infty} y_n = k$. Then for any $\epsilon_1 > 0$, $\epsilon_2 > 0$, there are N_1, N_2, such that

$$\begin{aligned} n > N_1 &\implies |x_n - l| < \epsilon_1, \\ n > N_2 &\implies |y_n - k| < \epsilon_2. \end{aligned}$$

We expect to choose ϵ_1, ϵ_2 as some modification of ϵ, as demonstrated in Example 1.2.17.

Let $N = \max\{N_1, N_2\}$. Then

$$\begin{aligned} n > N &\implies n > N_1,\ n > N_2 \\ &\implies |x_n - l| < \epsilon_1,\ |y_n - k| < \epsilon_2 \\ &\implies |(x_n + y_n) - (l + k)| \le |x_n - l| + |y_n - k| < \epsilon_1 + \epsilon_2. \end{aligned}$$

If $\epsilon_1 + \epsilon_2 \le \epsilon$, then this rigorously proves $\lim_{n\to\infty}(x_n + y_n) = l + k$. Of course this means that we may choose $\epsilon_1 = \epsilon_2 = \dfrac{\epsilon}{2}$ at the beginning of the argument.

The analysis above leads to the following formal proof. For any $\epsilon > 0$, apply the definition of $\lim_{n\to\infty} x_n = l$ and $\lim_{n\to\infty} y_n = k$ to $\dfrac{\epsilon}{2} > 0$. We find N_1 and N_2, such that

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

Then

$$\begin{aligned} n > \max\{N_1, N_2\} &\implies |x_n - l| < \frac{\epsilon}{2},\ |y_n - k| < \frac{\epsilon}{2} \\ &\implies |(x_n + y_n) - (l + k)| \le |x_n - l| + |y_n - k| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \end{aligned}$$

Example 1.2.19. The arithmetic rule $\lim_{n\to\infty} x_n y_n = \lim_{n\to\infty} x_n \ \lim_{n\to\infty} y_n$ in Proposition 1.1.3 means that, if we know the approximate values of the width and height of a rectangle, then multiplying the width and height approximates the area of the rectangle. The rigorous proof requires us to estimate how the approximation of the area is affected by the approximations of the width and height. Example 1.2.11 gives the key idea for such estimation.

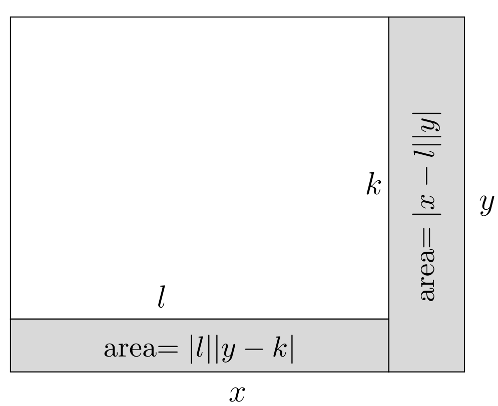


Figure 1.2.2: The error in product.

Let $\lim_{n\to\infty} x_n = l$ and $\lim_{n\to\infty} y_n = k$. Then for any $\epsilon_1 > 0$, $\epsilon_2 > 0$, there are N_1, N_2, such that

$$\begin{aligned} n > N_1 &\implies |x_n - l| < \epsilon_1, \\ n > N_2 &\implies |y_n - k| < \epsilon_2. \end{aligned}$$

Then for $n > N = \max\{N_1, N_2\}$, we have (see Figure 1.2.2)

$$\begin{aligned} |x_n y_n - lk| &= |(x_n - l)y_n + l(y_n - k)| \\ &\le |x_n - l||y_n| + |l||y_n - k| \\ &< \epsilon_1(|k| + \epsilon_2) + |l|\epsilon_2, \end{aligned}$$

where we use $|y_n - k| < \epsilon_2$ implying $|y_n| < |k| + \epsilon_2$. The proof of $\lim_{n\to\infty} x_n y_n = lk$ will be complete if, for any $\epsilon > 0$, we can choose $\epsilon_1 > 0$ and $\epsilon_2 > 0$, such that

$$\epsilon_1(|k| + \epsilon_2) + |l|\epsilon_2 \le \epsilon.$$

This can be achieved by choosing ϵ_1, ϵ_2 satisfying

$$\epsilon_2 \le 1, \quad \epsilon_1(|k| + 1) \le \frac{\epsilon}{2}, \quad |l|\epsilon_2 \le \frac{\epsilon}{2}.$$

In other words, if we choose

$$\epsilon_1 = \frac{\epsilon}{2(|k| + 1)}, \quad \epsilon_2 = \min\left\{1, \frac{\epsilon}{2|l|}\right\}$$

at the very beginning of the proof, then we get a rigorous proof of the arithmetic rule. The formal writing of the proof is left to the reader.

Example 1.2.20. The sandwich rule in Proposition 1.1.4 reflects the intuition that, if x and z are within ϵ of 5, then any number y between x and z is also within ϵ of 5

$$|x - 5| < \epsilon,\ |z - 5| < \epsilon,\ x \le y \le z \implies |y - 5| < \epsilon.$$

Geometrically, this means that if x and z lies inside an interval, say $(5 - \epsilon, 5 + \epsilon)$, then any number y between x and z also lies in the interval.

Suppose $x_n \le y_n \le z_n$ and $\lim_{n\to\infty} x_n = \lim_{n\to\infty} z_n = l$. For any $\epsilon > 0$, there are N_1 and N_2, such that

$$\begin{aligned} n > N_1 &\implies |x_n - l| < \epsilon, \\ n > N_2 &\implies |z_n - l| < \epsilon. \end{aligned}$$

Then

$$\begin{aligned} n > N = \max\{N_1, N_2\} &\implies |x_n - l| < \epsilon,\ |z_n - l| < \epsilon \\ &\implies l - \epsilon < x_n,\ z_n < l + \epsilon \\ &\implies l - \epsilon < x_n \le y_n \le z_n < l + \epsilon \\ &\iff |y_n - l| < \epsilon. \end{aligned}$$

Example 1.2.21. The order rule in Proposition 1.1.5 reflects the intuition that, if x is very close to 3 and y is very close to 5, then x must be less than y. More specifically, we know $x < y$ when x and y are within ± 1 of 3 and 5. Here 1 is half of the distance between 3 and 5.

Suppose $x_n \le y_n$, $\lim_{n\to\infty} x_n = l$, $\lim_{n\to\infty} y_n = k$. For any $\epsilon > 0$, there is N, such that (you should know from earlier examples how to find this N)

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

Picking any $n > N$, we get

$$l - \epsilon < x_n \le y_n < k + \epsilon.$$

Therefore we proved that $l - \epsilon < k + \epsilon$ for any $\epsilon > 0$. It is easy to see that the property is the same as $l \le k$.

Conversely, we assume $\lim_{n\to\infty} x_n = l$, $\lim_{n\to\infty} y_n = k$, and $l < k$. For any $\epsilon > 0$, there is N, such that $n > N$ implies $|x_n - l| < \epsilon$ and $|y_n - k| < \epsilon$. Then

$$n > N \implies x_n < l + \epsilon,\ y_n > k - \epsilon \implies y_n - x_n > (k - \epsilon) - (l + \epsilon) = k - l - 2\epsilon.$$

By choosing $\epsilon = \dfrac{k - l}{2} > 0$ at the beginning of the argument, we conclude that $y_n > x_n$ for $n > N$.

Exercise 1.2.12. Prove that if $\lim_{n\to\infty} x_n = l$, then $\lim_{n\to\infty} |x_n| = |l|$.

Exercise 1.2.13. Prove that $\lim_{n\to\infty} |x_n - l| = 0$ if and only if $\lim_{n\to\infty} x_n = l$.

Exercise 1.2.14. Prove that if $\lim_{n\to\infty} x_n = l$, then $\lim_{n\to\infty} cx_n = cl$.

Exercise 1.2.15. Prove that a sequence x_n converges if and only if the subsequences x_{2n} and x_{2n+1} converge to the same limit. This is a special case of Proposition 1.1.6.

Exercise 1.2.16. Suppose $x_n \ge 0$ for sufficiently big n and $\lim_{n\to\infty} x_n = 0$. Prove that $\lim_{n\to\infty} x_n^p = 0$ for any $p > 0$.

Exercise 1.2.17. Suppose $x_n \ge 0$ for sufficiently big n and $\lim_{n\to\infty} x_n = 0$. Suppose $y_n \ge c$ for sufficiently big n and some constant $c > 0$. Prove that $\lim_{n\to\infty} x_n^{y_n} = 0$.

1.3 Criterion for Convergence

Any number close to 3 must be between 2 and 4, and in particular have the absolute value no more than 4. The intuition leads to the following result.

Theorem 1.3.1. *If x_n converges, then $|x_n| \le B$ for a constant B and all n.*

The theorem basically says that any convergent sequence is bounded. The number B is a *bound* for the sequence.

If $x_n \le B$ for all n, then we say x_n is *bounded above*, and B is an *upper bound*. If $x_n \ge B$ for all n, then we say x_n is *bounded below*, and B is a *lower bound*. A sequence is bounded if and only if it is bounded above and bounded below.

The sequences n, $\dfrac{n^2+(-1)^n}{n+1}$ diverge because they are not bounded. On the other hand, the sequence $1, -1, 1, -1, \dots$ is bounded but diverges. Therefore the converse of Theorem 1.3.1 is not true in general.

Exercise 1.3.1. Prove that if x_n is bounded for sufficiently big n, i.e., $|x_n| \le B$ for $n \ge N$, then x_n is still bounded.

Exercise 1.3.2. Suppose x_n is the union of two subsequences x'_k and x''_k. Prove that x_n is bounded if and only if both x'_k and x''_k are bounded.

1.3.1 Monotone Sequence

The converse of Theorem 1.3.1 holds under some additional assumption. A sequence x_n is *increasing* if

$$x_1 \le x_2 \le x_3 \le \cdots \le x_n \le x_{n+1} \le \cdots .$$

It is *strictly increasing* if

$$x_1 < x_2 < x_3 < \cdots < x_n < x_{n+1} < \cdots .$$

The concepts of *decreasing* and *strictly decreasing* can be similarly defined. Moreover, a sequence is *monotone* if it is either increasing or decreasing.

The sequences $\dfrac{1}{n}$, $\dfrac{1}{2^n}$, $\sqrt[n]{2}$ are (strictly) decreasing. The sequences $-\dfrac{1}{n}$, n are increasing.

Theorem 1.3.2. *A monotone sequence converges if and only if it is bounded.*

An increasing sequence x_n is always bounded below by its first term x_1. Therefore x_n is bounded if and only if it is bounded above. Similarly, a decreasing sequence is bounded if and only if it is bounded below.

The world record for 100 meter dash is a decreasing sequence bounded below by 0. The proposition reflects the intuition that there is a limit on how fast human being can run. We note that the proposition does not tells us the exact value of the limit, just like we do not know the exact limit of the human ability.

Example 1.3.1. Consider the sequence

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

The sequence is clearly increasing. Moreover, the sequence is bounded above by

$$\begin{aligned} x_n &\leq 1 + \frac{1}{1\cdot 2} + \frac{1}{2\cdot 3} + \cdots + \frac{1}{(n-1)n} \\ &= 1 + \left(1 - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \cdots + \left(\frac{1}{n-1} - \frac{1}{n}\right) \\ &= 2 - \frac{1}{n} < 2. \end{aligned}$$

Therefore the sequence converges.

The limit of the sequence is the sum of the *infinite series*

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2} + \cdots.$$

We will see that the sum is actually $\dfrac{\pi^2}{6}$.

Exercise 1.3.3. Show the convergence of sequences.

1. $x_n = \dfrac{1}{1^3} + \dfrac{1}{2^3} + \dfrac{1}{3^3} + \cdots + \dfrac{1}{n^3}$.
2. $x_n = \dfrac{1}{1^{2.4}} + \dfrac{1}{2^{2.4}} + \dfrac{1}{3^{2.4}} + \cdots + \dfrac{1}{n^{2.4}}$.
3. $x_n = \dfrac{1}{1\cdot 3} + \dfrac{1}{3\cdot 5} + \dfrac{1}{5\cdot 7} + \cdots + \dfrac{1}{(2n-1)(2n+1)}$.
4. $x_n = \dfrac{1}{1!} + \dfrac{1}{2!} + \cdots + \dfrac{1}{n!}$.

Example 1.3.2. The number $\sqrt{2+\sqrt{2+\sqrt{2+\cdots}}}$ is the limit of the sequence x_n inductively given by

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

After trying first couple of terms, we expect the sequence to be increasing. This can be verified by induction. We have $x_2 = \sqrt{2+\sqrt{2}} > x_1 = \sqrt{2}$. Moreover, if we assume $x_n > x_{n-1}$, then

$$x_{n+1} = \sqrt{2+x_n} > \sqrt{2+x_{n-1}} = x_n.$$

This proves inductively that x_n is indeed increasing.

Next we claim that x_n is bounded above. For an increasing sequence, we expect its limit to be the upper bound. So we find the hypothetical limit first. Taking the limit on both sides of the equality $x_{n+1}^2 = 2 + x_n$ and applying the arithmetic rule, we get $l^2 = 2 + l$. The solution is $l = 2$ or -1. Since $x_n > 0$, by the order rule, we must have $l \geq 0$. Therefore we conclude that $l = 2$.

The hypothetical limit value suggests that $x_n < 2$ for all n. Again we verify this by induction. We already have $x_1 = \sqrt{2} < 2$. If we assume $x_n < 2$, then

$$x_{n+1} = \sqrt{2 + x_n} < \sqrt{2 + 2} = 2.$$

This proves inductively that $x_n < 2$ for all n.

We conclude that x_n is increasing and bounded above. By Theorem 1.3.2, the sequence converges, and the hypothetical limit value 2 is the real limit value.

Figure 1.3.1 suggests that our conclusion actually depends only on the general shape of the graph of the function, and has little to do with the exact formula $\sqrt{2 + x}$.

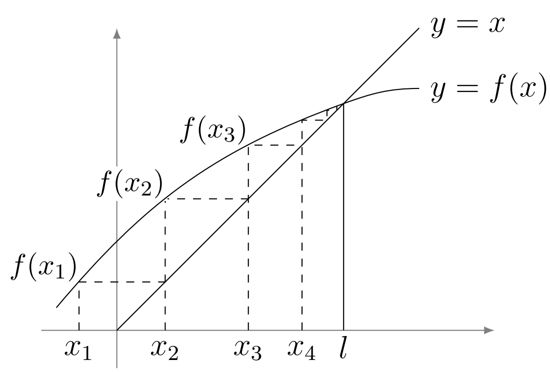


Figure 1.3.1: Limit of inductively defined sequence.

Exercise 1.3.4. Suppose a sequence x_n satisfies $x_{n+1} = \sqrt{2 + x_n}$.

1. Prove that if $-2 < x_1 < 2$, then x_n is increasing and converges to 2.
2. Prove that if $x_1 > 2$, then x_n is decreasing and converges to 2.

Exercise 1.3.5. For the three functions $f(x)$ in Figure 1.3.2, study the convergence of the sequences x_n defined by $x_{n+1} = f(x_n)$. Your answer depends on the initial value x_1.

Exercise 1.3.6. Suppose a sequence x_n satisfies $x_{n+1} = \dfrac{1}{2}(x_n^2 + x_n)$. Prove the following statements.

1. If $x_1 > 1$, then the sequence is increasing and diverges.
2. If $0 < x_1 < 1$, then the sequence is decreasing and converges to 0.
3. If $-1 < x_1 < 0$, then the sequence is increasing and converges to 0.
4. If $-2 < x_1 < -1$, then the sequence is decreasing for $n \geq 2$ and converges to 0.

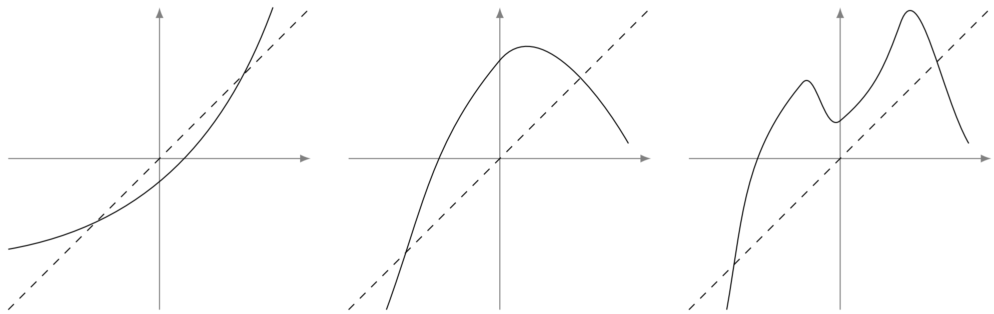


Figure 1.3.2: Three functions

5. If $x_1 < -2$, then the sequence is increasing for $n \geq 2$ and diverges.

Exercise 1.3.7. Determine the convergence of inductively defined sequences. Your answer may depend on the initial value x_1.

1. $x_{n+1} = x_n^2$.
2. $x_{n+1} = \dfrac{x_n^2 + 1}{2}$.
3. $x_{n+1} = 2x_n^2 - 1$.
4. $x_{n+1} = \dfrac{1}{x_n}$.
5. $x_{n+1} = 1 + \dfrac{1}{x_n}$.
6. $x_{n+1} = 2 - \dfrac{1}{x_n}$.

Exercise 1.3.8. Determine the convergence of inductively defined sequences, $a > 0$. In some cases, the sequence may not be defined after certain number of terms.

1. $x_{n+1} = \sqrt{a + x_n}$.
2. $x_{n+1} = \sqrt{x_n - a}$.
3. $x_{n+1} = \sqrt{a - x_n}$.
4. $x_{n+1} = \sqrt[3]{a + x_n}$.
5. $x_{n+1} = \sqrt[3]{x_n - a}$.
6. $x_{n+1} = \sqrt[3]{a - x_n}$.

Exercise 1.3.9. Explain the *continued fraction expansion*

$$\sqrt{2} = 1 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{2 + \cdots}}}.$$

What if 2 on the right side is changed to some other positive number?

Exercise 1.3.10. For any $a, b > 0$, define a sequence by

$$x_1 = a, \quad x_2 = b, \quad x_n = \frac{x_{n-1} + x_{n-2}}{2}.$$

Prove that the sequence converges.

Exercise 1.3.11. The arithmetic and the geometric means of $a, b > 0$ are $\dfrac{a+b}{2}$ and $\sqrt{ab}$. By repeating the process, we get two sequences defined by

$$x_1 = a, \quad y_1 = b, \quad x_{n+1} = \frac{x_n + y_n}{2}, \quad y_{n+1} = \sqrt{x_n y_n}.$$

Prove that $x_n \geq x_{n+1} \geq y_{n+1} \geq y_n$ for $n \geq 2$, and the two sequences converge to the same limit.

Exercise 1.3.12. The Fibonacci sequence

$$1, 1, 2, 3, 5, 8, 13, 21, 34, \dots$$

is defined by $x_0 = x_1 = 1$ and $x_{n+1} = x_n + x_{n-1}$. Consider the sequence $y_n = \dfrac{x_{n+1}}{x_n}$.

1. Find the relation between y_{n+1} and y_n.
2. Assume y_n converges, find the limit l.
3. Use the relation between y_{n+2} and y_n to prove that l is the upper bound of y_{2k} and the lower bound of y_{2k+1}.
4. Prove that the subsequence y_{2k} is increasing and the subsequence y_{2k+1} is decreasing.
5. Prove that the sequence y_n converges to l.

Exercise 1.3.13. To find $\sqrt{a}$ for $a > 0$, we start with a guess $x_1 > 0$ of the value of $\sqrt{a}$. Noting that x_1 and $\dfrac{a}{x_1}$ are on the two sides of $\sqrt{a}$, it is reasonable to choose the average $x_2 = \dfrac{1}{2}\left(x_1 + \dfrac{a}{x_1}\right)$ as the next guess. This leads to the inductive formula

$$x_{n+1} = \frac{1}{2}\left(x_n + \frac{a}{x_n}\right)$$

as a way of numerically computing better and better approximate values of $\sqrt{a}$.

1. Prove that $\lim_{n\to\infty} x_n = \sqrt{a}$.
2. We may also use weighted average $x_{n+1} = \dfrac{1}{3}\left(x_n + 2\dfrac{a}{x_n}\right)$ as the next guess. Do we still have $\lim_{n\to\infty} x_n = \sqrt{a}$ for the weighted average?
3. Compare the two methods for specific values of a and b (say $a = 4$, $b = 1$). Which way is faster?
4. Can you come up with a similar scheme for numerically computing $\sqrt[3]{a}$? What choice of the weight gives you the fastest method?

1.3.2 Application of Monotone Sequence

We use Theorem 1.3.2 to prove some limits and define a special number e.

Example 1.3.3. We give another argument for $\lim_{n\to\infty} a^n = 0$ in Example 1.1.11.

First assume $0 < a < 1$. Then the sequence a^n is decreasing and satisfies $0 < a^n < 1$. Therefore the sequence converges to a limit l. By the remark in Example 1.1.1, we also have $\lim_{n\to\infty} a^{n-1} = l$. Then by the arithmetic rule, we have

$$l = \lim_{n\to\infty} a^n = \lim_{n\to\infty} a \cdot a^{n-1} = a \lim_{n\to\infty} a^{n-1} = al.$$

Since $a \neq 1$, we get $l = 0$.

For the case $-1 < a < 0$, we may consider the even and odd subsequences of a^n and apply Proposition 1.1.6. Another way is to apply the sandwich rule to $-|a|^n \leq a^n \leq |a|^n$.

Example 1.3.4. We give another argument that the sequence $x_n = \dfrac{3^n(n!)^2}{(2n)!}$ in Example 1.1.16 converges to 0. By $\lim_{n\to\infty} \dfrac{x_n}{x_{n-1}} = 0.75 < 1$ and the order rule, we have $\dfrac{x_n}{x_{n-1}} < 1$ for sufficiently big n. Since x_n is always positive, we have $x_n < x_{n-1}$ for sufficiently big n. Therefore after finitely many terms, the sequence is decreasing. Moreover, 0 is the lower bound of the sequence, so that the sequence converges.

Let $\lim_{n\to\infty} x_n = l$. Then we also have $\lim_{n\to\infty} x_{n-1} = l$. If $l \neq 0$, then

$$\lim_{n\to\infty} \frac{x_n}{x_{n-1}} = \frac{\lim_{n\to\infty} x_n}{\lim_{n\to\infty} x_{n-1}} = \frac{l}{l} = 1.$$

But the limit is actually 0.75. The contradiction shows that $l = 0$.

Exercise 1.3.14. Extend Example 1.3.3 to a proof of $\lim_{n\to\infty} na^n = 0$ for $|a| < 1$.

Exercise 1.3.15. Extend Example 1.3.4 to prove that, if $\lim_{n\to\infty} \dfrac{|x_n|}{|x_{n-1}|} = l < 1$, then $\lim_{n\to\infty} x_n = 0$.

Example 1.3.5. For the sequence $\left(1 + \dfrac{1}{n}\right)^n$, we compare two consecutive terms by their binomial expansions

$$\begin{aligned}
\left(1+\frac{1}{n}\right)^n &= 1+n\frac{1}{n}+\frac{n(n-1)}{2!}\frac{1}{n^2}+\cdots+\frac{n(n-1)\cdots 3\cdot 2\cdot 1}{n!}\frac{1}{n^n}\\
&= 1+\frac{1}{1!}+\frac{1}{2!}\left(1-\frac{1}{n}\right)+\cdots\\
&\quad+\frac{1}{n!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{n-1}{n}\right),\\
\left(1+\frac{1}{n+1}\right)^{n+1} &= 1+\frac{1}{1!}+\frac{1}{2!}\left(1-\frac{1}{n+1}\right)+\cdots\\
&\quad+\frac{1}{n!}\left(1-\frac{1}{n+1}\right)\left(1-\frac{2}{n+1}\right)\cdots\left(1-\frac{n-1}{n+1}\right)\\
&\quad+\frac{1}{(n+1)!}\left(1-\frac{1}{n+1}\right)\left(1-\frac{2}{n+1}\right)\cdots\left(1-\frac{n}{n+1}\right).
\end{aligned}$$

A close examination shows that the sequence is increasing. Moreover, by the computation in Example 1.3.1, the first expansion gives

$$\begin{aligned}
\left(1+\frac{1}{n}\right)^n &< 1+\frac{1}{1!}+\frac{1}{2!}+\cdots+\frac{1}{n!}\\
&< 1+1+\frac{1}{1\cdot 2}+\frac{1}{2\cdot 3}+\cdots+\frac{1}{(n-1)n}<3.
\end{aligned}$$

By Theorem 1.3.2, the sequence converges. We denote the limit by e

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

Exercise 1.3.16. Find the limit.

1. $\left(\frac{n+1}{n}\right)^{n+1}$.
2. $\left(1-\frac{1}{n}\right)^n$.
3. $\left(1+\frac{1}{2n}\right)^n$.
4. $\left(\frac{2n+1}{2n-1}\right)^n$.

Exercise 1.3.17. Let $x_n=\left(1+\frac{1}{n}\right)^{n+1}$.

1. Use induction to prove $(1+x)^n\ge 1+nx$ for $x>-1$ and any natural number n.
2. Use the first part to prove $\frac{x_{n-1}}{x_n}>1$. This shows that x_n is decreasing.
3. Prove that $\lim_{n\to\infty}x_n=e$.
4. Prove that $\left(1-\frac{1}{n}\right)^n$ is increasing and converges to e^{-1}.

Exercise 1.3.18. Prove that for $n > k$, we have

$$\left(1+\frac{1}{n}\right)^n \geq 1+\frac{1}{1!}+\frac{1}{2!}\left(1-\frac{1}{n}\right)+\cdots+\frac{1}{k!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k}{n}\right).$$

Then use Proposition 1.1.5 to show that

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

Finally, prove

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

1.3.3 Cauchy Criterion

Theorem 1.3.2 gives a special case that we know the convergence of a sequence without knowing the actual limit value. Note that the definition of limit makes explicit use of the limit value and therefore cannot be used to derive the convergence here. The following provides the criterion for the convergence in general, again without referring to the actual limit value.

Theorem 1.3.3 (Cauchy Criterion). *A sequence x_n converges if and only if for any $\epsilon > 0$, there is N, such that*

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

Sequences satisfying the property in the theorem are called *Cauchy sequences*. The theorem says that a sequence converges if and only if it is a Cauchy sequence.

The necessity is easy to see. If $\lim_{n\to\infty} x_n = l$, then for big m, n, both x_m and x_n are very close to l (say within $\frac{\epsilon}{2}$). This implies that x_m and x_n are very close (within $\frac{\epsilon}{2}+\frac{\epsilon}{2}=\epsilon$).

The proof of sufficiency is much more difficult and relies on the following deep result that touches the essential difference between the real and rational numbers.

Theorem 1.3.4 (Bolzano-Weierstrass). *Any bounded sequence has a convergent subsequence.*

Using the theorem, the converse may be proved by the following steps.

1. A Cauchy sequence is bounded.

2. By Bolzano-Weierstrass Theorem, the sequence has a convergent subsequence.

3. If a Cauchy sequence has a subsequence converging to l, then the whole sequence converges to l.

Example 1.3.6. In Example 1.1.19, we argued that the sequence $(-1)^n$ diverges because two subsequences converge to different limits. Alternatively, we may apply the Cauchy criterion. For $\epsilon = 1$ and any N, we pick any $n > N$ and pick $m = n+1$. Then $m, n > N$ and $|x_m - x_n| = |(-1)^{n+1} - (-1)^n| = 2 > \epsilon$. This means that the Cauchy criterion fails, and therefore the sequence diverges.

Example 1.3.7. In Example 1.3.1, we argued the convergence of

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

by increasing and bounded property. Alternatively, for $m > n$, we have

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

Therefore for any ϵ, we have

$$m > n > N = \frac{1}{\epsilon} \implies |x_n - x_m| < \frac{1}{n} < \frac{1}{N} = \epsilon.$$

By the Cauchy criterion, x_n converges.

We note that the same argument can be used to show the convergence of

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

The method in Example 1.3.7 cannot be used here.

Example 1.3.8. The following is the partial sum of the *harmonic series*

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

For any n, we have

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

For $\epsilon = \frac{1}{2}$ and any N, we choose a natural number $n > N$ and also choose $m = 2n > N$. Then

$$|x_m - x_n| = x_{2n} - x_n \ge \epsilon.$$

This shows that the sequence fails the Cauchy criterion and diverges.

Exercise 1.3.19. If x_n is a Cauchy sequence, is $|x_n|$ also a Cauchy sequence? What about the converse?

Exercise 1.3.20. Suppose $|x_{n+1} - x_n| \le \dfrac{1}{n^2}$. Prove that x_n converges.

Exercise 1.3.21. Suppose c_n is bounded and $|r| < 1$. Prove that the sequence

$$x_n = c_0 + c_1 r + c_2 r^2 + \cdots + c_n r^n$$

converges.

Exercise 1.3.22. Use Cauchy criterion to determine convergence.

1. $1 + \dfrac{1}{\sqrt{2}} + \dfrac{1}{\sqrt{3}} + \cdots + \dfrac{1}{\sqrt{n}}$.
2. $1 - \dfrac{1}{2^3} + \dfrac{1}{3^3} - \cdots + \dfrac{(-1)^{n+1}}{n^3}$.
3. $\dfrac{1}{2} + \dfrac{2}{3} + \dfrac{3}{4} + \cdots + \dfrac{n-1}{n}$.
4. $1 + \dfrac{1}{3} + \dfrac{1}{5} + \cdots + \dfrac{1}{2n+1}$.
5. $1 + \dfrac{1}{2!} + \dfrac{1}{3!} + \cdots + \dfrac{1}{n!}$.
6. $1 + \dfrac{2}{1!} + \dfrac{3}{2!} + \cdots + \dfrac{n}{(n-1)!}$.

Example 1.3.9. Theorem 1.3.4 shows the existence of converging subsequences of a bounded sequence. How about the limit values of such subsequences?

Let us list all finite decimal expressions in $(0, 1)$ as a sequence

$$x_n\colon \quad 0.1,\ 0.2,\ \ldots,\ 0.9,\ 0.01,\ 0.02,\ \ldots, 0.99,\ 0.001,\ 0.002,\ \ldots,\ 0.999,\ \ldots.$$

The number $0.318309\cdots$ is the limit of the following subsequence

$$0.3,\ 0.31,\ 0.318,\ 0.3183,\ 0.31830,\ 0.318309,\ \ldots.$$

It is easy to see that any number in $[0, 1]$ is the limit of a convergent subsequence of x_n.

Exercise 1.3.23. Construct a sequence such that the limits of convergent subsequences are exactly $\dfrac{1}{n}$, $n \in \mathbb{N}$ and 0.

Exercise 1.3.24. Construct a sequence such that any number is the limit of some convergent subsequence.

Exercise 1.3.25. Use any suitable method or theorem to determine convergence.

1. $\dfrac{1}{2}, \dfrac{2}{1}, \dfrac{2}{3}, \dfrac{3}{2}, \ldots, \dfrac{n}{n+1}, \dfrac{n+1}{n}, \ldots$.
2. $\dfrac{1}{2}, -\dfrac{2}{1}, \dfrac{2}{3}, -\dfrac{3}{2}, \ldots, \dfrac{n}{n+1}, -\dfrac{n+1}{n}, \ldots$.

3. $1+\frac{1}{2^2}+\frac{2}{3^2}+\cdots+\frac{n-1}{n^2}$.

4. $\frac{1}{1\cdot 2}+\frac{1}{3\cdot 4}+\cdots+\frac{1}{(2n-1)2n}$.

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

6. $\frac{2}{1\cdot 3}+\frac{3}{2\cdot 4}+\cdots+\frac{n}{(n-1)(n+1)}$.

7. $\frac{1\cdot 3\cdot 5\cdots(2n-1)}{2\cdot 4\cdot 6\cdots(2n)}$.

8. $\frac{1\cdot 3\cdot 5\cdots(2n-1)}{n!}$.

9. $\sqrt{1+\sqrt{\frac{1}{2}+\sqrt{\frac{1}{3}+\cdots+\sqrt{\frac{1}{n}}}}}$.

10. $\sqrt{1+\sqrt{2+\sqrt{3+\cdots+\sqrt{n}}}}$.

1.4 Infinity

1.4.1 Divergence to Infinity

A sequence may diverge for various reasons. For example, the sequence n diverges because it can become arbitrarily big. On the other hand, the bounded sequence $(-1)^n$ diverges because it has two subsequences with different limits. The first example may be summarized by the following definition.

Definition 1.4.1. A sequence *diverges to infinity*, denoted $\lim_{n\to\infty} x_n = \infty$, if for any B, there is N, such that $n > N$ implies $|x_n| > B$.

In the definition, the infinity means that the absolute value (or the magnitude) of the sequence can become arbitrarily big. If we further take into account of the signs, then we get the following definitions.

Definition 1.4.2. A sequence diverges to $+\infty$, denoted $\lim_{n\to\infty} x_n = +\infty$, if for any B, there is N, such that $n > N$ implies $x_n > B$. A sequence diverges to $-\infty$, denoted $\lim_{n\to\infty} x_n = -\infty$, if for any B, there is N, such that $n > N$ implies $x_n < B$.

The meaning of $\lim_{n\to\infty} x_n = +\infty$ is illustrated in Figure 1.4.1. For example, we have $\lim_{n\to\infty} n = +\infty$. We also note that, in the definition of $\lim_{n\to\infty} x_n = +\infty$, we may additionally assume $B > 0$ (or $B > 100$) without loss of generality.

Example 1.4.1. We have

$$\lim_{n\to\infty} n^p = +\infty, \text{ for } p > 0.$$

For the rigorous proof, for any $B > 0$, we have

$$n > B^{\frac{1}{p}} \implies n^p > B.$$

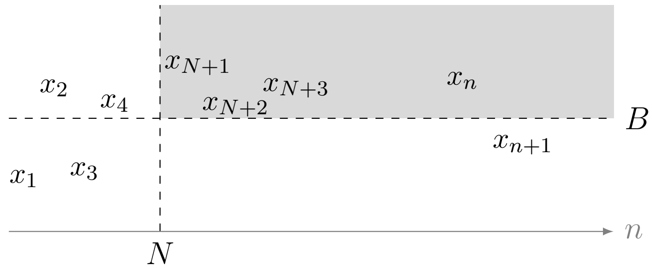


Figure 1.4.1: $n > N$ implies $x_n > B$.

Example 1.4.2. Example 1.1.11 may be extended to show that $\lim_{n\to\infty} a^n = \infty$ for $|a| > 1$. Specifically, let $|a| = 1 + b$. Then $|a| > 1$ implies $b > 0$, and we have

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

For any B, we then have

$$n > \frac{B}{b} \implies |a^n| > nb > B.$$

This proves $\lim_{n\to\infty} a^n = \infty$ for $|a| > 1$. If we take the sign into account, this also proves $\lim_{n\to\infty} a^n = +\infty$ for $a > 1$.

Example 1.4.3. Suppose $x_n \neq 0$. We prove that $\lim_{n\to\infty} x_n = 0$ implies $\lim_{n\to\infty} \dfrac{1}{x_n} = \infty$. Actually the converse is also true and the proof is left to the reader.

If $\lim_{n\to\infty} x_n = 0$, then for any $B > 0$, we apply the definition of the limit to $\dfrac{1}{B} > 0$ to get N, such that

$$n > N \implies |x_n| < \frac{1}{B} \implies \left|\frac{1}{x_n}\right| > B.$$

This proves $\lim_{n\to\infty} \dfrac{1}{x_n} = \infty$.

Applying what we just proved to the limit in Example 1.1.11, we get another proof of $\lim_{n\to\infty} a^n = \infty$ for $|a| > 1$.

Exercise 1.4.1. Let $x_n \neq 0$. Prove that $\lim_{n\to\infty} x_n = 0$ if and only if $\lim_{n\to\infty} \dfrac{1}{x_n} = \infty$.

Exercise 1.4.2. Prove that $\lim_{n\to\infty} x_n = +\infty$ if and only if $x_n > 0$ for sufficiently big n and $\lim_{n\to\infty} \dfrac{1}{x_n} = 0$.

Exercise 1.4.3. Rigorously prove divergence to infinity. Determine $\pm\infty$ if possible.

1. $\dfrac{n^2-n+1}{n+1}$.
2. $\dfrac{n}{\sqrt{n}+1}$.
3. $\dfrac{a^n}{n}$, $|a|>1$.
4. $n^p a^n$, $|a|>1$.
5. $\dfrac{n!}{4^n}$.
6. $n!a^n$.

1.4.2 Arithmetic Rule for Infinity

A sequence x_n is an *infinitesimal* if $\lim_{n\to\infty} x_n = 0$. Example 1.4.3 and Exercise 1.4.1 show that a sequence is an infinetesimal if and only if its reciprocal is an infinity.

Many properties of the finite limit can be extended to infinity. For example, we have $\dfrac{l}{0}=\infty$ for $l\neq 0$. This means that, if $\lim_{n\to\infty} x_n = l \neq 0$ and $\lim_{n\to\infty} y_n = 0$, then $\lim_{n\to\infty} \dfrac{x_n}{y_n}=\infty$. For example,

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

Note that l in $\dfrac{l}{0}$ can represent any sequence converging to l, and is not necessarily a constant.

The following are more extensions of the arithmetic rules to infinity. The rules are symbolically denoted by "arithmetic equalities", and the exact meaning of the rules are also given.

- $\dfrac{l}{\infty}=0$: If $\lim_{n\to\infty} x_n = l$ and $\lim_{n\to\infty} y_n=\infty$, then $\lim_{n\to\infty}\dfrac{x_n}{y_n}=0$.
- $(+\infty)+(+\infty)=+\infty$: If $\lim_{n\to\infty} x_n=+\infty$ and $\lim_{n\to\infty} y_n=+\infty$, then $\lim_{n\to\infty}(x_n+y_n)=+\infty$.
- $(-\infty)+l=-\infty$: If $\lim_{n\to\infty} x_n=-\infty$ and $\lim_{n\to\infty} y_n=l$, then $\lim_{n\to\infty}(x_n+y_n)=-\infty$.
- $(+\infty)\cdot l=-\infty$ for $l<0$: If $\lim_{n\to\infty} x_n=+\infty$ and $\lim_{n\to\infty} y_n=l<0$, then $\lim_{n\to\infty} x_n y_n=-\infty$.
- $\dfrac{l}{0^+}=+\infty$ for $l>0$: If $\lim_{n\to\infty} x_n=l>0$, $\lim_{n\to\infty} y_n=0$ and $y_n>0$, then $\lim_{n\to\infty}\dfrac{x_n}{y_n}=+\infty$.

On the other hand, we should always be cautious not to overextend the arithmetic properties. For example, the following "arithmetic rules" are actually wrong

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

A counterexample for the first equality is $x_n = n$ and $y_n = -n$, for which we have $\lim_{n\to\infty} x_n = \infty$, $\lim_{n\to\infty} y_n = \infty$ and $\lim_{n\to\infty}(x_n+y_n) = 0$. In general, one needs to use common sense to decide whether certain extended arithmetic rules make sense.

Example 1.4.4. By Example 1.4.1 and the extended arithmetic rule, we have

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

In general, any non-constant polynomial of n diverges to ∞, and for rational functions, we have

$$\lim_{n\to\infty}\frac{a_pn^p + a_{p-1}n^{p-1} + \cdots + a_1n + a_0}{b_qn^q + b_{q-1}n^{q-1} + \cdots + b_1n + b_0} = \begin{cases} +\infty, & \text{if } p > q,\ a_pb_q > 0, \\ -\infty, & \text{if } p > q,\ a_pb_q < 0, \\ \dfrac{a_p}{b_q}, & \text{if } p = q,\ b_q \neq 0, \\ 0, & \text{if } p < q,\ b_q \neq 0. \end{cases}$$

Exercise 1.4.4. Prove the extended arithmetic rules

$$\frac{l}{0} = \infty, \quad l + (+\infty) = +\infty, \quad (+\infty)\cdot(-\infty) = -\infty, \quad \frac{l}{0^-} = -\infty \text{ for } l > 0.$$

Exercise 1.4.5. Construct sequences x_n and y_n, such that both diverge to infinity, but $x_n + y_n$ can have any of the following behaviors.

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

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

3. $x_n + y_n$ is bounded but does not converge.

4. $x_n + y_n$ is not bounded and does not diverge to infinity.

The exercise shows that $\infty + \infty$ has no definite meaning.

Exercise 1.4.6. Prove that if $p > 0$, then $\lim_{n\to\infty} x_n = +\infty$ implies $\lim_{n\to\infty} x_n^p = +\infty$. What about the case $p < 0$?

Exercise 1.4.7. Prove the extended sandwich rule: If $x_n \le y_n$ for sufficiently big n, then $\lim_{n\to\infty} x_n = +\infty$ implies $\lim_{n\to\infty} y_n = +\infty$.

Exercise 1.4.8. Prove the extended order rule: If $\lim_{n\to\infty} x_n = l$ is finite and $\lim_{n\to\infty} y_n = +\infty$, then $x_n < y_n$ for sufficiently big n.

Exercise 1.4.9. Suppose $\lim_{n\to\infty} x_n = l > 1$. Prove that $\lim_{n\to\infty} x_n^n = +\infty$.

Exercise 1.4.10. Prove that if $\lim_{n\to\infty}\dfrac{x_n}{x_{n-1}}=l$ and $|l|>1$, then x_n diverges to infinity.

Exercise 1.4.11. Explain the infinities. Determine the sign of infinity if possible.

1. $\dfrac{n+\sin 2n}{\sqrt{n}-\cos n}$.
2. $\dfrac{n!}{a^n+b^n}$, $a+b\neq 0$.
3. $\dfrac{1}{\sqrt[n]{n}-1}$.
4. $\dfrac{1}{\sqrt[n]{n}-\sqrt[n]{2n}}$.
5. $n(\sqrt{n+2}-\sqrt{n})$.
6. $\dfrac{(-1)^n n^2}{n-1}$.
7. $\dfrac{3^n-2^n}{n}$.
8. $\dfrac{3^n-2^n}{n^3+n^2}$.
9. $\left(1+\dfrac{1}{n}\right)^{n^2}$.

1.4.3 Unbounded Monotone Sequence

The following complements Theorem 1.3.2.

Theorem 1.4.3. *Any unbounded monotone sequence diverges to infinity.*

If an increasing sequence x_n is bounded, then by Theorem 1.3.2, it converges to a finite limit. If the sequence is not bounded, then it is not bounded above. This means that any number B is not an upper bound, or some $x_N>B$. Then by x_n increasing, we have

$$n>N \implies x_n\geq x_N>B.$$

This proves that $\lim_{n\to\infty}x_n=+\infty$. Similarly, an unbounded decreasing sequence diverges to $-\infty$.

Example 1.4.5. In Example 1.3.8, we showed that the increasing sequence

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

diverges. By Theorem 1.4.3, we know that it diverges to $+\infty$. Therefore the sum of the harmonic series

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

Example 1.4.6. For $a>1$, the sequence a^n is increasing. If the sequence converges to a finite limit l, then

$$l=\lim_{n\to\infty}a^n=a\lim_{n\to\infty}a^{n-1}=al.$$

Since the sequence is increasing, we have $l\geq a>1$, which contradicts to $l=al$. By Theorem 1.4.3, therefore, we conclude that $\lim_{n\to\infty}a^n=+\infty$.

Exercise 1.4.12. Prove $\lim_{n\to\infty}\dfrac{a^n}{n^2}=+\infty$ for $a>1$.

1.5 Limit of Function

Similar to the sequence limit, we say a function $f(x)$ *converges* to l at a, and write

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

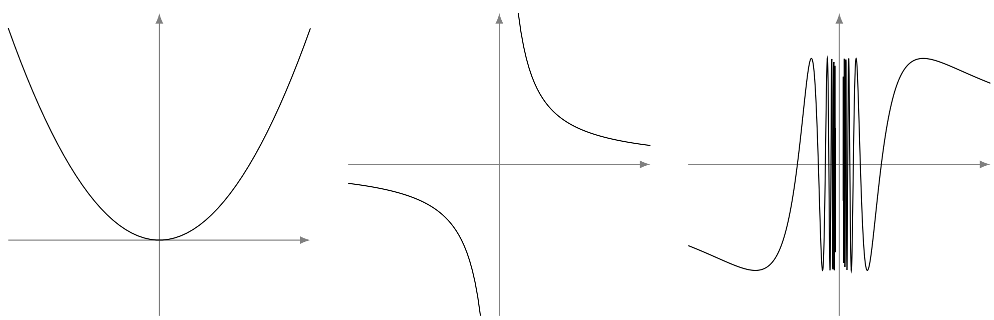
if $f(x)$ approaches l when x approaches a

$$x \to a,\ x \neq a \implies f(x) \to l.$$

Note that we include $x \neq a$ because $f(x)$ is not required to be defined at a.

The definition allows a or l to be ∞ (or $\pm\infty$ if the sign can be determined). When $l = \infty$, we should say that $f(x)$ *diverges* to ∞ at a.

In Figure 1.5.1, as x approaches 0, we see that x^2 approaches 0, $\dfrac{1}{x}$ gets arbitrarily big, and $\sin\dfrac{1}{x}$ swings between -1 and 1, never approaching any one specific finite number. We write $\lim_{x\to 0} x^2 = 0$, $\lim_{x\to 0} \dfrac{1}{x} = \infty$, and say that $\lim_{x\to 0} \sin\dfrac{1}{x}$ diverges.



$y = x^2$ $\qquad$ $y = \dfrac{1}{x}$ $\qquad$ $y = \sin\dfrac{1}{x}$

Figure 1.5.1: Behavior of functions near 0.

On the other hand, as x approaches the infinity, we find that x^2 gets arbitrarily big, and $\dfrac{1}{x}$ and $\sin\dfrac{1}{x}$ approach 0. Therefore $\lim_{x\to\infty} x^2 = \infty$, $\lim_{x\to\infty} \dfrac{1}{x} = \lim_{x\to\infty} \sin\dfrac{1}{x} = 0$. Moreover, $\sin x$ swings between -1 and 1, and $\lim_{x\to\infty} \sin x$ diverges.

1.5.1 Properties of Function Limit

The function limit shares similar properties with the sequence limit.

Proposition 1.5.1. *If $f(x) = g(x)$ for x sufficiently near a and $x \neq a$, then $\lim_{x\to a} f(x) = \lim_{x\to a} g(x)$.*

Proposition 1.5.2 (Arithmetic Rule)**.** *Suppose* $\lim_{x\to a} f(x) = l$ *and* $\lim_{x\to a} g(x) = k$. *Then*

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

where c *is a constant and* $k \neq 0$ *in the last equality.*

Proposition 1.5.3 (Sandwich Rule)**.** *If* $f(x) \le g(x) \le h(x)$ *and* $\lim_{x\to a} f(x) = \lim_{x\to a} h(x) = l$, *then* $\lim_{x\to a} g(x) = l$.

Proposition 1.5.4 (Order Rule)**.** *Suppose* $\lim_{x\to a} f(x) = l$ *and* $\lim_{x\to a} g(x) = k$.

1. *If* $f(x) \le g(x)$ *for* x *near* a *and* $x \neq a$, *then* $l \le k$.
2. *If* $l < k$, *then* $f(x) < g(x)$ *for* x *near* a *and* $x \neq a$.

Proposition 1.1.6 will be extended to Proposition 1.5.5 (composition rule) and Proposition 1.6.2 (relation between sequence and function limits).

Since

$$x \to a,\ x \neq a \implies c \to c,$$

and

$$x \to a,\ x \neq a \implies x \to a,$$

we have

$$\lim_{x\to a} c = c, \quad \lim_{x\to a} x = a.$$

It is also intuitively clear that

$$\lim_{x\to 0} |x| = 0, \quad \lim_{x\to\infty} \frac{1}{|x|^p} = 0 \text{ for } p > 0.$$

The subsequent examples are based on these limits.

Example 1.5.1. For $a > 0$, we have $|x| = x$ near a. By Proposition 1.5.1, we have $\lim_{x\to a} |x| = \lim_{x\to a} x = a = |a|$.

For $a < 0$, we have $|x| = -x$ near a. By Propositions 1.5.1 and 1.5.2, we have $\lim_{x\to a} |x| = \lim_{x\to a} -x = -\lim_{x\to a} x = -a = |a|$.

Combining the two cases with $\lim_{x\to 0} |x| = 0$, we get $\lim_{x\to a} |x| = |a|$.

Example 1.5.2. Let

$$f(x) = \begin{cases} y, & \text{if } y \neq 0, \\ A, & \text{if } y = 0. \end{cases}$$

Then $f(x) = x$ for $x \neq 0$, and by Proposition 1.5.1, we have $\lim_{x\to 0} f(x) = \lim_{x\to 0} x = 0$. Note that the limit is independent of the value $f(0) = A$. In fact, the limit does not even require the function to be defined at 0.

Example 1.5.3. The rational function $\dfrac{x^3-1}{x-1}$ is not defined at $x=1$. Yet the function converges at 1

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

Example 1.5.4. By the arithmetic rule, we have

$$\lim_{x\to a}(x^3-3x+1)=\left(\lim_{x\to a}x\right)^3-\lim_{x\to a}3\lim_{x\to a}x+\lim_{x\to a}1=a^3-3a+1.$$

More generally, for any polynomial $p(x)=c_nx^n+c_{n-1}x^{n-1}+\cdots+c_1x+c_0$ and finite a, we have

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

A rational function $r(x)=\dfrac{p(x)}{q(x)}$ is the quotient of two polynomials $p(x)$ and $q(x)$, and is defined at a if $q(a)\neq 0$. Further by the arithmetic rule, we also have

$$\lim_{x\to a}r(x)=\frac{\lim_{x\to a}p(x)}{\lim_{x\to a}q(x)}=\frac{p(a)}{q(a)}=r(a)$$

whenever $r(x)$ is defined at a.

Example 1.5.5. By the arithmetic rule, we have

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

This is comparable to Example 1.1.2. In general, Example 1.4.4 can be extended to the function limit

$$\lim_{x\to\infty}\frac{a_px^p+a_{p-1}x^{p-1}+\cdots+a_1x+a_0}{b_qx^q+b_{q-1}x^{q-1}+\cdots+b_1x+b_0}=\begin{cases}\infty, & \text{if } p>q,\\ \dfrac{a_p}{b_q}, & \text{if } p=q,\ b_q\neq 0,\\ 0, & \text{if } p<q,\ b_q\neq 0.\end{cases}$$

Example 1.5.6. Similar to Example 1.1.5, the function $\sqrt{|x|+2}-\sqrt{|x|}$ satisfies

$$\begin{aligned}0<\sqrt{|x|+2}-\sqrt{|x|}&=\frac{(\sqrt{|x|+2}-\sqrt{|x|})(\sqrt{|x|+2}+\sqrt{|x|})}{\sqrt{|x|+2}+\sqrt{|x|}}\\&=\frac{1}{\sqrt{|x|+2}+\sqrt{|x|}}<\frac{2}{\sqrt{|x|}}.\end{aligned}$$

By $\lim_{x\to\infty}\dfrac{2}{\sqrt{|x|}} = 2\lim_{x\to\infty}\dfrac{1}{\sqrt{|x|}} = 0$ and the sandwich rule, we get

$$\lim_{x\to+\infty}(\sqrt{|x|+2}-\sqrt{|x|}) = 0.$$

Example 1.5.7. The following computation

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

is wrong because $\lim_{x\to 0}\sin\frac{1}{x}$ diverges. However, if we use

$$-|x| \le x\sin\frac{1}{x} \le |x|, \quad \lim_{x\to 0}|x| = 0,$$

and the sandwich rule, then we get $\lim_{x\to 0} x\sin\dfrac{1}{x} = 0$.

Exercise 1.5.1. Explain that $\lim_{x\to a} f(x) = l$ if and only if $\lim_{x\to a}(f(x)-l) = 0$.

Exercise 1.5.2. Use the sandwich rule to prove that $\lim_{x\to a}|f(x)| = 0$ implies $\lim_{x\to a} f(x) = 0$.

Exercise 1.5.3. Find the limits.

1. $\lim_{x\to\infty}\dfrac{x}{x+2}$.
2. $\lim_{x\to 2}\dfrac{x}{x+2}$.
3. $\lim_{x\to -2}\dfrac{x}{x+2}$.
4. $\lim_{x\to\infty}\dfrac{x^2-1}{x^2+x-2}$.
5. $\lim_{x\to 0}\dfrac{x^2-1}{x^2+x-2}$.
6. $\lim_{x\to 1}\dfrac{x^2-1}{x^2+x-2}$.

Exercise 1.5.4. Find the limits.

1. $\lim_{x\to 0} x\cos\dfrac{1}{x^2}$.
2. $\lim_{x\to\infty}\dfrac{1}{\sqrt{|x|}}\sin x^2$.
3. $\lim_{x\to\infty}(\sqrt{|x|+a}-\sqrt{|x|+b})$.
4. $\lim_{x\to\infty}(\sqrt{(x+a)(x+b)}-x)$.
5. $\lim_{x\to\infty}\sqrt{\dfrac{x+a}{x+b}}$.
6. $\lim_{x\to\infty}\left(\dfrac{\sqrt[3]{x^2}}{\sqrt[3]{x}+a}-\dfrac{\sqrt[3]{x^2}}{\sqrt[3]{x}+b}\right)$.

1.5.2 Limit of Composition Function

Suppose three variables x, y, z are related by

$$x \xrightarrow{f} y = f(x) \xrightarrow{g} z = g(y).$$

Then z and x are related by $z = g(f(x))$, the composition of two functions. Suppose both f and g have limits

$$\lim_{x\to a} f(x) = b, \quad \lim_{y\to b} g(y) = c,$$

where b is the value of the first limit as well as the location of the second limit. The problem is whether the composition has limit

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

What we want is to combine two implications

$$\begin{aligned} x \to a,\ x \neq a &\implies y \to b, \\ y \to b,\ y \neq b &\implies z \to c, \end{aligned}$$

to get the third implication

$$x \to a,\ x \neq a \implies z \to c.$$

However, the two implications cannot be combined as is, because "$y \to b$" does not imply "$y \to b,\ y \neq b$". There are two ways to save this. The first is to strengthen the first implication to

$$x \to a,\ x \neq a \implies y \to b,\ y \neq b.$$

Here the extra condition is $f(x) \neq b$ for x near a and $x \neq a$. The second is to strengthen the second implication to

$$y \to b \implies z \to c.$$

Here the extra condition is $y = b \implies z \to c$. Since $y = b$ implies $z = g(b)$, the extra condition is simply $c = g(b)$, so the strengthened second implication becomes

$$y \to b \implies z \to c = g(b).$$

Proposition 1.5.5 (Composition Rule). *Suppose* $\lim_{x\to a} f(x) = b$ *and* $\lim_{y\to b} g(y) = c$. *Then we have* $\lim_{x\to a} g(f(x)) = c$, *provided one of the following extra conditions is satisfied*

1. $f(x) \neq b$ *for* x *near* a *and* $x \neq a$.
2. $c = g(b)$.

Note that the second condition means

$$\lim_{y\to b} g(y) = g(b).$$

Later on, this will become the definition of the continuity of $g(y)$ at b. Moreover, the composition rule in this case means

$$\lim_{x\to a} g(f(x)) = c = g(b) = g\left(\lim_{x\to a} f(x)\right).$$

So the continuity of $g(y)$ is the same as the exchangeability of the function g and the limit.

The composition rule extends Proposition 1.1.6 because a sequence x_n can be considered as a function

$$x_n \colon n \mapsto x(n) = x_n.$$

Then a subsequence can be considered as a composition with a function $n_k \colon \mathbb{N} \to \mathbb{N}$ that satisfies $\lim_{k\to\infty} n_k = \infty$

$$k \mapsto n(k) = n_k \mapsto x(n(k)) = x_{n_k}.$$

Example 1.5.8. We have

$$0 \le |\sqrt{x} - 1| = \frac{|x-1|}{\sqrt{x}+1} \le |x-1|.$$

Note that $|x-1|$ is the composition of $z = |y|$ and $y = x - 1$. By $\lim_{x\to 1}(x-1) = 1 - 1 = 0$, $\lim_{y\to 0}|y| = 0 = |0|$, and the composition rule (both extra conditions are satisfied), we get $\lim_{x\to 1}|x-1| = 0$. Then by the sandwich rule, we get $\lim_{x\to 1}|\sqrt{x} - 1| = 0$. This implies $\lim_{x\to 1}(\sqrt{x} - 1) = 0$ (see Exercise 1.5.2). Finally, by the arithmetic rule, we get $\lim_{x\to 1}\sqrt{x} = 1$.

Example 1.5.9. We have $\lim_{x\to 1}(3x^3 - 2) = 1$ from the arithmetic rule. We also know $\lim_{y\to 1}\sqrt{y} = 1$ from Example 1.5.8. The composition

$$x \mapsto y = f(x) = 3x^3 - 2 \mapsto z = g(y) = \sqrt{y} = g(f(x)) = \sqrt{3x^3 - 2}$$

should give us $\lim_{x\to 1}\sqrt{3x^3 - 2} = 1$.

We need to verify one of the extra conditions in the composition rule. If x is close to 1 and $x < 1$, then we have $x^3 < 1^3 = 1$, so that $3x^3 - 2 < 1$. Similarly, if x is close to 1 and $x > 1$, then $3x^3 - 2 > 1$. Therefore for x close to 1 and $\neq 1$, we indeed have $3x^3 - 2 \neq 1$. This verifies the first condition.

Although the validity of the first condition already allows us to apply the composition rule, the second condition $\lim_{y\to 1}\sqrt{y} = 1 = \sqrt{1}$ is also valid.

Note that it is rather tempting to write

$$\lim_{x\to 1}\sqrt{3x^3 - 2} = \lim_{3x^3-2\to 1}\sqrt{3x^3 - 2} = \lim_{y\to 1}\sqrt{y}.$$

In other words, the composition rule appears simply as a change of variable. However, one needs to be careful because hidden in the definition of $\lim_{x\to 1}$ is $x \neq 0$.

Similarly, the assumption $3x^3-2\neq 1$ is implicit in writing $\lim_{3x^3-2\to 1}$, and the first equality above actually requires you to establish

$$x\to 1,\ x\neq 1 \iff 3x^3-2\to 1,\ 3x^3-2\neq 1.$$

This turns out to be true in our specific example, but might fail for other examples.

Example 1.5.10. Example 1.5.8 can be extended to show that $\lim_{x\to a}\sqrt{x}=\sqrt{a}$ for any $a>0$ (see Exercise 1.5.5). This actually means that $\sqrt{x}$ is continuous, and therefore we may apply the composition rule to get

$$\begin{aligned}\lim_{x\to 1}\sqrt{\sqrt{3x^3-2}+7x} &= \sqrt{\lim_{x\to 1}\left(\sqrt{3x^3-2}+7x\right)}\\ &= \sqrt{\lim_{x\to 1}\sqrt{3x^3-2}+\lim_{x\to 1}7x}\\ &= \sqrt{\sqrt{\lim_{x\to 1}(3x^3-2)}+\lim_{x\to 1}7x}\\ &= \sqrt{\sqrt{3\cdot 1^2+1}+7\cdot 1}.\end{aligned}$$

Here is the detailed reason. The last equality is by the arithmetic rule. The third equality makes use of the continuity of the function $\sqrt{x}$ and the composition rule to move the limit from outside the square root to inside the square root. The second equality is by the arithmetic rule. Once we know $\lim_{x\to 1}(\sqrt{3x^3-2}+7x)$ converges to a positive number, the first equality then follows from the continuity of $\sqrt{x}$ and the composition rule.

Exercise 1.5.5. Show that $\lim_{x\to a}\sqrt{x}=\sqrt{a}$ for any $a>0$.

Exercise 1.5.6. Show that $\lim_{x\to a}\sqrt[3]{x}=\sqrt[3]{a}$ for any $a\neq 0$.

Exercise 1.5.7. Find the limits, $a,b>0$.

1. $\lim_{x\to 0}\sqrt{a+x}$.
2. $\lim_{x\to 0}\dfrac{1}{x}(\sqrt{a+x}-\sqrt{a-x})$.
3. $\lim_{x\to 0}\left(\sqrt{\dfrac{a+x}{a}}-\sqrt{\dfrac{a}{a+x}}\right)$.
4. $\lim_{x\to 0}\dfrac{1}{x}\left(\sqrt{\dfrac{a+x}{b}}-\sqrt{\dfrac{a}{b+x}}\right)$.

Example 1.5.11. A change of variable can often be applied to limits. For example, we have

$$\lim_{x\to a}f(x^2)=\lim_{y\to a^2}f(y),\ \text{for } a\neq 0.$$

The equality means that the two limits are equivalent.

Suppose $\lim_{y\to a^2} f(y) = l$. Then $f(x^2)$ is the composition of $f(y)$ with $y = x^2$, and by $a \neq 0$, the first condition is satisfied

$$x \to a,\ x \neq a \implies y \to a^2,\ y \neq a^2.$$

Therefore the composition rule can be applied to give $\lim_{x\to a} f(x^2) = l$.

Conversely, suppose $\lim_{x\to a} f(x^2) = l$. For x near $a > 0$, $f(y)$ is the composition of $f(x^2)$ with $x = \sqrt{y}$, and by $a \neq 0$, the first condition is satisfied

$$y \to a^2,\ y \neq a^2 \implies x \to a,\ x \neq a.$$

Therefore the composition rule can be applied to give $\lim_{y\to a^2} f(y) = l$. For the case $a < 0$, the similar argument with $x = -\sqrt{y}$ works, and the composition rule still gives $\lim_{y\to a^2} f(y) = l$.

We note that we cannot verify the second condition in the problem above because not much is assumed about f. In particular, f is not assumed to be continuous.

Here are more examples of equivalent limits

$$\lim_{x\to a} f(x) = \lim_{y\to 0} f(y-a), \quad \lim_{x\to a} f(x^3) = \lim_{y\to a^3} f(y), \quad \lim_{x\to\infty} f(x) = \lim_{y\to 0} f\left(\frac{1}{y}\right),$$

where the first condition for the composition rule is satisfied in both directions

$$\begin{aligned} x \to a,\ x \neq a &\iff x - a \to 0,\ x - a \neq 0; \\ x \to a,\ x \neq a &\iff x^3 \to a^3,\ x^3 \neq a^3; \\ x \to \infty &\iff \frac{1}{x} \to 0. \end{aligned}$$

In the last equivalence, we automatically have $x \neq \infty$ and $\dfrac{1}{x} \neq 0$.

Example 1.5.12. The composition rule fails when neither conditions are satisfied, which means that $f(x) = b$ for some $x \neq a$ arbitrarily close to a, and $c \neq g(b)$.

For a concrete example, consider

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

We have $\lim_{x\to 0} f(x) = 0$ (see Example 1.5.7) and $\lim_{y\to 0} g(y) = 0$ (see Example 1.5.2). This means $a = b = c = 0$. However, the composition is

$$g(f(x)) = \begin{cases} x \sin\dfrac{1}{x}, & \text{if } x \neq (n\pi)^{-1}, \\ A, & \text{if } x = (n\pi)^{-1}. \end{cases}$$

and $\lim_{x\to 0} g(f(x))$ converges if and only if $g(0) = A = 0 = c$.

Exercise 1.5.8. Rewrite the limits as $\lim_{x\to c} f(x)$ for suitable c.

1. $\lim_{x\to b} f(x-a)$.
2. $\lim_{x\to\infty} f(x+a)$.
3. $\lim_{x\to 0} f(ax+b)$.
4. $\lim_{x\to 1} f(\sqrt{x})$.
5. $\lim_{x\to -1} f(x^2+x)$.
6. $\lim_{x\to a} f(x^2+x)$.

1.5.3 One Sided Limit

In the sequence limit $\lim_{n\to\infty} x_n$, the subscript n has only the positive infinity direction to go. The function limit can have various directions. For example, $\lim_{x\to+\infty} f(x) = l$ means

$$x \to \infty,\ x > 0 \implies f(x) \to l.$$

Moreover, the *left limit* $\lim_{x\to a^+} f(x) = l$ means

$$x \to a,\ x > a \implies f(x) \to l.$$

Similarly, the *right limit* $\lim_{x\to a^-} f(x) = l$ means

$$x \to a,\ x < a \implies f(x) \to l.$$

All the properties of the usual (two sided) limits still hold for one sided limits. Moreover, we have the following relation (for $a = \infty$, $a^{\pm}$ means $\pm\infty$).

Proposition 1.5.6. $\lim_{x\to a} f(x) = l$ *if and only if* $\lim_{x\to a^+} f(x) = \lim_{x\to a^-} f(x) = l$.

Example 1.5.13. For the sign function

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

we have $\lim_{x\to 0^+} f(x) = 1$ and $\lim_{x\to 0^-} f(x) = -1$. Since the two limits are not equal, $\lim_{x\to 0} f(x)$ diverges.

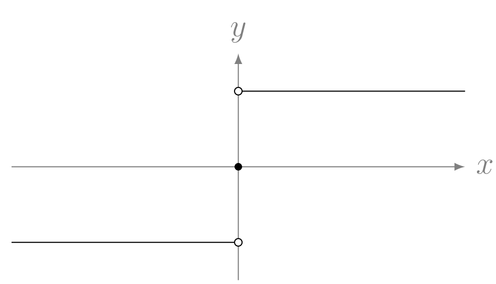


Figure 1.5.2: Sign function.

Example 1.5.14. To find $\lim_{x\to\infty}\dfrac{|x+4|}{x}$, we consider the limit at $+\infty$ and $-\infty$. For $x>0$, we have $\dfrac{|x+4|}{x}=\dfrac{x+4}{x}=1+4\dfrac{1}{x}$. By the arithmetic rule, we have

$$\lim_{x\to+\infty}\frac{|x+4|}{x}=\lim_{x\to+\infty}\left(1+4\frac{1}{x}\right)=1+4\lim_{x\to+\infty}\frac{1}{x}=1.$$

For $x<-4$, we have $\dfrac{|x+4|}{x}=-\dfrac{x+4}{x}=-1-4\dfrac{1}{x}$. By the arithmetic rule, we have

$$\lim_{x\to-\infty}\frac{|x+4|}{x}=\lim_{x\to-\infty}\left(-1-4\frac{1}{x}\right)=-1-4\lim_{x\to-\infty}\frac{1}{x}=-1.$$

Since the two limits are different, $\lim_{x\to\infty}\dfrac{|x+4|}{x}$ diverges.

Example 1.5.15. If we apply the argument in Example 1.5.6 to $x>0$, then we get

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

If we apply the argument to $x<0$, then we get

$$\lim_{x\to-\infty}(\sqrt{2-x}-\sqrt{-x})=0.$$

Exercise 1.5.9. Find the limits at 0.

1. $\begin{cases}1, & \text{if } x<0,\\ 2, & \text{if } x>0.\end{cases}$

2. $\begin{cases}1, & \text{if } x\neq 0,\\ 2, & \text{if } x=0.\end{cases}$

3. $\begin{cases}x, & \text{if } x<0,\\ -x^2, & \text{if } x>0.\end{cases}$

Exercise 1.5.10. Find the limits.

1. $\lim_{x\to+\infty}\dfrac{1}{\sqrt{x+a}}\sin x^2$.
2. $\lim_{x\to+\infty}(\sqrt{x+a}-\sqrt{x+b})$.
3. $\lim_{x\to+\infty}\sqrt{x}(\sqrt{x+a}-\sqrt{x+b})$.
4. $\lim_{x\to a^+}\dfrac{\sqrt{x}-\sqrt{a}+\sqrt{x-a}}{\sqrt{x^2-a^2}}$, $a>0$.

Example 1.5.16. The composition rule can also be applied to one sided limit. For example, we have $\lim_{x\to0^+}f(x^2)=\lim_{x\to0^+}f(x)$ by introducing $y=x^2$ and $x=\sqrt{y}$, and the first condition for the composition rule is satisfied in both directions

$$x\to0,\ x>0 \iff y\to0,\ y>0.$$

We also have $\lim_{x\to0^-}f(x^2)=\lim_{x\to0^+}f(x)$ by

$$x=-\sqrt{y}\to0,\ x<0 \iff y=x^2\to0,\ y>0.$$

Then we may use Proposition 1.5.6 to conclude that $\lim_{x\to0}f(x^2)=\lim_{x\to0^+}f(x)$.

Exercise 1.5.11. Rewrite the limits as $\lim_{x\to c}f(x)$ for suitable c.

1. $\lim_{x\to b^+} f(a-x)$.
2. $\lim_{x\to -\infty} f(x+a)$.
3. $\lim_{x\to 0^+} f(ax+b)$.
4. $\lim_{x\to 0^+} f(\sqrt{x})$.
5. $\lim_{x\to (-1)^-} f(x^2+x)$.
6. $\lim_{x\to a^-} f(x^2+x)$.

Example 1.5.17. We show that

$$\lim_{x\to a} x^p = a^p, \text{ for any } a>0 \text{ and any } p.$$

First consider the special case $a=1$. We find integers M and N satisfying $M<p<N$. Then for $x>1$, we have $x^M<x^p<x^N$. By the arithmetic rule, we have $\lim_{x\to 1^+} x^M = (\lim_{x\to 1^+} x)^M = 1^M = 1$. Similarly, we have $\lim_{x\to 1^+} x^N = 1$. Then by the sandwich rule, we get $\lim_{x\to 1^+} x^p = 1$.

For $0<x<1$, we have $x^M>x^p>x^N$. Again, we have $\lim_{x\to 1^-} x^M = \lim_{x\to 1^-} x^N = 1$ by the arithmetic rule. Then we get $\lim_{x\to 1^-} x^p = 1$ by the sandwich rule.

Combining $\lim_{x\to 1^+} x^p = 1$ and $\lim_{x\to 1^-} x^p = 1$, we get $\lim_{x\to 1} x^p = 1$.

For general $a>0$, we move the problem from a to 1 by introducing $y=\dfrac{x}{a}$

$$\lim_{x\to a} x^p = \lim_{y\to 1} a^p y^p = a^p \lim_{y\to 1} y^p = a^p 1 = a^p.$$

Specifically, we first use $\lim_{x\to 1} x^p = 1$, which we just proved, in the third equality. Then we use the arithmetic rule to get the second equality. Finally, the first equality is obtained by a change of variable, which is essentially the composition rule.

Exercise 1.5.12. Suppose $\lim_{x\to a} f(x) = l > 0$. Prove that $\lim_{x\to a} f(x)^p = l^p$. This extends Exercise 1.1.59 to function limit.

1.5.4 Limit of Trigonometric Function

The sine and tangent functions are defined for $0<x<\dfrac{\pi}{2}$ by Figure 1.5.3. The lengths of line AB, arc BC, and line CD are respectively $\sin x$, x, and $\tan x$. Since the length of line AB is smaller than the length of line BC, which is further smaller than the length of arc BC, we get

$$0<\sin x<x, \text{ for } 0<x<\frac{\pi}{2}.$$

By $\lim_{x\to 0^+} x = 0$ and the sandwich rule, we get

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

Changing x to $y=-x$ (i.e., applying the composition rule) gives the left limit

$$\lim_{x\to 0^-} \sin x = \lim_{y\to 0^+} \sin(-y) = \lim_{y\to 0^+} (-\sin y) = -\lim_{y\to 0^+} \sin y = 0.$$

Combining the two one sided limits at 0, we conclude

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

Then we further get

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

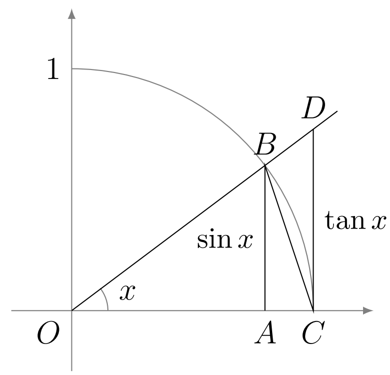


Figure 1.5.3: Trigonometric function.

Note that the area of fan OBC is $\frac{1}{2}x$, and the area of triangle ODC is $\frac{1}{2}\tan x$. Since the fan is contained in the triangle, we get $\frac{1}{2}x < \frac{1}{2}\tan x$, which is the same as $\cos x < \frac{\sin x}{x}$. Combined with $0 < \sin x < x$ obtained before, we get

$$\cos x < \frac{\sin x}{x} < 1, \text{ for } 0 < x < \frac{\pi}{2}.$$

By $\lim_{x\to 0^+}\cos x = 1$ and the sandwich rule, we get $\lim_{x\to 0^+}\frac{\sin x}{x} = 1$. Then changing x to $y = -x$ gives the left limit

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

Therefore we conclude

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

This further implies

$$\begin{aligned}
\lim_{x\to 0}\frac{\cos x - 1}{x^2} &= \lim_{x\to 0}\frac{-2\sin^2\frac{x}{2}}{x^2} = \lim_{y\to 0}\frac{-2\sin^2 y}{(2y)^2} = -\frac{1}{2}\lim_{y\to 0}\left(\frac{\sin y}{y}\right)^2 = -\frac{1}{2},\\
\lim_{x\to 0}\frac{\cos x - 1}{x} &= \lim_{x\to 0}\frac{\cos x - 1}{x^2}x = -\frac{1}{2}\cdot 0 = 0,\\
\lim_{x\to 0}\frac{\tan x}{x} &= \lim_{x\to 0}\frac{\sin x}{x}\frac{1}{\cos x} = 1\cdot\frac{1}{1} = 1.
\end{aligned}$$

We may get the limit of trigonometric functions at any a by using $y = x + a$ to move the limit to be at 0.

$$\begin{aligned}
\lim_{x\to a}\sin x &= \lim_{y\to 0}\sin(a+y) = \lim_{y\to 0}(\sin a\cos y + \cos a\sin y)\\
&= \sin a\lim_{y\to 0}\cos y + \cos a\lim_{y\to 0}\sin y = (\sin a)1 + (\cos a)0 = \sin a,\\
\lim_{x\to a}\cos x &= \lim_{y\to 0}\cos(a+y) = \lim_{y\to 0}(\cos a\cos y - \sin a\sin y)\\
&= \cos a\lim_{y\to 0}\cos y - \sin a\lim_{y\to 0}\sin y = (\cos a)1 + (\sin a)0 = \cos a,\\
\lim_{x\to a}\tan x &= \frac{\lim_{x\to a}\sin x}{\lim_{x\to a}\cos x} = \frac{\sin a}{\cos a} = \tan a, \text{ if } \cos a \neq 0.
\end{aligned}$$

Example 1.5.18. By the arithmetic rule and the composition rule, we have

$$\begin{aligned}
\lim_{x\to\frac{\pi}{2}}\frac{\cos x}{x-\dfrac{\pi}{2}} &= \lim_{y\to 0}\frac{\cos\left(y+\dfrac{\pi}{2}\right)}{y} = \lim_{y\to 0}\frac{-\sin y}{y} = -1,\\
\lim_{x\to\infty} x\sin\frac{1}{x} &= \lim_{y\to 0}\frac{1}{y}\sin y = 1,\\
\lim_{x\to a}\frac{\sin x - \sin a}{x-a} &= \lim_{y\to 0}\frac{\sin(a+y)-\sin a}{y}\\
&= \lim_{y\to 0}\left(\sin a\frac{\cos y - 1}{y} + \cos a\frac{\sin y}{y}\right) = \cos a.
\end{aligned}$$

Exercise 1.5.13. Find the limit.

1. $\lim_{x\to -1}\dfrac{\sin\pi x}{x+1}$.
2. $\lim_{x\to\frac{\pi}{4}}\dfrac{\tan x - 1}{4x-\pi}$.
3. $\lim_{x\to\frac{\pi}{4}}\dfrac{\sin x - \cos x}{4x-\pi}$.
4. $\lim_{x\to\frac{1}{3}}\dfrac{\sin\pi x - \sqrt{3}\cos\pi x}{3x-1}$.
5. $\lim_{x\to\infty} x\tan\dfrac{1}{x}$.
6. $\lim_{x\to\infty} x\left(\cos\dfrac{1}{x} - 1\right)$.
7. $\lim_{x\to a}\dfrac{\cos x - \cos a}{x-a}$.
8. $\lim_{x\to a}\dfrac{\tan x - \tan a}{x-a}$.
9. $\lim_{x\to 0}\dfrac{\cos(2x+1)-\cos(2x-1)}{x^2}$.
10. $\lim_{x\to+\infty} x(\sin\sqrt{x+2} - \sin\sqrt{x})$.
11. $\lim_{x\to 0}\dfrac{\tan(\sin x)}{x}$.
12. $\lim_{x\to 0}\dfrac{\tan(\sin x)}{\sin(\tan x)}$.
13. $\lim_{x\to 0^+}\dfrac{\tan(\sin x)}{\sin\sqrt{x}}$.

Exercise 1.5.14. Study the limit of the sequences $\sin(\sin(\sin\dots a)))$ and $\cos(\cos(\cos\dots a)))$, where the trigonometric functions are applied n times.

1.6 Rigorous Definition of Function Limit

The sequence limit is the behavior as n approaches infinity, which is described by $n > N$ (N is the measurement of bigness). The function limit is the behavior as x approaches a but not equal to a, which may be described by $0 < |x - a| < \delta$ (δ is the measurement of smallness). By replacing $n > N$ with $0 < |x - a| < \delta$ in the rigorous definition of $\lim_{n\to\infty} x_n$, we get the rigorous definition of $\lim_{x\to a} f(x) = l$ for finite a and l.

Definition 1.6.1. A function $f(x)$ converges to a finite number l at a if for any $\epsilon > 0$, there is $\delta > 0$, such that $0 < |x - a| < \delta$ implies $|f(x) - l| < \epsilon$.

The meaning of the definition is given by Figure 1.6. The other limits can be similarly defined. For example, $\lim_{x\to a^+} f(x) = l$ means that for any $\epsilon > 0$, there is $\delta > 0$, such that

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

The limit $\lim_{x\to\infty} f(x) = l$ means that for any $\epsilon > 0$, there is N, such that

$$|x| > N \implies |f(x) - l| < \epsilon.$$

Moreover, $\lim_{x\to a^-} f(x) = +\infty$ means that for any B, there is $\delta > 0$, such that

$$-\delta < x - a < 0 \implies f(x) > B.$$

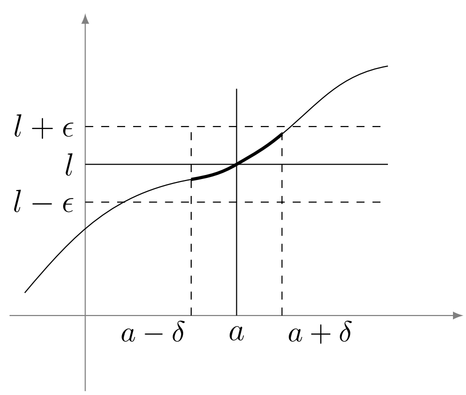


Figure 1.6.1: $0 < |x - a| < \delta$ implies $|f(x) - l| < \epsilon$.

Exercise 1.6.1. Write down the rigorous definitions of $\lim_{x\to a} f(x) = -\infty$, $\lim_{x\to a^-} f(x) = l$, and $\lim_{x\to+\infty} f(x) = -\infty$.

1.6.1 Rigorous Proof of Basic Limits

Example 1.6.1. Here is the rigorous reason for $\lim_{x\to a} c = c$. For any $\epsilon > 0$, take $\delta = 1$. Then

$$0 < |x-a| < \delta = 1 \implies |c-c| = 0 < \epsilon.$$

Here is the rigorous reason for $\lim_{x\to a} x = a$. For any $\epsilon > 0$, take $\delta = \epsilon$. Then

$$0 < |x-a| < \delta = \epsilon \implies |x-a| < \epsilon.$$

Example 1.6.2. To prove $\lim_{x\to 1} x^2 = 1$ rigorously means that, for any $\epsilon > 0$, we need to find suitable $\delta > 0$, such that

$$0 < |x-1| < \delta \implies |x^2-1| < \epsilon.$$

We have

$$0 < |x-1| < \delta \implies |x^2-1| = |x+1||x-1| \le |x+1|\delta.$$

Moreover, when x is close to 1, we expect $x+1$ to be close to 2. Such intuition can be made rigorous by

$$0 < |x-1| < 1 \implies |x+1| \le |x-1| + 2 < 3.$$

Therefore

$$0 < |x-1| < \delta,\ 0 < |x-1| < 1 \implies |x^2-1| \le |x+1|\delta < 3\delta.$$

To complete the rigorous proof, we only need to make sure $\delta \le 1$ and $3\delta \le \epsilon$.

The analysis above suggests the following rigorous proof. For any $\epsilon > 0$, choose $\delta = \min\left\{1, \dfrac{\epsilon}{3}\right\}$. Then

$$\begin{aligned} 0 < |x-1| < \delta &\implies |x-1| < 1,\ 3|x-1| < \epsilon \\ &\implies |x+1| \le |x-1| + 2 < 3,\ 3|x-1| < \epsilon \\ &\implies |x^2-1| = |(x+1)(x-1)| \le 3|x-1| < \epsilon. \end{aligned}$$

Example 1.6.3. To rigorously prove $\lim_{x\to 1} \dfrac{1}{x} = 1$, we note

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

When x is close to 1, we know $|x-1|$ is very small. We also know that $|x|$ is close to 1, so that $\dfrac{1}{|x|}$ can be controlled by a specific bound. Combining the two facts, we see that $\dfrac{1}{|x|}|x-1|$ can be very small. Of course concrete and specific estimation is needed to get a rigorous proof. If $|x-1|<\dfrac{1}{2}$, then $x>\dfrac{1}{2}$ and $\dfrac{1}{|x|}<2$. Therefore if we also have $|x-1|<\dfrac{\epsilon}{2}$, then we get $\dfrac{1}{|x|}|x-1|<\epsilon$.

The analysis above suggests the following rigorous proof. For any $\epsilon>0$, choose $\delta=\min\left\{\dfrac{1}{2},\dfrac{\epsilon}{2}\right\}$. Then

$$\begin{aligned}0<|x-1|<\delta &\implies |x-1|<\frac{1}{2},\ |x-1|<\frac{\epsilon}{2}\\ &\implies x>1-\frac{1}{2}=\frac{1}{2},\ |x-1|<\frac{\epsilon}{2}\\ &\implies \left|\frac{1}{x}-1\right|=\frac{1}{|x|}|x-1|\le 2|x-1|<\epsilon.\end{aligned}$$

Example 1.6.4. We prove

$$\lim_{x\to 0^+} x^p=0,\ \text{for any } p>0.$$

For any $\epsilon>0$, choose $\delta=\epsilon^{\frac{1}{p}}$. Then by $p>0$, we have

$$0<x<\delta \implies |x^p-0|=x^p<\delta^p=\epsilon.$$

We also prove

$$\lim_{x\to +\infty} x^p=+\infty,\ \text{for any } p>0.$$

For any $B>0$, choose $N=B^{\frac{1}{p}}$. Then by $p>0$, we have

$$x>N \implies x^p>N^p=B.$$

Combined with Example 1.5.17, we get

$$\lim_{x\to a}x^p=\begin{cases}a^p, & \text{if } p>0,\ 0<a<+\infty \text{ or } a=0^+,\\ +\infty, & \text{if } p>0,\ a=+\infty.\end{cases}$$

Changing x to $\dfrac{1}{x}$, we get the similar conclusions for $p<0$

$$\lim_{x\to a}x^p=\begin{cases}a^p, & \text{if } p<0,\ 0<a<+\infty,\\ +\infty, & \text{if } p<0,\ a=0^+,\\ 0, & \text{if } p<0,\ a=+\infty.\end{cases}$$

Example 1.6.5. We try to rigorously prove the limit in Example 1.5.5. Example 1.2.6 gives the rigorous proof for the limit of the similar sequence. Instead of copying the proof, we make a slightly different estimation of $|f(x)-l|$

$$\left|\frac{2x^2+x}{x^2-x+1}-2\right| = \left|\frac{3x-2}{x^2-x+1}\right| < \frac{4|x|}{\frac{x^2}{3}} = \frac{12}{|x|}.$$

The crucial inequality is based on the intuition that $|3x-2|<4|x|$ and $|x^2-x+1|>\frac{x^2}{2}$ for sufficiently big x. The first inequality is satisfied when $2<|x|$, and the second is satisfied when $|x|<\frac{x^2}{3}$ and $1<\frac{x^2}{3}$. It is easy to see that $2<|x|$, $|x|<\frac{x^2}{3}$ and $1<\frac{x^2}{3}$ are all satisfied when $|x|>4$. Therefore $|f(x)-l|<\frac{12}{|x|}$ when $|x|>4$.

Formally, for any $\epsilon>0$, choose $N=\max\left\{4, \frac{12}{\epsilon}\right\}$. Then

$$\begin{aligned}
|x|>N &\implies |x|>4,\ |x|>\frac{12}{\epsilon} \\
&\implies |3x-2|<4|x|,\ |x^2-x+1|>\frac{x^2}{3},\ \frac{12}{|x|}<\epsilon \\
&\implies \left|\frac{2x^2+x}{x^2-x+1}-2\right| = \left|\frac{3x-2}{x^2-x+1}\right| < \frac{4|x|}{\frac{x^2}{3}} = \frac{12}{|x|} < \epsilon.
\end{aligned}$$

Exercise 1.6.2. Extend the proof of $\lim_{x\to a} c = c$ and $\lim_{x\to a} x = a$ to the case $a=\pm\infty$.

Exercise 1.6.3. Rigorously prove the limits.

1. $\lim_{x\to 4}\sqrt{x}=2$.
2. $\lim_{x\to 2}\frac{1}{x}=\frac{1}{2}$.
3. $\lim_{x\to\infty}\frac{3x^2-2x+1}{x^2+3x-1}=3$.
4. $\lim_{x\to-\infty}\frac{x+1}{\sqrt{x^2+1}}=-1$.
5. $\lim_{x\to\infty}\frac{x+\sin x}{x+\cos x}=1$.
6. $\lim_{x\to\infty}\frac{\sin x}{x}=0$.

Exercise 1.6.4. Rigorously prove the limits. For the first three limits, give direct proof instead of using Example 1.5.17.

1. $\lim_{x\to a} x^2=a^2$.
2. $\lim_{x\to a}\sqrt{x}=\sqrt{a}$, $a>0$.
3. $\lim_{x\to a}\frac{1}{x}=\frac{1}{a}$, $a\neq 0$.
4. $\lim_{x\to\infty}\frac{x}{x+a}=1$.
5. $\lim_{x\to+\infty}(\sqrt{x+a}-\sqrt{x+b})=0$.
6. $\lim_{x\to\infty}\sqrt{\frac{x+a}{x+b}}=1$.

Exercise 1.6.5. Suppose $f(x) \geq 0$ for x near a and $\lim_{x\to a} f(x) = 0$. Suppose $g(x) \geq c$ for x near a and constant $c > 0$. Prove that $\lim_{x\to a} f(x)^{g(x)} = 0$. This extends Exercise 1.1.58 to the function limit.

1.6.2 Rigorous Proof of Properties of Limit

Example 1.6.6. The arithmetic rule $\lim_{x\to a}(f(x)+g(x)) = \lim_{x\to a} f(x)+\lim_{x\to a} g(x)$ in Proposition 1.5.2 can be proved in the same way as Example 1.2.18. For any $\epsilon > 0$, apply the definition of $\lim_{x\to a} f(x) = k$ and $\lim_{x\to a} g(x) = l$ to $\dfrac{\epsilon}{2} > 0$. We find δ_1 and δ_2, such that

$$\begin{aligned} 0 < |x-a| < \delta_1 &\implies |f(x)-l| < \frac{\epsilon}{2}, \\ 0 < |x-a| < \delta_2 &\implies |g(x)-k| < \frac{\epsilon}{2}. \end{aligned}$$

Then

$$\begin{aligned} &0 < |x-a| < \delta = \min\{\delta_1, \delta_2\} \\ &\implies |f(x)-l| < \frac{\epsilon}{2},\ |g(x)-k| < \frac{\epsilon}{2} \\ &\implies |(f(x)+g(x))-(l+k)| \leq |f(x)-l| + |g(x)-k| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \end{aligned}$$

Example 1.6.7. We prove the extended arithmetic rule $(+\infty)\cdot l = +\infty$ for $l > 0$. This means that, if $\lim_{x\to a} f(x) = +\infty$ and $\lim_{x\to a} g(x) = l > 0$, then $\lim_{x\to a} f(x)g(x) = +\infty$.

For any B, there is $\delta_1 > 0$, such that

$$0 < |x-a| < \delta_1 \implies f(x) > \frac{2}{l}B.$$

For $\epsilon = \dfrac{l}{2} > 0$, there is $\delta_2 > 0$, such that

$$0 < |x-a| < \delta_2 \implies |g(x)-l| < \frac{l}{2} \implies g(x) > \frac{l}{2}.$$

Combining the two implications, we get

$$0 < |x-a| < \delta = \min\{\delta_1, \delta_2\} \implies f(x)g(x) > \frac{2}{l}B \cdot \frac{l}{2} = B.$$

This completes the proof of $(+\infty)\cdot l = +\infty$.

As an application of the extended arithmetic rule, we have

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

In general, we have

$$\lim_{x\to+\infty}(a_nx^n+a_{n-1}x^{n-1}+\cdots+a_1x+a_0)=\begin{cases}+\infty, & \text{if } a_n>0,\\ -\infty, & \text{if } a_n<0.\end{cases}$$

Example 1.6.8. We prove Proposition 1.5.6 about one sided limits.

Assume $\lim_{x\to a} f(x)=l$. Then for any $\epsilon>0$, there is $\delta>0$, such that

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

The implication is the same as the following two implications

$$\begin{aligned} 0<x-a<\delta &\implies |f(x)-l|<\epsilon,\\ -\delta<x-a<0 &\implies |f(x)-l|<\epsilon. \end{aligned}$$

These are exactly the definitions of $\lim_{x\to a^+} f(x)=l$ and $\lim_{x\to a^-} f(x)=l$.

Conversely, assume $\lim_{x\to a^+} f(x)=\lim_{x\to a^-} f(x)=l$. Then for any $\epsilon>0$, there are $\delta_+,\delta_->0$, such that

$$\begin{aligned} 0<x-a<\delta_+ &\implies |f(x)-l|<\epsilon,\\ -\delta_-<x-a<0 &\implies |f(x)-l|<\epsilon. \end{aligned}$$

Therefore

$$\begin{aligned} 0<|x-a|<\min\{\delta_+,\delta_-\} &\implies 0<x-a<\delta_+ \text{ or } -\delta_-<x-a<0\\ &\implies |f(x)-l|<\epsilon. \end{aligned}$$

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

Example 1.6.9. We prove the second case of the composition rule in Proposition 1.5.5. In other words, $\lim_{x\to a} f(x)=b$ and $\lim_{y\to b} g(y)=g(b)$ imply $\lim_{x\to a} g(f(x))=g(b)$.

By $\lim_{y\to b} g(y)=g(b)$, for any $\epsilon>0$, there is $\mu>0$, such that

$$0<|b-y|<\mu \implies |g(y)-g(b)|<\epsilon.$$

Since the right side also holds when $y=b$, we actually have

$$|y-b|<\mu \implies |g(y)-g(b)|<\epsilon. \tag{1.6.1}$$

On the other hand, by $\lim_{x\to a} f(x)=b$, for the $\mu>0$ just found above, there is $\delta>0$, such that

$$0<|x-a|<\delta \implies |f(x)-b|<\mu.$$

Then we get

$$\begin{aligned} 0<|x-a|<\delta &\implies |f(x)-b|<\mu\\ &\implies |g(f(x))-g(b)|<\epsilon. \end{aligned}$$

In the second step, we apply the implication (1.6.1) to $y=f(x)$. This completes the proof that $\lim_{x\to a} g(f(x))=g(b)$.

Example 1.6.10. In Example 1.5.16, we argued that $\lim_{x\to 0} f(x^2) = \lim_{x\to 0^+} f(x)$. Now we give rigorous proof.

Suppose $\lim_{x\to 0} f(x^2) = l$. Then for any $\epsilon > 0$, there is $\delta > 0$, such that

$$0 < |x| < \delta \implies |f(x^2) - l| < \epsilon.$$

Now for any $y > 0$, let $x = \sqrt{y}$. Then

$$0 < y < \delta^2 \implies 0 < x = \sqrt{y} < \delta \implies |f(y) - l| = |f(x^2) - l| < \epsilon.$$

This proves $\lim_{y\to 0^+} f(y) = l$.

On the other hand, suppose $\lim_{x\to 0^+} f(x) = l$. Then for any $\epsilon > 0$, there is $\delta > 0$, such that

$$0 < x < \delta \implies |f(x) - l| < \epsilon.$$

Now for any y, let $x = y^2$. Then

$$0 < |y| < \sqrt{\delta} \implies 0 < x = y^2 < \delta \implies |f(y^2) - l| = |f(x) - l| < \epsilon.$$

This proves $\lim_{y\to 0} f(y^2) = l$.

Exercise 1.6.6. Prove the arithmetic rule $\lim_{x\to a} f(x)g(x) = \lim_{x\to a} f(x) \lim_{x\to a} g(x)$ in Proposition 1.5.2.

Exercise 1.6.7. Prove the sandwich rule in Proposition 1.5.3.

Exercise 1.6.8. Prove the order rule in Proposition 1.5.4.

Exercise 1.6.9. Prove the first case of the composition rule in Proposition 1.5.5.

Exercise 1.6.10. Prove the extended arithmetic rules $(-\infty) + l = -\infty$ and $\dfrac{l}{\infty} = 0$ for function limit.

Exercise 1.6.11. For a subset A of $\mathbb{R}$, define $\lim_{x\in A, x\to a} f(x) = l$ if for any $\epsilon > 0$, there is $\delta > 0$, such that

$$x \in A,\ 0 < |x - a| < \delta \implies |f(x) - l| < \epsilon.$$

Suppose $A \cup B$ contains all the points near a and $\neq a$. Prove that $\lim_{x\to a} f(x) = l$ if and only if $\lim_{x\in A, x\to a} f(x) = l = \lim_{x\in B, x\to a} f(x)$.

Exercise 1.6.12. Prove that $\lim_{x\to a} f(x) = l$ implies $\lim_{x\to a} \max\{f(x), l\} = l$ and $\lim_{x\to a} \min\{f(x), l\} = l$. Can you state and prove the sequence version of the result?

Exercise 1.6.13. Suppose $f(x) \leq 1$ for x near a and $\lim_{x\to a} f(x) = 1$. Suppose $g(x)$ is bounded near a. Prove that $\lim_{x\to a} f(x)^{g(x)} = 1$. What about the case $f(x) \geq 1$?

Exercise 1.6.14. Use Exercises 1.6.12, 1.6.13 and the sandwich rule to prove that, if $\lim_{x\to a} f(x) = 1$ and $g(x)$ is bounded near a, then $\lim_{x\to a} f(x)^{g(x)} = 1$. This is the function version of Exercise 1.1.58. An alternative method for doing the exercise is by extending Proposition 1.1.6.

1.6.3 Relation to Sequence Limit

The sequence limit and the function limit are related.

Proposition 1.6.2. $\lim_{x\to a} f(x) = l$ *if and only if*

$$\lim_{n\to\infty} x_n = a,\ x_n \neq a \implies \lim_{n\to\infty} f(x_n) = l.$$

The necessary direction means that if the whole function converges to l at a, then the restriction of the function to any sequence converging to (but not equal to) a also converges to l. The sufficiency direction means that if all such restrictions converge to l, then the original function converges to l.

The necessary direction can also be considered as a version of the composition rule because x_n can be considered as a function $x(n)$ with n as variable, and $f(x_n)$ is a composition

$$n \mapsto x = x_n \mapsto f(x) = f(x_n).$$

By analogy with the first case of Proposition 1.5.5, we have

$$\lim_{n\to\infty} x_n = a,\ x_n \neq a \text{ (at least for big } n\text{), and } \lim_{x\to a} f(x) = l \implies \lim_{n\to\infty} f(x_n) = l.$$

Example 1.6.11. By taking $a = +\infty$ in Proposition 1.6.2, the limit in Example 1.5.5 implies the limit in Example 1.1.2.

Example 1.6.12. From $\lim_{x\to 0} \dfrac{\sin x}{x} = 1$ and $\lim_{n\to\infty} \dfrac{1}{n} = \lim_{n\to\infty} \dfrac{1}{\sqrt{n}} = 0$, we get

$$\lim_{n\to\infty} n \sin\frac{1}{n} = \lim_{n\to\infty} \frac{\sin\dfrac{1}{n}}{\dfrac{1}{n}} = 1, \quad \lim_{n\to\infty} \sqrt{n}\sin\frac{1}{\sqrt{n}} = \lim_{n\to\infty} \frac{\sin\dfrac{1}{\sqrt{n}}}{\dfrac{1}{\sqrt{n}}} = 1.$$

Example 1.6.13. A consequence of Proposition 1.6.2 is that, if the restrictions of $f(x)$ to two sequences converging to a converge to different limits, then $\lim_{x\to a} f(x)$ diverges.

The *Dirichlet* function is

$$D(x) = \begin{cases} 1, & \text{if } x \in \mathbb{Q}, \\ 0, & \text{if } x \notin \mathbb{Q}. \end{cases}$$

For any a, we have a rational sequence x_n converging to a, and the restriction of $D(x)$ to the sequence is $D(x_n) = 1$, converging to 1. We also have an irrational sequence converging to a, and the restriction of $D(x)$ converges to 0. Therefore $D(x)$ diverges everywhere.

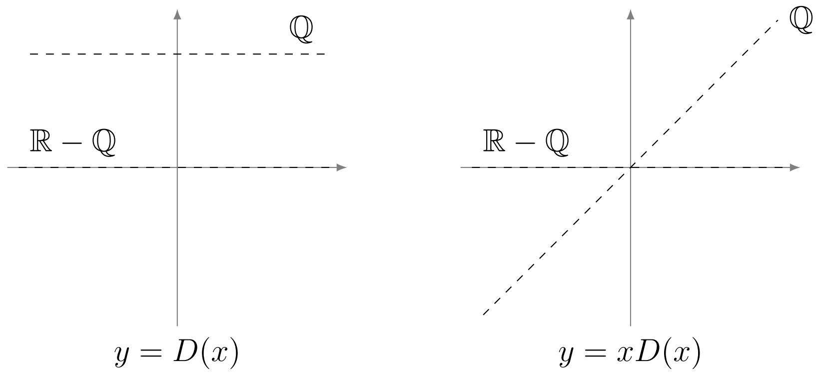


Figure 1.6.2: Dirichlet function.

By the same reason, the function $xD(x)$ diverges at everywhere except 0. On the other hand, we have $-|x| \le |xD(x)| \le |x|$. By $\lim_{x\to 0} |x| = 0$ and the sandwich rule, we have $\lim_{x\to 0} xD(x) = 0$. So $xD(x)$ converges only at 0.

If $\lim_{n\to\infty} f(x_n) = l$ for one (instead of all) sequence x_n converging to a, then it suggests (but does not necessarily imply) that $\lim_{x\to+\infty} f(x) = l$. Sometimes we can "fill the gap" between x_n and derive $\lim_{x\to+\infty} f(x) = l$ in general.

Example 1.6.14. The limit $\lim_{n\to\infty} a^n = 0$ in Example 1.1.11 suggests

$$\lim_{x\to+\infty} a^x = 0, \text{ for } 0 \le a < 1.$$

Here we do not consider $-1 < a < 0$ because a^x is not always defined.

For rigorous proof, we compare a^x with a^n for a natural number n near x. Specifically, for any $x > 1$, we have $n \le x \le n + 1$ for some natural number n. Then $0 \le a < 1$ implies

$$0 \le a^x < a^n.$$

By $\lim_{n\to\infty} a^n = 0$, the sandwich rule should imply $\lim_{x\to+\infty} a^x = 0$. However, we cannot quote the sandwich rule directly because we have a function sandwiched by a sequence. We have to repeat the proof of the sandwich rule.

For any $\epsilon > 0$, by $\lim_{n\to\infty} a^n = 0$, there is N, such that

$$n > N \implies a^n < \epsilon.$$

For $x > N + 1$, let n be a natural number satisfying $n \le x \le n + 1$. Then

$$x > N + 1 \implies n \ge x - 1 > N \implies 0 \le a^x \le a^n < \epsilon.$$

This rigorously proves $\lim_{x\to+\infty} a^x = 0$.

For $a > 1$, let $b = \dfrac{1}{a}$. Then $0 < b < 1$ and by the arithmetic rule, we have

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

Thus we have

$$\lim_{x\to+\infty} a^x = \begin{cases} +\infty, & \text{if } a > 1, \\ 1, & \text{if } a = 1, \\ 0, & \text{if } 0 < a < 1. \end{cases}$$

By using the composition rule to change x to $-x$, we also get

$$\lim_{x\to-\infty} a^x = \begin{cases} 0, & \text{if } a > 1, \\ 1, & \text{if } a = 1, \\ +\infty, & \text{if } 0 < a < 1. \end{cases}$$

We emphasize that we cannot use the following argument at the moment

$$x > N = \frac{\log\epsilon}{\log a} = \log_a \epsilon \implies 0 < a^x < \epsilon.$$

The reason is that our logical foundation only assumes the arithmetic operations and the exponential operation of real numbers. The concept of logarithm must be defined as the inverse operation of the exponential, and will be developed only after we have a theory of inverse functions. See Example 1.7.15.

Exercise 1.6.15. Use the limit in Example 1.2.15 and the idea of Example 1.6.14 to prove that $\lim_{x\to+\infty} x^2a^x = 0$ for $0 < a < 1$. Then use the sandwich rule to prove that $\lim_{x\to+\infty} x^pa^x = 0$ for any p and $0 < a < 1$.

Exercise 1.6.16. State and prove the extended exponential rule

$$a^{+\infty} = \begin{cases} +\infty, & \text{if } a > 1, \\ 0, & \text{if } 0 < a < 1. \end{cases}$$

Example 1.6.15. The limit $\lim_{n\to\infty} \sqrt[n]{n} = 1$ in Example 1.1.8 suggests that

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

By the composition rule, this is the same as

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

We will rigorously prove the second limit.

For any $0 < x < 1$, we have $\dfrac{1}{n+1} \le x \le \dfrac{1}{n}$ for some natural number n. Then

$$\frac{1}{(n+1)^{\frac{1}{n}}} \le \frac{1}{(n+1)^x} \le x^x \le \frac{1}{n^x} \le \frac{1}{n^{\frac{1}{n+1}}}.$$

By Example 1.1.9 (not quite Example 1.1.8), we know

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

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

$$\begin{aligned} n > N &\implies \left|\frac{1}{(n+1)^{\frac{1}{n}}} - 1\right| < \epsilon, \quad \left|\frac{1}{n^{\frac{1}{n+1}}} - 1\right| < \epsilon \\ &\implies \frac{1}{(n+1)^{\frac{1}{n}}} > 1-\epsilon, \quad \frac{1}{n^{\frac{1}{n+1}}} < 1+\epsilon. \end{aligned}$$

Then

$$\begin{aligned} 0 < x < \frac{1}{N+1} &\implies \frac{1}{n+1} \le x \le \frac{1}{n} \text{ for some natural number } n > \frac{1}{x} - 1 > N \\ &\implies 1-\epsilon < \frac{1}{(n+1)^{\frac{1}{n}}} \le x^x \le \frac{1}{n^{\frac{1}{n+1}}} < 1+\epsilon \\ &\implies |x^x - 1| < \epsilon. \end{aligned}$$

Example 1.6.16. We show that

$$\lim_{x\to b} a^x = a^b.$$

For the special case $b = 0$, the limit is $\lim_{x\to 0} a^x = 1$, and is closely related to $\lim_{n\to\infty} \sqrt[n]{a} = 1$ in Example 1.1.7. This suggests us to prove the special case using the idea in Example 1.6.15. See Exercise 1.6.17. Alternatively, we may prove the special case by comparing with $\lim_{x\to 0^+} x^x = 1$ in Example 1.6.15. For sufficiently small $x > 0$, we have $x < a < \dfrac{1}{x}$. This implies

$$x^x < a^x < \frac{1}{x^x} \text{ for sufficiently small } x > 0.$$

Since Example 1.6.15 tells us that $\lim_{x\to 0^+} x^x = 1 = \lim_{x\to 0^+} \dfrac{1}{x^x}$, by the sandwich rule, we get $\lim_{x\to 0^+} a^x = 1$. For $x < 0$, we may use the change of variable $y = -x$ to get

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

Combing the left and right limits gives the special case $\lim_{x\to 0} a^x = 1$.

The general case can be derived from the special case by introducing $x = y + b$

$$\lim_{x\to b} a^x = \lim_{y\to 0} a^{y+b} = \lim_{y\to 0} a^y a^b = a^b \lim_{y\to 0} a^y = a^b \cdot 1 = a^b.$$

Exercise 1.6.17. Prove $\lim_{x\to 0} a^x = 1$ by comparing with $\lim_{n\to\infty} \sqrt[n]{a} = 1$.

Exercise 1.6.18. Prove that $\lim_{n\to\infty} x_n = l$ implies $\lim_{n\to\infty} a^{x_n} = a^l$.

Exercise 1.6.19. Use Exercises 1.1.58 and 1.6.18 to prove that $\lim_{n\to\infty} x_n = l > 0$ and $\lim_{n\to\infty} y_n = k$ imply $\lim_{n\to\infty} x_n^{y_n} = l^k$.

Exercise 1.6.20. Prove that $\lim_{x\to a} f(x) = l$ implies $\lim_{x\to a} b^{f(x)} = b^l$.

Exercise 1.6.21. Use Exercises 1.6.14 and 1.6.20 to prove that $\lim_{x\to a} f(x) = l > 0$ and $\lim_{x\to a} g(x) = k$ imply $\lim_{x\to a} f(x)^{g(x)} = l^k$.

Example 1.6.17. Example 1.3.5 suggests that

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

This is the same as

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

Suppose $n \le x \le n+1$. Then

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

By Example 1.3.5, we have

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

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

$$\begin{aligned} n > N &\implies \left|\left(1+\frac{1}{n+1}\right)^n - e\right| < \epsilon, \quad \left|\left(1+\frac{1}{n}\right)^{n+1} - e\right| < \epsilon, \\ &\implies \left(1+\frac{1}{n+1}\right)^n > e-\epsilon, \quad \left(1+\frac{1}{n}\right)^{n+1} < e+\epsilon. \end{aligned}$$

Then

$$\begin{aligned} x > N+1 &\implies n \le x \le n+1 \text{ for some natural number } n \ge x-1 > N \\ &\implies e-\epsilon < \left(1+\frac{1}{n+1}\right)^n < \left(1+\frac{1}{x}\right)^x < \left(1+\frac{1}{n}\right)^{n+1} < e+\epsilon \\ &\implies \left|\left(1+\frac{1}{x}\right)^x - e\right| < \epsilon. \end{aligned}$$

This proves $\lim_{x\to+\infty}\left(1+\frac{1}{x}\right)^x = e$. Then we use the composition rule to further get

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

Exercise 1.6.22. Find the limits.

1. $\lim_{x\to\infty}\left(1+\frac{a}{x}\right)^x$.
2. $\lim_{x\to\infty}\left(1+\frac{a}{x}\right)^{bx}$.
3. $\lim_{x\to\infty}\left(1+\frac{1}{x^2}\right)^x$.
4. $\lim_{x\to+\infty}\left(1+\frac{1}{x}\right)^{x^2}$.
5. $\lim_{x\to-\infty}\left(1-\frac{1}{x}\right)^{x^2}$.
6. $\lim_{x\to\infty}\left(\frac{x+a}{x+b}\right)^x$.

Exercise 1.6.23. Find the limits.

1. $\left(1-\frac{1}{n}\right)^n$.
2. $\left(1+\frac{1}{2n}\right)^n$.
3. $\left(1+\frac{2}{n}\right)^n$.
4. $\left(1+\frac{a}{n}\right)^n$.
5. $\left(1+\frac{n}{n^2-1}\right)^{n+1}$.
6. $\left(1+\frac{n}{n^2+1}\right)^{n+1}$.
7. $\left(1+\frac{n}{n^2+(-1)^n}\right)^{n+1}$.
8. $\left(1+\frac{1}{n}\right)^{\frac{n^2}{n+1}}$.
9. $\left(1+\frac{2}{n}\right)^{\frac{n^2}{n-1}}$.
10. $\left(1+\frac{(-1)^n}{n^2-1}\right)^n$.
11. $\left(\frac{n+1}{n}\right)^{n+1}$.
12. $\left(\frac{n}{n+(-1)^n}\right)^{(-1)^n n}$.
13. $\left(\frac{n+2}{n-2}\right)^n$.
14. $\left(\frac{n-1}{n}\right)^{\frac{n^2}{n+(-1)^n}}$.
15. $\left(\frac{n-1}{n+2}\right)^{\frac{n^2+(-1)^n n}{n+1}}$.

1.6.4 More Properties of Function Limit

The following is the function version of Theorem 1.3.1.

Proposition 1.6.3. *If $f(x)$ converges at a, then $|f(x)| \le B$ for a constant B and all x near a.*

Like the sequence limit, the example of $\sin\dfrac{1}{x}$ at 0 shows that a bounded function does not necessarily converge. However, Theorem 1.3.2 suggests that a bounded monotone function should converge.

Definition 1.6.4. A function $f(x)$ is *increasing* if

$$x_1 < x_2 \implies f(x_1) \le f(x_2).$$

It is *strictly increasing* if

$$x_1 < x_2 \implies f(x_1) < f(x_2).$$

The concepts of *decreasing* and *strictly decreasing* are similar, and a function is *(strictly) monotone* if it is either (strictly) increasing or (strictly) decreasing.

Theorem 1.6.5. *If $f(x)$ is monotone and bounded on $(a, a+\delta)$, then $\lim_{x\to a^+} f(x)$ converges.*

The theorem also holds for the left limit. The theorem can be proved by choosing a decreasing sequence x_n converging to a. Then $f(x_n)$ is a bounded decreasing sequence. By Theorem 1.3.2, we have $\lim_{n\to\infty} f(x_n) = l$. Then $\lim_{x\to a^+} f(x) = l$ can be proved by comparing with the sequence limit, similar to (actually simpler than) Examples 1.6.15 and 1.6.17.

The Cauchy criterion in Theorem 1.3.3 can also be extended to functions.

Theorem 1.6.6 (Cauchy Criterion)**.** *The limit $\lim_{x\to a} f(x)$ converges if and only if for any $\epsilon > 0$, there is $\delta > 0$, such that*

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

The criterion also holds for one sided limit. Again the proof starts by choosing a sequence x_n converging to a, then applying the sequence version of Cauchy criterion (Theorem 1.3.3) to $f(x_n)$, and then comparing $\lim_{n\to\infty} f(x_n)$ and $\lim_{x\to a} f(x)$.

Exercise 1.6.24. If $f(x)$ is monotone and bounded on $(a-\delta, a)\cup(a, a+\delta)$, does $\lim_{x\to a} f(x)$ converge?

Exercise 1.6.25. Prove that if $\lim_{x\to a} f(x)$ converges, then $f(x)$ satisfies the Cauchy criterion.

1.7 Continuity

A function is continuous at a if its graph is "not broken" at a. For example, the function in Figure 1.7.1 is continuous at a_2 and a_5, and we have $\lim_{x\to a_2} f(x) = f(a_2)$ and $\lim_{x\to a_5} f(x) = f(a_5)$. It is not continuous at the other a_i for various reasons. The function diverges at a_1 and a_7 because it has different left and right limits. The function converges at a_3 but the limit is not $f(a_3)$. The function diverges to infinity at a_4. The function diverges at a_6 because the left limit diverges.

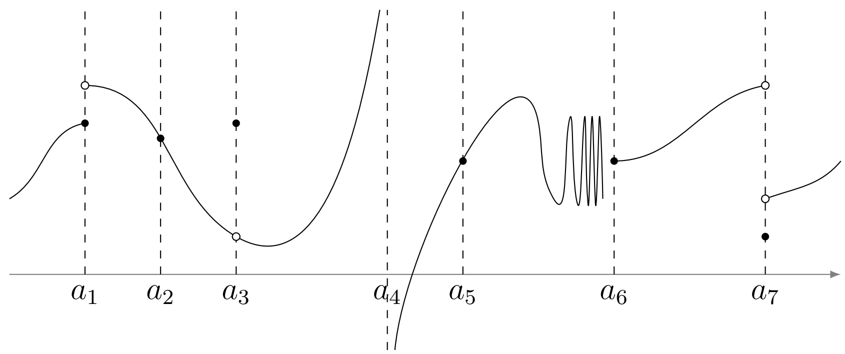


Figure 1.7.1: Continuity and discontinuity.

Definition 1.7.1. A function f is *continuous* at a if $\lim_{x\to a} f(x) = f(a)$.

The left side $\lim_{x\to a} f(x)$ implies that the function should be defined for x near a and $x \neq a$. The right side $f(a)$ implies that the function is also defined at a. Therefore the concept of continuity can only be applied to functions that are defined near a and *including* a, which means all x satisfying $|x-a| < \delta$ for some $\delta > 0$. Then by the definition of $\lim_{x\to a} f(x)$, it is easy to see that $f(x)$ is continuous at a if and only if for any $\epsilon > 0$, there is $\delta > 0$, such that

$$|x-a| < \delta \implies |f(x) - f(a)| < \epsilon.$$

A function $f(x)$ is *right continuous* at a if $\lim_{x\to a^+} f(x) = f(a)$. The definition can be applied to functions defined for x satisfying $a \leq x < a + \delta$ for some $\delta > 0$. Similar remark can be made for the left continuity. A function is continuous at a if and only if it is left and right continuous at a.

The function in Figure 1.7.1 is left continuous at a_1 and right continuous at a_6, but is not continuous at the two points.

A function is continuous on an interval if it is continuous at every point of the interval. For example, a function is continuous on $[0, 1)$ if it is continuous at every $0 < a < 1$ and is also right continuous at 0.

1.7.1 Meaning of Continuity

By Example 1.5.4, polynomials are continuous and rational functions are continuous wherever it is defined. By Examples 1.5.17, the power function x^p is continuous $(0, \infty)$ for all p and, by Example 1.6.4, is right continuous at 0 for $p > 0$. In Section 1.5.4, we find that all trigonometric functions are continuous wherever they are defined. By Example 1.6.16, the exponential function a^x is continuous.

By the properties of limit, we know that the arithmetic combinations and compositions of continuous functions are continuous, wherever the new function is defined.

As remarked after Proposition 1.5.5, if $f(x)$ is continuous at b and $\lim_{x\to a} g(x) = b$, then

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

In other words, the continuity of f means that the limit and the evaluation of f can be exchanged. By using Proposition 1.6.2 (another variant of the composition rule), the same remark can be applied to a sequence limit $\lim_{n\to\infty} x_n = b$ instead of the function limit $\lim_{x\to a} g(x) = b$, and we get

$$\lim_{n\to\infty} f(x_n) = f(b) = f\left(\lim_{n\to\infty} x_n\right).$$

Example 1.7.1. The sign function in Example 1.5.13 is continuous everywhere except at 0. The Dirichlet function $xD(x)$ in Example 1.6.13 is not continuous anywhere. The function $xD(x)$ is continuous at 0 and not continuous at all the other places.

Example 1.7.2. The function $\dfrac{x^3-1}{x-1}$ in Example 1.5.3 is not defined at $x = 1$, and we cannot talk about its continuity at the point. In order to make the function continuous at 1, we need to assign the value of the function at 1 to be $\lim_{x\to 1} \dfrac{x^3-1}{x-1} = 3$, and we get a continuous function

$$f(x) = \begin{cases} \dfrac{x^3-1}{x-1}, & \text{if } x \neq 1 \\ 3, & \text{if } x = 1 \end{cases} = x^2 + x + 1.$$

Example 1.7.3. By the composition rule and the continuity of $\sqrt{x}$, a^x and $\sin x$, we have

$$\begin{aligned} \lim_{x\to 2} \sqrt{x^3+1} &= \sqrt{\lim_{x\to 2} x^3 + 1} = 3, \\ \lim_{n\to\infty} a^{\frac{n}{n^2-1}} &= a^{\lim_{n\to\infty} \frac{n}{n^2-1}} = 1, \\ \lim_{x\to 1^-} \sin\sqrt{1-x^2} &= \sin\left(\lim_{x\to 1^-} \sqrt{1-x^2}\right) = \sin\sqrt{\lim_{x\to 1^-}(1-x^2)} = 0. \end{aligned}$$

Example 1.7.4. The continuity of $\sqrt{x}$ implies that, if $\lim_{n\to\infty} x_n = l > 0$, then $\lim_{n\to\infty} \sqrt{x_n} = \sqrt{l}$. This is Example 1.2.17.

The continuity of a^x implies that, if $\lim_{n\to\infty} x_n = l$, then $\lim_{n\to\infty} a^{x_n} = a^l$. This implies the limit $\lim_{n\to\infty} \sqrt[n]{a} = 1$ in Example 1.2.14, and also implies the limits such as $\lim_{n\to\infty} a^{\frac{1}{\sqrt{n}}} = 1$ and $\lim_{n\to\infty} a^{\frac{\sqrt{n}+1}{\sqrt{n}-\sin n}} = a$.

Exercise 1.7.1. Determine the intervals on which the function is continuous. Is it possible to extend to a continuous function at more points?

1. $\dfrac{x^2-3x+2}{x^2-1}$.
2. $\dfrac{x^2-1}{x-1}$.
3. $\operatorname{sign}(x)$.
4. $x\sin\dfrac{1}{x}$.
5. x^x.
6. $\dfrac{\cos x}{2x-\pi}$.

Exercise 1.7.2. Find a function on $\mathbb{R}$ that is continuous at $1, 2, 3$ and is not continuous at all the other places.

Exercise 1.7.3. Find a function on $\mathbb{R}$ that is not continuous at $1, 2, 3$ and is continuous at all the other places.

Exercise 1.7.4. Find two continuous functions $f(x)$ and $g(x)$, such that $\lim_{x\to 0} \dfrac{1+f(0)g(x)}{1+f(x)g(0)}$ converges but the value is not 1.

1.7.2 Intermediate Value Theorem

If we start at the sea level and climb to the mountain top of 1000 meters, then we will be at 500 meters somewhere along the way, and will be at 700 meter some other place. This is the intuition behind the following result.

Theorem 1.7.2 (Intermediate Value Theorem). *If $f(x)$ is continuous on $[a, b]$, then for any number γ between $f(a)$ and $f(b)$, there is $c \in [a, b]$ satisfying $f(c) = \gamma$.*

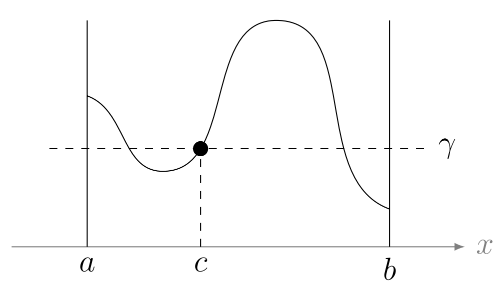


Figure 1.7.2: Intermediate value theorem.

Example 1.7.5. The polynomial $f(x) = x^3 - 3x + 1$ is continuous and satisfies $f(0) = 1, f(1) = -1$. Therefore $f(x)$ must attain value 0 somewhere on the interval $[0, 1]$. In other words, the polynomial has at least one root on $(0, 1)$.

To find more precise location of the root, we may try to evaluate the function at $0.1, 0.2, \dots, 0.9$ and find $f(0.3) = 0.727, f(0.4) = -0.136$. This tells us that $f(x)$ has a root on $(0.3, 0.4)$.

The discussion can be summarized as follows: If f is a continuous function on $[a, b]$, such that $f(a)$ and $f(b)$ have opposite signs, then f has at least one root in (a, b).

Example 1.7.6. By Example 1.6.7, we have $\lim_{x\to+\infty}(x^3 - 3x + 1) = +\infty$ and $\lim_{x\to-\infty}(x^3 - 3x + 1) = -\infty$. Therefore for sufficiently big $b > 0$, we have $f(b) > 0$ and $f(-b) < 0$. By the Intermediate Value Theorem, the polynomial has a root on $(-b, b)$.

In general, any odd order polynomial has at least one real root.

Back to $f(x) = x^3-3x+1$. In Example 1.7.5, we actually already know that $f(x)$ has at least one root on $(0, 1)$. In fact, by $f(-b) < 0, f(0) > 0, f(1) < 0, f(b) > 0$, we know f has at least one root on each of the intervals $(-b, 0)$, $(0, 1)$, $(1, b)$. Since a polynomial of order 3 has at most three roots, we conclude that $f(x)$ has exactly one root on each of the three intervals.

Example 1.7.7. We know

$$\lim_{x\to\frac{\pi}{2}^-} \tan x = \frac{\lim_{x\to\frac{\pi}{2}^-} \sin x}{\lim_{x\to\frac{\pi}{2}^-} \cos x} = \frac{1}{0^+} = +\infty,$$

$$\lim_{x\to-\frac{\pi}{2}^+} \tan x = \frac{\lim_{x\to\frac{\pi}{2}^+} \sin x}{\lim_{x\to\frac{\pi}{2}^+} \cos x} = \frac{-1}{0^+} = -\infty.$$

Therefore for any number γ, we can find $a > -\dfrac{\pi}{2}$ and very close to $-\dfrac{\pi}{2}$, such that $\tan a < \gamma$. We can also find $b < \dfrac{\pi}{2}$ and very close to $\dfrac{\pi}{2}$, such that $\tan b > \gamma$. Then $\tan x$ is continuous on $[a, b]$ and $\tan a < \gamma < \tan b$. This implies that $\gamma = \tan c$ for some $c \in (a, b)$. Therefore any number is the tangent of some angle between $-\dfrac{\pi}{2}$ and $\dfrac{\pi}{2}$.

The example shows that, if $f(x)$ is continuous on (a, b) satisfies $\lim_{x\to a^+} f(x) = -\infty$ and $\lim_{x\to b^-} f(x) = +\infty$, then $f(x)$ can take any number as value on (a, b). Note that the interval (a, b) here does not even have to be bounded.

Example 1.7.8. The function

$$f(x) = \begin{cases} x, & \text{if } -1 \le x \le 0, \\ x^2 + 1, & \text{if } 0 < x \le 1, \end{cases}$$

satisfies $f(-1) = -1, f(1) = 2$, but does not take any number in $(0, 1]$ as value. The problem is that the function is not continuous at 0, where a jump in value misses the numbers in $(0, 1]$. Therefore the Intermediate Value Theorem cannot be applied.

Exercise 1.7.5. Let $f(x) : [0, 1] \to [0, 1]$ be a continuous function. Prove that there exists at least one $c \in [0, 1]$ satisfying $f(c) = c$.

Exercise 1.7.6. Find all the possible values.

1. $\dfrac{x^2 - 3x + 2}{x^2 - 1}$.
2. x^x.
3. e^x.
4. $\sin x$.
5. $\sin \dfrac{1}{x}$.
6. $\begin{cases} 2^x, & \text{if } -1 \le x \le 0, \\ x^2 + 3, & \text{if } 0 < x \le 1. \end{cases}$

Exercise 1.7.7. $\lim_{x \to +\infty} \cos(\sqrt{x+2} + \sqrt{x})$ and $\lim_{x \to +\infty} \sqrt{x}(\sin \sqrt{x+2} - \sin \sqrt{x})$ diverge.

1.7.3 Continuous Inverse Function

Given a function $f(x)$, its *inverse function* $f^{-1}(y)$ is obtained by solving $f(x) = y$ for x.

Example 1.7.9. To find the inverse of $f(x) = 3x - 2$, we solve $3x - 2 = y$ and get $x = \frac{1}{3}y + \frac{2}{3}$. Therefore the inverse function is $f^{-1}(y) = \frac{1}{3}y + \frac{2}{3}$.

Example 1.7.10. To find the inverse of $f(x) = x^2$, we try to solve $x^2 = y$.

The problem is that the equation has no solution for $y < 0$ and has two solutions for $y > 0$. The ambiguity on which of the two solutions to choose can be removed if we additionally specify $x \ge 0$ or $x \le 0$. In other words, in order to unambiguously specify an inverse of $f(x) = x^2$, we must specify the ranges for x and for y.

If we consider

$$f_1(x) = x^2 \colon [0, +\infty) \to [0, +\infty),$$

which means that we specify $x \ge 0$ and $y \ge 0$, then $x^2 = y$ always has unique non-negative solution, which is usually denoted $x = f_1^{-1}(y) = \sqrt{y}$.

If instead we consider

$$f_2(x) = x^2 \colon (-\infty, 0] \to [0, +\infty),$$

then $x^2 = y$ also has a unique non-positive solution and gives the inverse

$$f_2^{-1}(y) = -\sqrt{y} \colon [0, +\infty) \to (-\infty, 0].$$

As a more elaborate example, the function

$$f_3(x) = x^2 \colon (-\infty, -1] \cup [0, 1) \to [0, +\infty)$$

has inverse

$$f_3^{-1}(y) = \begin{cases} \sqrt{y}, & \text{if } 0 \le y < 1 \\ -\sqrt{y}, & \text{if } y \ge 1 \end{cases} \colon [0, +\infty) \to (-\infty, -1] \cup [0, 1).$$

The examples show that, for the concept of inverse function to be unambiguous, we have to specify the ranges for the variable and the value. In this regard, if two functions have the same formula but different ranges, then we should really think of them as different functions.

In general, if a function $f(x)$ is defined for all $x \in D$, then D is the *domain* of the function, and all the values of $f(x)$ is the *range*

$$R = \{f(x) \colon x \in D\}.$$

With the domain and range explicitly specified, we express the function as a *map*

$$f(x) \colon D \to R.$$

Now the equation $f(x) = y$ has solution only when $y \in R$. Moreover, we need to make sure that the solution is unique in order for the inverse to be unambiguous. This means that the function is *one-to-one*

$$x_1, x_2 \in D,\ x_1 \ne x_2 \implies f(x_1) \ne f(x_2).$$

The condition is the same as

$$x_1, x_2 \in D,\ f(x_1) = f(x_2) \implies x_1 = x_2.$$

Example 1.7.11. Consider the function $f(x) = x^5 + 3x^3 + 1$ defined for all x. By the remark in Example 1.7.7, together with (see Example 1.6.7)

$$\lim_{x \to -\infty} f(x) = -\infty, \quad \lim_{x \to +\infty} f(x) = +\infty,$$

we see that any number can be the value of $f(x)$. This shows that the range of f is $\mathbb{R}$.

Is the function one-to-one? This can be established as follows. If $x_1 \ne x_2$, then either $x_1 < x_2$ or $x_1 > x_2$. In the first case, we have

$$\begin{aligned} x_1 < x_2 &\implies x_1^5 < x_2^5,\ x_1^3 < x_2^3 \\ &\implies f(x_1) = x_1^5 + 3x_1^3 + 1 < f(x_2) = x_2^5 + 3x_2^3 + 1. \end{aligned}$$

By switching the roles of x_1 and x_2, we get $x_1 > x_2$ implying $f(x_1) > f(x_2)$. Either way, we get $x_1 \ne x_2$ implying $f(x_1) \ne f(x_2)$.

We conclude that $f(x) = x^5 + 3x^3 + 1 \colon \mathbb{R} \to \mathbb{R}$ is invertible.

The argument in the example can be generalized. Suppose $f(x)$ is a *continuous* function defined on an *interval* domain D. Then the Intermediate Value Theorem implies that the range R is also an interval. Moreover, the uniqueness of the solution to $f(x) = y$ can be obtained if $f(x)$ is strictly increasing or strictly decreasing.

Theorem 1.7.3. *A strictly increasing and continuous function on an interval is invertible, and the inverse function is also strictly increasing and continuous. The same holds for strictly decreasing and continuous functions.*

It is remarkable that, according to the theorem, we get the continuity of the inverse function for free. For example, although it is impossible to find the formula for the solution of $x^5 + 3x^3 + 1 = y$, we already know the solution is a continuous function of y,

Example 1.7.12. The sine function can take any value in $[-1, 1]$. To make sure it is one-to-one, we specify the domain and range

$$\sin x\colon \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \to [-1, 1].$$

This is strictly increasing and continuous, and takes any number in $[-1, 1]$ as value. Therefore we get the *inverse sine function*

$$\arcsin y\colon [-1, 1] \to \left[-\frac{\pi}{2}, \frac{\pi}{2}\right].$$

The inverse sine function is also strictly increasing and continuous.

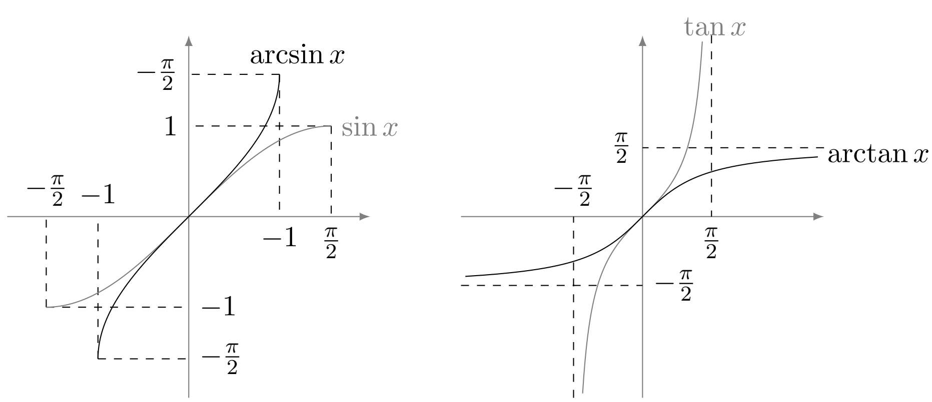


Figure 1.7.3: Inverse trigonometric functions.

Example 1.7.13. The cosine function

$$\cos x\colon [0, \pi] \to [-1, 1]$$

is strictly decreasing, continuous, and takes any number in $[-1,1]$ as value. Therefore we get the *inverse cosine function*

$$\arccos x\colon [-1,1] \to [0,\pi],$$

which is also strictly decreasing and continuous.

However, the inverse cosine function is not much a new function. If $x = \arccos y$, $x \in [0,\pi]$, $y \in [-1,1]$, then

$$\sin\left(\frac{\pi}{2} - x\right) = \cos x = y, \quad \frac{\pi}{2} - x \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$$

shows that $\frac{\pi}{2} - x = \arcsin y$. Therefore we have the equality

$$\arcsin y + \arccos y = \frac{\pi}{2}.$$

Example 1.7.14. The tangent function

$$\tan x\colon \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \to (-\infty, +\infty)$$

is strictly increasing, continuous, and by Example 1.7.7, takes any number as the value. Therefore we have the *inverse tangent function*

$$\arctan y\colon (-\infty, +\infty) \to \left(-\frac{\pi}{2}, \frac{\pi}{2}\right),$$

which is also strictly increasing and continuous.

We claim that

$$\lim_{y\to+\infty} \arctan y = \frac{\pi}{2}, \quad \lim_{y\to-\infty} \arctan y = -\frac{\pi}{2}.$$

For any $\frac{\pi}{2} > \epsilon > 0$, let $N = \tan\left(\frac{\pi}{2} - \epsilon\right)$. Then $\frac{\pi}{2} - \epsilon = \arctan N$, and the strict increasing property of arctan implies

$$y > N \implies \frac{\pi}{2} > \arctan y > \arctan N = \frac{\pi}{2} - \epsilon \implies \left|\arctan y - \frac{\pi}{2}\right| < \epsilon.$$

This proves the first limit. The second limit can be proved similarly.

Example 1.7.15. For $a > 1$, the exponential function

$$a^x\colon \mathbb{R} \to (0, +\infty)$$

is strictly increasing and continuous. Moreover, by Example 1.6.14, we have

$$\lim_{x\to-\infty} a^x = 0, \quad \lim_{x\to+\infty} a^x = +\infty.$$

Then by an argument similar to Example 1.7.7, any number in $(0, +\infty)$ is a value of a^x. Therefore the exponential function has an inverse

$$\log_a x : (0, +\infty) \to \mathbb{R},$$

called the *logarithmic function* with base a. Like the exponential function, the logarithmic function is strictly increasing and continuous. We can also show

$$\lim_{x\to+\infty} \log_a x = +\infty, \quad \lim_{x\to 0^+} \log_a x = -\infty, \quad \text{for } a > 1$$

by method similar to Example 1.7.13.

The logarithm $\log_e x$ based on the special value e is called the *natural logarithm*. We will denote the natural logarithm simply by $\log x$.

For $0 < a < 1$, the exponential a^x is strictly decreasing and continuous. The corresponding logarithm can be similarly defined and is also strictly decreasing and continuous. Moreover, we have

$$\lim_{x\to+\infty} \log_a x = -\infty, \quad \lim_{x\to 0^+} \log_a x = +\infty, \quad \text{for } 0 < a < 1.$$

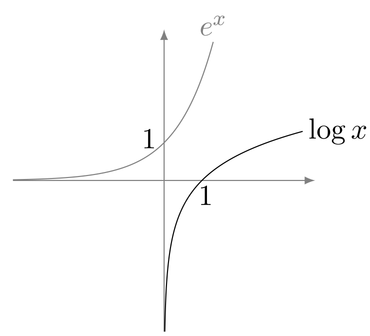


Figure 1.7.4: Exponential and logarithm.

Exercise 1.7.8. Let f be a strictly increasing function. Show that its inverse is also strictly increasing.

1.7.4 Continuous Change of Variable

Suppose $f(x)$ is continuous and strictly increasing near a. Then by Theorem 1.7.3, $f(x)$ can be inverted near a, and the inverse $f^{-1}(y)$ is also continuous and strictly increasing near $b = f(a)$. The continuity of f and f^{-1} means that

$$x \to a \iff y \to b.$$

The one-to-one property implies that

$$x \neq a \iff y \neq b.$$

Therefore the composition rule can be applied in both directions, and we have

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

Here the equality means that the convergence of both sides are equivalent, and the limits have the same value.

Example 1.7.16. Since $\sin x$ is strictly increasing and continuous near 0, we have

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

The same argument also tells us

$$\lim_{y\to 0}\frac{\arctan y}{y} = \lim_{x\to 0}\frac{x}{\tan x} = 1.$$

Example 1.7.17. In Examples 1.5.11, 1.5.16, 1.6.10, we find $\lim_{x\to a} f(x^2) = \lim_{x\to a^2} f(x)$ for $a \neq 0$ and $\lim_{x\to 0} f(x^2) = \lim_{x\to 0^+} f(x)$. Here we explain the two equalities from the viewpoint of continuous change of variable.

The function

$$x^2\colon [0,+\infty) \to [0,+\infty)$$

is strictly increasing and continuous, and is therefore invertible, with strictly increasing and continuous inverse

$$\sqrt{x}\colon [0,+\infty) \to [0,+\infty).$$

The continuous change of variable implies that $\lim_{x\to a} f(x^2) = \lim_{x\to a^2} f(x)$ for $a > 0$ and $\lim_{x\to 0^+} f(x^2) = \lim_{x\to 0^+} f(x)$. Note that the second equality makes use of the right continuity of x^2 and $\sqrt{x}$ at 0.

Similarly, the function

$$x^2\colon (-\infty, 0] \to [0,+\infty)$$

is strictly decreasing and continuous, and is therefore invertible, with strictly decreasing and continuous inverse. This implies that $\lim_{x\to a} f(x^2) = \lim_{x\to a^2} f(x)$ for $a < 0$ and $\lim_{x\to 0^-} f(x^2) = \lim_{x\to 0^+} f(x)$.

By

$$\lim_{x\to 0^+} f(x^2) = \lim_{x\to 0^+} f(x) = \lim_{x\to 0^-} f(x^2),$$

we also conclude that $\lim_{x\to 0} f(x^2) = \lim_{x\to 0^+} f(x)$.

Example 1.7.18. Since the natural logarithm $\log x$ is continuous, we may move the limit from outside the logarithm to inside the logarithm

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

Here the third equality follows from Example 1.6.17.

Exercise 1.7.9. Find the limits.

1. $\lim_{x\to 0} \dfrac{\log_a(x+1)}{x}$.
2. $\lim_{x\to a} \dfrac{\log x - \log a}{x-a}$.
3. $\lim_{x\to a} \dfrac{\log_b x - \log_b a}{x-a}$.

Exercise 1.7.10. Find the limits.

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

Exercise 1.7.11. Find the limits.

1. $\lim_{x\to 0} \dfrac{1}{x}\log\dfrac{ax+c}{bx+c}$.
2. $\lim_{x\to 0} \dfrac{1}{x}\log\dfrac{ax+b}{cx+d}$.
3. $\lim_{x\to 0} \dfrac{1}{x}\log\dfrac{a_2x^2+a_1x+a_0}{b_2x^2+b_1x+b_0}$.
4. $\lim_{x\to\infty} x\log\dfrac{ax+b}{ax+c}$.
5. $\lim_{x\to\infty} x\log\dfrac{ax+b}{cx+d}$.
6. $\lim_{x\to\infty} x\log\dfrac{a_2x^2+a_1x+a_0}{b_2x^2+b_1x+b_0}$.

Exercise 1.7.12. Find the limits.

1. $\lim_{x\to 0} \dfrac{x}{\log(ax+1)}$.
2. $\lim_{x\to 1} \dfrac{x^2-1}{\log x}$.
3. $\lim_{x\to 0} \dfrac{\log(x^2+3x+2)-\log 2}{x}$.
4. $\lim_{x\to 1} \dfrac{\log x}{\sin \pi x}$.

Example 1.7.19. Since $y = e^x - 1$ is strictly increasing and continuous, with inverse $x = \log(y+1)$, we may change the variable and use Example 1.7.18 to get

$$\lim_{x\to 0} \frac{e^x-1}{x} = \lim_{y\to 0} \frac{y}{\log(y+1)} = 1.$$

Exercise 1.7.13. Find the limits.

1. $\lim_{x\to 0} \dfrac{e^{2x}-1}{x}$.
2. $\lim_{x\to 0} \dfrac{a^x-1}{x}$.
3. $\lim_{x\to a} \dfrac{e^x-e^a}{x-a}$.
4. $\lim_{x\to b} \dfrac{a^x-a^b}{x-b}$.
5. $\lim_{x\to\infty} \dfrac{e^x-1}{x}$.
6. $\lim_{x\to 0} \dfrac{a^x-b^x}{x}$.
7. $\lim_{x\to 0} \dfrac{e^{ax}-e^{bx}}{x}$.
8. $\lim_{x\to+\infty} \dfrac{e^{ax}-e^{bx}}{x}$.
9. $\lim_{x\to 0} \dfrac{a^x-b^x}{c^x-d^x}$.

Example 1.7.20. The continuity of the logarithm implies the continuity of the function $x^x = e^{x\log x}$ on $(0,\infty)$. In general, if $f(x)$ and $g(x)$ are continuous and $f(x) > 0$, then $f(x)^{g(x)} = e^{g(x)\log f(x)}$ is continuous.

The continuity of the exponential and the logarithm can also be used to prove that

$$\lim_{n\to\infty} a_n = l > 0,\ \lim_{n\to\infty} b_n = k \implies \lim_{n\to\infty} a_n^{b_n} = l^k.$$

The reason is that the continuity of log implies $\lim_{n\to\infty} \log a_n = \log l$. Then the arithmetic rule implies $\lim_{n\to\infty} b_n \log a_n = k \log l$. Finally, the continuity of the exponential implies

$$\lim_{n\to\infty} a_n^{b_n} = \lim_{n\to\infty} e^{b_n \log a_n} = e^{\lim_{n\to\infty} b_n \log a_n} = e^{k\log l} = l^k.$$

In Exercises 1.6.19 and 1.6.21, we took a number of steps to prove the same property, without using the logarithmic function.

Example 1.7.21. For $p \neq 0$, $y = p\log(x+1)$ is strictly increasing and continuous. Using change the variable and Examples 1.7.18 and 1.7.19, we get

$$\begin{aligned}
\lim_{x\to 0} \frac{(x+1)^p - 1}{x} &= \lim_{x\to 0} \frac{e^{p\log(x+1)} - 1}{x} = \lim_{x\to 0} \frac{e^{p\log(x+1)} - 1}{p\log(x+1)}\, p\, \frac{\log(x+1)}{x} \\
&= p \lim_{x\to 0} \frac{e^{p\log(x+1)} - 1}{p\log(x+1)} \lim_{x\to 0} \frac{\log(x+1)}{x} \\
&= p \lim_{y\to 0} \frac{e^y - 1}{y} \lim_{x\to 0} \frac{\log(x+1)}{x} = p\cdot 1\cdot 1 = p.
\end{aligned}$$

Exercise 1.7.14. Find the limits.

1. $\lim_{x\to 1} \dfrac{x^p - 1}{x-1}$.
2. $\lim_{x\to 1} \dfrac{x^p - 1}{x^q - 1}$.
3. $\lim_{x\to a} \dfrac{x^p - a^p}{x-a}$.
4. $\lim_{x\to a} \dfrac{x^p - a^p}{x^q - a^q}$.
5. $\lim_{x\to a} \dfrac{\sin x^p - \sin a^p}{x-a}$.
6. $\lim_{x\to a} \dfrac{e^{x^p} - e^{a^p}}{b^x - b^a}$.

Exercise 1.7.15. Let

$$x_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n} - \log n, \quad y_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n} - \log(n+1).$$

1. Use Exercise 1.3.17 to prove that $\dfrac{1}{1+n} < \log\left(1 + \dfrac{1}{n}\right) < \dfrac{1}{n}$.
2. Prove that x_n is strictly decreasing and y_n is strictly increasing.
3. Prove that both x_n and y_n converge to the same limit

$$\lim_{n\to\infty} \left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} - \log n\right) = 0.577215669015328\cdots.$$

The number is called the *Euler*[1]*-Mascheroni*[2] *constant*.

[1]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.

[2]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.

Chapter 2

Differentiation

2.1 Linear Approximation

The basic idea of differentiation is solving problems by using simple functions to approximate general complicated functions. The simplest functions are the constant functions, which are usually too primitive to be useful. More effective approximations are given by linear functions $A + Bx$.

Definition 2.1.1. A *linear approximation* of a function $f(x)$ at x_0 is a linear function $L(x) = a + b(x - x_0)$, such that for any $\epsilon > 0$, there is $\delta > 0$, such that

$$|x - x_0| < \delta \implies |f(x) - L(x)| = |f(x) - a - b(x - x_0)| \le \epsilon |x - x_0|.$$

A function is *differentiable* if it has a linear approximation.

The differentiability at x_0 requires the function to be defined on a neighborhood of x_0, and the linear approximation depends only on the function near x_0.

In everyday life, we use approximations all the time. For example, when we measure certain distance and get 7 meters and 5 centimeters, we really mean give or take some millimeters. So the real distance might be 7.052 meters or 7.046 meters. The function $f(x)$ is like the real distance (7.052 meters or 7.046 meters), and the linear function $L(x)$ is like the reading (7.05 meters) from the ruler.

The accuracy of the measurement depends on how refined the ruler is. We often use the rulers with two units m and cm. The centimeter cm is smaller among the two units and is therefore the "basic unit" that gives the accuracy of the ruler. The *error* $|7m5.2cm - 7m5cm| = 0.2cm$ between the real distance and the measurement should be significantly smaller than the basic unit $1cm$.

Analogously, the linear function $L(x) = a \cdot 1 + b \cdot (x - x_0)$ is a combination of two units 1 and $x - x_0$. Since the approximation happens for x near x_0, $x - x_0$ is much smaller than 1 and is therefore the "basic unit". The *error* $|f(x) - L(x)|$ of the approximation should be significantly smaller than the size $|x - x_0|$ of the basic unit, which exactly means $\le \epsilon|x - x_0|$ on the right side of the definition.

2.1.1 Derivative

Geometrically, a function may be represented by its graph. The graph of a linear function is a straight line. Therefore a linear approximation at x_0 is a straight line that "best fits" the graph of the given function near x_0. This is the *tangent line* of the function.

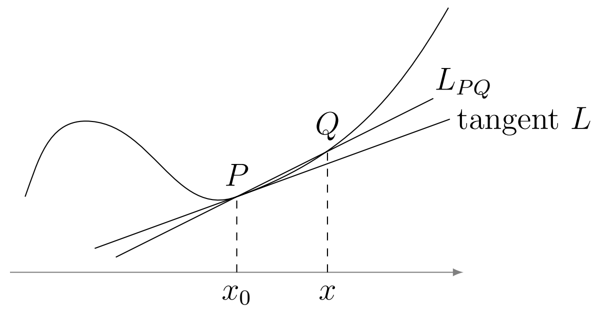


Figure 2.1.1: The linear approximation is the tangent line.

Specifically, the point P in Figure 2.1.1 is the point $(x_0, f(x_0))$ on the graph of $f(x)$. We pick a nearby point $Q = (x, f(x))$ on the graph, for x near x_0. The straight line connecting P and Q is the linear function (the variable in L_{PQ} is t because x is already used for Q)

$$L_{PQ}(t) = f(x_0) + \frac{f(x) - f(x_0)}{x - x_0}(t - x_0).$$

As $x \to x_0$, Q approaches P, and the linear function approaches $L(t) = a + b(t - x_0)$. Therefore we have $a = f(x_0)$, and b is given below.

Definition 2.1.2. The *derivative* of a function $f(x)$ at x_0 is

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

We emphasize that the linear approximation is the concept. As the coefficient b of the first order term, the derivative $f'(x_0)$ is the computation of the concept. The following says that the concept and its computation are equivalent.

Proposition 2.1.3. *A function $f(x)$ is differentiable at x_0 if and only if the derivative $f'(x_0)$ exists. Moreover, the linear approximation is given by $f(x_0) + f'(x_0)(x - x_0)$.*

The notation f' for the derivative is due to Joseph Louis Lagrange. It is simple and convenient, but could become ambiguous when there are several variables related in more complicated ways. Another notation $\dfrac{df}{dx}$, due to Gottfried Wilhelm Leibniz, is less ambiguous. So we also write

$$\left.\frac{df}{dx}\right|_{x=x_0} = \lim_{x\to x_0}\frac{f(x)-f(x_0)}{x-x_0} = \lim_{\Delta x\to 0}\frac{\Delta f}{\Delta x}, \quad \Delta f = f(x)-f(x_0),\ \Delta x = x-x_0.$$

We emphasize that Leibniz's notation is not the "quotient" of two quantities df and dx. It is an integrated notation that alludes to the fact that the derivative is the limit of the quotient $\dfrac{\Delta f}{\Delta x}$ of differences.

Example 2.1.1. The function $f(x) = 3x - 2$ is already linear. So its linear approximation must be $L(x) = f(x) = 3x - 2$. This reflects the intuition that, if the distance is exactly $7m5cm$, then the measure by the ruler in centimeters should be $7m5cm$. In particular, the derivative $f'(x) = 3$, or $\dfrac{d(3x-2)}{dx} = 3$. In general, we have $(A + Bx)' = B$.

Example 2.1.2. To find the linear approximation of x^2 at 1, we rewrite the function in terms of $x - 1$

$$x^2 = (1 + (x-1))^2 = 1 + 2(x-1) + (x-1)^2.$$

Note that $L(x) = 1 + 2(x-1)$ is linear, and the error $|x^2 - L(x)| = (x-1)^2$ is significantly smaller than $|x-1|$ when x is near 1

$$|x-1| < \delta = \epsilon \implies |x^2 - L(x)| \le \epsilon|x-1|.$$

Therefore $1 + 2(x-1)$ is the linear approximation of x^2 at 1, and the derivative $(x^2)'|_{x=1} = \left.\dfrac{d(x^2)}{dx}\right|_{x=1} = 2.$

Exercise 2.1.1. Find the linear approximations and then the derivatives.

1. $5x + 3$ at x_0.
2. $x^3 - 2x + 1$ at 1.
3. x^2 at x_0.
4. x^3 at x_0.
5. x^n at 1.
6. x^n at x_0.

Exercise 2.1.2. Interpret the limits as derivatives.

1. $\lim_{x\to 0}\dfrac{(1+x)^p - 1}{x}$.
2. $\lim_{x\to 0}\dfrac{\sqrt{x+9} - 3}{x}$.
3. $\lim_{x\to \frac{\pi}{2}}\dfrac{\cos x}{x - \frac{\pi}{2}}$.
4. $\lim_{x\to \pi}\dfrac{\sin x}{x - \pi}$.
5. $\lim_{x\to 0}\dfrac{\arcsin x}{x}$.
6. $\lim_{x\to 2}\dfrac{1}{x-2}\log\dfrac{x}{2}$.

2.1.2 Basic Derivative

We derive the derivatives of the power function, the exponential function, the logarithmic function, and the trigonometric functions.

Example 2.1.3. For $x_0 \neq 0$, we have

$$\frac{d}{dx}\left(\frac{1}{x}\right)\bigg|_{x=x_0} = \lim_{x\to x_0} \frac{\dfrac{1}{x} - \dfrac{1}{x_0}}{x - x_0} = \lim_{x\to x_0} \frac{x_0 - x}{x_0 x(x - x_0)} = \lim_{x\to x_0} -\frac{1}{x_0 x} = -\frac{1}{x_0^2}.$$

Therefore $\dfrac{1}{x}$ is differentiable at x_0, and the linear approximation is $\dfrac{1}{x_0} - \dfrac{1}{x_0^2}(x - x_0)$. We express the derivative as

$$\left(\frac{1}{x}\right)' = -\frac{1}{x^2}.$$

Example 2.1.4. For $x_0 > 0$, we have

$$\frac{d\sqrt{x}}{dx}\bigg|_{x=x_0} = \lim_{x\to x_0} \frac{\sqrt{x} - \sqrt{x_0}}{x - x_0} = \lim_{x\to x_0} \frac{\sqrt{x} - \sqrt{x_0}}{(\sqrt{x} - \sqrt{x_0})(\sqrt{x} + \sqrt{x_0})} = \frac{1}{2\sqrt{x_0}}.$$

Therefore $\sqrt{x}$ is differentiable, and the linear approximation is $\sqrt{x_0} - \dfrac{1}{2\sqrt{x_0}}(x - x_0)$. We express the derivative as

$$(\sqrt{x})' = \frac{d\sqrt{x}}{dx} = \frac{1}{2\sqrt{x}}.$$

Example 2.1.5. By Example 1.7.21, we have

$$\frac{d(x^p)}{dx}\bigg|_{x=1} = \lim_{h\to 0} \frac{(1+h)^p - 1}{h} = p.$$

Therefore x^p is differentiable at 1 and has linear approximation $1 + p(x - 1)$.

Examples 2.1.3 and 2.1.4 are the derivatives of x^p for $p = -1$ and $\dfrac{1}{2}$ at general $x_0 > 0$. For general p, we take $h = x_0 y$ and get

$$\begin{aligned}\frac{d(x^p)}{dx}\bigg|_{x=x_0} &= \lim_{h\to 0} \frac{(x_0+h)^p - x_0^p}{h} = \lim_{y\to 0} \frac{(x_0 + x_0 y)^p - x_0^p}{x_0 y} \\ &= \lim_{y\to 0} x_0^{p-1} \frac{(1+y)^p - 1}{y} = p x_0^{p-1}.\end{aligned}$$

We express the derivative as

$$(x^p)' = p x^{p-1}.$$

Example 2.1.6. By Example 1.7.18, we have

$$\frac{d\log x}{dx}\bigg|_{x=1} = \lim_{x\to 1} \frac{\log x - \log 1}{x - 1} = \lim_{y\to 0} \frac{\log(y+1)}{y} = 1.$$

Therefore $\log x$ is differentiable at 1 and has linear approximation $x-1$.

In general, at any $x_0>0$, by taking $h=x_0y$, we have

$$\left.\frac{d\log x}{dx}\right|_{x=x_0}=\lim_{h\to 0}\frac{\log(x_0+h)-\log x_0}{h}=\lim_{y\to 0}\frac{\log(y+1)}{x_0y}=\frac{1}{x_0}.$$

We express the derivative as

$$(\log x)'=\frac{1}{x}.$$

Example 2.1.7. By Example 1.7.19, we have

$$\left.\frac{de^x}{dx}\right|_{x=x_0}=\lim_{h\to 0}\frac{e^{x_0+h}-e^{x_0}}{h}=\lim_{h\to 0}e^{x_0}\frac{e^h-1}{h}=e^{x_0}.$$

We express the derivative as

$$(e^x)'=e^x.$$

Example 2.1.8. In Section 1.5.4, we find

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

Therefore $\sin x$ is differentiable at $x=0$, and the linear approximation at 0 is x.

More generally, we have

$$\begin{aligned}\left.\frac{d\sin x}{dx}\right|_{x=x_0}&=\lim_{h\to 0}\frac{\sin(x_0+h)-\sin x_0}{h}\\&=\lim_{h\to 0}\frac{\sin h\cos x_0+\cos h\sin x_0-\sin x_0}{h}\\&=\lim_{h\to 0}\left(\frac{\cos h-1}{h}\sin x_0+\frac{\sin h}{h}\cos x_0\right)\\&=0\cdot\sin x_0+1\cdot\cos x_0=\cos x_0.\end{aligned}$$

We express the result as

$$(\sin x)'=\cos x.$$

By similar method, we have

$$(\cos x)'=-\sin x.$$

Example 2.1.9. In Section 1.5.4, we find

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

Therefore $\sin x$ is differentiable at $x=0$, and the linear approximation at 0 is x.

More generally, we have

$$\begin{aligned}
\left.\frac{d\sin x}{dx}\right|_{x=x_0} &= \lim_{h\to 0}\frac{\sin(x_0+h)-\sin x_0}{h} \\
&= \lim_{h\to 0}\frac{\sin h\cos x_0+\cos h\sin x_0-\sin x_0}{h} \\
&= \lim_{h\to 0}\left(\frac{\cos h-1}{h}\sin x_0+\frac{\sin h}{h}\cos x_0\right) \\
&= 0\cdot\sin x_0+1\cdot\cos x_0=\cos x_0.
\end{aligned}$$

We express the result as

$$(\sin x)'=\cos x.$$

By similar method, we have

$$(\cos x)'=-\sin x.$$

Exercise 2.1.3. Find the derivatives and then the linear approximations.

1. $\sqrt[3]{x}$ at 1.
2. $(\log x)^2$ at 1.
3. $\cos x^2$ at 0.
4. $\tan x$ at 0.
5. $\arcsin x$ at 0.
6. $\arctan x$ at 0.
7. $\sin\sin x$ at 0.
8. $x^2D(x)$ at 0.

Exercise 2.1.4. Find the derivatives, $a>0$.

1. $\log_a x$.
2. a^x.
3. $\tan x$.
4. $\arcsin x$.

Exercise 2.1.5. We have $\log|x|=\log(-x)$ for $x<0$. Show that the derivative of $\log(-x)$ at $x_0<0$ is $\dfrac{1}{x_0}$. The interpret your result as

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

Exercise 2.1.6. What is the derivative of $\log_a|x|$?

Exercise 2.1.7. Suppose p is an odd integer. Then x^p is defined for $x<0$. Do we still have $(x^p)'=px^{p-1}$ for $x<0$?

2.1.3 Constant Approximation

If a measurement of distance by a ruler in centimeters gives 7 meters and 5 centimeters, then the measurement by another (more primitive) ruler in meters should give 7 meters.

Analogously, if $a+b(x-x_0)$ is a linear approximation of $f(x)$ at x_0, then a is a *constant approximation* of $f(x)$ at x_0. Since the "basic unit" for constant functions is 1, the constant approximation means that, for any $\epsilon > 0$, there is $\delta > 0$, such that

$$|x - x_0| < \delta \implies |f(x) - a| \le \epsilon.$$

This means exactly that $f(x)$ is continuous at x_0, and the approximating constant is $a = f(x_0)$. Therefore the fact of linear approximation implying constant approximation means the following.

Theorem 2.1.4. *If a function is differentiable at a, then it is continuous at a.*

We do not expect the continuity to imply differentiability, because we do not expect the measurement in meters can tell us the measurement in centimeters.

Example 2.1.10. The sign function

$$\mathrm{sign}(x) = \begin{cases} 1, & \text{if } x > 0, \\ 0, & \text{if } x = 0, \\ -1, & \text{if } x < 0, \end{cases}$$

is not continuous at 0, and is therefore not differentiable at 0. Of course, we have $(\mathrm{sign}(x))' = 0$ away from 0.

Example 2.1.11. The absolute value function $|x|$ is continuous everywhere. Yet the derivative

$$(|x|)'|_{x=0} = \lim_{x\to 0} \frac{|x| - |0|}{x - 0} = \lim_{x\to 0} \mathrm{sign}(x)$$

diverges. Therefore the continuous function is not differentiable at 0.

Example 2.1.12. The Dirichlet function $D(x)$ in Example 1.6.13 is not continuous anywhere and is therefore not differentiable anywhere.

On the other hand, the function $xD(x)$ is continuous at 0. Yet the derivative

$$(xD(x))'|_{x=0} = \lim_{x\to 0} \frac{xD(x)}{x} = \lim_{x\to 0} D(x)$$

diverges. Therefore $xD(x)$ is not differentiable at 0, despite the continuity.

Exercise 2.1.8. Find the derivative of $|x|$ at $x_0 \neq 0$.

Exercise 2.1.9. Determine the differentiability of $|x|^p$ at 0.

Exercise 2.1.10. Is $xD(x)$ differentiable at $x_0 \neq 0$?

Exercise 2.1.11. Determine the differentiability of $|x|^p D(x)$ at 0.

Exercise 2.1.12. Determine the differentiability of

$$f(x) = \begin{cases} |x|^p \sin \dfrac{1}{x}, & \text{if } x \neq 0, \\ 0, & \text{if } x = 0, \end{cases}$$

at 0.

Exercise 2.1.13. Let $[x]$ be the greatest integer $\leq x$. Study the differentiability of $[x]$.

2.1.4 One Sided Derivative

Like one sided limits, we have *one sided derivatives*

$$f'_+(x_0) = \lim_{x \to x_0^+} \frac{f(x) - f(x_0)}{x - x_0}, \quad f'_-(x_0) = \lim_{x \to x_0^-} \frac{f(x) - f(x_0)}{x - x_0}.$$

The derivative $f'(x_0)$ exists if and only if both $f'_+(x_0)$ and $f'_-(x_0)$ exist and are equal.

Example 2.1.13. We have

$$(|x|)'|_{\text{at } 0^+} = \lim_{x \to 0^+} \frac{|x|}{x} = 1, \quad (|x|)'|_{\text{at } 0^-} = \lim_{x \to 0^-} \frac{|x|}{x} = -1.$$

Therefore $|x|$ has left and right derivatives. Since the two one sided derivatives are different, the function is not differentiable at 0.

Example 2.1.14. Consider the function

$$f(x) = \begin{cases} e^x, & \text{if } x \geq 0, \\ x + 1, & \text{if } x < 0. \end{cases}$$

We have (note that $f(0) = e^0 = 0 + 1$)

$$f'_+(0) = (e^x)'|_{x=0} = 1, \quad f'_-(0) = (x+1)'|_{x=0} = 1.$$

Therefore $f'(0) = 1$ and has linear approximation $1 + x$ at 0.

Exercise 2.1.14. Determine the differentiability.

1. $|x^2 - 3x + 2|$ at $0, 1, 2$.
2. $\sqrt{1 - \cos x}$ at 0.
3. $|\sin x|$ at 0.
4. $|\pi^2 - x^2| \sin x$ at π.

Exercise 2.1.15. Determine the differentiability at 0.

1. $\begin{cases} x^2, & \text{if } x \ge 0, \\ x, & \text{if } x < 0. \end{cases}$

2. $\begin{cases} \dfrac{1}{x+1}, & \text{if } x \ge 0, \\ x, & \text{if } x < 0. \end{cases}$

3. $\begin{cases} xe^{-\frac{1}{x}}, & \text{if } x \ge 0, \\ 0, & \text{if } x < 0. \end{cases}$

4. $\begin{cases} \log(1+x), & \text{if } x \ge 0, \\ e^x - 1, & \text{if } x < 0. \end{cases}$

Exercise 2.1.16. Determine the differentiability, $p, q > 0$.

1. $\begin{cases} (x-a)^p(b-x)^q, & \text{if } a \le x \le b, \\ 0, & \text{otherwise.} \end{cases}$

2. $\begin{cases} \arctan x, & \text{if } |x| \le 1, \\ \dfrac{\pi}{4}x, & \text{if } |x| > 1. \end{cases}$

3. $\begin{cases} x^2 e^{-x^2}, & \text{if } |x| \le 1, \\ e^{-1}, & \text{if } |x| > 1. \end{cases}$

4. $\begin{cases} \log |x|, & \text{if } |x| \ge 1, \\ x, & \text{if } |x| < 1. \end{cases}$

Exercise 2.1.17. Find a, b, c, such that the function

$$f(x) = \begin{cases} \dfrac{a}{x}, & \text{if } x > 1, \\ bx + c, & \text{if } x \le 1, \end{cases}$$

is differentiable at 1.

Exercise 2.1.18. For $p \ge 0$, x^p is defined on $[0, \delta)$. What is the right derivative of x^p at 0?

Exercise 2.1.19. For some p (see Exercises 2.1.7 and 2.1.18), x^p is defined on $(-\delta, \delta)$. What is the derivative of x^p at 0?

Exercise 2.1.20. Suppose $g(x)$ is continuous at x_0. Show that $f(x) = |x - x_0|g(x)$ is differentiable at x_0 if and only if $g(x_0) = 0$.

2.2 Property of Derivative

2.2.1 Arithmetic Combination of Linear Approximation

Suppose $f(x)$ and $g(x)$ are linearly approximated respectively by $a + b(x - x_0)$ and $c + d(x - x_0)$ at x_0. Then $f(x) + g(x)$ is approximated by

$$(a + b(x - x_0)) + (c + d(x - x_0)) = (a + c) + (b + d)(x - x_0).$$

Therefore $f + g$ is differentiable and $(f + g)'(x_0) = b + d$. Since $b = f'(x_0)$ and $d = g'(x_0)$, we conclude that

$$(f + g)'(x) = f'(x) + g'(x), \text{ or } \frac{d(f+g)}{dx} = \frac{df}{dx} + \frac{dg}{dx}.$$

Similarly, $Cf(x)$ is approximated by

$$C(a+b(x-x_0)) = Ca + Cb(x-x_0).$$

Therefore $Cf(x)$ is differentiable and $(Cf)'(x_0) = Cb$, which means

$$(Cf)'(x) = Cf'(x), \text{ or } \frac{d(Cf)}{dx} = C\frac{df}{dx}.$$

We also have $f(x)g(x)$ approximated by

$$(a+b(x-x_0))(c+d(x-x_0)) = ac + (bc+ad)(x-x_0) + bd(x-x_0)^2.$$

Although the approximation is not linear, the square unit $(x-x_0)^2$ is much smaller than $x-x_0$ when x is close to x_0. Therefore $f(x)g(x)$ is differentiable and has linear approximation $ac+(bc+ad)(x-x_0)$, and we get $(fg)'(x_0) = bc+ad$. By $a = f(x_0), b = f'(x_0), c = g(x_0), d = g'(x_0)$, we get the *Leibniz rule*

$$(fg)'(x) = f'(x)g(x) + f(x)g'(x), \text{ or } \frac{d(fg)}{dx} = \frac{df}{dx}g + f\frac{dg}{dx}.$$

The explanation above on the derivatives of arithmetic combinations are analogous to the arithmetic properties of limits.

Exercise 2.2.1. Find the derivative of the polynomial $p(x) = c_nx^n + \cdots + c_1x + c_0$.

Exercise 2.2.2. Compute the derivatives.

1. $e^x \sin x$.
2. $\sin^2 x$.
3. e^{2x}.
4. $\sin^2 x \cos x$.
5. $\sin x - x\cos x$.
6. $\cos x + x\sin x$.
7. $(x-1)e^x$.
8. x^2e^x.
9. $x\log x - x$.
10. $2x^2\log x - x^2$.
11. $xe^x\cos x$.
12. $xe^x\cos x\log x$.

Exercise 2.2.3. Find a polynomial $p(x)$, such that $(p(x)e^x)' = x^2e^x$. In general, suppose $(p_n(x)e^x)' = x^ne^x$. Find the relation between polynomials $p_n(x)$.

Exercise 2.2.4. Find polynomials $p(x)$ and $q(x)$, such that $(p(x)\sin x + q(x)\cos x)' = x^2\sin x$. Moreover, find a function with derivative $x^2\cos x$?

Exercise 2.2.5. Find constants A and B, such that $(Ae^x\sin x + Be^x\cos x)' = e^x\sin x$. What about $(Ae^x\sin x + Be^x\cos x)' = e^x\cos x$?

2.2.2 Composition of Linear Approximation

Consider a composition $g \circ f$

$$x \mapsto y = f(x) \mapsto z = g(y) = g(f(x)).$$

Suppose $f(x)$ is linearly approximated by $a+b(x-x_0)$ at x_0 and $g(y)$ is linearly approximated by $c+d(y-y_0)$ at $y_0=f(x_0)$. Then

$$a=f(x_0)=y_0,\quad b=f'(x_0),\quad c=g(y_0)=g(f(x_0)),\quad d=g'(y_0)=g'(f(x_0)),$$

and the composition $g\circ f$ is approximated by the composition of linear approximations (recall $a=y_0$)

$$c+d[(a+b(x-x_0))-y_0]=c+db(x-x_0).$$

Therefore the composition is also differentiable, with

$$(g\circ f)'(x_0)=db=g'(y_0)f'(x_0)=g'(f(x_0))f'(x_0)$$

This gives us the *chain rule*

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

or

$$\frac{dz}{dx}=\frac{dz}{dy}\frac{dy}{dx}.$$

Example 2.2.1. We know $(\log x)'=\dfrac{1}{x}$ for $x>0$. For $x<0$, we have

$$(\log(-x))'=(\log y)'|_{y=-x}(-x)'=\left.\frac{1}{y}\right|_{y=-x}(-1)=\frac{1}{-x}(-1)=\frac{1}{x}.$$

Therefore we conclude

$$(\log|x|)'=\frac{1}{x},\ \text{for } x\neq 0.$$

Example 2.2.2. In Example 2.1.5, we use the definition to derive $x^p=px^{p-1}$. Alternatively, we may also derive the derivative of x^p at general $x_0>0$ from the derivative $(x^p)'_{x=1}=p$ at a special place.

To move from x_0 to 1, we introduce $y=\dfrac{x}{x_0}$. Then x^p is the composition

$$x\mapsto y=\frac{x}{x_0}\mapsto z=x^p=x_0^py^p.$$

Then $x=x_0$ corresponds to $y=1$, and we have

$$\begin{aligned}\left.\frac{d(x^p)}{dx}\right|_{x=x_0}&=\left.\frac{dz}{dy}\right|_{y=1}\left.\frac{dy}{dx}\right|_{x=x_0}=\left.\frac{d(x_0^py^p)}{dy}\right|_{y=1}\left.\frac{d}{dx}\left(\frac{x}{x_0}\right)\right|_{x=x_0}\\&=x_0^p\cdot\left.\frac{d(y^p)}{dy}\right|_{y=1}\cdot\frac{1}{x_0}=x_0^p\cdot p\cdot\frac{1}{x_0}=px_0^{p-1}.\end{aligned}$$

Exercise 2.2.6. Use the derivative at a special place to find the derivative at other places.

1. $\log x$.
2. e^x.
3. $\sin x$.
4. $\cos x$.

Exercise 2.2.7. Use $\cos x = \sin\left(\frac{\pi}{2} - x\right)$ and the derivative of sine to derive the derivative of cosine. Use the similar method to find the derivatives of $\cot x$ and $\csc x$.

Exercise 2.2.8. A function $f(x)$ is odd if $f(-x) = -f(x)$, and is even if $f(-x) = f(x)$. What can you say about the derivative of an odd function and the derivative of an even function?

Example 2.2.3. By Example 2.1.3 and the chain rule, we have

$$\left(\frac{1}{f(x)}\right)' = \left(\frac{1}{y}\right)'\Big|_{y=f(x)} f'(x) = -\frac{1}{y^2}\Big|_{y=f(x)} f'(x) = -\frac{f'(x)}{f(x)^2}.$$

Then we may use the Leibniz rule to get the derivative of quotient

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

Exercise 2.2.9. Derive the derivatives.

$$(\tan x)' = \sec^2 x, \quad (\sec x)' = \sec x \tan x, \quad (e^{-x})' = -e^{-x}.$$

Exercise 2.2.10. Compute the derivatives.

1. $\dfrac{1}{x+2}$.
2. $\dfrac{x+1}{x-2}$.
3. $\dfrac{x^2-x+1}{x^3+1}$.
4. $\dfrac{x^3+1}{x^2-x+1}$.
5. $\dfrac{1}{ax+b}$.
6. $\dfrac{ax+b}{cx+d}$.
7. $\dfrac{x}{x^2+ax+b}$.
8. $\dfrac{1}{(x+a)(x+b)}$.

Exercise 2.2.11. Compute the derivatives.

1. $\dfrac{\log x}{x}$.
2. $\dfrac{\log x}{x^p}$.
3. $\dfrac{x^p}{\log x}$.
4. $\dfrac{e^x}{x\log x}$.

Exercise 2.2.12. Compute the derivatives.

1. $\dfrac{\sin x}{a+\cos x}$.
2. $\dfrac{1}{a+\tan x}$.
3. $\dfrac{1+x\tan x}{\tan x - x}$.
4. $\dfrac{\cos x + x\sin x}{\sin x - x\cos x}$.

Exercise 2.2.13. The *hyperbolic trigonometric functions* are

$$\sinh x = \frac{e^x - e^{-x}}{2}, \quad \cosh x = \frac{e^x + e^{-x}}{2},$$

and

$$\tanh x = \frac{\sinh x}{\cosh x}, \quad \coth x = \frac{\cosh x}{\sinh x}, \quad \text{sech} x = \frac{1}{\cosh x}, \quad \text{csch} x = \frac{1}{\sinh x}.$$

Find their derivatives and express them in hyperbolic trigonometric functions.

Example 2.2.4. The function $(x^2-1)^{10}$ is the composition of $z = y^{10}$ and $y = x^2 - 1$. Therefore

$$\left((x^2-1)^{10}\right)' = \frac{d(y^{10})}{dy}\frac{d(x^2-1)}{dx} = 10y^9 \cdot 2x = 20x(x^2-1)^9.$$

Example 2.2.5. The function $a^x = e^{bx}$, $b = \log a$, is the composition of $z = e^y$ and $y = bx$. Therefore

$$(a^x)' = (e^y)'|_{y=bx}(bx)' = e^{bx}b = a^x \log a.$$

Exercise 2.2.14. Compute the derivatives.

1. $(1-x)^{10}$.
2. $(3x+2)^{10}$.
3. $(x^3-1)^{10}(1-x^2)^9$.
4. $(1+(1-x^2)^{10})^9$.
5. $((x^3-1)^8+(1-x^2)^{10})^9$.
6. $((1-(3x+2)^3)^8+1)^9$.
7. $\dfrac{(x+1)^9}{(3x+5)^8}$.
8. $\dfrac{x(x+1)}{(x+2)(x+3)}$.

Exercise 2.2.15. Compute the derivatives.

1. $\cos(x^5+3x^2+1)$.
2. $\tan^{10}\left(x(x+1)^9\right)$.
3. $\sin(\sqrt{x}+3)$.
4. $\sin(\sqrt{x-2}+3)$.
5. $(\sin x + \cos x)^{10}$.
6. $\sqrt{\sin x + \cos x}$.
7. $\left(\dfrac{\sin^3 x}{\cos^4 x}\right)^{10}$.
8. $\sin(\cos x)$.
9. $\sin(\cos(\tan x))$.
10. $\dfrac{\sin^2 x}{\sin x^2}$.
11. $\dfrac{\sin 2x + 2\cos 2x}{2\sin x - \cos 2x}$.
12. $\dfrac{\sin^8 \sqrt{x}}{1+\cos^{10}(\sin^9 x)}$.

Exercise 2.2.16. Compute the derivatives.

1. e^{x^2}.
2. $(x^2-1)e^{x^2}$.
3. $e^{(e^x)}$.
4. $e^{\log x}$.
5. $\log_x e$.
6. $\log(\log x)$.
7. $\log\left(\dfrac{1}{\log x}\right)$.
8. $\log(\log(\log x))$.
9. $\log|\cos x|$.
10. $\log|\tan x|$.
11. $\log|\sec x - \tan x|$.
12. $\log\dfrac{1-\sin x}{1+\sin x}$.

Exercise 2.2.17. Compute the derivatives.

1. $(ax+b)^p$.
2. $(ax^2+bx+c)^p$.
3. $(a+(bx^2+c)^p)^q$.
4. e^{ax}.
5. $\log(ax+b)$.
6. $\sin(ax+b)$.
7. $\cos(ax+b)$.
8. $\tan(ax+b)$.
9. $\sec(ax+b)$.

Exercise 2.2.18. Compute the derivatives.

1. $\dfrac{1}{\sqrt{a^2-x^2}}$.
2. $\dfrac{1}{\sqrt{a^2+x^2}}$.
3. $\dfrac{1}{\sqrt{x^2-a^2}}$.
4. $\dfrac{x}{\sqrt{a^2-x^2}}$.
5. $\dfrac{x}{\sqrt{a^2+x^2}}$.
6. $\dfrac{x}{\sqrt{x^2-a^2}}$.

Exercise 2.2.19. Compute the derivatives.

1. $\sqrt{1+\sqrt{x}}$.
2. $\dfrac{1}{\sqrt{1+\sqrt{x}}}$.
3. $\sqrt{1+\sqrt{1+\sqrt{x}}}$.
4. $\sqrt{x+\sqrt{x+\sqrt{x}}}$.
5. $(1+2\sqrt{x+1})^{10}$.
6. $(1+2\sqrt{x+1})^{-10}$.
7. $\dfrac{\sqrt{x}+1}{x-2}$.
8. $\dfrac{\sqrt{x}+1}{(1-\sqrt{x}+x)^3}$.
9. $\dfrac{(\sqrt{x}+1)^4}{(1-\sqrt{x}+x)^3}$.
10. $\sqrt{\dfrac{1+x^2}{1-x^2}}$.
11. $\sqrt[3]{\dfrac{1+x^2}{1-x^2}}$.
12. $\dfrac{\sqrt{1+\sqrt{x}}}{\sqrt{1-\sqrt{x}}}$.
13. $\sqrt{\dfrac{1+\sqrt{x}}{1-\sqrt{x}}}$.
14. $\left(\dfrac{1+\sqrt{x}}{1-\sqrt{x}}\right)^{10}$.
15. $\dfrac{\sqrt{x+1}-\sqrt{x-1}}{\sqrt{x+1}+\sqrt{x-1}}$.
16. $\dfrac{\sqrt{1+x}-\sqrt{1-x}}{\sqrt{1+x}+\sqrt{1-x}}$.
17. $\left(\dfrac{1}{1+\sqrt{x}}+\dfrac{1}{1-\sqrt{x}}\right)^{10}$.
18. $\sqrt{1+\dfrac{1}{\sqrt{x^2+1}}}$.

Exercise 2.2.20. Compute the derivatives.

1. $|x^2(x+2)^3|$.
2. $|\sin^3 x|$.
3. $|x(e^x-1)|$.
4. $|(x-1)^2\log x|$.

Exercise 2.2.21. Compute the derivatives.

1. $\sqrt{x}-\log|\sqrt{x}+a|$.
2. $\dfrac{b}{a^2(ax+b)}+\dfrac{1}{a^2}\log|ax+b|$.
3. $-\dfrac{1}{b}\log\dfrac{ax+b}{x}$.
4. $\dfrac{1}{b(ax+b)}-\dfrac{1}{b^2}\log\dfrac{ax+b}{x}$.
5. $\dfrac{1}{2}\log(a^2+x^2)$.
6. $\sqrt{x(a+x)}-a\log(\sqrt{x}+\sqrt{x+a})$.
7. $\dfrac{1}{\sqrt{b}}\log\dfrac{\sqrt{ax+b}-\sqrt{b}}{\sqrt{ax+b}+\sqrt{b}}$.
8. $x\log(x+\sqrt{x^2+a})-\sqrt{x^2+a}$.
9. $-\dfrac{1}{a}\log\dfrac{a+\sqrt{a^2+x^2}}{x}$.

10. $-\dfrac{1}{a}\log\dfrac{a+\sqrt{a^2-x^2}}{x}$.

11. $\dfrac{b}{2a}x-\dfrac{1}{4}x^2+\dfrac{1}{2}\left(x^2-\dfrac{b^2}{a^2}\right)\log(ax+b)$.

12. $-\dfrac{1}{2}x^2+\dfrac{1}{2}\left(x^2-\dfrac{a^2}{b^2}\right)\log(a^2-b^2x^2)$.

Example 2.2.6. By the chain rule, we have

$$\begin{aligned}\left(\log\left|x+\sqrt{x^2+a}\right|\right)' &= (\log|y|)'|_{y=x+\sqrt{x^2+a}}\left[x'+(\sqrt{z})'|_{z=x^2+a}(x^2+a)'\right]\\ &= \frac{1}{x+\sqrt{x^2+a}}\left[1+\frac{1}{2\sqrt{x^2+a}}2x\right]=\frac{1}{\sqrt{x^2+a}}.\end{aligned}$$

Now suppose we wish to find a function $f(x)$ with derivative

$$f'(x)=\frac{1}{\sqrt{x^2+ax+b}}.$$

By

$$f(x)=\frac{1}{\sqrt{\left(x+\dfrac{a}{2}\right)^2+b-\dfrac{a^2}{4}}}=\frac{1}{\sqrt{y^2+c}},\quad y=x+\frac{a}{2},\ c=b-\frac{a^2}{4},$$

we may substitute x by $x+\dfrac{a}{2}$ and substitute a by $c=b-\dfrac{a^2}{4}$. Then we get

$$\begin{aligned}\left(\log\left|x+\frac{a}{2}+\sqrt{x^2+ax+b}\right|\right)' &= \left(\log\left|y+\sqrt{y^2+c}\right|\right)'\Big|_{y=x+\frac{a}{2}}\left(x+\frac{a}{2}\right)'\\ &= \frac{1}{\sqrt{y^2+c}}\bigg|_{y=x+\frac{a}{2}}=\frac{1}{\sqrt{x^2+ax+b}}.\end{aligned}$$

Exercise 2.2.22. Find constants A and B, such that $(Ae^{ax}\sin bx+Be^{ax}\cos bx)'=e^{ax}\cos bx$. What about $(Ae^{ax}\sin bx+Be^{ax}\cos bx)'=e^{ax}\sin bx$?

Exercise 2.2.23. Use Example 2.2.6 to compute the derivatives.

1. $\log(e^x+\sqrt{1+e^{2x}})$.
2. $\log\left|x-\sqrt{x^2+a}\right|$.
3. $\log(\tan x+\sec x)$.

Exercise 2.2.24. Compute the derivative of $\log\dfrac{x}{x+1}$. Then find a function with derivative $\dfrac{1}{(x+a)(x+b)}$. In case $a^2\ge 4b$, can you find a function with derivative $\dfrac{1}{x^2+ax+b}$?

Exercise 2.2.25. Compute the derivative of $x\sqrt{x^2+a}+a\log(x+\sqrt{x^2+a})$. Then find a function with derivative $\sqrt{x^2+ax+b}$.

Exercise 2.2.26. Use Example 2.2.6 to compute the derivatives.

1. $\log(e^x + \sqrt{1+e^{2x}})$. 2. $\log(x - \sqrt{x^2+a})$. 3. $\log(\tan x + \sec x)$.

Example 2.2.7. By viewing $x^p = e^{p\log x}$ as a composition of $z = e^y$ and $y = p\log x$, we have

$$\frac{d(x^p)}{dx} = \frac{(e^{p\log x})}{dx} = \frac{d(e^y)}{dy}\bigg|_{y=p\log x}\frac{d(p\log x)}{dx} = e^{p\log x}\frac{p}{x} = x^p\frac{p}{x} = px^{p-1}.$$

This derives the derivative of x^p by using the derivatives of e^x and $\log x$.

Example 2.2.8. Suppose $u(x)$ and $v(x)$ are differentiable and $u(x) > 0$. Then $u(x)^{v(x)} = e^{u(x)\log v(x)}$, and

$$(u(x)^{v(x)})' = (e^{v(x)\log u(x)})' = (e^y)'|_{y=v(x)\log u(x)}(v(x)\log u(x))'.$$

By

$$(e^y)'|_{y=v(x)\log u(x)} = e^y|_{y=v(x)\log u(x)} = e^{v(x)\log u(x)} = u(x)^{v(x)},$$

and

$$(v(x)\log u(x))' = v'(x)\log u(x) + v(x)(\log u)'|_{u=u(x)}u'(x) = v'(x)\log u(x) + \frac{v(x)u'(x)}{u(x)},$$

We get

$$\begin{aligned}(u(x)^{v(x)})' &= u(x)^{v(x)}\left(v'(x)\log u(x) + \frac{v(x)u'(x)}{u(x)}\right)\\ &= u(x)^{v(x)-1}(u(x)v'(x)\log u(x) + u'(x)v(x)).\end{aligned}$$

Exercise 2.2.27. Compute the derivatives.

1. x^x.
2. x^{x^2}.
3. $(x^2)^x$.
4. $(a^x)^x$.
5. $(x^a)^x$.
6. $(x^x)^a$.
7. $(x^x)^x$.
8. $a^{(x^x)}$.
9. $x^{(a^x)}$.
10. $x^{(x^a)}$.
11. $x^{(x^x)}$.
12. $(x^x)^{(x^x)}$.

Exercise 2.2.28. Compute the derivatives.

1. $x^{\sin x}$.
2. $(\sin x)^x$.
3. $(\sin x)^{\cos x}$.
4. $\sqrt[x]{x}$.
5. $\sqrt[x]{\log x}$.
6. $x^{\log x}$.
7. $(e^x + e^{-x})^x$.
8. $(\log|x^2-1|)^x$.

Exercise 2.2.29. Let $f(x) = u(x)^{v(x)}$. Then $\log f(x) = u(x)\log v(x)$. By taking the derivative on both sides of the equality, derive the formula for $f'(x)$.

Exercise 2.2.30. Use the idea of Exercise 2.2.29 to compute the derivatives.

1. $\dfrac{x+a}{x+b}$.
2. $\dfrac{1}{(x+a)(x+b)}$.
3. $\dfrac{(x+c)(x+d)}{(x+a)(x+b)}$.
4. $\dfrac{(x+3)^7\sqrt{2x-1}}{(2x+1)^3}$.
5. $\dfrac{(x^2+x+1)^7}{(x^2-x+1)^3}$.
6. $\dfrac{e^{x^2+1}\sqrt{\sin x}}{(x^2-x+1)^3\log x}$.

2.2.3 Implicit Linear Approximation

The chain rule can be used to compute the derivatives of functions that are "implicitly" given. Such functions often do not have explicit formula expressions.

Strictly speaking, we need to know that implicitly given functions are differentiable before taking their derivatives. There are general theorems confirming such differentiability. In the subsequent examples, we will always assume the differentiability of implicitly defined functions.

Example 2.2.9. The unit circle $x^2+y^2=1$ on the plane is made up of the graphs of two functions $y=\sqrt{1-x^2}$ and $y=-\sqrt{1-x^2}$. We may certainly compute the derivative of each one explicitly

$$(\sqrt{1-x^2})' = \frac{1}{2}(1-x^2)^{-\frac{1}{2}}(-2x) = \frac{-x}{\sqrt{1-x^2}}, \quad (-\sqrt{1-x^2})' = \frac{x}{\sqrt{1-x^2}}.$$

On the other hand, we may use the fact that both functions $y=y(x)$ satisfy the equation $x^2+y(x)^2=1$. Taking the derivatives in x of both sides, we get $2x+2yy'=0$. Solving the equation, we get

$$y' = -\frac{x}{y}.$$

This is consistent with the two derivatives computed above.

There is yet another way of computing the derivative $y'(x)$. The circle can be parametrized as $x=\cos t$, $y=\sin t$. In this view, the function $y=y(x)$ satisfies $\sin t = y(\cos t)$. By the chain rule, we have

$$\cos t = y'(x)(-\sin t).$$

Therefore

$$y'(x) = -\frac{\cos t}{\sin t} = -\frac{x}{y}.$$

In general, the derivative of a function $y=y(x)$ given by a parametrized curve $x=x(t)$, $y=y(t)$ is

$$y'(x) = \frac{y'(t)}{x'(t)}.$$

Note that the formula is ambiguous, in that $y'(x) = -\cot t$ and $y'(t)=\cos t$ are not the same functions. The primes in the two functions refer to $\dfrac{d}{dx}$ and $\dfrac{d}{dt}$ respectively.

So it is better to keep track of the variables by using Leibniz's notation. The formula above becomes

$$\frac{dy}{dx} = \frac{\frac{dy}{dt}}{\frac{dx}{dt}}.$$

This is just another way of expressing the chain rule $\dfrac{dy}{dt} = \dfrac{dy}{dx}\dfrac{dx}{dt}$.

Exercise 2.2.31. Compute the derivatives of the functions $y = y(x)$ given by curves.

1. $x = \sin^2 t$, $y = \cos^2 t$.

2. $x = a(t - \sin t)$, $y = a(1 - \cos t)$.

3. $x = e^t \cos 2t$, $y = e^t \sin 2t$.

4. $x = (1 + \cos t)\cos t$, $y = (1 + \cos t)\sin t$.

Example 2.2.10. Like the unit circle, the equation $2y-2x^2-\sin y+1 = 0$ is a curve on the plane, made up of the graphs of several functions $y = y(x)$. Although we cannot find an explicit formula for the functions, we can still compute their derivatives.

Taking the derivative of both sides of the equation $2y - 2x^2 - \sin y + 1 = 0$ with respect to x and keeping in mind that y is a function of x, we get $2y'-4x-y'\cos y = 0$. Therefore

$$y' = \frac{4x}{2 - \cos y}.$$

The point $P = \left(\sqrt{\dfrac{\pi}{2}}, \dfrac{\pi}{2}\right)$ satisfies the equation and lies on the curve. The tangent line of the curve at the point has slope

$$y'|_P = \frac{4\sqrt{\frac{\pi}{2}}}{2 - \cos\frac{\pi}{2}} = \sqrt{2\pi}.$$

Therefore the tangent line at P is given by the equation

$$y - \frac{\pi}{2} = \sqrt{2\pi}\left(x - \sqrt{\frac{\pi}{2}}\right),$$

or

$$y = \sqrt{2\pi}x - \frac{\pi}{2}.$$

Example 2.2.11. The equations $x^2+y^2+z^2=2$ and $x+y+z=0$ specify a circle in the Euclidean space $\mathbb{R}^3$ and define functions $y=y(x)$ and $z=z(x)$. To find the derivatives of the functions, we take the derivatives of the two equations in x

$$2x+2yy'+2zz'=0, \quad 1+y'+z'=0.$$

Solving for y' and z', we get

$$y'=\frac{z-x}{y-z}, \quad z'=\frac{y-x}{z-y}.$$

Exercise 2.2.32. Compute the derivatives of implicitly defined functions.

1. $y^2+3y^3+1=x$.
2. $\sin y=x$.
3. $\dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}=1$.
4. $\sqrt{x}+\sqrt{y}=\sqrt{a}$.
5. $e^{x+y}=xy$.
6. $x^2+2xy-y^2-2x=0$.

Exercise 2.2.33. Find the derivative of the implicitly defined functions of x.

1. $x^p+y^p=2$ at $x=1, y=1$.
2. $xy=\sin(x+y)$ at $x=0, y=\pi$.
3. $\dfrac{x+y}{z}=\dfrac{y+z}{x}=\dfrac{z+x}{y}$ at $x=y=z=1$.

Exercise 2.2.34. If $f(\sin x)=x$, what can you say about the derivative of $f(x)$? What if $\sin f(x)=x$?

Example 2.2.12. In Example 1.7.11, we argued that the function $f(x)=x^5+3x^3+1$ is invertible. The inverse $g(x)$ satisfies $g(x)^5+3g(x)^3+1=x$. Taking the derivative in x on both sides, we get $5g(x)^4g'(x)+9g(x)^2g'(x)=1$. This implies

$$g'(x)=\frac{1}{5g(x)+9g(x)^2}.$$

Example 2.2.13. In Example 2.2.12, we interpreted the derivative of an inverse function as an implicit differentiation problem. In general, the inverse function $g(x)=f^{-1}(x)$ satisfies $f(g(x))=x$. Taking the derivative of both sides, we get $f'(g(x))g'(x)=1$. Therefore

$$(g(x))'=\frac{1}{f'(g(x))}, \text{ or } (f^{-1}(x))'=\frac{1}{f'(f^{-1}(x))}.$$

For example, the derivative of $\arcsin x$ is

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

In the last step, we have positive square root because $y \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$.

The computation can also be explained by considering two variables related by $y = y(x)$ and $x = x(y)$, with $x(y)$ being the inverse function of $y(x)$. The chain rule tells us $\frac{dx}{dy}\frac{dy}{dx} = \frac{dx}{dx} = 1$. Then we get

$$\frac{dx}{dy} = \frac{1}{\frac{dy}{dx}}, \text{ or } x'(y) = \frac{1}{y'(x)|_{x=x(y)}} = \frac{1}{y'(x(y))}.$$

Specifically, for $y = \arcsin x$, we have $x = \sin y$. then

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

Exercise 2.2.35. Derive the derivatives of the inverse trigonometric functions

$$(\arctan x)' = \frac{1}{1+x^2}, \quad (\arccos x)' = -\frac{1}{\sqrt{1-x^2}}, \quad (\mathrm{arcsec} x)' = \frac{1}{x\sqrt{x^2-1}}.$$

Exercise 2.2.36. Compute the derivatives.

1. $\arcsin\sqrt{x}$.
2. $\arcsin\sqrt{1-x^2}$.
3. $\arctan\frac{1+x}{1-x}$.
4. $\frac{\arccos x}{x} + \frac{1}{2}\log\frac{1-\sqrt{1-x^2}}{1+\sqrt{1-x^2}}$.
5. $\frac{\arcsin x}{\sqrt{1-x^2}} + \frac{1}{2}\log(1-x^2)$.
6. $\arctan\frac{1-x}{\sqrt{2x-x^2}}$.

Exercise 2.2.37. Compute the derivatives.

1. $2\arcsin\sqrt{\frac{x-a}{b-a}}$.
2. $\frac{1}{a}\arcsin\frac{a}{x}$.
3. $\frac{1}{a}\mathrm{arcsec}\frac{x}{a}$.
4. $\frac{2}{\sqrt{4ac-b^2}}\arctan\frac{2ax+b}{\sqrt{4ac-b^2}}$.
5. $-\sqrt{x(a-x)} - a\arctan\frac{\sqrt{x(a-x)}}{x-a}$.
6. $\frac{1}{2}x\sqrt{a^2-x^2} + \frac{1}{2}a^2\arctan\frac{x}{\sqrt{a^2-x^2}}$.
7. $x\log(x^2+a^2) + 2a\arctan\frac{x}{a} - 2x$.

Exercise 2.2.38. Compute the derivative of $\arcsin\frac{x}{a}$. Then use the idea and result of Example 2.2.6 to find a function with derivative $\frac{1}{\sqrt{ax^2+bx+c}}$.

Exercise 2.2.39. Compute the derivative of $x\sqrt{a^2-x^2}+a^2\arcsin\frac{x}{a}$. Then combine with the result of Exercise 2.2.25 to find a function with derivative $\sqrt{ax^2+bx+c}$.

Exercise 2.2.40. Compute the derivative of $\frac{1}{a}\arctan\frac{x}{a}$. Then for the case $a^2\le 4b$, find a function with derivative $\frac{1}{x^2+ax+b}$. This complements Exercise 2.2.24.

Exercise 2.2.41. Compute the derivative of $\log\frac{\sqrt{x+1}-1}{\sqrt{x+1}+1}$ and $\arctan\sqrt{x}$. Then find a function with derivative $\frac{1}{x\sqrt{ax+b}}$.

Exercise 2.2.42. Suppose $f(x)$ is invertible, with $f(1)=1$, $f'(1)=a$. Find the derivative of the functions $f\left(\frac{1}{f^{-1}\left(\frac{x}{f(x)}\right)}\right)$ and $f^{-1}(f^{-1}(x))$ at 1.

Exercise 2.2.43. Explain the formula for the derivative of the inverse function by considering the inverse of the linear approximation.

Exercise 2.2.44. Find the place on the curve $y=x^2$ where the tangent line is parallel to the straight line $x+y=1$.

Exercise 2.2.45. Show that the area enclosed by the tangent line on the curve $xy=a^2$ and the coordinate axes is a constant.

Exercise 2.2.46. Let P be a point on the curve $y=x^3$. The tangent at P meets the curve again at Q. Prove that the slope of the curve at Q is four times the slope at P.

2.3 Application of Linear Approximation

The linear approximation can be used to determine behaviors of functions. The idea is that, if the linear approximation of a function has certain behavior, then the function is likely to have the similar behavior.

2.3.1 Monotone Property and Extrema

We say a function $f(x)$ has *local maximum* at x_0, if

$$x\in \text{domain},\ |x-x_0|<\delta \implies f(x)\le f(x_0).$$

Similarly, $f(x)$ has *local minimum* at x_0, if

$$x\in \text{domain},\ |x-x_0|<\delta \implies f(x)\ge f(x_0).$$

The function has a *(global) maximum* at x_0 if $f(x_0) \geq f(x)$ for *all* x in the domain

$$x \in \text{domain} \implies f(x) \leq f(x_0).$$

The concepts of *(global) minimum* can be similarly defined. The maximum and minimum are *extrema* of the function. A global extreme is also a local extreme.

The local maxima are like the peaks in a mountain, and the global maximum is like the highest peak.

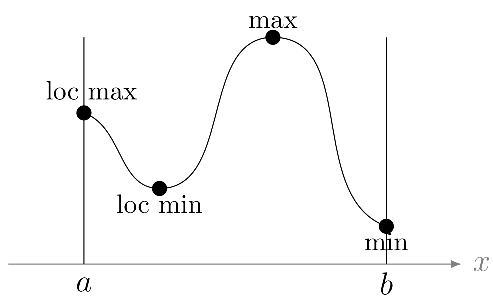


Figure 2.3.1: Local and global extrema.

The following result shows the existence of global extrema in certain case.

Theorem 2.3.1. *Any continuous function on a bounded closed interval has global maximum and global minimum.*

If a function f is increasing on $(x_0-\delta, x_0]$ (i.e., on the left of x_0 and including x_0), then $f(x_0)$ is the biggest value on $(x_0-\delta, x_0]$. If f is also decreasing on $[x_0, x_0+\delta)$ (i.e., on the right of x_0 and including x_0), then $f(x_0)$ is the biggest value on $[x_0, x_0+\delta)$. In other words, if f changes from increasing to decreasing as we pass x_0 from left to right, then x_0 is a local maximum of f. Similarly, if f changes from decreasing to increasing at x_0, then x_0 is a local minimum.

Example 2.3.1. The square function x^2 is strictly decreasing on $(-\infty, 0]$ because

$$x_1 < x_2 \leq 0 \implies x_1^2 > x_2^2.$$

By the same reason, the function is strictly increasing on $[0, +\infty)$. This leads to the local minimum at 0. In fact, by $x^2 \geq 0 = |0|$ for all x, we know x^2 has a global minimum at 0. The function has no local maximum and therefore no global maximum on $\mathbb{R}$.

On the other hand, if we restrict x^2 to $[-1, 1]$, then x^2 has global minimum at 0 and global maxima at -1 and 1. If we restrict to $[-1, 2]$, then x^2 has global minimum at 0, global maximum at 2, and local (but not global) maximum at -1. If we restrict to $(-1, 2)$, then x^2 has global minimum at 0, and has no local maximum.

Example 2.3.2. The sine function is strictly increasing on $\left[2n\pi - \frac{\pi}{2}, 2n\pi + \frac{\pi}{2}\right]$ and is strictly decreasing on $\left[2n\pi + \frac{\pi}{2}, 2n\pi + \frac{3\pi}{2}\right]$. This implies that $2n\pi + \frac{\pi}{2}$ are local maxima and $2n\pi - \frac{\pi}{2}$ are local minima. In fact, by $\sin\left(2n\pi - \frac{\pi}{2}\right) = -1 \le \sin x \le 1 = \sin\left(2n\pi + \frac{\pi}{2}\right)$, these local extrema are also global extrema.

Exercise 2.3.1. Determine the monotone property and find the extrema for $|x|$

1. on $[-1, 1]$.
2. on $(-1, 1]$.
3. on $[-2, 1]$.
4. on $(-\infty, 1]$.

Exercise 2.3.2. Determine the monotone property and find the extrema on $\mathbb{R}$.

1. $|x|$.
2. $x^2 + 2x$.
3. $|x^2 + 2x|$.
4. x^3.
5. x^6.
6. $\frac{1}{x}$.
7. $\sqrt{|x|}$.
8. $x + \frac{1}{x}$.
9. $\frac{1}{x^2+1}$.
10. $\cos x$.
11. $\sin^2 x$.
12. $\sin x^2$.
13. e^x.
14. e^{-x}.
15. $\log x$.
16. x^x.

Exercise 2.3.3. How are the extrema of the function related to the extrema of $f(x)$?

1. $f(x) + a$.
2. $af(x)$.
3. $f(x)^a$.
4. $a^{f(x)}$.

Exercise 2.3.4. How are the extrema of the function related to the extrema of $f(x)$?

1. $f(x + a)$.
2. $f(ax)$.
3. $f(x^2)$.
4. $f(\sin x)$.

Exercise 2.3.5. Is local maximum always the place where the function changes from increasing to decreasing? In other words, can you construct a function $f(x)$ with local maximum at 0, but $f(x)$ is *not* increasing on $(-\delta, 0]$ for any $\delta > 0$?

Exercise 2.3.6. Compare the global extrema on various intervals in Example 2.3.1 with Theorem 2.3.1.

2.3.2 Detect the Monotone Property

Suppose $f(x)$ is approximated by the linear function $L(x) = f(x_0) + f'(x_0)(x - x_0)$ near x_0. The linear function $L(x)$ is increasing if and only if the slope $f'(x_0) \ge 0$. Since $L(x)$ is very close to $f(x)$, we expect $f(x)$ to be also increasing. The expectation is true if the linear approximation is increasing *everywhere*.

Theorem 2.3.2. *If $f'(x) \ge 0$ on an interval, then $f(x)$ is increasing on the interval. If $f'(x) > 0$ on the interval, then $f(x)$ is strictly increasing on the interval.*

Similar statements hold for decreasing functions. Moreover, for a function on a closed interval $[a,b]$, we just need the derivative criterion to be satisfied on (a,b) and the function to be continuous on $[a,b]$.

Example 2.3.3. We have $(x^2)' = 2x < 0$ on $(-\infty, 0)$ and x^2 continuous on $(-\infty, 0]$. Therefore x^2 is strictly decreasing on $(-\infty, 0]$. By the similar reason, x^2 is strictly increasing on $[0, +\infty)$. This implies that 0 is a local minimum. The conclusion is consistent with the observation in Example 2.3.1 obtained by direct inspection.

x	$(-\infty, 0)$	0	$(0, +\infty)$
$f = x^2$	$\searrow$	loc min 0	$\nearrow$
$f' = 2x$	$-$	0	$+$

Example 2.3.4. The function $f(x) = x^3 - 3x + 1$ has derivative $f'(x) = 3(x+1)(x-1)$. The sign of the derivative implies that the function is strictly increasing on $(-\infty, -1]$ and $[1, +\infty)$, and is strictly decreasing on $[-1, 1]$. This implies that -1 is a local maximum and 1 is a local minimum.

x	$(-\infty, -1)$	-1	$(-1, 1)$	1	$(1, +\infty)$
$f = x^3 - 3x + 1$	$\nearrow$	loc max 3	$\searrow$	loc min -1	$\nearrow$
$f' = 3(x+1)(x-1)$	$+$	0	$-$	0	$+$

Example 2.3.5. The function $f(x) = \sin x - x\cos x$ has derivative $f'(x) = x \sin x$. The sign of the derivative determines the strict monotone property on the interval $[-5, 5]$ as described in the picture. The strict monotone property implies that $-\pi, 5$ are local minima, and $-5, \pi$ are local maxima.

x	-5	$(-5, -\pi)$	$-\pi$	$(-\pi, 0)$	0	$(0, \pi)$	π	$(\pi, 5)$	5
f	max	$\searrow$	min	$\nearrow$		$\nearrow$	max	$\searrow$	min
f'		$-$	0	$+$	0	$+$	0	$-$	

Example 2.3.6. The function $f(x) = \sqrt[3]{x^2}(x+1)$ has derivative $f'(x) = \dfrac{(5x+2)}{3\sqrt[3]{x}}$ for $x \neq 0$. Using the sign of the derivative and the continuity, we get the strict monotone property of the function, which implies that $-\dfrac{2}{5}$ is a local maximum, and 0 is a local minimum.

x	$(-\infty, -\frac{2}{5})$	$-\frac{2}{5}$	$(-\frac{2}{5}, 0)$	0	$(0, +\infty)$
f	$\nearrow$	max	$\searrow$	min	$\nearrow$
f'	$+$	0	$-$	no	$+$

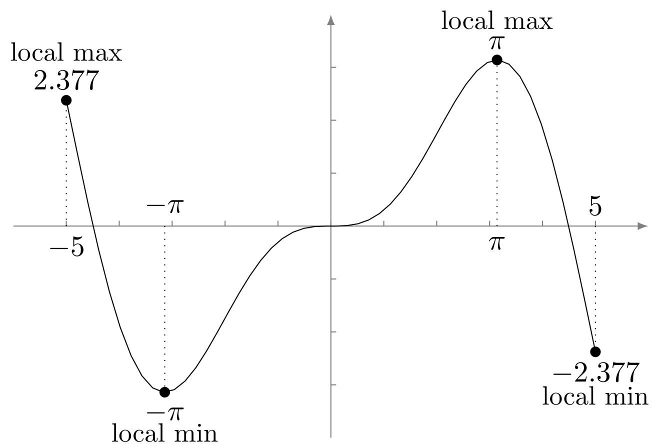


Figure 2.3.2: Graph of $\sin x - x\cos x$.

Example 2.3.7. The function $f(x) = \dfrac{x^3}{x^2-1}$ has derivative $f'(x) = \dfrac{x^2(x^2-3)}{(x^2-1)^2}$ for $x \neq \pm 1$. The sign of the derivative determines the strict monotone property away from ± 1. The strict monotone property implies that $-\sqrt{3}$ is a local minimum, and $\sqrt{3}$ is a local maximum.

x		$-\sqrt{3}$		-1		0		1		$\sqrt{3}$	
f	$\nearrow$	max	$\searrow$		$\searrow$		$\searrow$		$\searrow$	min	$\nearrow$
f'	$+$	0	$-$	no	$-$	0	$-$	no	$-$	0	$+$

Exercise 2.3.7. Determine the monotone property and find extrema.

1. $x^3 - 3x + 2$ on $\mathbb{R}$.
2. $x^3 - 3x + 2$ on $[-1, 2]$.
3. $x^3 - 3x + 2$ on $(-1, 2)$.
4. $|x^3 - 3x + 2|$ on $\mathbb{R}$.
5. $|x^3 - 3x + 2|$ on $[-1, 2]$.
6. $|x^3 - 3x + 2|$ on $(-1, 2)$.
7. $\sqrt{|x^3 - 3x + 2|}$.
8. $\sqrt[7]{(x^3 - 3x + 2)^2}$.
9. $\dfrac{1}{x^3 - 3x + 2}$.

Exercise 2.3.8. Determine the monotone property and find extrema.

1. $\dfrac{x}{1+x^2}$ on $[-1, 1]$.
2. $\dfrac{\sin x}{1+\sin^2 x}$ on $[0, 2\pi]$.
3. $\dfrac{\cos x}{1+\cos^2 x}$ on $[0, 2\pi]$.
4. $\dfrac{1+x^2}{x}$ on $[-1, 0) \cup (0, 1]$.
5. $\tan x + \cot x$ on $\left[-\dfrac{\pi}{4}, 0\right) \cup \left(0, \dfrac{\pi}{4}\right]$.
6. $e^x + e^{-x}$ on $\mathbb{R}$.

Exercise 2.3.9. Determine the monotone property and find extrema.

1. x^4 on $[-1,1]$.
2. $\cos^4 x$ on $\mathbb{R}$.
3. $\sin^2 x$ on $\mathbb{R}$.

Exercise 2.3.10. Determine the monotone property and find extrema.

1. $-x^4+2x^2-1$ on $[-2,2]$.
2. $\sqrt{3+2x-x^2}$ on $(-1,3]$.
3. $|x|^p(x+1)$ on $\mathbb{R}$.
4. x^2e^x on $\mathbb{R}$.
5. x^pa^x on $(0,+\infty)$.
6. $|x|e^{-|x-1|}$ on $[-2,2]$.
7. $x\log x$ on $(0,+\infty)$.
8. $x^2\log^3 x$ on $(0,+\infty)$.
9. $x^p\log x$ on $(0,+\infty)$.
10. $x^3+3\log x$ on $(0,+\infty)$.
11. $x-\log(1+x)$ on $(-1,+\infty)$.
12. $e^{-x}\sin x$ on $\mathbb{R}$.
13. $x-\sin x$ on $[0,2\pi]$.
14. $|x-\sin x|$ on $[-\pi,\pi]$.
15. $|x-\sin x|$ on $\left[-\dfrac{\pi}{2},\dfrac{\pi}{2}\right]$.
16. $2\sin x+\sin 2x$ on $[0,2\pi]$.
17. $2x-4\sin x+\sin 2x$ on $[0,\pi]$.

Exercise 2.3.11. Show that $2x+\sin x=c$ has only one solution. Show that $x^4+x=c$ has at most two solutions.

Exercise 2.3.12. If f is differentiable and has 9 roots on (a,b), how many roots does f' have on (a,b)? If f also has second order derivative, how many roots does f'' have on (a,b)?

Exercise 2.3.13. Find smallest $A>0$, such that $\log x\le A\sqrt{x}$. Find smallest $B>0$, such that $\log x\ge -\dfrac{B}{\sqrt{x}}$.

Exercise 2.3.14. A quantity is measured n times, yielding the measurements $x_1,\dots,x_n$. Find the estimate value $\hat{x}$ of x that minimizes the squared error $(x-x_1)^2+\cdots+(x-x_n)^2$.

Exercise 2.3.15. Find the biggest term in the sequence $\sqrt[n]{n}$.

2.3.3 Compare Functions

If we apply Theorem 2.3.2 to $f(x)-g(x)$, then we get the following comparison of two functions.

Theorem 2.3.3. *Suppose $f(x)$ and $g(x)$ are continuous for $x\ge a$ and differentiable for $x>a$. If $f(a)\ge g(a)$ and $f'(x)\ge g'(x)$ for $x>a$, then $f(x)\ge g(x)$ for $x>a$. If $f(a)\ge g(a)$ and $f'(x)>g'(x)$ for $x>a$, then $f(x)>g(x)$ for $x>a$.*

There is a similar statement for the case $x<a$.

Example 2.3.8. We have $e^x > 1$ for $x > 0$ and $e^x < 1$ for $x < 0$. This is the comparison of e^x with the constant term of the Taylor expansion (or the 0th Taylor expansion) in Example 2.5.5. How do we compare e^x with the first order Taylor expansion $1 + x$?

We have $e^0 = 1+0$. For $x > 0$, we have $(e^x)' = e^x > (1+x)' = 1$. Therefore we get $e^x > 1+x$ for $x > 0$. On the other hand, for $x < 0$, we have $(e^x)' = e^x < (1+x)' = 1$. Therefore we also get $e^x > 1 + x$ for $x < 0$. We conclude that

$$e^x > 1 + x \text{ for } x \neq 0.$$

Example 2.3.9. We claim that

$$\frac{x}{1+x} < \log(1+x) < x \text{ for } x > -1,\ x \neq 0.$$

The three functions have the same value 0 at 0. Then we compare their derivatives

$$\left(\frac{x}{1+x}\right)' = \frac{1}{(1+x)^2}, \quad (\log(1+x))' = \frac{1}{1+x}, \quad (x)' = 1.$$

We have

$$\frac{1}{(1+x)^2} < \frac{1}{1+x} < 1 \text{ for } x > 0,$$

and

$$\frac{1}{(1+x)^2} > \frac{1}{1+x} > 1 \text{ for } -1 < x < 0.$$

The inequalities then follow from Theorem 2.3.3.

Example 2.3.10. For $a, b > 0$ and $p > q > 0$, we claim that

$$(a^p + b^p)^{\frac{1}{p}} < (a^q + b^q)^{\frac{1}{q}}.$$

By symmetry, we may assume $a \leq b$. Then $c = \dfrac{b}{a} > 1$, and the inequality means that $f(x) = (1 + c^x)^{\frac{1}{x}}$ is strictly decreasing for $x > 0$.

By Example 2.2.8, we have

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

So we study the monotone property of the function $g(t) = t \log t$. By

$$g'(t) = \log t + 1 > 0, \text{ for } t > e^{-1},$$

we see that $g(t)$ is increasing for $t > e^{-1}$. Since $c \geq 1$ and $x > 0$ implies $1 + c^x > c^x \geq 1 > e^{-1}$, we get $c^x \log c^x \leq (1 + c^x)\log(1 + c^x)$. Therefore $f'(x) < 0$ and $f(x)$ is decreasing.

Exercise 2.3.16. State Theorem 2.3.3 for the case $x < a$.

Exercise 2.3.17. Prove the inequality.

1. $\sin x > \frac{2}{\pi}x$, for $0 < x < \frac{\pi}{2}$.
2. $\frac{1}{2^{p-1}} \le x^p + (1-x)^p \le 1$, for $0 \le x \le 1$, $p > 1$.
3. $\frac{\sqrt{3}}{6+2\sqrt{3}} \le \frac{1+x}{2+x^2} \le \frac{\sqrt{3}}{6-2\sqrt{3}}$.
4. $\left(1+\frac{1}{x}\right)^x < e < \left(1+\frac{1}{x}\right)^{x+1}$, for $x > 0$.
5. $\arctan x - \arctan y \le 2\arctan\frac{x-y}{2}$, for $x > y > 0$.
6. $\frac{x^2}{2(1+x)} < x - \log(1+x) < \frac{x^2}{2}$, for $x > 0$. What about $-1 < x < 0$?

Exercise 2.3.18. For natural number n and $0 < a < 1$, prove that the equation

$$1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} = ae^x$$

has only one solution on $(0, +\infty)$.

2.3.4 First Derivative Test

We saw that the local extrema are often the places where the function changes between increasing and decreasing. If the function is differentiable, then these are the places where the derivative changes the sign. In particular, we expect the derivatives at these places to become 0. This leads to the following criterion for the candidates of local extrema.

Theorem 2.3.4. *If $f(x)$ is differentiable at a local extreme x_0, then $f'(x_0) = 0$.*

If $f'(x_0) > 0$, then the linear approximation $L(x) = f(x_0) + f'(x_0)(x - x_0)$ of f near x_0 is strictly increasing. This means that $L(x) < L(x_0)$ for $x < x_0$ and $L(x) > L(x_0)$ for $x > x_0$. Since L is very close to f near x_0, we expect that f is also "lower" on the left of x_0 and "higher" on the right of x_0. In particular, this implies that x_0 is not a local extreme of f. By the similar argument, if $f'(x_0) < 0$, the x_0 is also not a local extreme. This is the reason behind the theorem.

Since our reason makes explicit use of both the left and right sides, the criterion does not work for one sided derivatives. Therefore for a function f defined on an interval, the candidates for the local extrema must be one of the following cases:

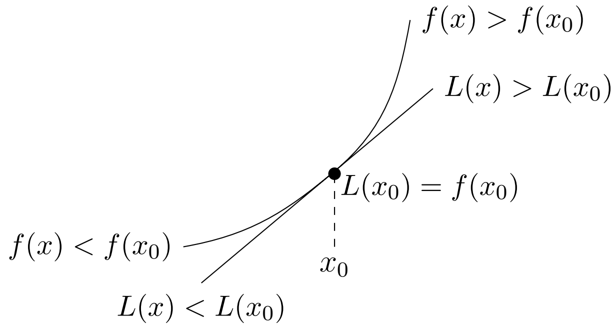


Figure 2.3.3: What happens near x_0 when $f'(x_0) > 0$.

1. End points of the interval.

2. Points inside the interval where f is not differentiable.

3. Points inside the interval where f is differentiable and has derivative 0.

Example 2.3.11. The derivative $(x^2)' = 2x$ vanishes only at 0. Therefore the only candidate for the local extrema of x^2 on $\mathbb{R}$ is 0. By $x^2 \geq 0^2$ for all x, 0 is a minimum.

If we restrict x^2 to the closed interval $[-1, 2]$, then the end points -1 and 2 are also candidates for the local extrema. By $x^2 \leq (-1)^2$ on $[-1, 0]$ and $x^2 \leq 2^2$ on $[-1, 2]$, -1 is a local maximum and 2 is a global maximum.

On the other hand, the restriction of x^2 on the open interval $(-1, 2)$ has no other candidates for local extrema besides 0. The function has global minimum at 0 and has no local maximum on $(-1, 2)$.

Example 2.3.12. Consider the function

$$f(x) = \begin{cases} x^2, & \text{if } x \neq 0, \\ 2, & \text{if } x = 0, \end{cases}$$

that modifies the square function by reassigning the value at 0. The function is not differentiable at 0 and has nonzero derivative away from 0. Therefore on $[-1, 2]$, the candidates for the local extrema are 0 and the end points -1 and 2. The end points are also local maxima, like the unmodified x^2. By $x^2 < f(0) = 2$ on $[-1, 1]$, 0 is a local maximum. The modified square function $f(x)$ has no local minimum on $[-1, 2]$.

Example 2.3.13. The function $f(x) = x^3 - 3x + 1$ in Example 2.3.4 has derivative $f'(x) = 3(x+1)(x-1)$. The possible local extrema on $\mathbb{R}$ are ± 1. These are not the global extrema on the whole line because $\lim_{x \to -\infty} f(x) = -\infty$ and $\lim_{x \to +\infty} f(x) = +\infty$.

If we restrict the function to $[-2,2]$, then ± 2 are also possible local extrema. By comparing the values

$$f(-2)=-1, \quad f(-1)=3, \quad f(1)=-1, \quad f(2)=3,$$

we get global minima at $-2, 1$, and global maxima at $2, -1$.

Example 2.3.14. By $(x^3)' = 3x^2$, 0 is the only candidate for the local extreme of x^3. However, we have $x^3 < 0^3$ for $x < 0$ and $x^3 > 0^3$ for $x > 0$. Therefore 0 is actually not a local extreme.

The example shows that the converse of Theorem 2.3.4 is not true.

Example 2.3.15. The function $f(x) = xe^{-x}$ has derivative $f'(x) = (x-1)e^x$. The only possible local extreme on $\mathbb{R}$ is at 1. We have $\lim_{x\to-\infty} f(x) = -\infty$ and $\lim_{x\to+\infty} f(x) = 0$. We claim that the limits imply that $f(1) = e^{-1}$ is a global maximum.

Since the limits at both infinity are $< f(1)$, there is N, such that $f(x) < f(1)$ for $|x| \geq N$. In particular, we have $f(\pm N) < f(1)$. Then consider the function on $[-N, N]$. On the bounded and closed interval, Theorem 2.3.1 says that the continuous function must reach its maximum, and the candidates for the maximum on $[-N, N]$ are $-N, 1, N$. Since $f(\pm N) < f(1)$, we see that $f(1)$ is the maximum on $[-N, N]$. Combined with $f(x) < f(1)$ for $|x| \geq N$, we conclude that $f(1)$ is the maximum on the whole real line.

Exercise 2.3.19. Find the global extrema.

1. $x^2(x-1)^3$ on $\mathbb{R}$.
2. $x^2(x-1)^3$ on $[-1,1]$.
3. $|x^2-1|$ on $[-2,1]$.
4. x^2+bx+c on $\mathbb{R}$.
5. $1-x^4+x^5$ on $\mathbb{R}$.
6. $\sin x^2$ on $[-1, \sqrt{\pi}]$.
7. $x \log x$ on $(0, +\infty)$.
8. $x \log x$ on $(0, 1]$.
9. x^x on $(0, 1]$.
10. $(x^2+1)e^x$ on $\mathbb{R}$.

2.3.5 Optimization Problem

Example 2.3.16. Given the circumference a of a rectangle, which rectangle has the largest area?

Let one side of the rectangle be x. Then the other side is $\dfrac{a}{2} - x$, and the area

$$A(x) = x\left(\frac{a}{2} - x\right).$$

The problem is to find the maximum of $A(x)$ on $\left[0, \dfrac{a}{2}\right]$.

By $A'(x) = \frac{a}{2} - 2x$, the candidates for the local extrema are $0, \frac{a}{4}, \frac{a}{2}$. The values of A at the three points are $0, \frac{a^2}{16}, 0$. Therefore the maximum is reached when $x = \frac{a}{4}$, which means the rectangle is a square.

Example 2.3.17. The distance from a point $P = (x_0, y_0)$ on the plane to a straight line $ax + by + c = 0$ is the minimum of the distance from P to a point (x, y) on the line. The distance is minimum when the square of the distance

$$f(x) = (x - x_0)^2 + (y - y_0)^2$$

is minimum. Note that y is a function of x given by the equation $ax + by + c = 0$ and satisfies $a + by' = 0$.

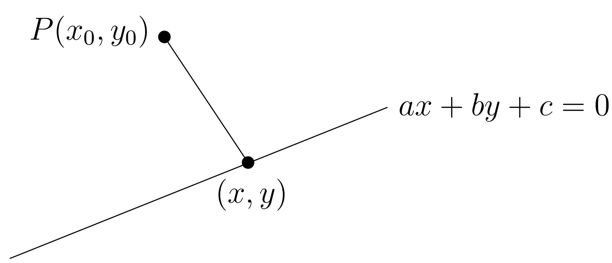


Figure 2.3.4: Distance from a point to a straight line.

From

$$f'(x) = 2(x - x_0) + 2(y - y_0)y' = \frac{2}{b}(b(x - x_0) - a(y - y_0)),$$

we know that $f(x)$ is minimized when

$$b(x - x_0) - a(y - y_0) = 0.$$

Moreover, recall that (x, y) must also be on the straight line

$$ax + by + c = 0.$$

Solving the system of two linear equations, we get

$$x - x_0 = -\frac{a(ax_0 + by_0 + c)}{a^2 + b^2}, \quad y - y_0 = -\frac{b(ax_0 + by_0 + c)}{a^2 + b^2}.$$

The minimum distance is

$$\sqrt{(x - x_0)^2 + (y - y_0)^2} = \frac{|ax_0 + by_0 + c|}{\sqrt{a^2 + b^2}}.$$

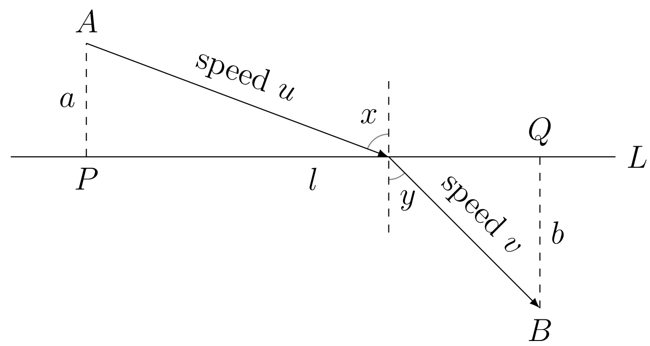


Figure 2.3.5: Snell's law.

Example 2.3.18. Consider light traveling from a point A in one medium to point B in another medium. Fermat's principle says that the path taken by the light is the path of shortest traveling time.

Let u and v be the speed of light in the respective medium. Let L be the place where two media meet. Draw lines AP and BQ perpendicular to L. Let the length of AP, BP, PQ be a, b, l. Let x be the angle by which the light from A hits L. Let y be the angle by which the light leaves L and reaches B.

The angles x and y are related by

$$a \tan x + b \tan y = l.$$

This can be considered as an equation that implicitly defines y as a function of x. The derivative of $y = y(x)$ can be obtained by implicit differentiation

$$y'(x) = -\frac{a \sec^2 x}{b \sec^2 y}.$$

The time it takes for the light to travel from A to B is

$$T = \frac{a \sec x}{u} + \frac{b \sec y}{v}.$$

By thinking of y as a function of x, the time T becomes a function of x. The time will be shortest when

$$\begin{aligned} 0 = \frac{dT}{dx} &= \frac{a \sec x \tan x}{u} + \frac{b \sec y \tan y}{v} y' \\ &= \frac{a \sec x \tan x}{u} - \frac{b \sec y \tan y}{v} \frac{a \sec^2 x}{b \sec^2 y} \\ &= a \sec^2 x \left(\frac{\sin x}{u} - \frac{\sin y}{v} \right). \end{aligned}$$

This means that the ratio between the sine of the angles x and y is the same as the ratio between the speeds of light

$$\frac{\sin x}{\sin y} = \frac{u}{v}.$$

This is *Snell's law of refraction*.

Exercise 2.3.20. A rectangle is inscribed in an isosceles triangle. Show that the biggest area possible is half of the area of the triangle.

Exercise 2.3.21. Among all the rectangles with area A, which one has the smallest perimeter?

Exercise 2.3.22. Among all the rectangles with perimeter L, which one has the biggest area?

Exercise 2.3.23. A rectangle is inscribed in a circle of radius R. When does the rectangle have the biggest area?

Exercise 2.3.24. Determine the dimensions of the biggest rectangle inscribed in the ellipse $\dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}=1$.

Exercise 2.3.25. Find the volume of the biggest right circular cone with a given slant height l.

Exercise 2.3.26. What is the shortest distance from the point $(2,1)$ to the parabola $y = 2x^2$?

2.4 Mean Value Theorem

If one travels between two cities at the average speed of 100 kilometers per hour, then we expect that the speed reaches exactly 100 kilometers per hour somewhere during the trip. Mathematically, let $f(t)$ be the distance traveled by the time t. Then the average speed from the time a to time b is

$$\frac{f(b)-f(a)}{b-a}.$$

Our expectation can be interpreted as

$$\frac{f(b)-f(a)}{b-a} = f'(c) \text{ for some } c \in (a,b).$$

2.4.1 Mean Value Theorem

Theorem 2.4.1 (Mean Value Theorem). *If $f(x)$ is continuous on $[a,b]$ and differentiable on (a,b), then there is $c \in (a,b)$, such that*

$$\frac{f(b)-f(a)}{b-a} = f'(c).$$

The conclusion can also be expressed as

$$f(b) - f(a) = f'(c)(b - a), \text{ for some } a < c < b,$$

or

$$f(a + h) - f(a) = f'(a + \theta h)h, \text{ for some } 0 < \theta < 1.$$

We also note that the conclusion is symmetric in a, b. Therefore there is no need to insist $a < b$.

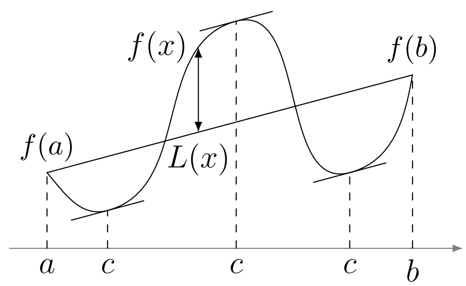


Figure 2.4.1: Mean value theorem.

Geometrically, the Mean Value Theorem means that the straight line L connecting the two ends $(a, f(a))$ and $(b, f(b))$ of the graph of f is parallel to the tangent of the function somewhere. Figure 2.4.1 suggests that c in the Mean Value Theorem is the place where the the distance between the graphs of f and L has local extrema. Since such local extrema for the distance $f(x) - L(x)$ always exists by Theorem 2.3.1, we get $(f - L)'(c) = 0$ for some c by Theorem 2.3.4. Therefore $f'(c) = L'(c) = \dfrac{f(b) - f(a)}{b - a}$.

Example 2.4.1. We try to verify the Mean Value Theorem for $f(x) = x^3 - 3x + 1$ on $[-1, 1]$. This means finding c, such that

$$f'(c) = 3(c^2 - 1) = \frac{f(1) - f(-1)}{1 - (-1)} = \frac{-1 - 3}{1 - (-1)} = -2.$$

We get $c = \pm\dfrac{1}{\sqrt{3}}$.

Example 2.4.2. By the Mean Value Theorem, we have

$$\log(1 + x) = \log(1 + x) - \log 1 = \frac{1}{1 + \theta x}x \text{ for some } 0 < \theta < 1.$$

Since

$$\frac{x}{1 + x} \leq \frac{1}{1 + \theta x}x \leq x \text{ for } x > -1,$$

we conclude that

$$\frac{x}{1+x} \le \log(1+x) \le x.$$

The inequality already appeared in Example 2.3.9.

Example 2.4.3. For the function $|x|$ on $[-1,1]$, there is no $c \in (-1,1)$ satisfying

$$f(1) - f(-1) = 0 = f'(c)(1-(-1)).$$

The Mean Value Theorem does not apply because $|x|$ is not differentiable at 0.

Exercise 2.4.1. Is the conclusion of the Mean Value Theorem true? If true, find c. If not, explain why.

1. x^3 on $[-1,1]$.
2. 2^x on $[0,1]$.
3. $\dfrac{1}{x}$ on $[1,2]$.
4. $|x^3-3x+1|$ on $[-1,1]$.
5. $\sqrt{|x|}$ on $[-1,1]$.
6. $\cos x$ on $[-a,a]$.
7. $\log x$ on $[1,2]$.
8. $\log|x|$ on $[-1,1]$.
9. $\arcsin x$ on $[0,1]$.

Exercise 2.4.2. Suppose $f(1) = 2$ and $f'(x) \le 3$ on $\mathbb{R}$. How large and how small can $f(4)$ be? What happens when the largest or the smallest value is reached? How about $f(-4)$?

Exercise 2.4.3. Prove inequality.

1. $|\sin x - \sin y| \le |x-y|$.
2. $\dfrac{x-y}{x} < \log\dfrac{x}{y} < \dfrac{x-y}{y}$, for $x > y > 0$.
3. $|\arctan x - \arctan y| \le |x-y|$.

Exercise 2.4.4. Find the biggest interval on which $|e^x - e^y| > |x-y|$? What about $|e^x - e^y| < |x-y|$?

Exercise 2.4.5. Suppose $f(x)$ is continuous at x_0 and differentiable on $(x_0-\delta, x_0)\cup(x_0, x_0+\delta)$. Prove that if $\lim_{x\to x_0} f'(x) = l$ converges, then $f(x)$ is differentiable at x_0 and $f'(x_0) = l$.

2.4.2 Criterion for Constant Function

The Mean Value Theorem can be used to prove Theorem 2.3.2: If $f' \ge 0$ on an interval, then for any $x_1 < x_2$ on the interval, by the Mean Value Theorem, we have

$$f(x_2) - f(x_1) = f'(c)(x_2 - x_1) \ge 0.$$

For the special case $f' = 0$ throughout the interval, the argument gives the following result.

Theorem 2.4.2. *If $f'(x) = 0$ for all $x \in (a, b)$, then $f(x)$ is a constant on (a, b).*

Applying the theorem to $f(x) - g(x)$, we get the following result.

Theorem 2.4.3. *If $f'(x) = g'(x)$ for all $x \in (a, b)$, then there is a constant C, such that $f(x) = g(x) + C$ on (a, b).*

Example 2.4.4. The function e^x satisfies $f(x)' = f(x)$. Are there any other functions satisfying the equation?

If $f(x)' = f(x)$. Then

$$(e^{-x}f(x))' = (e^{-x})'f(x) + e^{-x}(f(x))' = e^x(-f(x) + f'(x)) = 0.$$

Therefore $e^{-x}f(x) = C$ is a constant, and $f(x) = Ce^x$.

Example 2.4.5. Suppose $f'(x) = x$ and $f(1) = 2$. Then $f'(x) = \left(\frac{x^2}{2}\right)'$ implies $f(x) = \frac{x^2}{2} + C$ for some constant C. By taking $x = 1$, we get $2 = \frac{1}{2} + C$. Therefore $C = \frac{3}{2}$ and $f(x) = \frac{x^2 + 3}{2}$.

Example 2.4.6. By

$$(\arcsin x)' = \frac{1}{\sqrt{1 - x^2}}, \quad (\arccos x)' = -\frac{1}{\sqrt{1 - x^2}}.$$

we have $(\arcsin x + \arccos x)' = 0$, and we have $\arcsin x + \arccos x = C$. The constant can be determined by taking a special value $x = 0$

$$C = \arcsin 0 + \arccos 0 = 0 + \frac{\pi}{2} = \frac{\pi}{2}.$$

Therefore we have

$$\arcsin x + \arccos x = \frac{\pi}{2}.$$

Exercise 2.4.6. Prove that a differentiable function is linear on an interval if and only if its derivative is a constant.

Exercise 2.4.7. Find all functions on an interval satisfying the following equations.

1. $f'(x) = -2f(x)$.
2. $f'(x) = xf(x)$.
3. $f'(x) = f(x)^2$
4. $f'(x)f(x) = 1$.

Exercise 2.4.8. Prove equality.

1. $\arctan x + \arctan x^{-1} = \dfrac{\pi}{2}$, for $x \neq 0$.

2. $3\arccos x - \arccos(3x - 4x^3) = \pi$, for $|x| \leq \dfrac{1}{2}$.

3. $\arctan \dfrac{x+a}{1-ax} - \arctan x = \arctan a$, for $ax < 1$.

4. $\arctan \dfrac{x+a}{1-ax} - \arctan x = \arctan a - \pi$, for $ax > 1$.

2.4.3 L'Hospital's Rule

The following limits cannot be computed by simple arithmetic rules.

$$\begin{aligned}
\frac{0}{0}&\colon \lim_{x\to 1}\frac{x^3-1}{x-1},\ \lim_{x\to 0}\frac{\sin x}{x},\ \lim_{x\to 0}\frac{\log(1+x)}{x};\\
\frac{\infty}{\infty}&\colon \lim_{x\to 0}\frac{\log x}{x^{-1}},\ \lim_{x\to\infty}\frac{\log x}{x},\ \lim_{x\to\infty}\frac{x^2}{e^x};\\
1^{\infty}&\colon \lim_{x\to\infty}\left(1+\frac{1}{x}\right)^x,\ \lim_{x\to 0}(1+\sin x)^{\log x};\\
\infty-\infty&\colon \lim_{x\to 0}\left(\frac{1}{x}-\frac{1}{e^x-1}\right).
\end{aligned}$$

We say these limits are *indeterminate*. Other indeterminates include $0\cdot\infty$, $\infty+\infty$, 0^0, ∞^0. The derivative can help us computing such limits.

Theorem 2.4.4 (L'Hospital's Rule). *Suppose $f(x)$ and $g(x)$ are differentiable functions on (a,b), with $g'(x)\neq 0$. Suppose*

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. *$\lim_{x\to a^+} \dfrac{f'(x)}{g'(x)} = l$ converges.*

Then $\lim_{x\to a^+} \dfrac{f(x)}{g(x)} = l$.

The theorem computes the limits of the indeterminates of type $\dfrac{0}{0}$ or $\dfrac{\infty}{\infty}$. The conclusion is the equality

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

whenever the right side converges. It is possible that the left side converges but the right side diverges.

The theorem also has a similar left sided version, and the left and right sided versions may be combined to give the two sided version. Moreover, l'Hospital's rule also allows a or l to be any kind of infinity.

The reason behind l'Hospital's rule is the following version of the Mean Value Theorem, which can be proved similar to the Mean Value Theorem.

Theorem 2.4.5 (Cauchy's Mean Value Theorem). *If $f(x)$ and $g(x)$ are continuous on $[a,b]$ and differentiable on (a,b), such that $g'(x) \neq 0$ on (a,b), then there is $c \in (a,b)$, such that*

$$\frac{f(b)-f(a)}{g(b)-g(a)} = \frac{f'(c)}{g'(c)}.$$

Consider the parametrized curve $(g(t), f(t))$ for $t \in [a,b]$. The theorem says that the straight line connecting the two ends $(g(a), f(a))$ and $(g(b), f(b))$ of the curve is parallel to the tangent of the curve somewhere. The slope of the tangent is $\dfrac{f'(c)}{g'(c)}$.

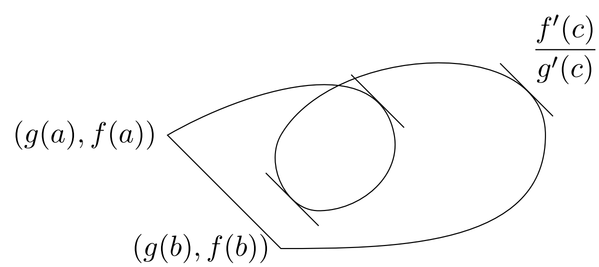


Figure 2.4.2: Cauchy's Mean Value Theorem.

For the case $\lim_{x\to a^+} f(x) = \lim_{x\to a^+} g(x) = 0$ of l'Hospital's rule, we may extend f and g to continuous functions on $[a,b)$ by assigning $f(a) = g(a) = 0$. Then for any $a < x < b$, we may apply Cauchy's Mean Value Theorem to the functions on $[a,x]$ and get

$$\frac{f(x)}{g(x)} = \frac{f(x)-f(a)}{g(x)-g(a)} = \frac{f'(c)}{g'(c)} \text{ for some } c \in (a,x).$$

Since $x \to a^+$ implies $c \to a^+$, we conclude that $\lim_{c\to a^+} \dfrac{f'(c)}{g'(c)} = l$ implies $\lim_{x\to a^+} \dfrac{f(x)}{g(x)} = l$.

By changing the variable x to $\dfrac{1}{x}$, it is not difficult to extend the proof to the case $a = \pm\infty$. The proof for the $\dfrac{\infty}{\infty}$ type is must more complicated and is omitted here.

Example 2.4.7. In Example 1.2.15, we proved that $\lim_{x\to+\infty} a^x = 0$ for $0 < a < 1$. Exercise 1.6.15 extended the limit to $\lim_{x\to+\infty} x^2 a^x = 0$. We derive the second limit from the first one by using l'Hospital's rule.

We have $b = \dfrac{1}{a} > 1$, and $\lim_{x\to+\infty} a^x = 0$ is the same as $\lim_{x\to+\infty} b^x = \infty$. We also have $\lim_{x\to+\infty} x^2 = \infty$. Therefore $\lim_{x\to+\infty} x^2a^x = \lim_{x\to+\infty} \dfrac{x^2}{b^x}$ is of type $\dfrac{\infty}{\infty}$, and we may apply l'Hospital's rule (twice)

$$\begin{aligned}\lim_{x\to+\infty}\frac{x^2}{b^x} &=_{(3)} \lim_{x\to+\infty}\frac{(x^2)'}{(b^x)'} = \lim_{x\to+\infty}\frac{2x}{b^x\log b}\\ &=_{(2)} \lim_{x\to+\infty}\frac{(2x)'}{(b^x\log b)'} = \lim_{x\to+\infty}\frac{2}{b^x(\log b)^2} =_{(1)} 0.\end{aligned}$$

Here is the precise reason behind the computation. The equality $=_{(1)}$ is from Example 1.2.15. Then by l'Hospital's rule, the convergence of the right side of $=_{(2)}$ implies the convergence of the left side of $=_{(2)}$ and the equality $=_{(2)}$ itself. The left of $=_{(2)}$ is the same as the right side of $=_{(3)}$. By l'Hospital's rule gain, the convergence of the right side of $=_{(3)}$ implies the convergence of the left side of $=_{(3)}$ and the equality $=_{(3)}$.

Example 2.4.8. Applying l'Hospital's rule to the limit $\lim_{x\to 0}\dfrac{\sin x}{x}$ of type $\dfrac{0}{0}$, we get

$$\lim_{x\to 0}\frac{\sin x}{x} = \lim_{x\to 0}\frac{(\sin x)'}{(x)'} = \lim_{x\to 0}\cos x = 1.$$

However, this argument is logically circular because it makes use of the formula $(\sin x)' = \cos x$. A special case of this formula is

$$\lim_{x\to 0}\frac{\sin x}{x} = (\sin x)'|_{x=0} = 1,$$

which is exactly the conclusion we try to get.

Exercise 2.4.9. Are the application of l'Hospital's rule logically circular?

1. $\lim_{x\to 0}\dfrac{x}{\sin x}$.
2. $\lim_{x\to 0}\dfrac{\cos x - 1}{x}$.
3. $\lim_{x\to 0}\dfrac{\cos x - 1}{x^2}$.
4. $\lim_{x\to 1}\dfrac{x^2 - 1}{x - 1}$.
5. $\lim_{x\to 1}\dfrac{x^2 - 1}{x^3 - 1}$.
6. $\lim_{x\to 0}\dfrac{\log(1+x)}{x}$.

Example 2.4.9. By blindly using l'Hospital's rule four times, we have

$$\begin{aligned}
\lim_{x\to 0}\frac{\sin^2 x-\sin x^2}{x^4} &= \lim_{x\to 0}\frac{2\sin x\cos x-2x\cos x^2}{4x^3}\\
&= \lim_{x\to 0}\frac{\cos^2 x-\sin^2 x-\cos x^2+2x^2\sin x^2}{6x^2}\\
&= \lim_{x\to 0}\frac{-4\cos x\sin x+6x\sin x^2+4x^3\cos x^2}{12x}\\
&= \lim_{x\to 0}\frac{4\sin^2 x-4\cos^2 x+6\sin x^2+24x^2\cos x^2-8x^4\sin x^2}{12}\\
&= -\frac{1}{3}.
\end{aligned}$$

We find that it is increasingly difficult to calculate the derivatives. The following compute the the limit after calculating the derivatives twice.

$$\begin{aligned}
\lim_{x\to 0}\frac{\sin^2 x-\sin x^2}{x^4} &= \lim_{x\to 0}\frac{2\sin x\cos x-2x\cos x^2}{4x^3}\\
&= \lim_{x\to 0}\frac{\cos^2 x-\sin^2 x-\cos x^2+2x^2\sin x^2}{6x^2}\\
&= \lim_{x\to 0}\left(-\frac{1}{3}\left(\frac{\sin x}{x}\right)^2+\frac{1-\cos x^2}{6x^2}+\frac{1}{3}\sin x^2\right)\\
&= -\frac{1}{3}\cdot 1^2+0+\frac{1}{3}\cdot 0=-\frac{1}{3}.
\end{aligned}$$

In fact, the smartest way is not to calculate the derivatives at all. See Example 2.5.11.

Exercise 2.4.10. Use l'Hospital's rule to compute the limits.

1. $\lim_{x\to 0}\dfrac{\sin x-x}{x^3}$.
2. $\lim_{x\to 0}\dfrac{\sin x-\tan x}{x^3}$.
3. $\lim_{x\to 0}\dfrac{e^x-e^{\sin x}}{x^3}$.
4. $\lim_{x\to 0}\dfrac{\cos(\sin x)-\cos x}{x^4}$.
5. $\lim_{x\to 0}\dfrac{1-\cos x^2}{x^3\sin x}$.
6. $\lim_{x\to 1}\dfrac{x^x-x}{\log x-x+1}$.
7. $\lim_{x\to 0}\dfrac{x-\tan x}{x-\sin x}$.
8. $\lim_{x\to 1}\dfrac{(x-1)\log x}{1+\cos \pi x}$.

Example 2.4.10. The limit $\lim_{x\to 0^+}x\log x$ is of type $0\cdot\infty$. We convert into type $\dfrac{\infty}{\infty}$ and apply l'Hospital's rule (in the second equality)

$$\lim_{x\to 0^+}x\log x=\lim_{x\to 0^+}\frac{\log x}{x^{-1}}=\lim_{x\to 0^+}\frac{x^{-1}}{-x^{-2}}=\lim_{x\to 0^+}-x=0.$$

The similar argument gives

$$\lim_{x\to 0^+} x^p \log x = 0, \text{ for } p > 0.$$

Taking a positive power of the limit, we further get

$$\lim_{x\to 0^+} x^p(-\log x)^q = 0, \text{ for } p, q > 0.$$

By converting x to $\frac{1}{x}$, we also have

$$\lim_{x\to+\infty} \frac{(\log x)^q}{x^p} = 0, \text{ for } p, q > 0.$$

Example 2.4.11. We compute the limit in Example 2.5.13 by first converting it to type $\frac{0}{0}$ and then applying l'Hospital's rule

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

Example 2.4.12. If we apply l'Hospital's rule to the limit $\lim_{x\to\infty}\frac{x+\sin x}{x}$ of type $\frac{\infty}{\infty}$, then we get

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

We find that the left converges and the right diverges. The reason for the l'Hospital's rule to fail is that the second condition is not satisfied.

2.5 High Order Approximation

Linear approximations can be used to solve many problems. When linear approximations are not enough, however, we may use high order approximations.

Definition 2.5.1. An *n-th order approximation* of $f(x)$ at x_0 is a degree n polynomial

$$P(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)^2 + \cdots + a_n(x - x_0)^n,$$

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

$$|x - x_0| < \delta \implies |f(x) - P(x)| \le \epsilon|x - x_0|^n.$$

A function is *n-th order differentiable* if it has n-order approximation.

The error $R_n(x) = f(x) - P(x)$ of the approximation is called the *remainder*. The definition means that

$$\begin{aligned} f(x) &= P(x) + R_n(x) \\ &= a_0 + a_1(x - x_0) + a_2(x - x_0)^2 + \cdots + a_n(x - x_0)^n + o((x - x_0)^n), \end{aligned}$$

where the "small o" notation means that the remainder term satisfies

$$\lim_{x\to a} \frac{f(x) - P(x)}{(x - x_0)^n} = \lim_{x\to a} \frac{R_n(x)}{(x - x_0)^n} = 0.$$

The n-th order approximation of a function is unique. See Exercise 2.5.7. Moreover, if $m < n$, then the *truncation* $a_0 + a_1(x - x_0) + a_2(x - x_0)^2 + \cdots + a_m(x - x_0)^m$ is the m-th order approximation of f at x_0. After all, if we have the 10th order approximation, then we should also have the 5th order approximation.

Example 2.5.1. We have

$$x^4 = (1 + (x - 1))^4 = 1 + 4(x - 1) + 6(x - 1)^2 + 4(x - 1)^3 + (x - 1)^4.$$

For any $\epsilon > 0$, we have $|x - 1| < \delta = \min\left\{1, \frac{\epsilon}{5}\right\}$ implying

$$\begin{aligned} |x^4 - 1 - 4(x - 1) - 6(x - 1)^2| &= |4(x - 1)^3 + (x - 1)^4| \le (4 + |x - 1|)|x - 1|^3 \\ &\le (4 + 1)\frac{\epsilon}{5}|x - 1|^2 = \epsilon|x - 1|^2. \end{aligned}$$

Therefore $1 + 4(x - 1) + 6(x - 1)^2$ is the quadratic approximation of x^4 at 1. By similar argument, we get approximations of other orders.

$$\begin{aligned} \text{linear}&: 1 + 4(x - 1), \\ \text{quadratic}&: 1 + 4(x - 1) + 6(x - 1)^2, \\ \text{cubic}&: 1 + 4(x - 1) + 6(x - 1)^2 + 4(x - 1)^3, \\ \text{quartic}&: 1 + 4(x - 1) + 6(x - 1)^2 + 4(x - 1)^3 + (x - 1)^4, \\ \text{quintic}&: 1 + 4(x - 1) + 6(x - 1)^2 + 4(x - 1)^3 + (x - 1)^4, \\ &\vdots \end{aligned}$$

Example 2.5.2. The limit

$$\lim_{x\to 0} \frac{\cos x - 1}{x^2} = -\frac{1}{2}$$

in Example 1.5.18 can be interpreted as

$$\lim_{x\to 0} \frac{\cos x - 1 + \frac{1}{2}x^2}{x^2} = 0.$$

This means that $\cos x$ is second order differentiable at 0, with quadratic approximation $1 - \frac{1}{2}x^2$.

Exercise 2.5.1. Prove that $\dfrac{1}{1-x} = 1 + x + x^2 + \cdots + x^n + \dfrac{x^{n+1}}{1-x}$. What does this tell you about the differentiability of $\dfrac{1}{1-x}$ at 0?

Exercise 2.5.2. Show that $\dfrac{1}{1+x}$ and $\dfrac{1}{1+x^2}$ are differentiable of arbitrary order. What are their high order approximations?

Exercise 2.5.3. What is the n-th order approximation of $1 + 2x + 3x^2 + \cdots + 100x^{100}$ at 0?

Exercise 2.5.4. Use l'Hospital's rule to compute the limits. Then interpret your results as high order approximations.

1. $\lim_{x\to 0} \dfrac{\sin x - x}{x^3}$.
2. $\lim_{x\to 0} \dfrac{\sin x^2 - x^2}{x^6}$.
3. $\lim_{x\to 0} \dfrac{\sin^2 x - x^2}{x^4}$.
4. $\lim_{x\to 0} \dfrac{1}{x^5}\left(\sin x - x + \dfrac{1}{6}x^3\right)$.
5. $\lim_{x\to 0} \dfrac{\sin x - \tan x}{x^3}$.
6. $\lim_{x\to 0} \dfrac{1}{x^3}\left(e^x - 1 - x - \dfrac{1}{2}x^2\right)$.
7. $\lim_{x\to 0} \dfrac{1}{x^3}\left(\log(1+x) - x + \dfrac{1}{2}x^2\right)$.
8. $\lim_{x\to 1} \dfrac{2\log x + (x-1)(x-3)}{(x-1)^3}$.

Exercise 2.5.5. For what choice of a, b, c is the function

$$\begin{cases} x^4, & \text{if } x \geq 1, \\ a + bx + cx^2, & \text{if } x < 1, \end{cases}$$

second order differentiable at 1? Is it possible for the function to be third order differentiable?

Exercise 2.5.6. Suppose $f(x)$ is second order differentiable at 0. Show that

$$3f(x) - 3f(2x) + f(3x) = f(0) + o(x^2).$$

Exercise 2.5.7. Suppose $P(x) = a_0 + a_1(x-a) + a_2(x-a)^2$ satisfies $\lim_{x\to a} \dfrac{P(x)}{(x-a)^2} = 0$. Prove that $a_0 = a_1 = a_2 = 0$. Then explain that the result means the uniqueness of quadratic approximation. Moreover, extend the result to high order approximation.

Exercise 2.5.8. Suppose $P(x)$ is the n-order approximation of $f(x)$. What is the n-order approximation of $f(-x)$? Then use Exercise 2.5.7 to explain that the high order approximation of an even function has not odd power terms. What about the high order approximation of an odd function?

2.5.1 Taylor Expansion

The linear approximation may be computed by the derivative. The high order approximation may be computed by repeatedly taking the derivative. The idea is suggested by the following example. By applying l'Hospital's rule three times, we get a more precise limit than the one in Example 2.5.2

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

Each application of the l'Hospital's rule means taking derivative once. Therefore we get the third order approximation of $\cos x$ at 0 by taking derivative three times.

If $f(x)$ is differentiable everywhere on an open interval, then the derivative $f'(x)$ is a function on the open interval. If the derivative function $f'(x)$ is also differentiable, then we get the *second order derivative* $f''(x) = (f'(x))'$. If the function $f''(x)$ is yet again differentiable, then taking the derivative one more time gives the *third order derivative* $f'''(x) = (f''(x))'$. The process may continue and we have the *n-th order derivative* $f^{(n)}(x)$. The Leibniz notation for the high order derivative $f^{(n)}(x)$ is $\dfrac{d^n f}{dx^n}$.

Let $f(x) = \cos x$ and $P(x) = 1 - \frac{1}{2}x^2$. The key to the repeated application of the l'Hospital's rule is that the numerator is always 0 at $x_0 = 0$. This means that

$$f(x_0) = P(x_0),\ f'(x_0) = P'(x_0),\ f''(x_0) = P''(x_0),\ f'''(x_0) = P'''(x_0).$$

In general, if $P(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)^2 + a_3(x - x_0)^3$, then the above equalities become

$$f(x_0) = a_0,\ f'(x_0) = a_1,\ f''(x_0) = 2a_2,\ f'''(x_0) = 3\cdot 2a_3.$$

Theorem 2.5.2. *If $f(x)$ has n-th order derivative at x_0, then f is n-th order differentiable, with n-th order approximation*

$$T_n(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n.$$

The polynomial T_n is called the n-th order *Taylor expansion* of f.

Note that the existence of the derivative $f^{(n)}(x_0)$ implicitly assumes that $f^{(k)}(x)$ exists for all x near x_0 and all $k < n$. The theorem gives one way (but not the only way!) to compute the high order approximation in case the function has high order derivative. However, we will show in Example 2.5.8 that it is possible to have high order approximation without the existence of the high order derivative. Here the concept (high order differentiability) is strictly weaker than the computation (high order derivative).

Example 2.5.3. The high order derivatives of the power function x^p are

$$\begin{aligned}
(x^p)' &= px^{p-1}, \\
(x^p)'' &= p(p-1)x^{p-2}, \\
&\vdots \\
(x^p)^{(n)} &= p(p-1)\cdots(p-n+1)x^{p-n}.
\end{aligned}$$

More generally, we have

$$((a+bx)^p)^{(n)} = p(p-1)\cdots(p-n+1)b^n(a+bx)^{p-n}.$$

For $a = b = 1$, we get the Taylor expansion at 0

$$(1+x)^p = 1 + px + \frac{p(p-1)}{2!}x^2 + \cdots + \frac{p(p-1)\cdots(p-n+1)}{n!}x^n + o(x^n).$$

For $a = 1$, $b = -1$ and $p = -1$, we get the Taylor expansion at 0

$$\frac{1}{1-x} = 1 + x + x^2 + \cdots + x^n + o(x^n).$$

You may compare with Exercise 2.5.1.

Example 2.5.4. By $(\log x)' = x^{-1}$ and the derivatives from Example 2.5.3, we have $(\log x)^{(n)} = (-1)^{n-1}\dfrac{(n-1)!}{x^n}$. This gives the Taylor expansion at 1

$$\log x = (x-1) - \frac{1}{2}(x-1)^2 + \frac{1}{3}(x-1)^3 - \cdots + (-1)^{n+1}\frac{1}{n}(x-1)^n + o((x-1)^n).$$

This can also be expressed as a Taylor expansion at 0

$$\log(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \cdots + (-1)^{n+1}\frac{1}{n}x^n + o(x^n).$$

Example 2.5.5. By $(e^x)' = e^x$, it is easy to see that $(e^x)^{(n)} = e^x$ for all n. This gives the Taylor expansion at 0

$$e^x = 1 + \frac{1}{1!}x + \frac{1}{2!}x^2 + \cdots + \frac{1}{n!}x^n + o(x^n).$$

Example 2.5.6. The high order derivatives of $\sin x$ and $\cos x$ are 4-periodic in the sense that $\sin^{(n+4)} x = \sin^{(n)} x$ and $\cos^{(n+4)} x = \cos^{(n)} x$, and are given by

$$\begin{aligned}
(\sin x)' &= \cos x, & (\cos x)' &= -\sin x, \\
(\sin x)'' &= -\sin x, & (\cos x)'' &= -\cos x, \\
(\sin x)''' &= -\cos x, & (\cos x)''' &= \sin x, \\
(\sin x)'''' &= \sin x, & (\cos x)'''' &= \cos x.
\end{aligned}$$

This gives the Taylor expansions at 0

$$\sin x = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \cdots + (-1)^{n+1}\frac{1}{(2n-1)!}x^{2n-1} + o(x^{2n}),$$
$$\cos x = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \cdots + (-1)^{n}\frac{1}{(2n)!}x^{2n} + o(x^{2n+1}).$$

Note that we have $o(x^{2n})$ for $\sin x$ at the end, which is more accurate than $o(x^{2n-1})$. The reason is that the $2n$-th term $0 \cdot x^{2n}$ is omitted from the expression, so that the approximation is actually of $2n$-th order. The similar remark applies to $\cos x$.

We also note that the Taylor expansions of e^x, $\sin x$, $\cos x$ are related by the equality

$$e^{ix} = \cos x + i \sin x, \quad i = \sqrt{-1}.$$

Exercise 2.5.9. Prove the following properties of high order derivative

$$(f+g)^{(n)} = f^{(n)} + g^{(n)},$$
$$(cf)^{(n)} = cf^{(n)},$$
$$(fg)^{(n)} = \sum_{i=0}^{n} \frac{n!}{i!(n-i)!} f^{(i)} g^{(j)}.$$

Exercise 2.5.10. Prove the chain rule for second order derivative

$$(g(f(x)))'' = g''(f(x))f'(x)^2 + g'(f(x))f''(x).$$

Exercise 2.5.11. Compute derivatives of all order.

1. a^x.
2. e^{ax+b}.
3. $\sin(ax+b)$.
4. $\cos(ax+b)$.
5. $\log(ax+b)$.
6. $\log \dfrac{ax+b}{cx+d}$.
7. $\dfrac{ax+b}{cx+d}$.
8. $\dfrac{1}{(ax+b)(cx+d)}$.

Exercise 2.5.12. Compute high order derivatives.

1. $(\tan x)'''$.
2. $(\sec x)'''$.
3. $(\sin x^2)'''$.
4. $(\arcsin x)''$.
5. $(\arctan x)''$.
6. $(x^x)''$.
7. $\dfrac{d^2}{dx^2}\left(1+\dfrac{1}{x}\right)^x$.

Exercise 2.5.13. Use high order derivatives to find high order approximations.

1. a^x, $n = 5$, at 0.
2. a^x, $n = 5$, at 1.
3. $\sin^2 x$, $n = 6$, at 0.
4. $\sin^2 x$, $n = 6$, at π.
5. e^{x^2}, $n = 6$, at 0.
6. e^{x^2}, $n = 6$, at 1.
7. $x^3 e^x$, $n = 5$, at 0.
8. $x^3 e^x$, $n = 5$, at 1.
9. $e^x \sin x$, $n = 5$ at 1.

Exercise 2.5.14. Compute high order derivatives.

1. $(x^2+1)e^x$.
2. $(x^2+1)\sin x$.
3. $x^2(x-1)^p$.
4. $x\log x$.

Exercise 2.5.15. Prove $(x^{n-1}e^{\frac{1}{x}})^{(n)} = \dfrac{(-1)^n}{x^{n+1}}e^{\frac{1}{x}}$.

Exercise 2.5.16. Prove $(e^{ax}\sin(bx+c))^{(n)} = (a^2+b^2)^{\frac{n}{2}}e^{ax}\sin(bx+c+n\theta)$, where $\sin\theta = \dfrac{b}{\sqrt{a^2+b^2}}$. What is the similar formula for $(e^{ax}\cos(bx+c))^{(n)}$?

Exercise 2.5.17. Suppose $f(x)$ has second order derivative near x_0. Prove that

$$f''(x_0) = \lim_{h\to 0}\frac{f(x_0+h)+f(x_0-h)-2f(x_0)}{h^2}.$$

Exercise 2.5.18. Compare $e^x, \sin x, \cos x, \log(1+x)$ with their Taylor expansions. For example, is e^x bigger than or smaller than $1+x+\dfrac{x^2}{2!}+\dfrac{x^3}{3!}+\cdots+\dfrac{x^n}{n!}$?

2.5.2 High Order Approximation by Substitution

The functions (and their variations) in Examples 2.5.3 through 2.5.11 are the only ones that we can compute all the high order derivative functions. These give the basic examples of high order approximations. We get other high order approximations by combining the basic ones.

Example 2.5.7. Substituting x by $\dfrac{b}{a}x$ in the Taylor expansion of $(1+x)^p$, we get

$$\begin{aligned}
(a+bx)^p &= a^p\left(1+\frac{b}{a}x\right)^p \\
&= a^p\left[1+p\frac{b}{a}x+\frac{p(p-1)}{2!}\frac{b^2}{a^2}x^2+\cdots\right. \\
&\quad \left.+\frac{p(p-1)\cdots(p-n+1)}{n!}\frac{b^n}{a^n}x^n+o\left(\frac{b^n}{a^n}x^n\right)\right] \\
&= a^p+pa^{p-1}bx+\frac{p(p-1)}{2!}a^{p-2}b^2x^2+\cdots \\
&\quad +\frac{p(p-1)\cdots(p-n+1)}{n!}a^{p-n}b^nx^n+o(x^n).
\end{aligned}$$

Note that we used $a^po\left(\dfrac{b^n}{a^n}x^n\right) = o(x^n)$ in the computation. The reason is that $o\left(\dfrac{b^n}{a^n}x^n\right)$ really means a function $R\left(\dfrac{b^n}{a^n}x^n\right)$, where $R(x)$ is the remainder of the

n-th order Taylor expansion of $(1+x)^p$. Since $\lim_{x\to 0}\dfrac{R(x)}{x^n}=0$, we get

$$\lim_{x\to 0}\frac{a^pR\left(\dfrac{b^n}{a^n}x^n\right)}{x^n}=\lim_{y\to 0}\frac{a^pR(y)}{\dfrac{a^n}{b^n}y^n}=a^{p-n}b^n\lim_{y\to 0}\frac{R(y)}{y^n}=0.$$

This means $a^pR\left(\dfrac{b^n}{a^n}x^n\right)=o(x^n)$.

Further substitution of a,b,x by $x_0,1,x-x_0$ gives the Taylor expansion of x^p at x_0

$$\begin{aligned}x^p &= (x_0+(x-x_0))^p\\ &= x_0^p+px_0^{p-1}(x-x_0)+\frac{p(p-1)}{2!}x_0^{p-2}(x-x_0)^2+\cdots\\ &\quad+\frac{p(p-1)\cdots(p-n+1)}{n!}x_0^{p-n}(x-x_0)^n+o((x-x_0)^n).\end{aligned}$$

The Taylor expansion can also be obtained from the high order derivative in Example 2.5.3.

Example 2.5.8. The Taylor expansion of $\dfrac{1}{1-x}$ at 0 in Example 2.5.3 induces the following approximations

$$\begin{aligned}\frac{1}{1+x}&=1-x+x^2-\cdots+(-1)^nx^n+o(x^n),\\ \frac{1}{1+x^2}&=1-x^2+x^4-\cdots+(-1)^nx^{2n}+o(x^{2n}).\end{aligned}$$

Similar to the Taylor expansions of $\sin x$ and $\cos x$, we expect that the odd power terms vanish in the Taylor expansion of $\dfrac{1}{1+x^2}$. Therefore the remainder should be improved to $o(x^{2n+1})$. To get the improved remainder, we consider the $2(n+1)$-th order Taylor expansion of $\dfrac{1}{1+x^2}$

$$\frac{1}{1+x^2}=1-x^2+x^4-\cdots+(-1)^nx^{2n}+(-1)^{n+1}x^{2(n+1)}+o(x^{2(n+1)}).$$

This shows that the remainder of the $2n$-th order Taylor expansion is $(-1)^{n+1}x^{2(n+1)}+R(x)$, where $R(x)$ satisfies $\lim_{x\to 0}\dfrac{R(x)}{x^{2(n+1)}}=0$. By

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

we get

$$(-1)^{n+1}x^{2(n+1)} + o(x^{2(n+1)}) = o(x^{2n+1}),$$

and the improved approximation

$$\frac{1}{1+x^2} = 1 - x^2 + x^4 - \cdots + (-1)^n x^{2n} + o(x^{2n+1}).$$

Finally, it is easy to see that $\dfrac{1}{1+x^2}$ has derivative of any order. From the coefficients in the Taylor expansion, we get

$$\frac{d^n}{dx^n}\bigg|_{x=0}\left(\frac{1}{1+x^2}\right) = n!a_n = \begin{cases} 0, & \text{if } n = 2k-1, \\ (-1)^k(2k)!, & \text{if } n = 2k. \end{cases}$$

It is practically impossible to get this by directly computing the high order derivatives (i.e., by repeatedly taking derivatives).

Exercise 2.5.19. Explain and justify the following claims about remainders.

1. $o(x^5) = o(x^3)$.
2. $o(x^5) + o(x^5) = o(x^5)$.
3. $o(x^4) + o(x^5) = o(x^3)$.
4. $x^3 o(x^5) = o(x^8)$.
5. $o(x^3)o(x^5) = o(x^8)$.
6. $o(x^3) + x^5 = o(x^3)$.

Exercise 2.5.20. Find the Taylor expansion of $\dfrac{1}{1-x^3}$ at 0, and the high order derivatives of the function at 0.

Exercise 2.5.21. Use the high order derivatives in Example 2.5.8 to find the Taylor expansion of $\arctan x$ at 0.

Exercise 2.5.22. Find the Taylor expansion of $\dfrac{1}{\sqrt{1-x^2}}$ at 0. Find the high order derivatives of the function at 0. Then find the Taylor expansion of $\arcsin x$ at 0.

Example 2.5.9. The Taylor expansion of e^x at 0 induces

$$\begin{aligned} e^{-x} &= 1 - \frac{1}{1!}x + \frac{1}{2!}x^2 - \cdots + (-1)^n\frac{1}{n!}x^n + o(x^n), \\ e^x = e^{x_0}e^{x-x_0} &= e^{x_0} - \frac{1}{1!}e^{x_0}(x-x_0) + \frac{1}{2!}e^{x_0}(x-x_0)^2 - \cdots \\ &\quad + (-1)^n\frac{1}{n!}e^{x_0}(x-x_0)^n + o((x-x_0)^n), \\ e^{x^2} &= 1 + \frac{1}{1!}x^2 + \frac{1}{2!}x^4 + \cdots + \frac{1}{n!}x^{2n} + o(x^{2n+1}). \end{aligned}$$

Note that we have the more accurate remainder $o(x^{2n+1})$ for e^{x^2} for the reason similar to Example 2.5.8. Moreover, the Taylor expansion of e^{x^2} gives

$$(e^{x^2})^{(n)}|_{x=0} = \begin{cases} 0, & \text{if } n = 2k-1, \\ \dfrac{(2k)!}{k!}, & \text{if } n = 2k. \end{cases}$$

Example 2.5.10. The high order approximation of x^2e^x at 0 is

$$\begin{aligned} x^2e^x &= x^2\left(1 + \frac{1}{1!}x + \frac{1}{2!}x^2 + \cdots + \frac{1}{n!}x^n + o(x^n)\right) \\ &= x^2 + \frac{1}{1!}x^3 + \frac{1}{2!}x^4 + \cdots + \frac{1}{n!}x^{n+2} + o(x^{n+2}). \end{aligned}$$

Here we use $x^2o(x^n) = o(x^{n+2})$. The n-th order approximation is

$$x^2e^x = x^2 + \frac{1}{1!}x^3 + \frac{1}{2!}x^4 + \cdots + \frac{1}{(n-2)!}x^n + o(x^n).$$

Can you find the n-th order derivative of x^2e^x at 0?

On the other hand, to find the high order approximation of x^2e^x at 1, we express the variables in terms of $x-1$ and get

$$\begin{aligned}
x^2e^x &= ((x-1)+1)^2e^{(x-1)+1} = e((x-1)^2+2(x-1)+1)e^{x-1}\\
&= e((x-1)^2+2(x-1)+1)\left(\sum_{i=0}^{n}\frac{1}{i!}(x-1)^i+o((x-1)^n)\right)\\
&= e\sum_{i=0}^{n}\frac{1}{i!}(x-1)^{i+2}+o((x-1)^{n+2})\\
&\quad + 2e\sum_{i=0}^{n}\frac{1}{i!}(x-1)^{i+1}+o((x-1)^{n+1})\\
&\quad + e\sum_{i=0}^{n}\frac{1}{i!}(x-1)^{i}+o((x-1)^{n})\\
&= e\sum_{i=2}^{n}\frac{1}{(i-2)!}(x-1)^i+o((x-1)^n)\\
&\quad + 2e(x-1)+2e\sum_{i=2}^{n}\frac{1}{(i-1)!}(x-1)^i+o((x-1)^n)\\
&\quad + e+e(x-1)+e\sum_{i=2}^{n}\frac{1}{i!}(x-1)^i+o((x-1)^n)\\
&= e+3e(x-1)+e\sum_{i=2}^{n}\left(\frac{1}{(i-2)!}+\frac{2}{(i-1)!}+\frac{1}{i!}\right)(x-1)^i+o((x-1)^n)\\
&= e+3e(x-1)+e\sum_{i=2}^{n}\frac{i^2+i+1}{i!}(x-1)^i+o((x-1)^n).
\end{aligned}$$

Exercise 2.5.23. Use the basic Taylor expansions to find the high order approximations and derivatives of functions in Exercise 2.5.11.

Exercise 2.5.24. Use the basic Taylor expansions to find the high order approximations and derivatives at 0.

1. $\dfrac{1}{x(x+1)(x+2)}$.
2. $\sqrt{1-x^2}$.
3. $\sqrt{1+x^3}$.
4. $\log(1+x^2)$.
5. $\log(1+3x+2x^2)$.
6. $\log\dfrac{1+x^2}{1-x^3}$.
7. e^{2x}.
8. a^{x^2}.
9. $\sin x\cos 2x$.
10. $\sin x\cos 2x\sin 3x$.
11. $\sin x^2$.
12. $\sin^2 x$.

Exercise 2.5.25. Use the basic Taylor expansions to find high order approximations and high order derivatives.

1. x^3+5x-1 at 1.
2. x^p at -3.
3. $\dfrac{x+3}{x+1}$ at 1.
4. $\sqrt{x+1}$ at 1.
5. e^{-2x} at 4.
6. $\log x$ at 2.
7. $\log(3-x)$ at 2.
8. $\sin x$ at $\dfrac{\pi}{2}$.
9. $\sin x$ at π.
10. $\cos x$ at π.
11. $\sin 2x$ at $\dfrac{\pi}{4}$.
12. $\sin^2 x$ at π.

Exercise 2.5.26. Use the basic Taylor expansions to find high order approximations and high order derivatives at x_0.

1. a^x.
2. x^2e^x.
3. $\log x$.
4. $\sin x$.
5. $\sin x\cos 2x$.
6. $\sin^2 x$.

2.5.3 Combination of High Order Approximations

So far we only used simple substitutions to get new approximations. In the subsequent examples, we compute more sophisticated combinations of approximations.

Example 2.5.11. By the Taylor expansion of $\sin x$, we have

$$\begin{aligned}\sin^2 x-\sin x^2 &= \left(x-\frac{1}{6}x^3+\frac{1}{120}x^5+o(x^6)\right)^2-\left(x^2-\frac{1}{6}x^6+o((x^2)^4)\right)\\ &= \left(x^2-\frac{1}{3}x^4+\frac{1}{36}x^6+\frac{1}{60}x^6+o(x^6)^2+2xo(x^6)+\cdots\right)\\ &\quad -\left(x^2-\frac{1}{6}x^6+o(x^8)\right)\\ &= -\frac{1}{3}x^4+\frac{19}{90}x^6+o(x^7).\end{aligned}$$

The term $o(x^7)$ at the end comes from

$$\lim_{x\to 0}\frac{R(x)}{x^6}=0 \implies \lim_{x\to 0}\frac{R(x)^2}{x^7}=0,\ \lim_{x\to 0}\frac{xR(x)}{x^7}=0,\ \ldots .$$

In particular, we get the limit in Example 2.4.9

$$\lim_{x\to 0}\frac{\sin^2 x-\sin x^2}{x^4}=-\frac{1}{3}.$$

We also get the high order derivatives of $f(x)=\sin^2 x-\sin x^2$ at 0

$$f'(0)=f''(0)=f'''(0)=f^{(5)}(0)=0,\quad f^{(4)}(0)=-8,\quad f^{(6)}(0)=152.$$

Example 2.5.12. We may compute the Taylor expansions of $\tan x$ and $\sec x$ from the Taylor expansions of $\sin x$, $\cos x$ and $\dfrac{1}{1-x}$

$$\begin{aligned}
\sec x = \frac{1}{\cos x} &= \frac{1}{1-\frac{1}{2}x^2+\frac{1}{24}x^4+o(x^5)} \\
&= 1+\left(\frac{1}{2}x^2-\frac{1}{24}x^4+o(x^5)\right)+\left(\frac{1}{2}x^2-\frac{1}{24}x^4+o(x^5)\right)^2 \\
&\quad +\left(\frac{1}{2}x^2-\frac{1}{24}x^4+o(x^5)\right)^3+o(x^6) \\
&= 1+\frac{1}{2}x^2-\frac{1}{24}x^4+\frac{1}{4}x^4+o(x^5) \\
&= 1+\frac{1}{2}x^2+\frac{5}{24}x^4+o(x^5), \\
\tan x = \sin x\sec x &= \left(x-\frac{1}{6}x^3+\frac{1}{120}x^5+o(x^6)\right)\left(1+\frac{1}{2}x^2+\frac{5}{24}x^4+o(x^5)\right) \\
&= x-\frac{1}{6}x^3+\frac{1}{120}x^5+\frac{1}{2}x^3-\frac{1}{12}x^5+\frac{5}{24}x^5+o(x^6) \\
&= x+\frac{1}{3}x^3+\frac{2}{15}x^5+o(x^6).
\end{aligned}$$

The expansions give $(\sec x)^{(4)}_{x=0} = 5$ and $(\tan x)^{(5)}_{x=0} = 16$.

Example 2.5.13. We computed $\lim_{x\to 0}\left(\dfrac{1}{x}-\dfrac{1}{e^x-1}\right)$ by using l'Hospital's rule in Example 2.4.11. Alternatively,, we use the Taylor expansions of e^x and $\dfrac{1}{1-x}$

$$\begin{aligned}
\frac{1}{x}-\frac{1}{e^x-1} = \frac{1}{x}-\frac{1}{x+\frac{x^2}{2}+o(x^2)} &= \frac{1}{x}\left(1-\frac{1}{1+\frac{x}{2}+o(x)}\right) \\
&= \frac{1}{x}\left(1-1+\left(\frac{x}{2}+o(x)\right)+o\left(\frac{x}{2}+o(x)\right)\right) = \frac{1}{2}+\frac{o(x)}{x}.
\end{aligned}$$

This implies that the limit is $\dfrac{1}{2}$.

Example 2.5.14. We know $\lim_{x\to\infty}\left(1+\dfrac{1}{x}\right)^x = e$ from Example 1.6.17. The next question is what the difference $\left(1+\dfrac{1}{x}\right)^x - e$ looks like. As x goes to infinity, does the difference approach 0 like $\dfrac{1}{x}$?

The question is the same as the behavior of $(1+x)^{\frac{1}{x}} - e$ near 0. By the Taylor expansions of $\log(1+x)$ and e^x at 0, we get

$$\begin{aligned}
(1+x)^{\frac{1}{x}} - e &= e^{\frac{1}{x}\log(1+x)} - e \\
&= e^{\frac{1}{x}\left(x-\frac{x^2}{2}+o(x^2)\right)} - e \\
&= e\left(e^{-\frac{x}{2}+o(x)} - 1\right) \\
&= e\left[\left(-\frac{x}{2}+o(x)\right) + o\left(-\frac{x}{2}+o(x)\right)\right] \\
&= -\frac{e}{2}x + o(x).
\end{aligned}$$

Translated back into x approaching infinity, we have

$$\left(1+\frac{1}{x}\right)^x - e = -\frac{e}{2x} + o\left(\frac{1}{x}\right).$$

Exercise 2.5.27. Find the 5-th order approximations at 0.

1. $e^x \sin x$.
2. $\sqrt{x+1}e^{x^2}\tan x$.
3. $(1+x)^x$.
4. $\log\dfrac{\sin x}{x}$.
5. $\log\cos x$.
6. $\dfrac{x}{e^x-1}$.

Exercise 2.5.28. Use approximations to compute limits.

1. $\lim_{x\to 0}\dfrac{1-\cos x^2}{x^3\sin x}$.
2. $\lim_{x\to 0}\dfrac{\sin x - \tan x}{x^3}$.
3. $\lim_{x\to 0}\left(\dfrac{1}{\sin^2 x} - \dfrac{1}{x^2}\right)$.
4. $\lim_{x\to 0}\left(\dfrac{1}{x^2} - \dfrac{1}{\tan^2 x}\right)$.
5. $\lim_{x\to 1}\left(\dfrac{x}{x-1} - \dfrac{1}{\log x}\right)$.
6. $\lim_{x\to 0}(\cos x + \sin x)^{\frac{1}{x(x+1)}}$.
7. $\lim_{x\to 0}\dfrac{e^x - e^{\sin x}}{x^3}$.
8. $\lim_{x\to 0}\log\dfrac{\cos ax}{\cos bx}$.
9. $\lim_{x\to 0}\dfrac{x-\tan x}{x-\sin x}$.
10. $\lim_{x\to\infty} x^2\left(e - \dfrac{1}{x} - \left(1+\dfrac{1}{x}\right)^x\right)$.
11. $\lim_{x\to\infty} x^2\log\left(x\sin\dfrac{1}{x}\right)$.
12. $\lim_{x\to 1}\dfrac{(x-1)\log x}{1+\cos\pi x}$.

Exercise 2.5.29. Use whatever method you prefer to compute limits, $p, q > 0$.

1. $\lim_{x\to 0^+} x^p e^{-x^q}$.
2. $\lim_{x\to+\infty} x^p e^{-x^q}$.
3. $\lim_{x\to+\infty} x^p \log x$.
4. $\lim_{x\to 1^+}(x-1)^p\log x$.
5. $\lim_{x\to 1^+}(x-1)^p(\log x)^q$.
6. $\lim_{x\to+\infty} x^p e^{-x^q}\log x$.

7. $\lim_{x\to+\infty} x^p \log(\log x)$.
8. $\lim_{x\to+\infty} (\log x)^p (\log(\log x))^q$.
9. $\lim_{x\to e^+} (x-e)^p \log(\log x)$.

Exercise 2.5.30. Use whatever method you prefer to compute limits, $p, q > 0$.

1. $\lim_{x\to 0^+} \dfrac{\tan^p x - x^p}{\sin^p x - x^p}$.
2. $\lim_{x\to 0^+} \dfrac{\sin^p x - \tan^p x}{x^q}$.
3. $\lim_{x\to \frac{\pi}{4}} \dfrac{\tan x - \cot x}{4x - \pi}$.
4. $\lim_{x\to 0} \dfrac{a\tan bx - b\tan ax}{a\sin bx - b\sin ax}$.

Exercise 2.5.31. Use whatever method you prefer to compute limits.

1. $\lim_{x\to\infty} x^3 \left(\sin\dfrac{1}{x} - \dfrac{1}{2}\sin\dfrac{2}{x}\right)$.
2. $\lim_{x\to 0} \dfrac{1}{x}\left(\dfrac{2}{x(2+x)} - \dfrac{1}{e^x - 1}\right)$.
3. $\lim_{x\to 0} \left(\dfrac{1}{\log(1+x)} - \dfrac{1}{x}\right)$.
4. $\lim_{x\to 0} \left(\dfrac{1}{\log(x+\sqrt{1+x^2})} - \dfrac{1}{\log(1+x)}\right)$.

Exercise 2.5.32. Use whatever method you prefer to compute limits.

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

Exercise 2.5.33. Use whatever method you prefer to compute limits.

1. $\lim_{x\to 0} \dfrac{(1+ax)^b - (1+bx)^a}{x^2}$.
2. $\lim_{x\to 0} \dfrac{(1+ax+cx^2)^b - (1+bx+dx^2)^a}{x^2}$.
3. $\lim_{x\to a} \dfrac{a^x - x^a}{x-a}$.
4. $\lim_{x\to 0} \dfrac{(a+x)^x - a^x}{x^2}$.
5. $\lim_{x\to 0^+} \dfrac{a^x - a^{\sin x}}{x^3}$, $a > 0$.
6. $\lim_{x\to 0^+} \dfrac{\log(\sin ax)}{\log(\sin bx)}$, $a, b > 0$.
7. $\lim_{x\to 0} \dfrac{\log(\cos ax)}{\log(\cos bx)}$.
8. $\lim_{x\to 0} \dfrac{\cos(\sin x) - \cos x}{x^4}$.
9. $\lim_{x\to 0} \dfrac{\arcsin 2x - 2\arcsin x}{x^3}$.
10. $\lim_{x\to 0} \dfrac{1-\cos x^2}{x^2\sin x^2}$.
11. $\lim_{x\to \frac{\pi}{6}} \dfrac{1-2\sin x}{\cos 3x}$.
12. $\lim_{x\to 1^+} \log x \tan\dfrac{\pi x}{2}$.
13. $\lim_{x\to 0^+} x^{x^x - 1}$.
14. $\lim_{x\to 0} x^{\sin x}$.
15. $\lim_{x\to 0^+} (-\log x)^x$.

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

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

18. $\lim_{x\to 0}\left(\frac{2}{\pi}\arccos x\right)^{\frac{1}{x}}$.

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

Exercise 2.5.34. In Example 2.4.2, we applied the Mean Value Theorem to get $\log(1+x) = \frac{x}{1+\theta x}$ for some $0<\theta<1$.

1. Find explicit formula for $\theta=\theta(x)$.
2. Compute $\lim_{x\to 0}\theta$ by using l'Hospital's rule.
3. Compute $\lim_{x\to 0}\theta$ by using high order approximation.

You may try the same for other functions such as $e^x-1=e^{\theta x}$. What can you say about $\lim_{x\to 0}\theta$ in general?

Exercise 2.5.35. Show that the limits converge but cannot be computed by L'Hospital's rule.

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

Exercise 2.5.36. Find a,b so that the following hold.

1. $x-(a+b\cos x)\sin x = o(x^4)$.
2. $x-a\sin x-b\tan x = o(x^4)$.

2.5.4 Implicit High Order Differentiation

Example 2.5.15. In Example 2.2.9, the function $y=y(x)$ implicitly given by the unit circle $x^2+y^2=1$ has derivative $y'=-\frac{x}{y}$. Then

$$y''=-\frac{y-xy'}{y^2}=-\frac{y+x\frac{x}{y}}{y^2}=-\frac{x^2+y^2}{y^3}=-\frac{1}{y^3}.$$

You may verify the result by directly computing the second order derivative of $y=\pm\sqrt{1-x^2}$.

Example 2.5.16. In Example 2.2.10, we computed the derivative of the function $y=y(x)$ implicitly given by the equation $2y-2x^2-\sin y+1=0$ and then obtained the linear approximation at $P=\left(\sqrt{\frac{\pi}{2}},\frac{\pi}{2}\right)$. We can certainly continue finding the formula for the second order derivative of $y(x)$ and then get the quadratic approximation at P.

Alternatively, we may compute the quadratic approximation at P by postulating the approximation to be

$$y = \frac{\pi}{2} + a_1\Delta x + a_2\Delta x^2 + o(\Delta x^2), \quad \Delta x = x - \sqrt{\frac{\pi}{2}}.$$

Substituting into the equation, we get

$$\begin{aligned}
0 &= 2\left(\frac{\pi}{2} + a_1\Delta x + a_2\Delta x^2 + o(\Delta x^2)\right) - 2\left(\sqrt{\frac{\pi}{2}} + \Delta x\right)^2 \\
&\quad - \sin\left(\frac{\pi}{2} + a_1\Delta x + a_2\Delta x^2 + o(\Delta x^2)\right) + 1 \\
&= 2a_1\Delta x + 2a_2\Delta x^2 - 4\sqrt{\frac{\pi}{2}}\Delta x - 2\Delta x^2 + o(\Delta x^2) - \cos(a_1\Delta x + a_2\Delta x^2 + o(\Delta x^2)) + 1 \\
&= 2a_1\Delta x + 2a_2\Delta x^2 - 2\sqrt{2\pi}\Delta x - 2\Delta x^2 + \frac{1}{2}(a_1\Delta x + a_2\Delta x^2)^2 + o(\Delta x^2)
\end{aligned}$$

The coefficients of Δx and Δx^2 on the right must vanish. Therefore

$$2a_1 - 2\sqrt{2\pi} = 0, \quad 2a_2 - 2 + \frac{1}{2}a_1^2 = 0.$$

The solution is $a_1 = \sqrt{2\pi}$, $a_2 = \dfrac{2-\pi}{2}$, and the quadratic approximation is

$$y(x) = \frac{\pi}{2} + \sqrt{2\pi}\Delta x + \frac{2-\pi}{2}\Delta x^2 + o(\Delta x^2), \quad \Delta x = x - \sqrt{\frac{\pi}{2}}.$$

Exercise 2.5.37. Compute quadratic approximations of implicitly defined functions.

1. $y^2 + 3y^3 + 1 = x$.
2. $\sin y = x$.
3. $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$.
4. $\sqrt{x} + \sqrt{y} = \sqrt{a}$.
5. $e^{x+y} = xy$.
6. $x^2 + 2xy - y^2 - 2x = 0$.

Exercise 2.5.38. Compute quadratic approximations of functions $y = y(x)$ given by the curves.

1. $x = \sin^2 t$, $y = \cos^2 t$.
2. $x = a(t - \sin t)$, $y = a(1 - \cos t)$.
3. $x = e^t \cos 2t$, $y = e^t \sin 2t$.
4. $x = (1 + \cos t)\cos t$, $y = (1 + \cos t)\sin t$.

2.5.5 Two Theoretical Examples

Theorem 2.5.2 tells us that the existence of high order derivative implies the high order differentiability. The following example shows that the converse is not true.

Example 2.5.17. The function $x^3D(x)$ satisfies

$$|x| < \delta = \epsilon \implies |x^3D(x) - 0| \le |x|^3 \le \epsilon |x|^2.$$

Therefore the function is second order differentiable, with $0 = 0 + 0x + 0x^2$ as the quadratic approximation.

On the other hand, we have $(x^3D(x))'|_{x=0} = 0$ and $x^3D(x)$ is not differentiable (because not even continuous) away from 0. Therefore the second order derivative is not defined.

Exercise 2.5.39. Show that for any n, there is a function that is n-th order differentiable at 0 but has no second order derivative at 0.

Exercise 2.5.40. The lack of high order derivatives for the function in Example 2.5.17 is due to discontinuity away from 0. Can you find a function with the following properties?

1. f that has first order derivative everywhere on $(-1, 1)$.
2. f has no second order derivative at 0.
3. f is second order differentiable at 0.

The next example deals with the following intuition from everyday life. Suppose we try to measure a length by more and more refined rulers. If our readings from meter ruler, centimeter ruler, millimeter ruler, micrometer ruler, etc, are all 0, then the real length should be 0. Similarly, the Taylor expansion of a function at 0 is the measurement by "x^n-ruler". The following example shows that, even if the readings by all the "x^n-ruler" are 0, the function does not have to be 0.

Example 2.5.18. The function

$$f(x) = \begin{cases} e^{-\frac{1}{|x|}}, & \text{if } x \ne 0, \\ 0, & \text{if } x = 0, \end{cases}$$

has derivative

$$f'(x) = \begin{cases} \dfrac{1}{x^2} e^{-\frac{1}{|x|}}, & \text{if } x > 0, \\ -\dfrac{1}{x^2} e^{-\frac{1}{|x|}}, & \text{if } x < 0, \\ 0, & \text{if } x = 0. \end{cases}$$

The derivative at $x \neq 0$ is computed by the usual chain rule, and the derivative at 0 is computed directly

$$f'(0) = \lim_{x\to 0} \frac{1}{x} e^{-\frac{1}{|x|}} = \lim_{y\to\infty} \frac{y}{e^{|y|}} = 0.$$

In general, it can be inductively proved that

$$f^{(n)}(x) = \begin{cases} p\left(\dfrac{1}{x}\right) e^{-\frac{1}{x}}, & \text{if } x > 0, \\ (-1)^n p\left(-\dfrac{1}{x}\right) e^{-\frac{1}{x}}, & \text{if } x < 0, \\ 0, & \text{if } x = 0, \end{cases}$$

where $p(t)$ is a polynomial depending only on n.

The function has the special property that the derivative of any order vanishes at 0. Therefore the function is differentiable of any order, and all the high order approximations are 0

$$f(x) = 0 + 0x + 0x^2 + \cdots + 0x^n + o(x^n).$$

However, the function is not 0.

The example can be understood in two ways. The first is that, for some functions, even more refined ruler is needed in order to measure "beyond all orders". The second is that the function above is not "measurable by polynomials". The functions that are measurable by polynomials are call *analytic*, and the function above is not analytic.

Exercise 2.5.41. Directly show (i.e., without calculating the high order derivatives) that the function in Example 2.5.18 is differentiable of any order, with 0 as the approximation of any order.

2.6 Application of High Order Approximation

2.6.1 High Derivative Test

Theorem 2.3.4 gives the first order derivative condition for a (two sided) differentiable function to have local extreme. As the subsequent examples show, the theorem only provides candidates for the local extrema. To find out whether such candidates are indeed local extrema, high order approximations are needed.

Let us consider the first non-trivial high order approximation at x_0

$$f(x) = f(x_0) + c(x - x_0)^n + o((x - x_0)^n) = f(x_0) + c(x - x_0)^n(1 + o(1)), \quad c \neq 0.$$

When x is close to x_0, we have $1 + o(1) > 0$ and therefore $f(x) > f(x_0)$ when $c(x - x_0)^n > 0$ and $f(x) < f(x_0)$ when $c(x - x_0)^n < 0$. Specifically, we have the following signs of $c(x - x_0)^n$ for various cases.

- If n is odd and $c > 0$, then $c(x - x_0)^n > 0$ for $x > x_0$ and $c(x - x_0)^n < 0$ for $x < x_0$.
- If n is odd and $c < 0$, then $c(x - x_0)^n < 0$ for $x > x_0$ and $c(x - x_0)^n > 0$ for $x < x_0$.
- If n is even and $c > 0$, then $c(x - x_0)^n > 0$ for $x \neq x_0$.
- If n is even and $c < 0$, then $c(x - x_0)^n < 0$ for $x \neq x_0$.

The sign of $c(x-x_0)^n$ then further determines whether $f(x) < f(x_0)$ or $f(x) > f(x_0)$, and we get the following result.

Theorem 2.6.1. *Suppose $f(x)$ has high order approximation $f(x_0) + c(x - x_0)^n$ at x_0.*

1. *If n is odd and $c \neq 0$, then x_0 is not a local extreme.*
2. *If n is even and $c > 0$, then x_0 is a local minimum.*
3. *If n is even and $c < 0$, then x_0 is a local maximum.*

If f has n-th order derivative at x_0, the condition of the theorem means

$$f'(x_0) = f''(x_0) = \cdots = f^{(n-1)}(x_0) = 0, \quad f^{(n)}(x_0) = n!c \neq 0.$$

The special case $n = 1$ is Theorem 2.3.4. For the special case $n = 2$, the theorem gives the *second derivative test*: Suppose $f'(x_0) = 0$ (i.e., the criterion in Theorem 2.3.4 is satisfied), and f has second order derivative at x_0.

1. If $f''(x_0) > 0$, then x_0 is a local minimum.
2. If $f''(x_0) < 0$, then x_0 is a local maximum.

Example 2.6.1. In Example 2.3.13, we found the candidates ± 1 for the local extrema of $f(x) = x^3 - 3x + 1$. The second order derivative $f''(x) = 6x$ at the two candidates are

$$f''(1) = 6 > 0, \quad f''(-1) = -6 < 0.$$

Therefore 1 is a local minimum and -1 is a local maximum.

Example 2.6.2. Consider the function $y = y(x)$ implicitly defined in Example 2.2.10. By $y'(x) = \dfrac{4x}{2-\cos y}$, we find a candidate $x = 0$ for the local extreme of $y(x)$. Then we have

$$y''(x) = \frac{4}{2-\cos y} + 4x\,\frac{d}{dy}\left(\frac{1}{2-\cos y}\right)\bigg|_{y=y(x)} y'.$$

At the candidate $x = 0$, we already have $y'(0) = 0$. Therefore $y''(0) = \dfrac{4}{2-\cos y(0)} > 0$. This shows that $x = 0$ is a local minimum of the implicitly defined function.

Example 2.6.3. The function $f(x) = x^2 - x^3 D(x)$ has no second order derivative at 0, but still has the quadratic approximation $f(x) = x^2 + o(x^2)$. The quadratic approximation tells us that 0 is a local minimum of $f(x)$.

Example 2.6.4. Let $f(x) = \dfrac{\sin x}{6x - x^3}$ for $x \neq 0$ and $f(0) = \dfrac{1}{6}$. Then for $x \neq 0$ close to 0, we have

$$\begin{aligned} f(x) &= \frac{1}{6x-x^3}\left(x - \frac{x^3}{6} + \frac{x^5}{120} + o(x^6)\right) = \frac{1}{6-x^2}\left(1 - \frac{x^2}{6} + \frac{x^4}{120} + o(x^5)\right) \\ &= \frac{1}{6}\left(1 + \frac{x^2}{6} + \frac{x^4}{36} + o(x^5)\right)\left(1 - \frac{x^2}{6} + \frac{x^4}{120} + o(x^5)\right) = \frac{1}{6} + \frac{x^4}{120} + o(x^5). \end{aligned}$$

We note that by $f(0) = \dfrac{1}{6}$, the 4-th order approximation also holds for $x = 0$. Then by Theorem 2.6.1, we find that $x = 0$ is a local minimum.

Alternatively, we may directly use the idea leading to Theorem 2.6.1. For $x \neq 0$ close to 0, we have

$$f(x) = \frac{1}{6x-x^3}\left(x - \frac{x^3}{6} + \frac{x^5}{120} + o(x^6)\right) = \frac{1}{6} + \frac{x^4(1+o(x))}{(6-x^2)\cdot 120}.$$

For small $x \neq 0$, we have $\dfrac{x^4(1+o(x))}{(6-x^2)\cdot 120} > 0$, which further implies $f(x) > \dfrac{1}{6} = f(0)$. Therefore 0 is a (strict) local minimum.

Exercise 2.6.1. Find the local extrema by using quadratic approximations.

1. $x^3 - 3x + 1$ on $\mathbb{R}$.
2. xe^{-x} on $\mathbb{R}$.
3. $x \log x$ on $(0, +\infty)$.
4. $(x^2+1)e^x$ on $\mathbb{R}$.

Exercise 2.6.2. Find the local extrema for the function $y = y(x)$ implicitly given by $x^3 + y^3 = 6xy$.

Exercise 2.6.3. For $p > 1$, determine whether 0 is a local extreme for the function

$$\begin{cases} x^2 + |x|^p \sin x \sin \dfrac{1}{x}, & \text{if } x \neq 0, \\ 0, & \text{if } x = 0. \end{cases}$$

Exercise 2.6.4. Determine whether 0 is a local extreme.

1. $x^3 + x^4$.
2. $\sin x - x \cos x$.
3. $\tan x - \sin x$.
4. $\left(1 - x + \dfrac{1}{2!}x^2\right) e^x$.
5. $\left(1 - x + \dfrac{1}{2!}x^2 - \dfrac{1}{3!}x^3\right) e^x$.

Exercise 2.6.5. Let $f(0) = 1$ and let $f(x)$ be given by the following for $x \neq 0$. Determine whether 0 is a local extreme.

1. $\dfrac{\sin x}{x + ax^2}$.
2. $\dfrac{\sin x}{x + bx^3}$.
3. $\dfrac{\sin x}{x + ax^2 + bx^3}$.

2.6.2 Convex Function

A function f is *convex* on an interval if for any x, y in the interval, the straight line $L_{x,y}$ connecting points $(x, f(x))$ and $(y, f(y))$ lies above the part of the graph of f over $[x, y]$. This means

$$L_{x,y}(z) = f(x) + \frac{f(y) - f(x)}{y - x}(z - x) \geq f(z) \text{ for any } x \leq z \leq y.$$

The function is *concave* if $L_{x,y}$ lies below the graph of f, which means changing $\geq f(z)$ above to $\leq f(z)$. A function f is convex if and only if $-f$ is concave.

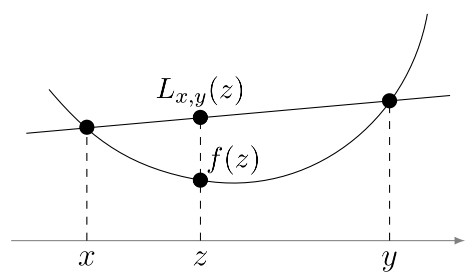


Figure 2.6.1: Convex function.

By the geometric intuition illustrated in Figure 2.6.2, the following are equivalent convexity conditions, for any $x \leq z \leq y$,

1. $L_{x,y}(z) \geq f(z)$.
2. $\text{slope}(L_{x,z}) \leq \text{slope}(L_{x,y})$.

3. $\text{slope}(L_{z,y}) \geq \text{slope}(L_{x,y})$.

4. $\text{slope}(L_{x,z}) \leq \text{slope}(L_{z,y})$.

Algebraically, it is not difficult to verify the equivalence.

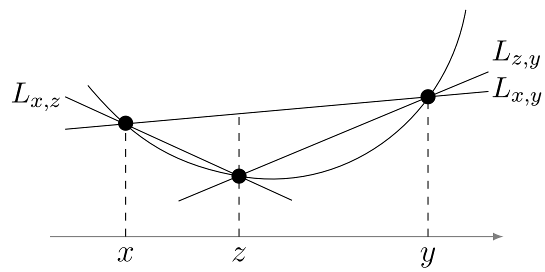


Figure 2.6.2: Convexity by comparing slopes.

If a convex function f is differentiable, then we may take $y \to x^+$ in the inequality $\text{slope}(L_{x,z}) \leq \text{slope}(L_{x,y})$, and get $f'(x) \leq \text{slope}(L_{x,y})$. Similarly, we may take $z \to y^-$ in the inequality $\text{slope}(L_{z,y}) \geq \text{slope}(L_{x,y})$, and get $f'(y) \geq \text{slope}(L_{x,y})$. Combining the two inequalities, we get

$$x < y \implies f'(x) \leq f'(y).$$

It turns out that the converse is also true.

Theorem 2.6.2. *A differentiable function f on an interval is convex if and only if f' is increasing. If f has second order derivative, then this is equivalent to $f'' \geq 0$.*

Similarly, a function f is concave if and only if f' is decreasing, and is also equivalent to $f'' \leq 0$ in case f'' exists.

The converse of Theorem 2.6.2 is explained by Figure 2.6.2. If f' is increasing, then by the Mean Value Theorem (for the two equalities), we have

$$\text{slope}(L_{x,z}) = f'(c) \leq f'(d) = \text{slope}(L_{z,y}).$$

A major application of the convexity is another interpretation of the convexity. In the setup above, a number z between x and y means $z = \lambda x + (1 - \lambda)y$ for some $\lambda \in [0, 1]$. Then the linear function $L_{x,y}(z) = az + b$ preserves the linear relation

$$\begin{aligned}
L_{x,y}(z) &= a(\lambda x + (1 - \lambda)y) + b \\
&= \lambda(ax + b) + (1 - \lambda)(ay + b) \\
&= \lambda L_{x,y}(x) + (1 - \lambda)L_{x,y}(y) \\
&= \lambda f(x) + (1 - \lambda)f(y).
\end{aligned}$$

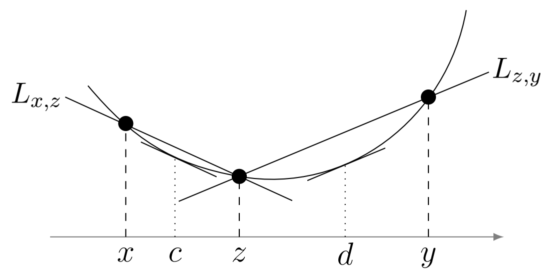


Figure 2.6.3: Increasing $f'(x)$ implies convexity.

Therefore the convexity means

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda)f(y) \text{ for any } 0 \le \lambda \le 1.$$

The following generalizes the inequality.

Theorem 2.6.3 (Jensen's inequality). *If f is convex, and*

$$\lambda_1 + \lambda_2 + \cdots + \lambda_n = 1, \quad 0 \le \lambda_1, \lambda_2, \ldots, \lambda_n \le 1,$$

then

$$f(\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_n x_n) \le \lambda_1 f(x_1) + \lambda_2 f(x_2) + \cdots + \lambda_n f(x_n).$$

By reversing the direction of inequality, we also get Jensen's inequality for concave functions.

In all the discussions about convexity, we may also consider the *strict* inequalities. So we have a concept of strict convexity, and a differentiable function is strictly convex on an interval if and only if its derivative is strictly increasing. Jensen's inequality can also be extended to the strict case.

Example 2.6.5. By $(x^p)'' = p(p-1)x^{p-2}$, we know x^p is convex on $(0, +\infty)$ when $p \ge 1$ or $p < 0$, and is concave when $0 < p \le 1$.

For $p \ge 1$, Jensen's inequality means that

$$(\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_n x_n)^p \le \lambda_1 x_1^p + \lambda_2 x_2^p + \cdots + \lambda_n x_n^p.$$

In particular, if all $\lambda_i = \dfrac{1}{n}$, then we get

$$\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right)^p \le \frac{x_1^p + x_2^p + \cdots + x_n^p}{n}, \text{ for } p \ge 1, x_i \ge 0.$$

This means that the p-th power of the average is smaller than the average of the p-th power.

We note that $(x^p)'' > 0$ for $p > 1$. Therefore all the inequalities are strict, provided some $x_i > 0$ and $0 < \lambda_i < 1$.

By replacing p with $\dfrac{p}{q}$ and replacing x_i with x_i^q, we get

$$\left(\frac{x_1^q + x_2^q + \cdots + x_n^q}{n}\right)^{\frac{1}{q}} \le \left(\frac{x_1^p + x_2^p + \cdots + x_n^p}{n}\right)^{\frac{1}{p}}, \text{ for } p > q > 0.$$

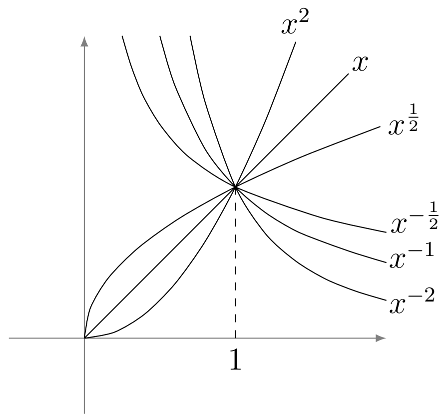


Figure 2.6.4: x^p for various p.

Example 2.6.6. By $(\log x)'' = -\dfrac{1}{x^2} < 0$, the logarithmic function is concave. Then Jensen's inequality tells us that

$$\log(\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_n x_n) \ge \lambda_1 \log x_1 + \lambda_2 \log x_2 + \cdots + \lambda_n \log x_n.$$

This is the same as

$$\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_n x_n \ge x_1^{\lambda_1} x_2^{\lambda_2} \cdots x_n^{\lambda_n}.$$

For the special case $\lambda_i = \dfrac{1}{n}$, we get

$$\frac{x_1 + x_2 + \cdots + x_n}{n} \ge \sqrt[n]{x_1 x_2 \cdots x_n}.$$

In other words, the arithmetic mean is bigger than the geometric mean.

Exercise 2.6.6. Study the convexity.

1. x^3+ax+b.
2. $x+x^{\frac{5}{3}}$.
3. $x+x^p$.
4. $(x^2+1)e^x$.
5. e^{-x^2}.
6. $\sqrt{x^2+1}$.
7. $\dfrac{1}{x^2+1}$.
8. $\dfrac{x^2-1}{x^2+1}$.
9. $\log(x^2+1)$.
10. x^x.
11. $x+2\sin x$.
12. $x\sin(\log x)$.

Exercise 2.6.7. Find the condition on A and B so that the function

$$f(x)=\begin{cases}Ax^p, & \text{if } x\ge 0,\\ B(-x)^q, & \text{if } x<0,\end{cases}$$

is convex. Is the function necessarily differentiable at 0?

Exercise 2.6.8. Let $p,q>0$ satisfy $\dfrac{1}{p}+\dfrac{1}{q}=1$. Use the concavity of $\log x$ to prove Young's inequality

$$\frac{1}{p}x^p+\frac{1}{q}y^q\ge xy.$$

Exercise 2.6.9. For the case $\lambda_1\ne 1$, Find suitable $\mu_2,\dots,\mu_n$ satisfying

$$\lambda_1x_1+\lambda_2x_2+\cdots+\lambda_nx_n=\lambda_1x_1+(1-\lambda_1)(\mu_2x_2+\cdots+\mu_nx_n).$$

Then prove Jensen's inequality by induction.

Exercise 2.6.10. Use the concavity of $\log x$ to prove that, for $x_i>0$, we have

$$\frac{x_1\log x_1+x_2\log x_2+\cdots+x_n\log x_n}{x_1+x_2+\cdots+x_n}\le\log\frac{x_1^2+x_2^2+\cdots+x_n^2}{x_1+x_2+\cdots+x_n}\le\log(x_1+x_2+\cdots+x_n).$$

Exercise 2.6.11. Use Exercise 2.6.10 to show that $f(p)=\log(x_1^p+x_2^p+\cdots+x_n^p)^{\frac{1}{p}}$ is decreasing. Then explain

$$(x_1^q+x_2^q+\cdots+x_n^q)^{\frac{1}{q}}\ge(x_1^p+x_2^p+\cdots+x_n^p)^{\frac{1}{p}},\ \text{for } p>q>0.$$

Note that the similar inequality in Example 2.6.5 is in reverse direction.

Exercise 2.6.12. Verify the convexity of $x\log x$ and then use the property to prove the inequality $(x+y)^{x+y}\le(2x)^x(2y)^y$. Can you extend the inequality to more variables?

2.6.3 Sketch of Graph

We have learned the increasing and decreasing properties, and the convex and concave properties. We may also pay attention to the symmetry properties such as even or odd function, and the periodic property. Moreover, we should pay attention to the following special points.

1. *Intercepts*, where the graph of function intersects the axes.

2. Disruptions, where the functions are not continuous, or not differentiable.

3. Local extrema, which is often (but not restricted to) the places where the function changes between increasing and decreasing.

4. *Points of inflection*, which is the place where the function changes between convex and concave.

5. Infinity, including the finite places where the function tends to infinity, and the behavior of the function at the infinity.

One infinity behavior is the *asymptotes* of a function. If a linear function $a+bx$ satisfies

$$\lim_{x\to+\infty}(f(x)-a-bx)=0,$$

then the linear function is an asymptote at $+\infty$. If $b=0$, then the line is a *horizontal asymptote*. We also have similar asymptote at $-\infty$ (perhaps with different a and b). Moreover, if $\lim_{x\to x_0} f(x)=\infty$ at a finite x_0, then the line $x=x_0$ is a *vertical asymptote*.

In subsequent examples, we sketch the graphs of functions and try to indicate the characteristics listed above as much as possible.

Example 2.6.7. In Example 2.3.4, we determined the monotone property and the local extrema of $f(x)=x^3-3x+1$. The second order derivative $f''(x)=6x$ also tells us that $f(x)$ is concave on $(-\infty,0]$ and convex on $[0,+\infty)$, which makes 0 into a point of inflection. Moreover, the function has no asymptotes. The function is also symmetric with respect to the point $(0,1)$ ($f(x)-1$ is an odd function). Based on these information, we may sketch the graph.

x	$(-\infty,-1)$	-1	$(-1,0)$	0	$(0,1)$	1	$(1,+\infty)$
$f=x^3-3x+1$	$-\infty\leftarrow$	3		0		-1	$\rightarrow+\infty$
$f'=3(x+1)(x-1)$	$+$	0		$-$		0	$+$
	$\nearrow$	max		$\searrow$		min	$\nearrow$
$f''=6x$		$+$		0		$-$	
		$\smile$		infl		$\frown$	

Example 2.6.8. In Example 2.3.6, we determined the monotone property of $f(x)=\sqrt[3]{x^2}(x+1)$. The second order derivative $f''(x)=\dfrac{2(5x-1)}{9\sqrt[3]{x^4}}$ implies that the function is concave on $(-\infty,0)$ and on $\left(0,\dfrac{1}{5}\right]$, convex on $\left[\dfrac{1}{5},+\infty\right)$, with $\dfrac{1}{5}$ as a point of inflection. Moreover, we have $\lim_{x\to-\infty} f(x)=-\infty$, $\lim_{x\to+\infty} f(x)=+\infty$. The

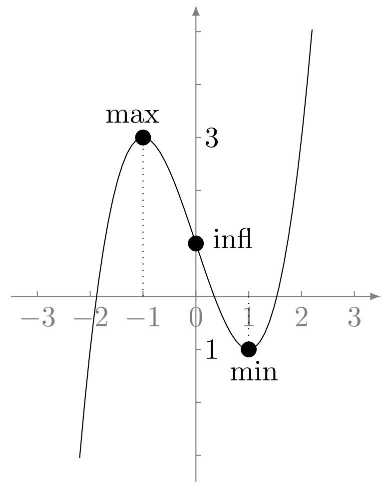


Figure 2.6.5: Graph of $x^3 - 3x + 1$.

function has no asymptote and no symmetry. We also note that $\lim_{x\to 0} \dfrac{f(x)}{x} = \lim_{x\to 0} \dfrac{x+1}{\sqrt[3]{x}} = \infty$. Therefore the tangent of $f(x)$ at 0 is vertical.

x	$(-\infty, -\frac{2}{5})$	$-\frac{2}{5}$	$(-\frac{2}{5}, 0)$	0	$(0, \frac{1}{5})$	$\frac{1}{5}$	$(\frac{1}{5}, +\infty)$
f	$-\infty \leftarrow$	0.1518		0		0.1073	$\rightarrow +\infty$
f'	$+$	0	$-$	no		$+$	
	$\nearrow$	max	$\searrow$	min		$\nearrow$	
f''		$-$		no	$-$	0	$+$
		⌢			⌢	infl	⌣

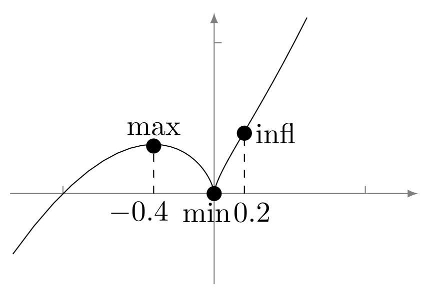


Figure 2.6.6: Graph of $\sqrt[3]{x^2}(x+1)$.

Example 2.6.9. In Example 2.3.15, we determined the monotone property of $f(x) = xe^{-x}$. From $f''(x) = (x-2)e^{-x}$, we also know the function is concave on $(-\infty, 2]$ and convex on $[2, +\infty)$, with $f(2) = e^{-2}$ as a point of inflection. Moreover, we have $\lim_{x\to-\infty} f(x) = -\infty$, $\lim_{x\to+\infty} f(x) = 0$, so that the x-axis is a horizontal asymptote. The function has no symmetry.

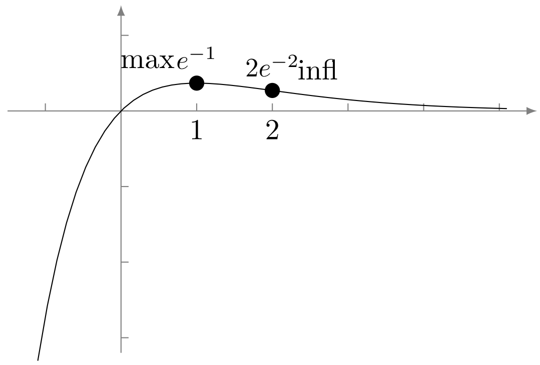


Figure 2.6.7: Graph of xe^{-x}.

Example 2.6.10. In Example 2.3.7, we determined the monotone property and the local extrema of $f(x) = \dfrac{x^3}{x^2-1}$. The function is not defined at ± 1, and has limits

$$\lim_{x\to 1^\pm} f(x) = \pm\infty, \quad \lim_{x\to -1^\pm} f(x) = \mp\infty.$$

These give vertical asymptotes at ± 1. Moreover, we have

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

Therefore $y = x$ is a slant asymptote at ∞.

The second order derivative $f''(x) = \dfrac{2x(x^2+3)}{(x^2-1)^3}$ shows that the function is convex on $(-1, 0)$, $(1, +\infty)$, and is concave at the other two intervals. Therefore 0 is a point of inflection.

We also know the function is odd. So the graph is symmetric with respect to the origin.

Exercise 2.6.13. Use the graph of $x^3 - 3x + 1$ in Example 2.6.7 to sketch the graphs.

1. $-x^3 + 3x^2 - 1$.
2. $x^3 - 3x + 2$.
3. $|x^3 - 3x + 1|$.
4. $x^3 - 3x$.
5. $x^3 - bx$, $b > 0$.
6. $ax^3 - bx + c$, $ab > 0$.

Exercise 2.6.14. Use the graph of xe^{-x} in Example 2.6.9 to sketch the graphs.

1. xe^x.
2. xe^{2x}.
3. $(x + 1)e^x$.
4. $(ax + b)e^x$.
5. xa^x.
6. $|x|e^{-|x-1|}$.

Exercise 2.6.15. Sketch the graphs.

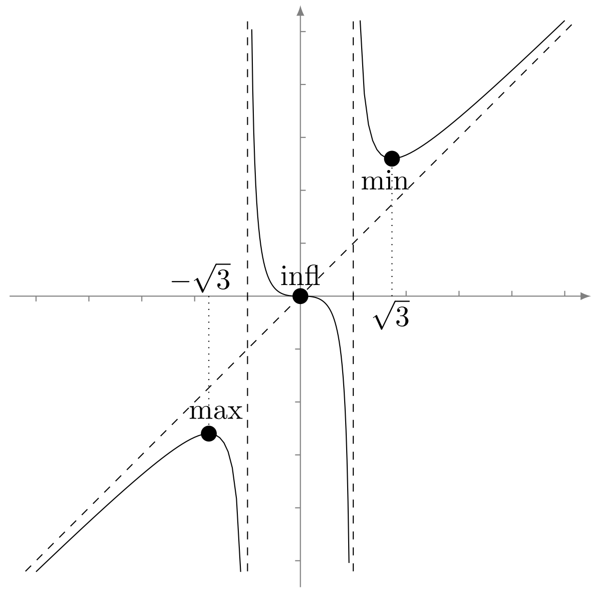


Figure 2.6.8: Graph of $\dfrac{x^3}{x^2-1}$.

1. $x - \sin x$.
2. $|x - \sin x|$.
3. $x - \cos x$.
4. $x + \sin x$.
5. $x + \cos x$.
6. $\dfrac{1}{2} + \sin x \cos x$.

Exercise 2.6.16. Sketch the graphs.

1. $\dfrac{1}{x^2}$.
2. $\dfrac{1}{x^2+1}$.
3. $\dfrac{1}{x^2-1}$.
4. $\dfrac{1}{x^2+bx+c}$.

Exercise 2.6.17. Sketch the graphs.

1. $(x^2+1)^p$.
2. $(x^2-1)^p$.
3. $\sqrt{1-x^2}$.
4. $\sqrt{ax^2+bx+c}$.

Exercise 2.6.18. Sketch the graph of $\dfrac{ax+b}{cx+d}$ by using the graph of $\dfrac{1}{x}$.

Exercise 2.6.19. Sketch the graphs on the natural domains of definitions.

1. $x^4 - 2x^2 + 1$.
2. $x + \dfrac{1}{x}$.
3. $\dfrac{x}{x^2+1}$.
4. $\dfrac{x^3}{(x-1)^2}$.
5. $x\sqrt{x-1}$.
6. $\sqrt{x^2+1} - x$.
7. $x^{\frac{1}{3}}(1-x)^{\frac{2}{3}}$.
8. x^2e^{-x}.
9. $\dfrac{1}{1+e^x}$.
10. $e^{\frac{1}{x}}$.
11. $x \log x$.
12. $x - \log(1+x)$.
13. $\log(1 - \log x)$.
14. $\log(1+x^4)$.
15. $e^{-x}\sin x$.

16. $x \tan x$.

17. $\frac{1}{1+\sin^2 x}$.

18. $2\sin x + \sin 2x$.

19. $2x - 4\sin x + \sin 2x$.

2.7 Numerical Application

The linear approximation can be used to find approximate values of functions.

Example 2.7.1. The linear approximation of $\sqrt{x}$ at $x = 4$ is

$$L(x) = \sqrt{4} + (\sqrt{x})'|_{x=4}(x-4) = 2 + \frac{1}{4}(x-4).$$

Therefore the value of the square root near 4 can be approximately computed

$$\sqrt{3.96} \approx 2 + \frac{1}{4}(-0.04) = 1.99, \quad \sqrt{4.05} \approx 2 + \frac{1}{4}(0.05) = 2.0125.$$

Example 2.7.2. Assume some metal balls of radius $r = 10$ are selected to make a ball bearing. If the radius is allowed to have 1% relative error, what is the maximal relative error of the weight?

The weight of the ball is

$$W = \frac{4}{3}\rho\pi r^3.$$

where ρ is the density. The error ΔW of the weight caused by the error Δr of the radius is

$$\Delta W \approx \frac{dW}{dr}\Delta r = 4\rho\pi r^2 \Delta r.$$

Therefore the relative error is

$$\frac{\Delta W}{W} \approx 3\frac{\Delta r}{r}.$$

Given the relative error of the radius is no more than 1%, we have $\left|\frac{\Delta r}{r}\right| \le 1\%$, so that the relative error of the weight is $\left|\frac{\Delta W}{W}\right| \le 3\%$.

Example 2.7.3. In Example 2.2.11, we computed the derivatives of the functions $y = y(x)$ and $z = z(x)$ given by the equations $x^2 + y^2 + z^2 = 2$ and $x + y + z = 0$, which is really a circle in $\mathbb{R}^3$. The point $P = (1, 0, -1)$ lies on the circle, where we have

$$y(1) = 0, \quad z(1) = -1, \quad y'(1) = \frac{1-(-1)}{(-1)-0} = -2, \quad z'(1) = \frac{1-0}{0-(-1)} = 1.$$

Therefore

$$y(1.01) \approx 0 - 2 \cdot 0.01 = -0.02, \qquad z(1.01) \approx -1 + 1 \cdot 0.01 = -0.99,$$
$$y(0.98) \approx 0 - 2 \cdot (-0.02) = 0.04, \qquad z(0.98) \approx -1 + 1 \cdot (-0.02) = -1.02.$$

In other words, the points $(1.01, -0.01, -0.99)$ and $(0.98, 0.04, -1.02)$ are near $(1, 0, -1)$ and almost on the circle.

Exercise 2.7.1. For $a > 0$ and small x, derive

$$\sqrt[n]{a^n + x} \approx a + \frac{x}{na^{n-1}}.$$

Then find the approximate values.

1. $\sqrt[4]{15}$.
2. $\sqrt{46}$.
3. $\sqrt[5]{39}$.
4. $\sqrt[7]{127}$.

Exercise 2.7.2. The period of a pendulum is $T = 2\pi\sqrt{\frac{L}{g}}$, where L the length of the pendulum and g is the gravitational constant. If the length of the pendulum is increased by 0.4%, what is the change in the period?

2.7.1 Remainder Formula

We may get more accurate values by using high order approximations. On the other hand, we have more confidence on the estimated values if we also know what the error is. The following result gives a formula for the error.

Theorem 2.7.1 (Lagrange Form of the Remainder). *If $f(x)$ has $(n+1)$-st order derivative on (a, x), then the remainder of the n-th order Taylor expansion of $f(x)$ at a is*

$$R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1} \text{ for some } c \in (a, x).$$

We only illustrate the argument for the case $n = 2$. We know that the remainder satisfies $R(a) = R'(a) = R''(a) = 0$. Therefore by Cauchy's Means Value Theorem (Theorem 2.4.5), we have

$$\begin{aligned}
\frac{R_2(x)}{(x-a)^3} &= \frac{R_2(x) - R_2(a)}{(x-a)^3 - (a-a)^3} = \frac{R_2'(c_1)}{3(c_1-a)^2} && (a < c_1 < x) \\
&= \frac{R_2'(c_1) - R_2'(a)}{3[(c_1-a)^2 - (a-a)^2]} = \frac{R_2''(c_2)}{3 \cdot 2(c_2 - a)} && (a < c_2 < c_1) \\
&= \frac{R_2''(c_2) - R_2''(a)}{3 \cdot 2(c_2 - a)} = \frac{R_2'''(c_3)}{3 \cdot 2 \cdot 1} = \frac{f'''(c_3)}{3!}. && (a < c_3 < c_2)
\end{aligned}$$

In the last step, we use $R_2''' = f'''$ because the $f - R_2$ is a quadratic function and has vanishing third order derivative.

A slight modification of the proof above actually gives a proof that the Taylor expansion is high order approximation (Theorem 2.5.2).

Example 2.7.4. The error for the linear approximation in Example 2.7.1 can be estimated by

$$R_1(x) = -\frac{\frac{1}{4c^{\frac{3}{2}}}}{2!}\Delta x^2 = \frac{1}{8c^{\frac{3}{2}}}\Delta x^3.$$

For both approximate values, we have

$$|R_1| \le \frac{1}{8\cdot 4^{\frac{3}{2}}}0.05^2 = 0.0000390625 < 4\times 10^{-5}.$$

If we use the quadratic approximation at 4

$$\sqrt{x} \approx 2 + \frac{1}{4}(x-4) - \frac{1}{64}(x-4)^2,$$

then we get better estimated values

$$\begin{aligned}\sqrt{3.96} &\approx 2 + \frac{1}{4}(-0.04) - \frac{1}{64}(-0.04)^2 = 1.989975,\\ \sqrt{4.05} &\approx 2 + \frac{1}{4}(0.05) - \frac{1}{64}(0.05)^2 = 2.0124609375.\end{aligned}$$

The error can be estimated by

$$R_2(x) = \frac{\frac{3}{8c^{\frac{5}{2}}}}{3!}\Delta x^3 = \frac{1}{16c^{\frac{5}{2}}}\Delta x^3.$$

For both computations, we have

$$|R_2| \le \frac{1}{16\cdot 4^{\frac{5}{2}}}0.05^3 = 0.00000025 = 2.5\times 10^{-7}.$$

The true values are $\sqrt{3.96} = 1.989974874213\cdots$ and $\sqrt{4.05} = 2.01246117975\cdots$.

Example 2.7.5. The Taylor expansion of e^x tells us

$$e = e^1 = 1 + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} + R_n(1).$$

By

$$|R_n(1)| = \frac{e^c}{(n+1)!}1^{n+1} \le \frac{e}{(n+1)!}, \quad 0 < c < 1,$$

we know

$$|R_{13}(1)| \le 0.000000000035 = 3.5\times 10^{-11}.$$

On the other hand,

$$1 + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{13!} = 2.718281828447\cdots,$$

Therefore $e = 2.7182818284\cdots$.

Exercise 2.7.3. Find approximate values by using Taylor expansions and estimate the errors.

1. $\sin 1$, 3rd order approximation.
2. $\log 2$, 5th order approximation.
3. e^{-1}, 5th order approximation.
4. $\arctan 2$, 3rd order approximation.

Exercise 2.7.4. Find approximate values accurate up to the 10-th digit.

1. $\sin 1$.
2. $\sqrt{4.05}$.
3. e^{-1}.
4. $\tan 46°$.

Exercise 2.7.5. Find the approximate value of $\tan 1$ accurate up to the 10-th digit by using the Taylor expansions of $\sin x$ and $\cos x$.

Exercise 2.7.6. If we use the Taylor expansion to calculate e accurate up to the 100-th digit, what is the order of the Taylor expansion we should use?

2.7.2 Newton's Method

The linear approximation can also be used to find approximate solutions of equations. To solve $f(x) = 0$, we start with a rough estimation x_0 and consider the linear approximation at x_0

$$L_0(x) = f(x_0) + f'(x_0)(x - x_0).$$

We expect the solution of the linear equation $L_0(x) = 0$ to be very close to the solution of $f(x) = 0$. The solution of the linear equation is easy to find

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

Although x_1 is not the exact solution of $f(x) = 0$, chances are it is an improvement of the initial estimation x_0.

To get an even better approximate solution, we repeat the process and consider the linear approximation at x_1

$$L_1(x) = f(x_1) + f'(x_1)(x - x_1)$$

The solution of the linear equation $L_1(x) = 0$

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}$$

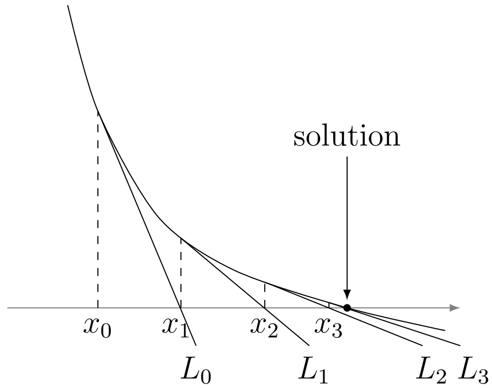


Figure 2.7.1: Newton's method.

is an even better estimation than x_1. The idea leads to an inductively constructed sequence

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

We expect the sequence to rapidly converge to the exact solution of $f(x) = 0$.

The scheme above for finding the approximate solution is called the *Newton's method*. The method may fail for various reasons. However, if the function is reasonably good and the initial estimation x_0 is sufficiently close to the exact solution, then the method indeed produces a sequence that rapidly converges to the exact solution. In fact, the error between x_n and the exact solution c satisfies

$$|x_{n+1} - c| \le M|x_n - c|^2$$

for some constant M that depends only on the function.

Example 2.7.6. By Example 1.7.5, the equation $x^3-3x+1 = 0$ should have a solution on $(0.3, 0.4)$. By Example 1.7.6, the equation should also have a second solution > 1 and a third solution < 0. More precisely, by $f(-2) = -1$, $f(-1) = 3$, $f(1) = -1$, $f(2) = 3$, the second solution is on $(-2, -1)$ and the third solution is on $(1, 2)$. Taking -2, 0.3, 2 as initial estimations, we apply Newton's method and compute the sequence

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

We find the three solutions

$$-1.87938524157182\cdots, \quad 0.347296355333861\cdots, \quad 1.53208888623796\cdots.$$

n	$x_0 = -2$	$x_0 = 0.3$	$x_0 = 2$
1	-1.88888888888889	0.346520146520147	1.66666666666667
2	-1.87945156695157	0.347296117887934	1.54861111111111
3	-1.87938524483667	0.347296355333838	1.53239016186538
4	-1.87938524157182	0.347296355333861	1.53208898939722
5	-1.87938524157182	0.347296355333861	1.53208888623797
6			1.53208888623796
7			1.53208888623796

Note that the initial estimation cannot be 1 or -1, because the derivative vanishes at the points. Moreover, if we start from $0.88, 0.89, 0.90$, we get very different sequences that respectively converge to the three sequences. We see that Newton's method can be very sensitive to the initial estimation, especially when the estimation is close to where the derivative vanishes.

n	$x_0 = 0.88$	$x_0 = 0.89$	$x_0 = 0.90$
1	-0.5362647754137	-0.657267917268	-0.80350877192983
2	0.6122033746535	0.920119732577	1.91655789116111
3	0.2884916149262	-1.212642459862	1.63097998546252
4	0.3461342508923	-3.235117846394	1.54150263235725
5	0.3472958236620	-2.419800571908	1.53218794505509
6	0.3472963553337	-2.014098301161	1.53208889739446
7	0.3472963553339	-1.891076746708	1.53208888623796
8		-1.879485375060	
9		-1.879385249013	
10		-1.879385241572	

Example 2.7.7. We solve the equation $\sin x + x\cos x = 0$ by starting with the estimation $x_0 = 1$. After five steps, we find the exact solution should be $0.325639452727856\cdots$.

n	x_n
0	1.000000000000000
1	0.471924667505487
2	0.330968826345873
3	0.325645312076542
4	0.325639452734876
5	0.325639452727856
6	0.325639452727856

Exercise 2.7.7. Applying Newton's method to solve $x^3 - x - 1 = 0$ with the initial estimations 1, 0.6 and 0.57. What lesson can you draw from the conclusion?

Exercise 2.7.8. Use Newton's method to find the unique positive root of $f(x) = e^x - x - 2$.

Exercise 2.7.9. Use Newton's method to find all the solutions of $x^2 - \cos x = 0$.

Exercise 2.7.10. Use Newton's method to find the approximate values of $\sqrt{4.05}$ and e^{-1} accurate up to the 10-th digit.

Exercise 2.7.11. Use Newton's method to find all solutions accurate up to the 6-th digit.

1. $x^4 = x + 3$.
2. $e^x = 3 - 2x$.
3. $\cos^2 x = x$.
4. $x + \tan x = 1$.

Note that one may rewrite the equation into another equivalent form and derive a simpler recursive relation.

Exercise 2.7.12. The ancient Babylonians used the recursive relation

$$x_{n+1} = \frac{1}{2}\left(x_n + \frac{a}{x_n}\right)$$

to get more an more accurate approximate values of $\sqrt{a}$. Explain the scheme by Newton's method.

Exercise 2.7.13. What approximate values does the recursive relation $x_{n+1} = 2x_n - ax_n^2$ give you? Explain by Newton's method.

Exercise 2.7.14. Explain why Newton's method does not work if we try to solve $x^3 - 3x + 1 = 0$ by starting at the estimation 1.

Exercise 2.7.15. Newton's method fails to solve the following equations by starting at any $x_0 \neq 0$. Why?

1. $\sqrt[3]{x} = 0$.
2. $\mathrm{sign}(x)\sqrt{|x|} = 0$.

Chapter 3

Integration

3.1 Area and Definite Integral

3.1.1 Area below Non-negative Function

Let $f(x)$ be a non-negative function on $[a, b]$. We wish to find the area of the region

$$G_{[a,b]}(f) = \{(x, y) \colon a \le x \le b,\ 0 \le y \le f(x)\}$$

between the graph of the function and the x-axis.

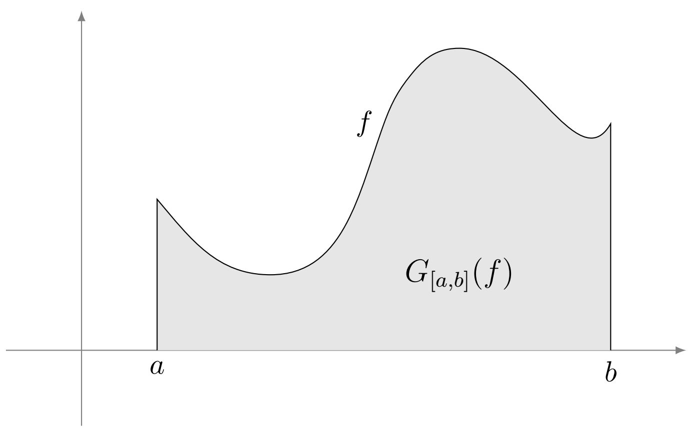


Figure 3.1.1: The region between the function and the x-axis.

Our strategy is the following. For any $x \in [a, b]$, let $A(x)$ be the area of $G_{[a,x]}(f)$, which is part of the region over $[a, x]$. We will find how the function $A(x)$ changes and recover $A(x)$ from its change. The area we wish to find is then the value $A(b)$. The subsequent argument assumes that $f(x)$ is continuous.

Consider an interval $[x, x+h] \subset [a, b]$, which implicitly assumes $h > 0$. Then the change $A(x+h)-A(x)$ is the area of $G_{[x,x+h]}(f)$. Note that $G_{[x,x+h]}(f)$ is sandwiched between two rectangles

$$[x, x+h] \times [0, m] \subset G_{[x,x+h]}(f) \subset [x, x+h] \times [0, M],$$

where

$$m = \min_{[x,x+h]} f, \quad M = \max_{[x,x+h]} f.$$

Since bigger region should have bigger area, we have

$$mh \le A(x+h) - A(x) \le Mh.$$

By $h > 0$, this is the same as

$$m \le \frac{A(x+h) - A(x)}{h} \le M. \tag{3.1.1}$$

Since the (right) continuity of $f(x)$ implies

$$\lim_{h\to 0^+} m = f(x) = \lim_{h\to 0^+} M,$$

by the sandwich rule, we further get the right derivative

$$A'_+(x) = \lim_{h\to 0^+} \frac{A(x+h) - A(x)}{h} = f(x). \tag{3.1.2}$$

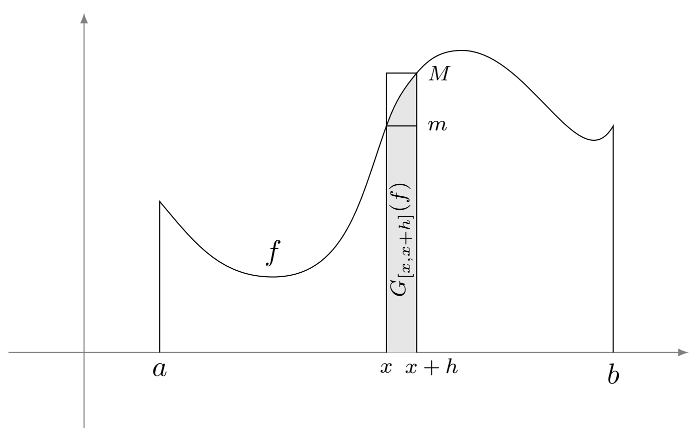


Figure 3.1.2: Estimate the change of area.

The argument above assumes $h > 0$. For $h < 0$, we consider $[x+h, x] \subset [a, b]$. Then the change $A(x+h) - A(x)$ is the *negative* of the area of $G_{[x+h,x]}(f)$, and the interval $[x+h, x]$ has length $-h$. By the same reason as before, we get

$$m(-h) \le -(A(x+h) - A(x)) \le M(-h).$$

By $-h > 0$, we still get the inequality (3.1.1), and further application of the sandwich rule gives the left derivative

$$A'_-(x) = \lim_{h\to 0^-} \frac{A(x+h) - A(x)}{h} = f(x).$$

We conclude that, for non-negative and continuous $f(x)$, we have

$$A'(x) = f(x). \tag{3.1.3}$$

Example 3.1.1. To find the area of the region below $f(x) = c$ over $[a, b]$, by (3.1.3), we have $A'(x) = c = (cx)'$. Then by Theorem 2.4.3, we get $A(x) = cx + C$. Further by $A(a) = 0$, we get $C = -ca$ and $A(x) = c(x - a)$. Therefore the area of the region is $A(b) = c(b - a)$.

The region $G_{[a,b]}(c)$ is actually a rectangle of base $b - a$ and height c. The computation of the area is consistent with the common sense.

Example 3.1.2. To find the area of the region below $f(x) = x$ over $[0, a]$, we start with $A'(x) = x = \left(\dfrac{1}{2}x^2\right)'$. This implies $A(x) = \dfrac{1}{2}x^2 + C$. By $A(0) = 0$, we further get $C = 0$ and $A(x) = \dfrac{1}{2}x^2$. Therefore the region has area $A(a) = \dfrac{1}{2}a^2$.

The region is actually a triangle, more precisely half of the square of side length a. The computation of the area is consistent with the common sense.

The pattern we see from the examples above is that, to find the area below a non-negative function and over an interval $[a, b]$, we first find a function $F(x)$ satisfying $f(x) = F'(x)$. Then by Theorem 2.4.3, $A'(x) = F'(x)$ implies $A(x) = F(x) + C$. Further, by $A(a) = 0$, we get $C = -F(a)$. Therefore $A(x) = F(x) - F(a)$, and the area we wish to find is

$$\text{Area}(G_{[a,b]}(f)) = F(b) - F(a).$$

This is the *Newton-Leibniz formula*. The function F is naturally called an *antiderivative* of f.

Example 3.1.3. To find the area of the region below x^2 and over $[0, a]$, we use $\left(\dfrac{1}{3}x^3\right)' = x^2$. The area is $\dfrac{1}{3}a^3 - \dfrac{1}{3}0^3 = \dfrac{1}{3}a^3$.

More generally, for any $p \neq -1$ and $0 < a < b$, by $\left(\dfrac{1}{p+1}x^p\right)' = x^p$, the area of the region below x^p and over $[a, b]$ is

$$\frac{1}{p+1}(b^{p+1} - a^{p+1}).$$

For example, the area of the region below $\sqrt{x}$ and over $[1, 2]$ is

$$\frac{2}{3}x^{\frac{3}{2}}\Big|_{x=1}^{x=2} = \frac{2}{3}(2^{\frac{3}{2}} - 1^{\frac{3}{2}}) = \frac{2}{3}(2\sqrt{2} - 1).$$

Exercise 3.1.1. Find the area of the region below the function over the given interval.

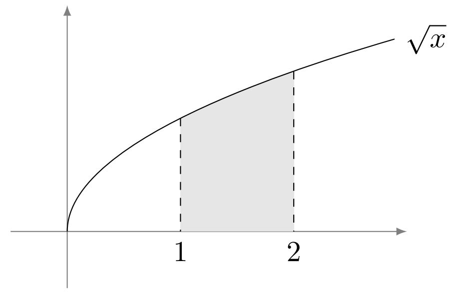


Figure 3.1.3: Area below parabola is $\dfrac{2}{3}(2\sqrt{2}-1)$.

1. x^p on $[0,1]$, $p>0$.
2. $\sin x$ on $[0,\frac{\pi}{2}]$.
3. e^x on $[0,a]$.
4. $\dfrac{1}{x}$ on $[1,a]$.
5. $\sqrt{1+x}$ on $[1,2]$.
6. $\log x$ on $[1,a]$.

Exercise 3.1.2. Find the area of the region bounded by $1-x^2$ and the x-axis.

3.1.2 Definite Integral of Continuous Function

What do we get if we apply the Newton-Leibniz formula to general continuous functions, which might become negative somewhere? The answer is the *signed area*. This means that we count the region between the non-negative part of f and the x-axis as having positive area and count the region between the non-positive part of f and the x-axis as having negative area. See Figure 3.1.4.

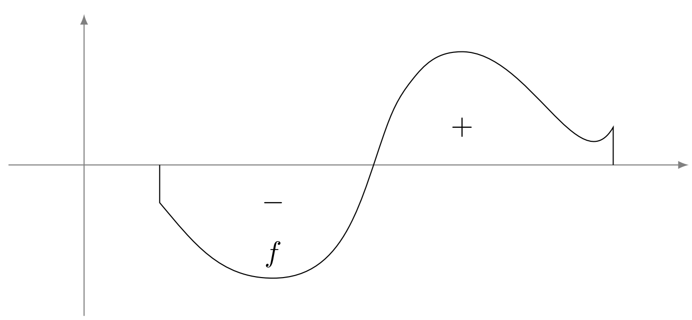


Figure 3.1.4: Computation by the Newton-Leibniz formula gives signed area.

To justify our claim, let $A(x)$ be the signed area for $f(x)$ over $[a,x]$. What we are really concerned with is the change of $A(x)$ where f is negative. So we consider $[x,x+h]\subset[a,b]$, with $h>0$ and $f<0$ on $[x,x+h]$. The change $A(x+h)-A(x)$ is the *negative* of the positive, "unsigned" area of the region

$$G_{[x,x+h]}(f)=\{(t,y)\colon x\le t\le x+h,\ 0\ge y\ge f(t)\}$$

between f and the x-axis along the interval $[x,x+h]$. We have the similar inclusion (see Figure 3.1.5, and note that $[M,0] \subset [m,0]$ because $0 \geq M \geq m$)

$$[x, x+h] \times [M,0] \subset G_{[x,x+h]}(f) \subset [x, x+h] \times [m,0], \quad m = \min_{[x,x+h]} f, M = \max_{[x,x+h]} f.$$

The heights of the rectangles are respectively $-M$ and $-m$, and we get (beware of the signs)

$$(-M)h \leq -(A(x+h) - A(x)) \leq (-m)h.$$

Again we get the inequality (3.1.1) and subsequently the limit (3.1.2).

The discussion for the case $h < 0$ is similar, and we conclude that $A'(x) = f(x)$.

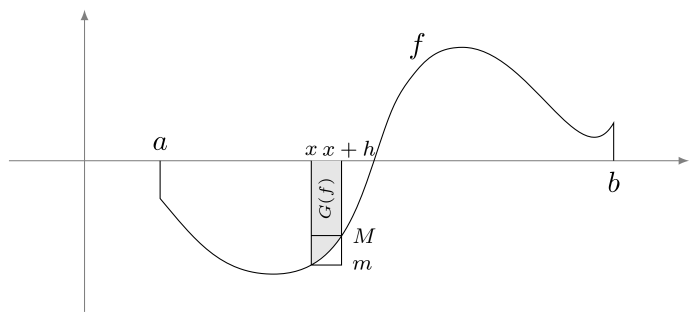


Figure 3.1.5: Estimate the change of negative area.

The signed area is the *definite integral* of $f(x)$ and is denoted $\int_a^b f(x)dx$. The function $f(x)$ is called the *integrand* and the ends a, b of the interval are called the *upper limit* and *lower limit*. The argument above and the explanation before Example 3.1.3 show that the definite integral can be computed by the Newton-Leibniz formula

$$\int_a^b f(x)dx = F(b) - F(a), \text{ where } F'(x) = f(x).$$

Example 3.1.4. By $(x)' = 1$ and $\left(\frac{1}{2}x^2\right)' = x$, we get

$$\int_a^b dx = b - a, \quad \int_a^b xdx = \frac{1}{2}(b^2 - a^2).$$

In general, for any integer $n \neq -1$, we have

$$\int_a^b x^n dx = \frac{1}{n+1}(b^{n+1} - a^{n+1}).$$

However, for $n < 0$, a and b need to have the same sign. The reason is that we derived the Newton-Leibniz formula under the assumption that the integrand is continuous. If a and b have different sign, then $0 \in [a,b]$, and x^n is not continuous on $[a,b]$ for $n<0$.

On the other hand, for any p, x^p is defined for $x>0$. Then for $b \geq a > 0$, we have

$$\int_a^b x^p dx = \frac{1}{p+1}(b^{p+1} - a^{p+1}).$$

We note that x^p is also defined at 0 for $p \geq 0$, and the formula above holds for $p \geq 0$ and $b \geq a \geq 0$. The reason is that x^p is right continuous at 0, and we derived Newton-Leibniz formula by one-sided derivatives.

Example 3.1.5. By $(e^x)' = e^x$ and $(\log x)' = \dfrac{1}{x}$, we get

$$\int_a^b e^x dx = e^b - e^a, \quad \int_a^b \frac{dx}{x} = \log b - \log a = \log \frac{b}{a}.$$

Note that the second integral requires $b \geq a > 0$.

Example 3.1.6. By $(\sin x)' = \cos x$ and $(\cos x)' = -\sin x$, we get

$$\int_a^b \cos x dx = \sin b - \sin a, \quad \int_a^b \sin x dx = \cos a - \cos b.$$

For example, we have

$$\int_{-\pi}^0 \sin x dx = \cos(-\pi) - \cos 0 = -2.$$

Example 3.1.7. From the derivatives of $\arcsin x$ and $\arctan x$, we get

$$\int_a^b \frac{dx}{\sqrt{1-x^2}} = \arcsin b - \arcsin a, \quad \int_a^b \frac{dx}{1+x^2} = \arctan b - \arctan a.$$

Of course, we need $|a|, |b| < 1$ in the first equality because the integrand is defined only on the open interval $(-1,1)$. In particular, the area of the region below $\dfrac{1}{1+x^2}$ and over $[0,1]$ is

$$\int_0^1 \frac{dx}{1+x^2} = \arctan 1 - \arctan 0 = \frac{\pi}{4}.$$

We also note that

$$\lim_{\substack{a\to-\infty \\ b\to+\infty}} \int_a^b \frac{dx}{1+x^2} = \lim_{b\to+\infty} \arctan b - \lim_{a\to-\infty} \arctan a = \frac{\pi}{2} - \left(-\frac{\pi}{2}\right) = \pi.$$

So the area of the unbounded region between $\dfrac{1}{1+x^2}$ and the x-axis is π.

Exercise 3.1.3. Use the area meaning of definite integral to directly find the value.

1. $\int_{-1}^{1}\sqrt{1-x^2}dx.$
2. $\int_{0}^{3}(x-2)dx.$
3. $\int_{a}^{b}|x-1|dx.$

Exercise 3.1.4. Compute and compare

$$\int_0^2 x^3dx,\quad \int_0^2 x^5dx,\quad \int_0^2 4x^3dx,\quad \int_0^2 6x^5dx,\quad \int_0^2 (4x^3+6x^5)dx.$$

What can you observe about the relation between

$$\int_a^b f(x)dx,\quad \int_a^b g(x)dx,\quad \int_a^b cf(x)dx,\quad \int_a^b (f(x)+g(x))dx.$$

Exercise 3.1.5. Compute definite integral.

1. $\int_{-1}^{2}(x^2-3x-4)dx.$
2. $\int_{0}^{2}(3x+1)^2dx.$
3. $\int_{0}^{2}(3x+1)(x-3)dx.$
4. $\int_{0}^{8}\sqrt{3x+1}dx.$
5. $\int_{0}^{1}(3+x\sqrt{x})dx.$
6. $\int_{1}^{2}\left(x+\frac{1}{x}\right)^2dx.$
7. $\int_{0}^{b}e^{x+a}dx.$
8. $\int_{0}^{1}(e^{-x}+\sin\pi x)dx.$
9. $\int_{0}^{\frac{\pi}{4}}\sec x\tan xdx.$

Exercise 3.1.6. For non-negative integers m and n, prove that

$$\int_0^{2\pi}\cos mx\sin nxdx = 0;$$

$$\int_0^{2\pi}\cos mx\cos nxdx = \begin{cases} 0, & \text{if } m\neq n,\\ \pi, & \text{if } m=n\neq 0,\\ 2\pi, & \text{if } m=n=0;\end{cases}$$

$$\int_0^{2\pi}\sin mx\sin nxdx = \begin{cases} 0, & \text{if } m\neq n \text{ or } m=n=0,\\ \pi, & \text{if } m=n\neq 0.\end{cases}$$

Exercise 3.1.7. Compute $\int_a^b \sqrt[3]{x}dx$ and $\int_a^b \frac{1}{\sqrt[3]{x}}dx$. Explain for what range of a,b are the formulae valid.

Exercise 3.1.8. What is wrong with the equality?

1. $\int_{-1}^{1}\frac{1}{x^2}dx = -\frac{1}{x}\Big|_{x=-1}^{x=1} = 2.$
2. $\int_{0}^{\pi}\sec^2 xdx = \tan\pi - \tan 0 = 0.$

Exercise 3.1.9. What is the area of the unbounded region between $\frac{1}{\sqrt{1-x^2}}$ and the x-axis, over the interval $(-1,1)$?

3.1.3 Property of Area and Definite Integral

Since the definite integral is the signed area, the usual properties of area is reflected as properties of the definite integral. An important property of area is the additivity. Specifically, if $X \cap Y$ has zero area, then the area of $X \cup Y$ should be the area of X plus the area of Y. Translated to definite integral, this means

$$\int_a^c f(x)dx = \int_a^b f(x)dx + \int_b^c f(x)dx. \tag{3.1.4}$$

The equality can be used to calculate the definite integral of "piecewise continuous" functions.

Example 3.1.8. The definite integral of the function (which is not continuous at 0)

$$f(x) = \begin{cases} -2x, & \text{if } -1 \le x < 0, \\ e^x, & \text{if } 0 \le x \le 1, \end{cases}$$

on $[-1, 1]$ is

$$\begin{aligned} \int_{-1}^1 f(x)dx &= \int_{-1}^0 f(x)dx + \int_0^1 f(x)dx = \int_{-1}^0 (-2x)dx + \int_0^1 e^x dx \\ &= -x^2|_{-1}^0 + e^x|_0^1 = e. \end{aligned}$$

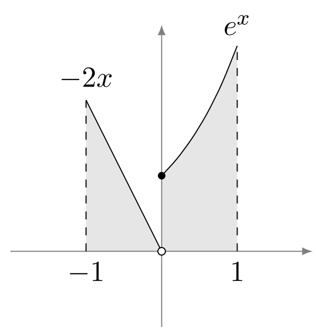


Figure 3.1.6: Definite integral of a piecewise continuous function.

We note that the computation of $\displaystyle\int_{-1}^0 f(x)dx$ actually reassigns the value $f(0) = 0$ to make the function continuous on $[-1, 0]$. The modification happens inside the vertical line $x = 0$. Since the vertical line has zero area, this does not affect the whole integral.

In general, changing the value of the integrand at finitely many places does not affect the integral.

Presumably, the definite integral $\int_a^b f(x)dx$ is defined only for the case $a \le b$, and the equality (3.1.4) implicitly assumes $a \le b \le c$. However, by

$$\int_a^a f(x)dx = 0,$$

and by taking $c = a$ in (3.1.4), we get

$$\int_b^a f(x)dx = -\int_a^b f(x)dx.$$

This extends the definite integral to the case the upper limit is smaller than the lower limit. With such extension, the equality (3.1.4) holds for any combination of a, b, c. Moreover, the extended definite integral is still computed by the antiderivative as before.

Another important property of area is positivity. Translated into definite integral, this means

$$f \ge 0 \implies \int_a^b f(x)dx \ge 0, \text{for } a < b. \tag{3.1.5}$$

Note that if $a > b$, then $\int_a^b f(x)dx \le 0$. The positivity is further extended to monotonicity in Example 3.5.5.

If we shift the graph under $f(x)$ over $[a, b]$ by d, we get the graph under $f(x-d)$ over $[a+d, b+d]$. Since the area is not changed by shifting, we get

$$\int_{a+d}^{b+d} f(x-d)dx = \int_a^b f(x)dx. \tag{3.1.6}$$

See Exercise for more examples of properties of area implying properties of definite integral.

In Section 3.5, we will introduce more properties from the the viewpoint of computation (i.e., Newton-Leibniz formula). Some of these properties cannot be easily explained by properties of area.

Exercise 3.1.10. Suppose $\int_0^2 f(x)dx = 3$, $\int_5^4 f(x)dx = 2$, $\int_5^0 f(x)dx = 0$. Find $\int_2^4 f(x)dx$.

Exercise 3.1.11. Use area to explain the equalities.

1. $\int_{-b}^{-a} f(-x)dx = \int_a^b f(x)dx$.

2. $\int_a^b cf(x)dx = c\int_a^b f(x)dx$.

3. $\displaystyle\int_a^b f(cx)dx = c^{-1}\int_{ca}^{cb} f(x)dx$.

Then express $\displaystyle\int_a^b f(Ax+B)dx$ as some multiple of the definite integral of $f(x)$ over some interval.

Exercise 3.1.12. Compute the integrals.

1. $\displaystyle\int_0^b a^x dx$.
2. $\displaystyle\int_a^b (1+2x)^n dx$.
3. $\displaystyle\int_{-1}^2 \mathrm{sign}(x)dx$.
4. $\displaystyle\int_{-1}^2 x^2\mathrm{sign}(x)dx$.
5. $\displaystyle\int_{-1}^2 |x|dx$.
6. $\displaystyle\int_2^{-1} |x|dx$.
7. $\displaystyle\int_0^2 |x^2-3x+2|dx$.
8. $\displaystyle\int_0^{2\pi} |\sin x|dx$.
9. $\displaystyle\int_{2\pi}^0 |\sin x|dx$.

Exercise 3.1.13. Compute the integrals.

1. $\displaystyle\int_0^2 f(x)dx,\ f(x)=\begin{cases} x^2, & \text{if } x<1,\\ x^{-2}, & \text{if } x\ge 2.\end{cases}$
2. $\displaystyle\int_{-2}^2 f(x)dx,\ f(x)=\begin{cases} e^{2|x|}, & \text{if } |x|<1,\\ e^{-x}, & \text{if } |x|\ge 1.\end{cases}$
3. $\displaystyle\int_0^\pi f(x)dx,\ f(x)=\begin{cases} \sin x, & \text{if } x<\frac{\pi}{2},\\ \cos x, & \text{if } x\ge \frac{\pi}{2}.\end{cases}$
4. $\displaystyle\int_{-\pi}^\pi f(x)dx,\ f(x)=\begin{cases} \sin x, & \text{if } |x|<\frac{\pi}{2},\\ \cos x, & \text{if } |x|\ge \frac{\pi}{2}.\end{cases}$

3.2 Rigorous Definition of Integral

3.2.1 What is Area?

The definite integral is defined as the signed area. Therefore the rigorous definition of integral relies on the rigorous definition of area. Any reasonable definition of area should have the following three properties (the area of a subset $X\subset\mathbb{R}^2$ is denoted $\mu(X)$):

1. Bigger subsets have bigger area: $X\subset Y$ implies $\mu(X)\le\mu(Y)$.
2. Areas can be added: If $\mu(X\cap Y)=0$, then $\mu(X\cup Y)=\mu(X)+\mu(Y)$.
3. Rectangles have the usual area: $\mu(\langle a,b\rangle\times\langle c,d\rangle)=(b-a)(d-c)$.

Here $\langle a, b\rangle$ can mean any of $[a, b]$, (a, b), $(a, b]$, or $[a, b)$. A carefully review of the argument in Section 3.1 shows that nothing beyond the three properties are used.

Suppose a plane region $A \subset \mathbb{R}^2$ is a union of finitely many rectangles, $A = \cup_{i=1}^n I_i$, such that the intersections between I_i are at most lines. Since lines have zero area by the third property, we may use the second property to further define $\mu(A) = \sum_{i=1}^n \mu(I_i)$. We give such a plane region the temporary name "good region", since we have definite idea about the area of a good region. (Strictly speaking, we still need to argue that $\sum_{i=1}^n \mu(I_i)$ is independent of the decomposition $A = \cup_{i=1}^n I_i$.)

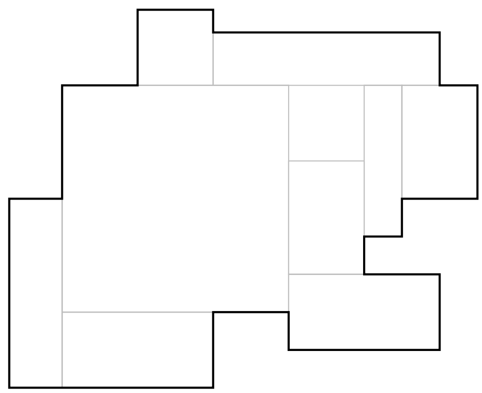


Figure 3.2.1: Good region.

For any (bounded) subset $X \subset \mathbb{R}^2$, we may try to approximate X by good regions, from inside as well as from outside. In other words, we consider good regions A and B satisfying $A \subset X \subset B$. Then by the first property of area, we must have

$$\mu(A) \subset \mu(X) \subset \mu(B).$$

Note that $\mu(A)$ and $\mu(B)$ have been defined, and $\mu(X)$ is yet to be defined. So we introduce the *inner area* (the maximum should really be the supremum)

$$\mu_*(X) = \max\{\mu(A) \colon A \subset X,\ A \text{ is a good region}\},$$

as the lower bound for $\mu(X)$, and the *outer area* (the minimum should really be the infimum)

$$\mu^*(X) = \min\{\mu(B) \colon B \supset X,\ B \text{ is a good region}\},$$

as the upper bound for $\mu(X)$. We say that the subset X *has area* (or *Jordan measurable*) if $\mu_*(X) = \mu^*(X)$, and the common value is the *area* $\mu(X)$ of X. If $\mu_*(X) \neq \mu^*(X)$, then we say X *has no area*.

The subset X has area if and only if for any $\epsilon > 0$, there are good regions A and B, such that $A \subset X \subset B$ and $\mu(B) - \mu(A) < \epsilon$. In other words, we can find good inner and outer approximations, such that the difference between the approximations can be arbitrarily small.

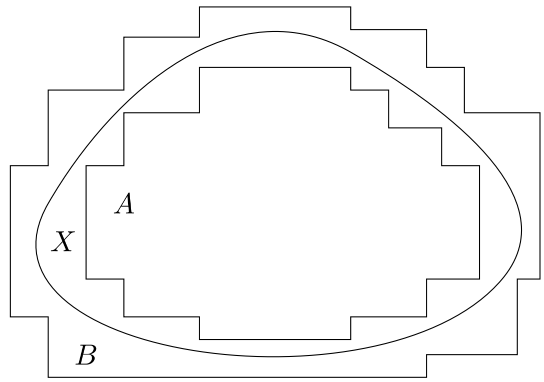


Figure 3.2.2: Approximation by good regions.

Example 3.2.1. A point can be considered as a reduced rectangle and has area 0. If X consists of finitely many points, then we can take $B = X$ be the union of all "point rectangles" in X. Since $\mu(B) = 0$, we get $\mu^*(X) = 0$. By $0 \le \mu_*(X) \le \mu^*(X)$, we also have $\mu_*(X) = 0$. Therefore finitely many points has area 0.

Example 3.2.2. Consider the triangle with vertices $(0,0)$, $(1,0)$ and $(1,1)$. We partition the interval $[0,1]$ into n parts of equal length

$$[0,1] = \cup_{i=1}^n \left[\frac{i-1}{n}, \frac{i}{n}\right].$$

Correspondingly, we have the inner and outer approximations of the triangle

$$A_n = \cup_{i=1}^n \left[\frac{i-1}{n}, \frac{i}{n}\right] \times \left[0, \frac{i-1}{n}\right], \quad B_n = \cup_{i=1}^n \left[\frac{i-1}{n}, \frac{i}{n}\right] \times \left[0, \frac{i}{n}\right].$$

They have area

$$\mu(A_n) = \sum_{i=1}^n \frac{1}{n}\frac{i-1}{n} = \frac{1}{2n}(n-1), \quad \mu(B_n) = \sum_{i=1}^n \frac{1}{n}\frac{i}{n} = \frac{1}{2n}(n+1).$$

By taking sufficiently big n, the difference $\mu(B_n) - \mu(A_n) = \dfrac{1}{n}$ can be arbitrarily small. Therefore the triangle has area, and the area is given by $\lim_{n\to\infty} \mu(A_n) = \lim_{n\to\infty} \mu(B_n) = \dfrac{1}{2}$. This justifies the conclusion of Example 3.1.2 for the case $a = 1$.

Example 3.2.3. For an example of subsets without area, i.e., satisfying $\mu_*(X) \neq \mu^*(X)$, let us consider the subset $X = (\mathbb{Q} \cap [0,1])^2$ of all rational pairs in the unit square.

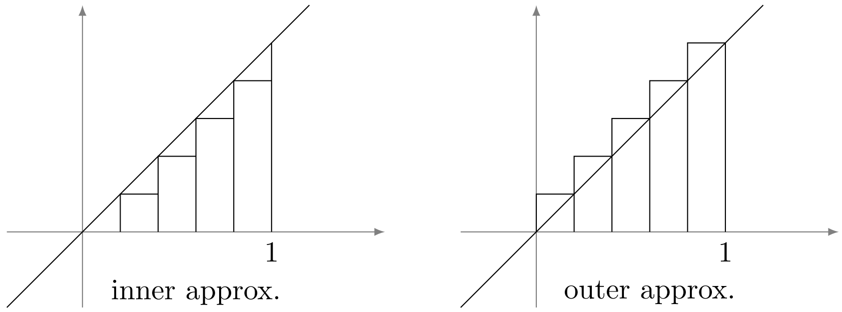


Figure 3.2.3: Approximating triangle.

Since the only rectangles contained in X are single points, any good region $A \subset X$ must be finitely many points. Therefore $\mu(A) = 0$ for any good region $A \subset X$, and $\mu_*(X) = 0$.

On the other hand, if B is a good region containing X, then B must almost contain the whole square $[0,1]^2$, with the only exception of finitely many horizontal or vertical (irrational) line segments. Therefore we have $\mu(B) \geq \mu([0,1]^2) = 1$. This implies $\mu^*(X) \geq 1$ (show that $\mu^*(X) = 1$!).

Exercise 3.2.1. Use inner and outer approximations to explain that any rectangle has area given by the multiplication of two sides. This justifies Example 3.1.1.

Exercise 3.2.2. Explain that a (not necessarily horizontal or vertical) straight line segment has area 0.

Exercise 3.2.3. Explain that the region between $y = x$ and the x-axis over $[0, a]$ has area $\frac{1}{2}a^2$. This fully justifies the computation in Example 3.1.2.

Exercise 3.2.4. Explain that the subset $X = (\mathbb{Q} \cap [0,1]) \times [0,1]$ of all vertical rational lines in the unit square has no area.

Exercise 3.2.5. Show that if $X \subset Y$, then $\mu_*(X) \leq \mu_*(Y)$ and $\mu^*(X) \leq \mu^*(Y)$. In particular, we have $\mu(X) \leq \mu(Y)$ in case both X and Y have areas. The property is used in deriving (3.1.1).

3.2.2 Darboux Sum

After the rigorous definition of area, we can give the rigorous definition of definite integral.

Definition 3.2.1. A function $f(x)$ is *Riemann integrable* if the region

$$G_{[a,b]}(f) = \{(x, y) \colon a \leq x \leq b,\ y \text{ is between } 0 \text{ and } f(x)\}$$

has area. Moreover, the *Riemann integral* is

$$\int_a^b f(x)dx = \mu\left(G_{[a,b]}(f)\cap H_+\right) - \mu\left(G_{[a,b]}(f)\cap H_-\right),$$

where

$$H_+ = \{(x,y)\colon y\geq 0\},\quad H_- = \{(x,y)\colon y\leq 0\}$$

are the upper and lower half planes.

Suppose $f\geq 0$ on $[a,b]$. As indicated by Figure 3.2.4, for any inner approximation of $G_{[a,b]}(f)$, we can always choose "full vertical strips" to get a better approximation for $G_{[a,b]}(f)$. Here better means closer to the expected value of $\mu(G_{[a,b]}(f))$. The outer approximations have similar improvements by full vertical strips. Therefore we only need to consider the approximations by full vertical strips.

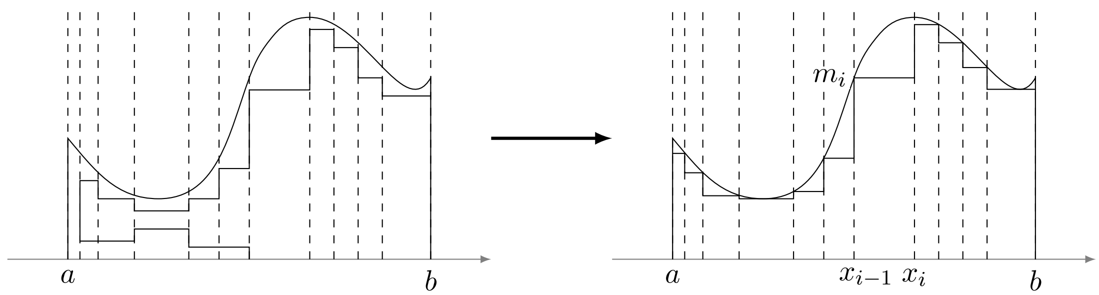


Figure 3.2.4: Better inner approximations by vertical strips.

An approximation by full vertical strips is determined by a *partition* of the interval

$$P\colon a = x_0 < x_1 < \cdots < x_n = b.$$

On the i-th interval $[x_{i-1},x_i]$, the inner strip has height $m_i = \min_{[x_{i-1},x_i]} f$ (the minimum should really be the *infimum*), and the outer strip has height $M_i = \max_{[x_{i-1},x_i]} f$ (the maximum should really be the *supremum*). Therefore the inner and outer approximations are

$$A_P = \cup_{i=1}^n [x_{i-1},x_i]\times[0,m_i) \subset X \subset B_P = \cup_{i=1}^n [x_{i-1},x_i]\times[0,M_i].$$

The areas of the two approximations are the *lower and upper Darboux sums*

$$L(P,f) = \mu(A_P) = \sum_{i=1}^n m_i(x_i - x_{i-1}),$$

$$U(P,f) = \mu(B_P) = \sum_{i=1}^n M_i(x_i - x_{i-1}).$$

The Riemann integrability of f means that $G_{[a,b]}(f)$ has area, which further means that the difference between inner and outer approximations can be arbitrarily small. Therefore we get the following criterion for the integrability.

Theorem 3.2.2 (Riemann Criterion). *A bounded function f on $[a,b]$ is Riemann integrable, if and only if for any $\epsilon > 0$, there is a partition P, such that*

$$U(P,f) - L(P,f) = \sum_{i=1}^{n} \left(\max_{[x_{i-1},x_i]} f - \min_{[x_{i-1},x_i]} f \right)(x_i - x_{i-1}) < \epsilon.$$

The quantity

$$\omega_{[x_{i-1},x_i]} f = \max_{[x_{i-1},x_i]} f - \min_{[x_{i-1},x_i]} f$$

measures how much the value of f fluctuates on $[x_{i-1}, x_i]$ and is called the *oscillation* of the function on the interval. Since the continuity of a function can imply that such oscillations are uniformly small, continuous functions are always Riemann integrable. The criterion also implies that monotone functions are Riemann integrable. On the other hand, there are functions that are not Riemann integrable.

Example 3.2.4. Consider the Dirichlet function $D(x)$ on $[0,1]$. We always have

$$\min_{[x_{i-1},x_i]} D = 0, \quad \max_{[x_{i-1},x_i]} D = 1.$$

Therefore $U(P,f) - L(P,f) = 1$ cannot be arbitrarily small. We conclude that the Dirichlet function is not Riemann integrable.

The example is closely related to Example 3.2.3. See Exercise 3.2.4.

We note that $U(P,f) - L(P,f)$ is the area of the good subset $B_P - A_P = \cup_{i=1}^{n} [x_{i-1}, x_i] \times [m_i, M_i]$. If we consider all the partitions P, the all such good subsets are essentially all the outer approximations of the graph $\{(x, f(x)) \colon a \le x \le b\}$ of f (a curve, not including the part below f). Therefore Theorem 3.2.2 basically says that a function is Riemann integrable if and only if the graph curve of the function has zero area.

The graph curve is part of the boundary of $G(f)$. In this viewpoint, Theorem 3.2.2 is a special case of the following.

Theorem 3.2.3. *A bounded subset $X \subset \mathbb{R}^2$ has area if and only if its boundary ∂X has zero area.*

We remark that the theory of area can be easily extended to the theory of volume for subsets in $\mathbb{R}^n$. We may then get the rigorous definition of multivariable Riemann integrals on subsets of Euclidean spaces, where the subsets should have volume themselves. The high dimensional versions of Theorems 3.2.2 and 3.2.3 are still valid.

Further extension of the area theory introduces countably many in place of finitely many. The result is the modern theory of Lebesgue measure and Lebesgue integral.

3.2.3 Riemann Sum

Suppose f is Riemann integrable. When the partition gets more and more refined, the upper and lower Darboux sums, as the areas of the outer and inner approximations, will become closer to the integral $\int_a^b f(x)dx$. Now choose ϕ_i satisfying

$$m_i \le \phi_i \le M_i.$$

Then we get the *Riemann sum*

$$S(P,f) = \sum_{i=1}^{n} \phi_i(x_i - x_{i-1})$$

sandwiched between the two Darboux sums

$$L(P,f) \le S(P,f) \le U(P,f).$$

We conclude that $S(P,f)$ will also become closer to the integral $\int_a^b f(x)dx$.

A useful case of the Riemann sum is obtained by taking $\phi_i = f(x_i^*)$ to be the values of some sample points in the partition intervals

$$S(P,f) = \sum_{i=1}^{n} f(x_i^*)(x_i - x_{i-1}), \quad x_i^* \in [x_{i-1}, x_i].$$

This is what is usually called the Riemann sum in most textbooks.

Theorem 3.2.4. *Suppose f is Riemann integrable on $[a,b]$. Then for any $\epsilon > 0$, there is a partition P_0 of $[a,b]$, such that for any partition P obtained by adding more partition points to P_0 (we say P is a refinement of P_0), we have*

$$\left| S(P,f) - \int_a^b f(x)dx \right| < \epsilon.$$

The statement above is very similar to the limit of sequences and functions, and we may write

$$\int_a^b f(x)dx = \lim_P S(P,f).$$

The subtlety here is that refinement of partitions replaces $n > N$ or $|x-a| < \delta$.

3.3 Numerical Calculation of Integral

3.3.1 Left and Right Rule

Although Riemann integrals can be computed by the Newton-Leibniz formula, it is often impossible to find the exact formula of a function F satisfying $F' = f$.

Moreover, even if we can find a formula for F, it might be too complicated to evaluate. For many practical applications, it is sufficient to find an approximate value of the integration. Many efficient numerical schemes have been invented for this purpose.

All the schemes are the extensions of the Riemann sum in Section 3.2.3. Usually one starts with a partition that evenly divides the interval

$$P_n\colon a = x_0 < x_1 = a + h < \cdots < x_i = a + ih < \cdots < x_n = b = a + nh,$$

where

$$h = \frac{b-a}{n} = x_i - x_{i-1}$$

is the *step size* of the partition. By taking all the sample points to be the left of the partition intervals, we get $x_i^* = x_{i-1} = a + (i-1)h$ and the *left rule*

$$L_n = h(f(x_0) + f(x_1) + \cdots + f(x_{n-1})).$$

By taking all the sample points to be the right of the partition interval, we get $x_i^* = x_i = a + ih$ and the *right rule*

$$R_n = h(f(x_1) + f(x_2) + \cdots + f(x_n)).$$

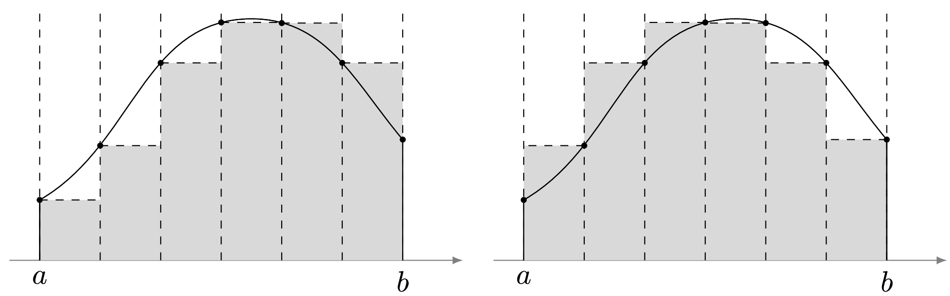


Figure 3.3.1: Left and right rules.

Example 3.3.1. For $f(x) = x$ on $[0,1]$, we have

$$\begin{aligned}
L_n &= \frac{1}{n}\left(\frac{0}{n} + \frac{1}{n} + \cdots + \frac{n-1}{n}\right) \\
&= \frac{1}{n^2}(0 + 1 + \cdots + (n-1)) = \frac{1}{n^2}\frac{1}{2}(n-1)n = \frac{n-1}{2n}, \\
R_n &= \frac{1}{n}\left(\frac{1}{n} + \frac{2}{n} + \cdots + \frac{n}{n}\right) \\
&= \frac{1}{n^2}(1 + 2 + \cdots + n) = \frac{1}{n^2}\frac{1}{2}n(n+1) = \frac{n+1}{2n}.
\end{aligned}$$

Both converge to $\dfrac{1}{2} = \displaystyle\int_0^1 x dx$ as $n \to \infty$.

Example 3.3.2. To compute the integral of $f(x) = \dfrac{1}{1+x^2}$ on $[0,1]$, we take $n = 4$. The partition is

$$P_4\colon 0 < 0.25 < 0.5 < 0.75 < 1, \quad h = 0.25.$$

The values of $f(x)$ at the five partition points are

$$1.000000, \quad 0.941176, \quad 0.800000, \quad 0.640000, \quad 0.500000.$$

Then we get the following approximate values of $\displaystyle\int_0^1 \frac{dx}{1+x^2}$

$$\begin{aligned} L_4 &= 0.25 \times (1.000000 + 0.941176 + 0.800000 + 0.640000) \approx 0.845294, \\ R_4 &= 0.25 \times (0.941176 + 0.800000 + 0.640000 + 0.500000) \approx 0.720294. \end{aligned}$$

By Example 3.1.7, the actual value is $\dfrac{\pi}{4} = 0.7853981634\cdots$.

Exercise 3.3.1. Find L_n and R_n and confirm the value of related integral.

1. $f(x) = x$ on $[a,b]$.
2. $f(x) = x^2$ on $[0,1]$.
3. $f(x) = 2^x$ on $[0,1]$.
4. $f(x) = a^x$ on $[0,1]$.

Exercise 3.3.2. Explain the identity.

1. $\lim_{n\to\infty} \left(\dfrac{1}{n^2} + \dfrac{2}{n^2} + \cdots + \dfrac{n-1}{n^2}\right) = \displaystyle\int_0^1 x dx$
2. $\lim_{n\to\infty} \left(\dfrac{1}{n+1} + \dfrac{1}{n+2} + \cdots + \dfrac{1}{2n}\right) = \displaystyle\int_0^1 \frac{dx}{1+x}$.
3. $\lim_{n\to\infty} \left(\dfrac{n}{n^2+1^2} + \dfrac{n}{n^2+2^2} + \cdots + \dfrac{n}{n^2+n^2}\right) = \displaystyle\int_0^1 \frac{dx}{1+x^2}$.
4. $\lim_{n\to\infty} \dfrac{1}{n}\left(\sin\dfrac{\pi}{n} + \sin\dfrac{2\pi}{n} + \cdots + \sin\dfrac{(n-1)\pi}{n}\right) = \displaystyle\int_0^1 \sin \pi x dx$.

Exercise 3.3.3. Interpret the limit as integration.

1. $\lim_{n\to\infty} \dfrac{1^p + 2^p + \cdots + n^p}{n^{p+1}}$.
2. $\lim_{n\to\infty} \dfrac{1}{n}\sum_{k=1}^n f\left(a + k\dfrac{b-a}{n}\right)$.

Exercise 3.3.4. By interpreting $\displaystyle\int_1^2 \log x dx$, find $\lim_{n\to\infty} \dfrac{1}{n}\sqrt[n]{\dfrac{(2n)!}{n!}}$.

3.3.2 Midpoint Rule and Trapezoidal Rule

The left and right rules are quite primitive approximations of the integral. A better choice is the middle points

$$\bar{x}_i = \frac{x_{i-1}+x_i}{2} = a + \frac{2i-1}{2}h$$

and the corresponding Riemann sum

$$M_n = h(f(\bar{x}_1) + f(\bar{x}_2) + \cdots + f(\bar{x}_n)).$$

Another choice is the average of the Riemann sums using the left and right points.

$$T_n = \frac{L_n + R_n}{2} = \frac{h}{2}(f(x_0) + 2f(x_1) + 2f(x_2) + \cdots + 2f(x_{n-1}) + f(x_n)).$$

The two approximation schemes are the *midpoint rule* and the *trapezoidal rule*.

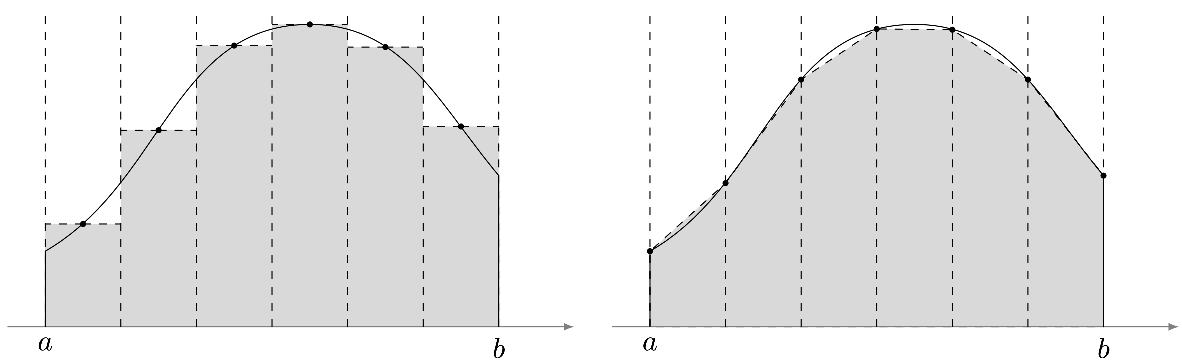


Figure 3.3.2: Midpoint and trapezoidal rules.

Example 3.3.3. For $f(x) = x$ on $[0,1]$, we have

$$\begin{aligned} M_n &= \frac{1}{n}\left(\frac{1}{2n} + \frac{3}{2n} + \cdots + \frac{2n-1}{2n}\right) \\ &= \frac{1}{2n^2}\left(1 + 3 + \cdots + (2n-1)\right) = \frac{1}{2n^2}n^2 = \frac{1}{2}, \\ T_n &= \frac{1}{2}\left(\frac{n-1}{2n} + \frac{n+1}{2n}\right) = \frac{1}{2}. \end{aligned}$$

Both happen to be equal to $\frac{1}{2} = \int_0^1 x dx$.

Example 3.3.4. For $f(x)=x^2$ on $[0,1]$, we have

$$\begin{aligned}
M_n &= \frac{1}{n}\left(\frac{1^2}{4n^2}+\frac{3^2}{4n^2}+\cdots+\frac{(2n-1)^2}{4n^2}\right)\\
&= \frac{1}{4n^3}[(1^2+2^2+\cdots+(2n)^2)-2^2(1^2+2^3+\cdots+n^2)]\\
&= \frac{1}{4n^3}\left(\frac{1}{6}2n(2n+1)(4n+1)-\frac{1}{6}n(n+1)(2n+1)\right)=\frac{4n^2-1}{12n^2},\\
T_n &= \frac{1}{2n}\left(\frac{0^2}{n^2}+2\frac{1^2}{n^2}+2\frac{2^2}{n^2}+\cdots+2\frac{(n-1)^2}{n^2}+\frac{n^2}{n^2}\right)\\
&= \frac{1}{2n^3}[2(1^2+2^2+\cdots+n^2)-n^2]\\
&= \frac{1}{2n^3}\left(2\frac{1}{6}n(n+1)(2n+1)-n^2\right)=\frac{2n^2+1}{6n^2}.
\end{aligned}$$

Compared with the actual value $\int_0^1 x^2dx=\frac{1}{3}$, the errors are $\frac{1}{12n^2}$ and $\frac{1}{6n^2}$.

Example 3.3.5. We apply the midpoint and trapezoidal rules to $\int_0^1 \frac{dx}{1+x^2}$. For $n=4$, we have

$$\begin{aligned}
M_4 &= 0.25\times(0.984615+0.876712+0.719101+0.566372)\approx 0.786700,\\
T_4 &= \frac{0.25}{2}\times(1.000000+2\times 0.941176+2\times 0.800000+2\times 0.640000+0.500000)\\
&\approx 0.782794.
\end{aligned}$$

For $n=8$, we have $h=0.125$ and the following values.

i	x_i	$\frac{1}{1+x_i^2}$	$\bar{x}_i$	$\frac{1}{1+\bar{x}_i^2}$
0	0	1.000000		
1	0.125	0.984615	0.0625	0.996109
2	0.25	0.941176	0.1875	0.966038
3	0.325	0.876712	0.3125	0.911032
4	0.5	0.800000	0.4375	0.839344
5	0.625	0.719101	0.5625	0.759644
6	0.75	0.640000	0.6875	0.679045
7	0.875	0.566372	0.8125	0.602353
8	1	0.500000	0.9375	0.532225

Then we get the approximations

$$M_8\approx 0.785721,\quad T_8\approx 0.784747.$$

Compared with the actual value $\displaystyle\int_0^1 \frac{dx}{1+x^2} = \frac{\pi}{4} = 0.7853981634\cdots$, the following are the errors of various schemes.

error	$n=4$	$n=8$
$\lvert R_n - I\rvert$	0.065104	0.031901
$\lvert L_n - I\rvert$	0.059896	0.030599
$\lvert M_n - I\rvert$	0.001302	0.000323
$\lvert T_n - I\rvert$	0.002604	0.000651

We observe that the midpoint and trapezoidal rules are much more accurate, and the error for the midpoint rule is about half of the error for the trapezoidal rule. Moreover, doubling the number of partition points improves the error by a factor of 4 for the two rules. The following gives an estimation of the errors.

Theorem 3.3.1. *Suppose $f''(x)$ is continuous and bounded by K_2 on $[a,b]$, then*

$$\left|\int_a^b f(x)dx - M_n\right| \le \frac{K_2(b-a)^3}{24n^2}, \quad \left|\int_a^b f(x)dx - T_n\right| \le \frac{K_2(b-a)^3}{12n^2}.$$

The estimations are derived in Exercises 3.5.9 and 3.5.17.

Exercise 3.3.5. To compute the integral $\displaystyle\int_a^b x^2dx$, for any partition of $[a,b]$, we take $x_i^* = \sqrt{\frac{1}{3}(x_{i-1}^2 + x_{i-1}x_i + x_i^2)} \in [x_{i-1}, x_i]$. Show that the Riemann sum is exactly the value of the integral. How can you generalize this to $\displaystyle\int_a^b x^n dx$?

Exercise 3.3.6. For the integral $\displaystyle\int_a^b x^2dx$, we take any partition of $[a,b]$, in which the intervals may not have the same length. Estimate the error of the various schemes in terms of the size $\delta = \max_{i=1}^n (x_i - x_{i-1})$ of the partition.

Exercise 3.3.7. Apply the midpoint and trapezoidal rules to the integral and compare with the actual value.

1. $\displaystyle\int_1^2 \frac{dx}{x}$, $n = 6, 12$.
2. $\displaystyle\int_0^\pi \sin x dx$, $n = 4, 12$.

Exercise 3.3.8. Apply the midpoint and trapezoidal rules to the integral. Moreover, estimate the number of partition points needed for the approximation to be accurate up to 10^{-6}.

1. $\displaystyle\int_0^{\pi} \cos x^2 dx$, $n = 5, 10$.
2. $\displaystyle\int_0^{\pi} \frac{\sin x}{x} dx$, $n = 5, 10$.
3. $\displaystyle\int_0^{2} \frac{1}{\sqrt{1+x^3}} dx$, $n = 5, 10$.
4. $\displaystyle\int_0^{1} e^{x^2} dx$, $n = 10$.
5. $\displaystyle\int_1^{2} e^{\frac{1}{x}} dx$, $n = 10$.
6. $\displaystyle\int_1^{2} \frac{\log x}{1+x} dx$, $n = 10$.

Exercise 3.3.9. Show that $T_{2n} = \dfrac{1}{2}(M_n + T_n)$.

Exercise 3.3.10. Prove that if f is a concave positive function, then $T_n \le \displaystyle\int_a^b f(x)dx < M_n$.

3.3.3 Simpson's Rule

The midpoint rule is based on the constant approximation, and the trapezoidal rule is based on linear approximation (actually not quite, as average of two constant approximations). We may expect better approximation by using quadratic curves.

Since a quadratic curve is determined by three points, we try to approximate $f(x)$ on the interval $[x_{i-1}, x_{i+1}]$ by the quadratic function $Q(x) = A(x-x_i)^2+B(x-x_i)+C$ satisfying

$$\begin{aligned} f(x_{i-1}) &= Q(x_{i-1}) = Ah^2 - Bh + C, \\ f(x_i) &= Q(x_i) = C, \\ f(x_{i+1}) &= Q(x_{i+1}) = Ah^2 + Bh + C. \end{aligned}$$

Then $\displaystyle\int_{x_{i-1}}^{x_{i+1}} f(x)dx$ is approximated by (the first equality uses (3.1.6))

$$\int_{x_{i-1}}^{x_{i+1}} Q(x)dx = \frac{2}{3}Ah^3 + 2Ch = \frac{h}{3}(f(x_{i-1}) + 4f(x_i) + f(x_{i+1})).$$

Suppose n is even. Then we may apply the quadratic approximations to $[x_0, x_2]$, $[x_2, x_4]$, ..., $[x_{n-2}, x_n]$. Adding such approximations together, we get an approximation of $\displaystyle\int_a^b f(x)dx$

$$S_n = \frac{h}{3}(f(x_0)+4f(x_1)+2f(x_2)+4f(x_3)+2f(x_4)+\cdots+2f(x_{n-2})+4f(x_{n-1})+f(x_n)).$$

This is *Simpson's rule*. Observe that $S_n = \dfrac{1}{3}(2T_n + M_{\frac{n}{2}})$ is the weighted average of the trapezoidal (with step size h) and midpoint (with step size $2h$) rules.

The errors in Simpson's rule can be estimated by the bound on the fourth order derivative.

Theorem 3.3.2. *Suppose $f^{(4)}(x)$ is continuous and bounded by K_4 on $[a, b]$, then*

$$\left|\int_a^b f(x)dx - S_n\right| \le \frac{K_4(b-a)^5}{180n^4}.$$

A consequence of the theorem is that doubling the partition improves the error by a factor of 16!

Example 3.3.6. Applying Simpson's rule to $\displaystyle\int_0^1 \frac{dx}{1+x^2}$ for $n = 4$, we use the same data from Example 3.3.1 to get

$$\begin{aligned}S_4 &= \frac{0.25}{3} \times (1.000000 + 4 \times 0.941176 + 2 \times 0.800000 + 4 \times 0.640000 + 0.500000)\\ &\approx 0.785392.\end{aligned}$$

The error $|S_n - I| = 0.000540$ is comparable to the midpoint and trepezoidal rule for $n = 8$.

How many partition points are needed in order to get the approximate value accurate up to the 6-th digit? To answer the question, we compute the derivatives

$$f^{(4)}(x) = \frac{24(5x^4 - 10x^2 + 1)}{(1+x^2)^5}, \quad f^{(5)}(x) = -\frac{240x(x^2-3)(3x^2-1)}{(1+x^2)^6}.$$

From $f^{(5)}(x)$, the extrema of $f^{(4)}(x)$ on $[0, 1]$ can only be at 0, $\dfrac{1}{\sqrt{3}}$ or 1. By

$$|f^{(4)}(0)| = 24, \quad \left|f^{(4)}\left(\frac{1}{\sqrt{3}}\right)\right| = \frac{81}{8}, \quad |f^{(4)}(1)| = 3,$$

we get $K_4 = 24$. Then the question becomes

$$\frac{24}{180n^4} \le 10^{-6}.$$

Therefore we need $n \ge 19.1$. Since n should be an even integer, this means $n \ge 20$.

We may carry out the similar estimation for the midpoint and trapezoidal rules. We find $K_2 = |f''(0)| = 2$, so that the estimations become

$$\frac{2}{24n^2} \le 10^{-6}, \quad \frac{2}{12n^2} \le 10^{-6}.$$

The answers are respectively $n \ge 289$ and $n \ge 409$.

Exercise 3.3.11. Repeat Exercise 3.3.7 for the Simpson's rule.

Exercise 3.3.12. Repeat Exercise 3.3.8 for the Simpson's rule.

Exercise 3.3.13. If we apply Simpson's rule to $\int_0^1 \frac{dx}{1+x^2}$ to get an approximate value for π accurate up to 10^{-6}, how many partition points do we need?

Exercise 3.3.14. Simpson's 3/8 rule is obtained by using cubic instead of quadratic approximation. Derive the formula for this rule.

3.4 Indefinite Integral

3.4.1 Fundamental Theorem of Calculus

The Newton-Leibniz formula is derived from $A'(x) = f(x)$, where $A(x) = \int_a^x f(t)dt$ is the signed area over $[a, x]$. The equality is summarized below.

Theorem 3.4.1 (Fundamental Theorem of Calculus). *If $f(x)$ is continuous at x, then*

$$\frac{d}{dx}\int_a^x f(t)dt = f(x).$$

Note that the continuity was used critically in our argument for $A'(x) = f(x)$.

Example 3.4.1. Let $f(x)$ be a continuous function. To find the derivative of $\int_a^{x^2} f(t)dt$, we note that the integral is a composition

$$\int_a^{x^2} f(t)dt = A(x^2), \quad A(x) = \int_a^x f(t)dt.$$

By the chain rule and the Fundamental Theorem of Calculus, we have

$$\frac{d}{dx}\int_a^{x^2} f(t)dt = \frac{dA(x^2)}{dx} = A'(x^2)2x = 2xf(x^2).$$

The Fundamental Theorem also implies the following derivatives

$$\begin{aligned} \frac{d}{dx}\int_a^x f(t^2)dt &= f(x^2), \\ \frac{d}{dx}\int_a^x f(t)^2dt &= f(x)^2, \\ \frac{d}{dx}\left(\int_a^x f(t)dt\right)^2 &= 2f(x)\int_a^x f(t)dt. \end{aligned}$$

Example 3.4.2. The function $f(x) = \int_0^x e^{t^2} dt$ cannot be expressed as combinations of the usual elementary functions. Still, we know $f'(x) = e^{x^2}$. We also have

$$\frac{d}{dx}\int_x^{x^2} e^{t^2} dt = \frac{d}{dx}\left(\int_0^{x^2} e^{t^2} dt - \int_0^x e^{t^2} dt\right) = 2xe^{x^4} - e^{x^2}.$$

Example 3.4.3. For the sign function

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

we have

$$A(x) = \int_0^x \text{sign}(x)dx = |x|, \quad A'(x) = \begin{cases} 1, & \text{if } x > 0, \\ \text{no}, & \text{if } x = 0, \\ -1, & \text{if } x < 0. \end{cases}$$

We note that $A(x)$ is not differentiable at 0, exactly the place where the sign function is not continuous. The example shows that the continuity assumption cannot be dropped from the Fundamental Theorem.

Example 3.4.4. The *sine integral function* is

$$\text{Si}(x) = \int_0^x \frac{\sin t}{t} dt.$$

Since the integrand can be made continuous by assigning value 1 at 0, we know

$$\text{Si}'(x) = \begin{cases} \dfrac{\sin x}{x}, & \text{if } x \neq 0, \\ 1, & \text{if } x = 0. \end{cases}$$

Therefore the function $\text{Si}(x)$ is strictly increasing on the following intervals

$$\dots,\ [-5\pi, -4\pi],\ [-3\pi, -2\pi],\ [-\pi, \pi],\ [2\pi, 3\pi],\ [4\pi, 5\pi],\ \dots,$$

and is strictly decreasing on the following intervals

$$\dots,\ [-4\pi, -3\pi],\ [-2\pi, -\pi],\ [\pi, 2\pi],\ [3\pi, 4\pi],\ \dots.$$

This implies that $\text{Si}(x)$ has local maxima at $\dots, -6\pi, -4\pi, -2\pi, \pi, 3\pi, 5\pi, \dots$, and has local minima at $\dots, -5\pi, -3\pi, -\pi, 2\pi, 4\pi, 6\pi, \dots$. Moreover, we can also calculate the second order derivative

$$\text{Si}''(x) = \begin{cases} \dfrac{x\cos x - \sin x}{x^2}, & \text{if } x \neq 0, \\ 0, & \text{if } x = 0, \end{cases}$$

and find the convexity property of $\text{Si}(x)$.

Example 3.4.5. Suppose f is a continuous function satisfying

$$2\int_0^x tf(t)dt = x\int_0^x f(t)dt.$$

Then by taking derivative on both sides, we get

$$2xf(x) = xf(x) + \int_0^x f(t)dt.$$

Let $F(x) = \int_0^x f(t)dt$. Then the equation is the same as $xF'(x) = F(x)$. This implies

$$\left(\frac{F(x)}{x}\right)' = \frac{xF'(x) - F(x)}{x^2} = 0.$$

Therefore $\int_0^x f(t)dt = F(x) = Cx$ for a constant C, and we further find that $f(x) = C$ is a constant.

Exercise 3.4.1. Find the derivative of function.

1. $\int_0^x t^3 dt$.
2. $\int_0^{x^2} t^3 dt$.
3. $\int_0^{x^3} t^2 dt$.
4. $\int_{x^2}^{x^3} t^2 dt$.

Exercise 3.4.2. Find the derivative of function.

1. $\int_1^x \frac{dt}{1+t^3}$.
2. $\int_1^{x^2} \log(1+t^2)dt$.
3. $\int_x^\pi \cos t^2 dt$.
4. $\int_{x^2}^1 \sqrt{1+\sqrt{t}}dt$.
5. $\int_{\tan x}^\pi \arctan t dt$.
6. $\int_{\tan x}^{\cot x} (1+t^2)^{\frac{3}{2}} dt$.

Exercise 3.4.3. Let $f(x)$ be a continuous function. Find the derivative.

1. $\int_{x^2}^b f(t)dt$.
2. $\int_x^{x^2} f(t)dt$.
3. $\int_a^x f(t^2)dt$.
4. $\int_a^x e^{f(t)}dt$.
5. $\int_x^b f(\sin t)dt$.
6. $\int_{\sin x}^{\cos x} f(t)dt$.
7. $\int_0^{f(x)} f(t)dt$.
8. $\int_0^{f^{-1}(x)} f(t)dt$.

In the 7-th problem, f is differentiable. In the 8-th problem, f is invertible and differentiable.

Exercise 3.4.4. Study the monotone and convex properties, including the extrema and the points of inflection.

1. $\displaystyle\int_0^x \frac{dt}{1+t+t^2}$.
2. $\displaystyle\int_0^x \sin\frac{\pi t^2}{2}dt$.
3. $\displaystyle\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$.

Exercise 3.4.5. Find continuous functions $f(x)$ satisfying the equality.

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 $(-\infty,+\infty)$.
4. $\displaystyle\int_0^x f(t)dt = (2x-1)e^{2x} + \int_0^x e^{-t}f(t)dt$ on $(-\infty,+\infty)$.

Exercise 3.4.6. Find the limit.

1. $\displaystyle \lim_{x\to 0}\frac{1}{x}\int_0^x \frac{\sin t}{t}dt$.
2. $\displaystyle \lim_{x\to 0}\frac{1}{x^3}\int_0^x \sin t^2 dt$.
3. $\displaystyle \lim_{x\to +\infty}\frac{\left(\int_0^x e^{t^2}dt\right)^2}{\int_0^x e^{2t^2}dt}$.

Exercise 3.4.7. Prove that for a positive continuous function $f(x)$ on $(0,+\infty)$, the function

$$g(x) = \frac{\left(\int_0^x tf(t)dt\right)^2}{\int_0^x f(t)dt}$$

is strictly increasing on $(0,+\infty)$.

Exercise 3.4.8. Discuss where $f(x)$ is not continuous and where $\displaystyle\int_0^x f(t)dt$ is not differentiable.

1. $f(x) = \begin{cases} x, & \text{if } x\neq 0, \\ 1, & \text{if } x = 0, \end{cases}$.
2. $f(x) = \begin{cases} x, & \text{if } x > 0, \\ 1, & \text{if } x \le 0, \end{cases}$.
3. $f(x) = \begin{cases} \left(x^2\sin\dfrac{1}{x}\right)', & \text{if } x\neq 0, \\ 0, & \text{if } x = 0, \end{cases}$.

3.4.2 Indefinite Integral

A function F is an *antiderivative* of f if $F' = f$. By Theorem 2.4.3, the antiderivative is unique up to adding constants. Therefore we denote all the antiderivatives of f

by

$$\int f(x)dx = F(x) + C, \text{ for some } F(x) \text{ satisfying } F'(x) = f(x).$$

This is the *indefinite integral* of f.

The Newton-Leibniz formula says that the definite integral of a continuous function can be calculated from the antiderivative

$$\int_a^b f(x)dx = F(b) - F(a).$$

The Fundamental Theorem of Calculus says that the signed area gives *one* antiderivative, and can be used as $F(x)$ above

$$\int f(x)dx = \int_a^x f(t)dt + C.$$

Example 3.4.6. By

$$(x^{p+1})' = (p+1)x^p, \quad (\log|x|)' = \frac{1}{x}, \quad (e^x)' = e^x,$$

we get

$$\int x^p dx = \begin{cases} \dfrac{x^{p+1}}{p+1} + C, & \text{for } p \neq -1, \\ \log|x| + C, & \text{for } p = -1; \end{cases} \quad \int e^x dx = e^x + C.$$

More generally, we have

$$\int (ax+b)^p dx \begin{cases} \dfrac{(ax+b)^{p+1}}{(p+1)a} + C, & \text{for } p \neq -1, \\ \dfrac{1}{a}\log|ax+b| + C, & \text{for } p = -1; \end{cases} \quad \int a^x dx = \frac{a^x}{\log a} + C.$$

Example 3.4.7. The antiderivative of the logarithmic function is more complicated

$$\int \log|x|dx = x\log|x| - x + C.$$

The equality can be verified by taking the derivative

$$(x\log|x| - x)' = \log|x| + x\frac{1}{x} - 1 = \log|x|.$$

Example 3.5.9 gives the systematic way of deriving $\displaystyle\int \log|x|dx$.

Example 3.4.8. The derivatives of the trigonometric functions give us

$$\int \cos x dx = \sin x + C, \qquad \int \sin x dx = -\cos x + C,$$
$$\int \sec^2 x dx = \tan x + C, \qquad \int \sec x \tan x dx = \sec x + C.$$

The antiderivatives of $\tan x$ and $\sec x$ are more complicated, and are given in Example 3.5.27.

Example 3.4.9. The derivative of the inverse sine function gives

$$\int \frac{dx}{\sqrt{1-x^2}} = \arcsin x + C, \quad \int \frac{dx}{\sqrt{a^2-x^2}} = \arcsin \frac{x}{a} + C, \quad a > 0.$$

Similarly, the derivative of the inverse tangent function gives

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

The similar integrals $\int \frac{dx}{x^2-a^2}$ and $\int \frac{dx}{\sqrt{x^2+a}}$ are given in Exercise 3.4.10 and Examples 3.5.2, 3.5.31, 3.5.32.

Exercise 3.4.9. Compute the integrals.

1. $\int \sqrt[4]{1-x} dx$.
2. $\int \frac{1}{\sqrt[3]{2x+1}} dx$.
3. $\int a^x dx$.
4. $\int \csc^2 x dx$.
5. $\int \csc x \cot x dx$.
6. $\int \frac{dx}{\cos^2 x}$.

Exercise 3.4.10. Verify the antiderivatives.

1. $\int \frac{\log |x|}{x} dx = \frac{1}{2}(\log |x|)^2 + C$.
2. $\int e^{ax} \cos bx dx = \frac{e^{ax}}{a^2+b^2}(a \cos bx + b \sin bx) + C$.
3. $\int \cos(ax+b) dx = \frac{1}{a} \sin(ax+b) + C$.
4. $\int \sqrt{a^2-x^2} dx = \frac{a^2}{2} \arcsin \frac{x}{a} + \frac{1}{2} x \sqrt{a^2-x^2} + C$, $a > 0$.
5. $\int \frac{dx}{x^2-a^2} = \frac{1}{2a} \log \left| \frac{x-a}{x+a} \right| + C$.
6. $\int \frac{dx}{\sqrt{x^2+a}} = \log \left| x + \sqrt{x^2+a} \right| + C$.

7. $\displaystyle\int \sqrt{x^2+a}dx = \frac{1}{2}x\sqrt{x^2+a} + \frac{a}{2}\log\left|x+\sqrt{x^2+a}\right| + C.$

Exercise 3.4.11. If $\int f(x)dx = F(x) + C$, then what is $\int f(ax+b)dx$? Apply your conclusion to compute the integrals.

1. $\displaystyle\int \log|a+bx|dx.$
2. $\displaystyle\int \sin(ax+b)dx.$
3. $\displaystyle\int \sec^2(3x-1)dx.$
4. $\displaystyle\int \frac{dx}{\sqrt{1-(x-1)^2}}.$
5. $\displaystyle\int \frac{dx}{\sqrt{x(1-x)}}.$
6. $\displaystyle\int \frac{dx}{x^2+2x+2}.$

Exercise 3.4.12. Find the antiderivative of $x(ax^2+b)^p$. Then compute the integrals.

1. $\displaystyle\int x\sqrt{x^2+3}dx.$
2. $\displaystyle\int \frac{xdx}{x^2+1}.$
3. $\displaystyle\int \frac{xdx}{\sqrt{4-x^2}}.$

Exercise 3.4.13. Compute the integrals.

1. $\displaystyle\int x\sin(ax^2+b)dx.$
2. $\displaystyle\int xe^{ax^2+b}dx.$
3. $\displaystyle\int x^2(ax^3+b)^p dx.$

One should not just mindlessly compute the antiderivative. Sometimes we need to consider the meaning of antiderivative and question whether the answer makes sense.

Example 3.4.10. Without much thinking, we may write

$$\int |x|dx = \begin{cases} \dfrac{1}{2}x^2 + C, & \text{if } x \ge 0, \\ -\dfrac{1}{2}x^2 + C, & \text{if } x < 0. \end{cases}$$

However, the constant C in the two cases cannot be independently chosen because the antiderivative must be differentiable and is therefore continuous at 0. The more sensible answer is

$$\int |x|dx = \begin{cases} \dfrac{1}{2}x^2, & \text{if } x \ge 0 \\ -\dfrac{1}{2}x^2, & \text{if } x < 0 \end{cases} + C.$$

In other words, the constant C is two cases must be equal.

For another example, instead of

$$f(x) = \begin{cases} e^x, & \text{if } x \ge 0, \\ 1, & \text{if } x < 0, \end{cases} \qquad \int f(x)dx = \begin{cases} e^x + C, & \text{if } x \ge 0, \\ x + C, & \text{if } x < 0, \end{cases}$$

we should have

$$\int f(x)dx = \begin{cases} e^x, & \text{if } x \ge 0 \\ x+1, & \text{if } x < 0 \end{cases} + C.$$

Exercise 3.4.14. Compute indefinite integral.

1. $\begin{cases} x^2, & \text{if } x \le 0, \\ \sin x, & \text{if } x > 0. \end{cases}$

2. $\begin{cases} 1 - x^2, & \text{if } |x| \le 1, \\ \sin(1 - |x|), & \text{if } |x| > 1. \end{cases}$

Exercise 3.4.15. Does the sign function

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

have antiderivative? Does the function have definite integral? What do you learn from the example?

3.5 Properties of Integration

Computationally, integration is the reverse of differentiation. Therefore properties of differentiation have corresponding properties of integration.

3.5.1 Linear Property

Suppose $F(x)$ and $G(x)$ are antiderivatives of $f(x)$ and $g(x)$. Then the linear property of the derivative

$$(F(x) + G(x))' = F'(x) + G'(x) = f(x) + g(x), \quad (cF(x))' = cF'(x) = cf(x),$$

implies the linear property of the antiderivative

$$\int (f(x) + g(x))dx = \int f(x)dx + \int g(x)dx, \quad \int cf(x)dx = c\int f(x)dx.$$

By the Newton-Leibniz formula, we get the linear property for the definite integral

$$\begin{aligned} \int_a^b (f(x) + g(x))dx &= (F(b) + G(b)) - (F(a) + G(a)) \\ &= (F(b) - F(a)) + (G(b) - G(a)) = \int_a^b f(x)dx + \int_a^b g(x)dx, \end{aligned}$$

$$\int_a^b cf(x)dx = cF(b) - cF(a) = c(F(b) - F(a)) = c\int_a^b f(x)dx.$$

Example 3.5.1. We have

$$\begin{aligned} \int x(1+x)^2 dx &= \int (x + 2x^2 + x^3)dx = \int xdx + 2\int x^2dx + \int x^3dx \\ &= \frac{1}{2}x^2 + \frac{2}{3}x^3 + \frac{1}{4}x^4 + C. \end{aligned}$$

This furthers gives the definite integral

$$\int_0^1 x(1+x)^2dx = \frac{1}{2}(1^2-0^2)+\frac{2}{3}(1^3-0^3)+\frac{1}{4}(1^4-0^4)=\frac{17}{12}.$$

On the other hand, it would be very complicated to compute $\int x(x+1)^{10}dx$ by the binomial expansion of $(1+x)^{10}$. The following is much simpler

$$\begin{aligned}\int x(x+1)^{10}dx &= \int((x+1)-1)(x+1)^{10}dx = \int(x+1)^{11}dx - \int(x+1)^{10}dx \\ &= \frac{1}{12}(x+1)^{12}-\frac{1}{11}(x+1)^{11}+C = \frac{1}{12\cdot 11}(11x-1)(x+1)^{11}+C.\end{aligned}$$

Exercise 3.5.1. Compute the integrals.

1. $\int x\sqrt{x+1}dx$.
2. $\int x(ax+b)^p dx$.
3. $\int x^2(ax+b)^p dx$.
4. $\int (x-1)(x+1)^{\frac{4}{3}}dx$.
5. $\int (x-1)(x+1)^p dx$.
6. $\int \frac{x}{(x+1)^{10}}dx$.
7. $\int \frac{x^2-x+1}{(x+1)^{10}}dx$.
8. $\int \frac{x-1}{\sqrt{x}}dx$.
9. $\int \left(\frac{x-1}{x}\right)^2 dx$.
10. $\int \left(\frac{x-1}{x^2}\right)^2 dx$.
11. $\int \left(\frac{x-1}{x+1}\right)^2 dx$.
12. $\int \frac{(x-1)^2}{(x+1)^4}dx$.

Exercise 3.5.2. Find A, B satisfying

$$\frac{ax+b}{cx+d} = A + \frac{B}{cx+d}.$$

Then compute the antiderivatives of $\frac{ax+b}{cx+d}$ and $\left(\frac{ax+b}{cx+d}\right)^2$.

Exercise 3.5.3. Compute the integrals.

1. $\int (e^x - e^{-x})^2 dx$.
2. $\int (2^x+3^x)^2 dx$.
3. $\int \frac{2^{x+1}-3^{x-1}}{6^x}dx$.

Example 3.5.2. To find the antiderivative of $\frac{1}{x^2-a^2}$, we use

$$\frac{1}{x^2-a^2} = \frac{1}{(x-a)(x+a)} = \frac{1}{2a}\left(\frac{1}{x-a}-\frac{1}{x+a}\right)$$

to get

$$\int \frac{dx}{x^2-a^2} = \frac{1}{2a}\log|x-a| - \frac{1}{2a}\log|x+a| + C = \frac{1}{2a}\log\left|\frac{x-a}{x+a}\right| + C.$$

This furthers gives the definite integral

$$\int_2^3 \frac{1}{x^2-1}dx = \frac{1}{2}\log\left|\frac{3-1}{3+1}\right| - \frac{1}{2}\log\left|\frac{2-1}{2+1}\right| = \frac{1}{2}\log\frac{3}{2}.$$

As noted in Example 3.1.4, we should not blindly use the Newton-Leibniz formula in competing the definite integral. For example, we cannot get

$$\int_0^2 \frac{1}{x^2-1}dx = \frac{1}{2}\log\left|\frac{2-1}{2+1}\right| - \frac{1}{2}\log\left|\frac{0-1}{0+1}\right| = -\frac{1}{2}\log 3,$$

because the interval $[0,2]$ contains 1, where the integrand approaches infinity.

Example 3.5.3. The idea in Example 3.5.2 can be extended

$$\begin{aligned}
\int \frac{dx}{x(x+1)(x+2)} &= \int \frac{1}{x}\left(\frac{1}{x+1} - \frac{1}{x+2}\right)dx = \int\left(\frac{1}{x(x+1)} - \frac{1}{x(x+2)}\right)dx \\
&= \int\left[\left(\frac{1}{x} - \frac{1}{x+1}\right) - \frac{1}{2}\left(\frac{1}{x} - \frac{1}{x+2}\right)\right]dx \\
&= \frac{1}{2}\log|x| - \log|x+1| + \frac{1}{2}\log|x+2| + C \\
&= \frac{1}{2}\log\left|\frac{x(x+2)}{(x+1)^2}\right| + C, \\
\int \frac{dx}{(x^2-1)^2} &= \frac{1}{4}\int\left(\frac{1}{x-1} - \frac{1}{x+1}\right)^2 dx \\
&= \frac{1}{4}\int\left(\frac{1}{(x-1)^2} + \frac{1}{(x+1)^2} - \frac{2}{(x+1)(x-1)}\right)dx \\
&= \frac{1}{4}\int\left[\frac{1}{(x-1)^2} + \frac{1}{(x+1)^2} - \left(\frac{1}{x-1} - \frac{1}{x+1}\right)\right]dx \\
&= \frac{1}{4}\left(-\frac{1}{x-1} - \frac{1}{x+1} + \log\left|\frac{x+1}{x-1}\right|\right) + C \\
&= -\frac{x}{2(x^2-1)} + \frac{1}{4}\log\left|\frac{x+1}{x-1}\right| + C.
\end{aligned}$$

Exercise 3.5.4. Compute the integrals.

1. $\int \frac{xdx}{x^2-1}$.
2. $\int \frac{x^2dx}{x^2-1}$.
3. $\int \frac{dx}{x^2+3x+2}$.
4. $\int \frac{(2x+1)dx}{x^2+3x+2}$.
5. $\int_0^1 \frac{(2x+1)dx}{x^2+3x+2}$.
6. $\int \frac{x^2dx}{x^2+3x+2}$.
7. $\int \frac{x^2dx}{x^2+1}$.
8. $\int \frac{dx}{(x^2+1)(x^2+4)}$.
9. $\int \frac{x^2dx}{(x^2+1)(x^2+4)}$.

Exercise 3.5.5. Compute the integrals.

1. $\displaystyle\int \frac{dx}{(x+a)(x+b)}$.
2. $\displaystyle\int \frac{xdx}{(x+a)(x+b)}$.
3. $\displaystyle\int \frac{dx}{(x+a)(x+b)(x+c)}$.

Example 3.5.4. By trigonometric formula, we have

$$\begin{aligned}\int_0^{\pi} \sin^2 xdx &= \int_0^{\pi} \frac{1}{2}(1-\cos 2x)dx = \frac{1}{2}\int_0^{\pi} xdx - \frac{1}{2}\int_0^{\pi} \cos 2xdx \\ &= \frac{1}{4}(\pi^2 - 0^2) - \frac{1}{4}(\sin 2\pi - \sin 0) = \frac{1}{4}\pi^2.\end{aligned}$$

Similar idea gives

$$\begin{aligned}\int \sin^2 xdx &= \frac{1}{2}\int (1-\cos 2x)dx = \frac{1}{2}x - \frac{1}{4}\sin 2x + C, \\ \int \sin x \cos 2xdx &= \frac{1}{2}\int (\sin 3x - \sin x)dx = -\frac{1}{6}\cos 3x + \frac{1}{2}\cos x + C, \\ \int \tan^2 xdx &= \int (\sec^2 x - 1)dx = \tan x - x + C.\end{aligned}$$

Exercise 3.5.6. Compute the integrals.

1. $\displaystyle\int \cos x \sin xdx$.
2. $\displaystyle\int_0^{\pi} \sin^2 x \cos xdx$.
3. $\displaystyle\int \cos^2 xdx$.
4. $\displaystyle\int \sin^3 xdx$.
5. $\displaystyle\int \cos x \sin 2xdx$.
6. $\displaystyle\int \cot^2 xdx$.
7. $\displaystyle\int_0^{\frac{\pi}{2}} \sin x \cos 2xdx$.
8. $\displaystyle\int_0^{\pi} |\sin x \cos 2x|dx$.
9. $\displaystyle\int_0^{\pi} |\sin x - \cos x|dx$.

Example 3.5.5. For $f \ge g$ and $a \le b$, by the inequality (3.1.5) and the linearity of definite integral, we have

$$\int_a^b f(x)dx - \int_a^b g(x)dx = \int_a^b (f(x)-g(x))dx \ge 0.$$

Therefore we have

$$f \ge g \implies \int_a^b f(x)dx \ge \int_a^b g(x)dx, \text{ for } a < b.$$

The inequality corresponds to Theorem 2.3.3 that uses the derivatives to compare functions. However, it is more direct to get the inequality by using the non-negativity of area.

If we apply the inequality to $-|f| \le f \le |f|$, then we get

$$\int_a^b |f(x)|dx \ge \left|\int_a^b f(x)dx\right|, \text{ for } a < b.$$

Example 3.5.6 (Average). The *average* of a function f on $[a,b]$ is $\dfrac{1}{b-a}\displaystyle\int_a^b f(x)dx$. If $m \le f \le M$ on $[a,b]$, then

$$m(b-a) = \int_a^b m dx \le \int_a^b f(x)dx \le \int_a^b M dx = M(b-a).$$

This implies that the average of f lies between m and M, which is consistent with our intuition.

For continuous f, we may take m and M to be the minimum and maximum of f on $[a,b]$. By the Intermediate Value Theorem, any value between m and M can be reached by the function. Therefore the average

$$\frac{1}{b-a}\int_a^b f(x)dx = f(c), \text{ for some } c \in (a,b).$$

This conclusion is the *Integral Mean Value Theorem*.

Example 3.5.7. Consider the function $F(x) = \displaystyle\int_0^x \frac{\sin t^2}{t}dt$. The 4-th order Taylor expansion $T(x) = x - \dfrac{x^3}{6}$ of $\sin x$ means that, for any $\epsilon > 0$, there is $\delta > 0$, such that

$$|x| < \delta \implies |\sin x - T(x)| \le \epsilon |x|^4.$$

Then for t between 0 and x, we have

$$|x| < \sqrt{\delta} \implies |t^2| < \delta \implies |\sin t^2 - T(t^2)| \le \epsilon |t|^8,$$

so that

$$|x| < \sqrt{\delta} \implies \left|\frac{\sin t^2}{t} - t - \frac{t^5}{6}\right| = \left|\frac{\sin t^2}{t} - \frac{1}{t}T(t^2)\right| \le \epsilon |t|^7.$$

Therefore

$$|x| < \sqrt{\delta} \implies \left|F(x) - \frac{x^2}{2} - \frac{x^6}{36}\right| = \left|\int_0^x \left(\frac{\sin t^2}{t} - t - \frac{t^5}{6}\right) dt\right| \le \epsilon \int_0^x |t|^7 dt = \frac{\epsilon}{8}|x|^8.$$

This means exactly the 7-th order approximation of $F(x)$

$$F(x) = \frac{1}{2}x^2 + \frac{1}{36}x^6 + o(x^8).$$

Example 3.5.8. Suppose $f(x)$ has second order derivative on $[a,b]$. We may take the linear approximation at the middle point $c = \dfrac{a+b}{2}$. By the Lagrange form of the remainder (Theorem 2.7.1), we get

$$f(x) = f(c) + f'(c)(x-c) + \frac{f''(\bar{x})}{2}(x-c)^2,$$

where $\bar{x}$ depends on x and lies between x and c. Then we have

$$\left|\int_a^b (f(x) - f(c) - f'(c)(x-c))dx\right| = \left|\int_a^b \frac{f''(\bar{x})}{2}(x-c)^2 dx\right|$$

By the linear property of the integral, we have the left side

$$\begin{aligned}\int_a^b (f(x) - f(c) - f'(c)(x-c))dx &= \int_a^b f(x)dx - f(c)\int_a^b dx - f'(c)\int_a^b (x-c)dx \\ &= \int_a^b f(x)dx - f(c)(b-a).\end{aligned}$$

Let K_2 be the bound for the second order derivative. In other words, $|f''| \le K_2$ on $[a,b]$. Then the right side

$$\left|\int_a^b \frac{f''(\bar{x})}{2}(x-c)^2 dx\right| \le \int_a^b \frac{|f''(\bar{x})|}{2}(x-c)^2 dx \le \frac{K_2}{2}\int_a^b (x-c)^2 dx = \frac{K_2}{24}(b-a)^3.$$

We conclude the inequality

$$\left|\int_a^b f(x)dx - f(c)(b-a)\right| \le \frac{K_2}{24}(b-a)^3.$$

Exercise 3.5.7. Show that the integration of n-th order approximation is $(n+1)$-st order approximation. Specifically, find high order approximation of function at 0.

1. $\displaystyle\int_0^x \frac{\cos t - 1}{t}dt$, order 5.
2. $\displaystyle\int_0^{\sqrt{x}} \frac{\sin t - t}{t^2}dt$, order 4.
3. $\displaystyle\int_{-x^2}^0 \frac{e^t - 1}{t}dt$, order 5.
4. $\displaystyle\int_{-x}^x \frac{\log(1+t)}{t}dt$, order 7.

Exercise 3.5.8. Derive an estimation for $\left|\int_a^b f(x)dx - f(a)(b-a)\right|$ in terms of the bound K_1 of f on $[a,b]$.

Exercise 3.5.9. Apply the estimation in Example 3.5.8 to each interval of a partition and derive the error formula for the midpoint rule in Theorem 3.3.1.

3.5.2 Integration by Parts

The Leibniz rule says that, if $F(x)$ and $G(x)$ are antiderivatives of $f(x)$ and $g(x)$, then

$$(F(x)G(x))' = F'(x)G(x) + F(x)G'(x) = f(x)G(x) + F(x)g(x).$$

In other words, $F(x)G(x)$ is an antiderivative of $f(x)G(x)+F(x)g(x)$, or

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

If we use the differential notation

$$dF(x)=F'(x)dx=f(x)dx,\quad dG(x)=G'(x)dx=g(x)dx,$$

then the equality becomes

$$\int F(x)dG(x)=F(x)G(x)-\int G(x)dF(x).$$

The equality can be used in the following way. To compute an integral $\int h(x)dx$, we separate the integrand into a product $h(x)=F(x)g(x)$ of two parts and integrate the second part to get $\int g(x)dx=G(x)+C$. Then $\int h(x)dx=\int F(x)dG(x)$, which by the equality above is converted into the computation of another integral $\int G(x)dF(x)$ that exchanges $F(x)$ and $G(x)$. This method of computing the integral is called the *integration by parts*.

By Newton-Leibniz formula, the integration by parts for indefinite integral implies the method for definite integral

$$\int_a^b F(x)dG(x)=F(b)G(b)-F(a)G(a)-\int_a^b G(x)dF(x).$$

The use of Newton-Leibniz formula requires that $f=F'$ and $g=G'$ to be continuous. Then it is not hard to extend the equality to the case that F and G are continuous on $[a,b]$ and have continuous derivatives at all but finitely many points on $[a,b]$.

Example 3.5.9. The antiderivative of the logarithmic function in Example 3.4.7 may be derived by using the integration by parts (taking $F(x)=\log|x|$ and $G(x)=x$)

$$\begin{aligned}\int\log|x|dx&=x\log|x|-\int xd\log|x|=x\log|x|-\int x(\log|x|)'dx\\&=x\log|x|-\int dx=x\log|x|-x+C.\end{aligned}$$

The antiderivative $x\log|x|-x$ just obtained can be further used

$$\begin{aligned}\int x\log|x|dx&=\int xd(x\log|x|-x)=x(x\log|x|-x)-\int(x\log|x|-x)dx\\&=x^2\log x-x^2-\int x\log|x|dx+\frac{1}{2}x^2.\end{aligned}$$

Solving the equation, we get

$$\int x\log|x|dx = \frac{1}{2}x^2\log x - \frac{1}{4}x^2 + C.$$

The following is an alternative way of applying the integration by parts to the same integral

$$\begin{aligned}\int x\log|x|dx &= \frac{1}{2}\int \log|x|d(x^2) = \frac{1}{2}x^2 - \frac{1}{2}\int x^2 d(\log|x|)\\ &= \frac{1}{2}x^2\log|x| - \frac{1}{2}\int x^2\frac{1}{x}dx = \frac{1}{2}x^2\log|x| - \frac{1}{4}x^2 + C.\end{aligned}$$

We may compute $\int x^p \log x dx$ by the similar idea.

Example 3.5.10. The integral in Example 3.5.1 can also be computed by using the integration by parts

$$\begin{aligned}\int x(x+1)^{10}dx &= \frac{1}{11}\int xd(x+1)^{11} && \text{(integrate } (x+1)^{10} \text{ part)}\\ &= \frac{1}{11}x(x+1)^{11} - \frac{1}{11}\int (x+1)^{11}dx && \text{(exchange two parts)}\\ &= \frac{1}{11}x(x+1)^{11} - \frac{1}{12\cdot 11}(x+1)^{12} + C.\end{aligned}$$

Example 3.5.11. Using integration by parts, we have

$$\begin{aligned}\int x^2e^{-x}dx &= -\int x^2de^{-x} = -x^2e^{-x} + 2\int xe^{-x}dx\\ &= -x^2e^{-x} - 2\int xde^{-x} = -x^2e^{-x} - 2xe^{-x} + 2\int e^{-x}dx\\ &= -(x^2+2x+2)e^{-x} + C.\end{aligned}$$

In general, we have the recursive formula

$$\int x^n a^x dx = \frac{1}{\log a}x^n a^x - \frac{n}{\log a}\int x^{n-1}a^x dx.$$

Exercise 3.5.10. Compute the integral.

1. $\int x(ax+b)^p dx$.
2. $\int x^2(ax+b)^p dx$.
3. $\int (x-1)(x+1)^p dx$.

Exercise 3.5.11. Compute the integral.

1. $\int (x^2-1)a^x dx$.
2. $\int (x+a^x)^2 dx$.
3. $\int \frac{xe^x dx}{(x+1)^2}$.

Exercise 3.5.12. Compute the integral.

1. $\int x^2 \log |x| dx$.
2. $\int x^p \log x dx$.
3. $\int (\log |x|)^2 dx$.
4. $\int x^p (\log |x|)^2 dx$.
5. $\int x \log(x+1) dx$.
6. $\int x \log \frac{1+x}{1-x} dx$.

Exercise 3.5.13. Derive the recursive formula for $\int (\log |x|)^n dx$. How about $\int x^p (\log x)^n dx$?

Exercise 3.5.14. Compute $\int_0^1 x^n a^x dx$.

Exercise 3.5.15. For natural numbers m, n, show that $\int_0^1 x^m (1-x)^n dx = \frac{m!n!}{(m+n+1)!}$.

Exercise 3.5.16. Compute the integral.

1. $\int \log(\sqrt{x+a}+\sqrt{x-a}) dx$.
2. $\int \log(\sqrt{a+x}-\sqrt{a-x}) dx$.
3. $\int \left(\frac{\log(x+a)}{x+b}+\frac{\log(x+b)}{x+a}\right) dx$.
4. $\int \frac{x}{\sqrt{1+x^2}} \log(x+\sqrt{1+x^2}) dx$.
5. $\int \frac{x \log(x+\sqrt{1+x^2})}{\sqrt{1+x^2}} dx$.
6. $\int \left(\log(x+\sqrt{1+x^2})\right)^2 dx$.

Exercise 3.5.17. Let f have second order derivative on $[a, b]$.

1. Show that for any constants A and B, we have
$$\begin{aligned}\int_a^b f(x)dx = &\left((x+A)f(x) - \frac{1}{2}((x+A)^2+B)f'(x)\right)_a^b \\ &+ \frac{1}{2}\int_a^b ((x+A)^2+B)f''(x)dx.\end{aligned}$$
2. By choosing suitable A and B in the first part, show that
$$\int_a^b f(x)dx = \frac{f(a)+f(b)}{2}(b-a) + \int_a^b ((x+A)^2+2B)f''(x)dx,$$
and
$$\int_a^b |(x+A)^2+2B| dx = \frac{1}{12}(b-a)^3.$$
3. Use the second part to derive the error formula for the trapezoidal rule in Theorem 3.3.1.

Example 3.5.12. Using the integration by parts, we have

$$\int x\cos x dx = \int x d\sin x = x\sin x - \int \sin x dx = x\sin x + \cos x + C.$$

The idea can be extended to product of x^n, $\sin ax$ and $\cos bx$ for various a and b

$$\begin{aligned}\int x\sin x\sin 2x dx &= \frac{1}{2}\int x(\cos 3x - \cos x)dx = \frac{1}{2}\int x d\left(\frac{1}{3}\sin 3x - \sin x\right)\\ &= \frac{1}{6}x(\sin 3x - 3\sin x) - \frac{1}{6}\int(\sin 3x - 3\sin x)dx\\ &= \frac{1}{6}x\sin 3x - \frac{1}{2}x\sin x + \frac{1}{18}\cos 3x - \frac{1}{2}\cos x + C.\end{aligned}$$

An example of the definite integral is

$$\begin{aligned}\int_0^{\frac{\pi}{2}} x^2\sin x dx &= -\int_0^{\frac{\pi}{2}} x^2 d\cos x = -\left(\frac{\pi}{2}\right)^2\cos\frac{\pi}{2} + 0^2\cos 0 + \int_0^{\frac{\pi}{2}} 2x\cos x dx\\ &= 2\int_0^{\frac{\pi}{2}} x d\sin x = 2\frac{\pi}{2}\sin\frac{\pi}{2} - 2\cdot 0\sin 0 - 2\int_0^{\frac{\pi}{2}}\sin x dx\\ &= \pi + 2\cos\frac{\pi}{2} - 2\cos 0 = \pi - 2.\end{aligned}$$

Example 3.5.13. Let

$$I_0 = \int e^{ax}\cos bx dx, \quad J_0 = \int e^{ax}\sin bx dx.$$

We have

$$\begin{aligned}I_0 &= a^{-1}\int \cos bx de^{ax} = a^{-1}e^{ax}\cos bx - a^{-1}\int e^{ax} d\cos bx\\ &= a^{-1}e^{ax}\cos bx + a^{-1}bJ_0,\\ J_0 &= a^{-1}\int \sin bx de^{ax} = a^{-1}e^{ax}\sin bx - a^{-1}\int e^{ax} d\sin bx\\ &= a^{-1}e^{ax}\sin bx - a^{-1}bI_0.\end{aligned}$$

Solving the system for I_0 and J_0, we get

$$\begin{aligned}\int e^{ax}\cos bx dx &= \frac{e^{ax}}{a^2+b^2}(a\cos bx + b\sin bx) + C,\\ \int e^{ax}\sin bx dx &= \frac{e^{ax}}{a^2+b^2}(-b\cos bx + a\sin bx) + C.\end{aligned}$$

Let

$$I_1 = \int xe^x\cos x dx, \quad J_1 = \int xe^x\sin x dx.$$

Using the earlier computation of I_0 and J_0, we have

$$\begin{aligned} I_1 &= \int x\cos x de^x = xe^x\cos x - \int(\cos x - x\sin x)e^x dx \\ &= xe^x\cos x - \frac{1}{2}e^x(\cos x+\sin x) + J_1. \end{aligned}$$

Similarly, we have

$$J_1 = xe^x\sin x - \frac{1}{2}e^x(-\cos x+\sin x) - I_1.$$

Solving the two equations, we get

$$\begin{aligned} \int xe^x\cos x dx &= \frac{1}{2}xe^x(\cos x+\sin x) - e^x\sin x + C, \\ \int xe^x\sin x dx &= \frac{1}{2}xe^x(-\cos x+\sin x) + e^x\cos x + C. \end{aligned}$$

Example 3.5.14. Let

$$I_{m,n} = \int \cos^m x\sin^n x dx.$$

If $n\neq 0$, then we may integrate a copy of $\sin x$ to get

$$\begin{aligned} I_{m,n} &= -\int \cos^m x\sin^{n-1} x d\cos x \\ &= -\cos^{m+1} x\sin^{n-1} x \\ &\quad + \int(-m\cos^{m-1} x\sin^n x + (n-1)\cos^{m+1} x\sin^{n-2} x)\cos x dx \\ &= -\cos^{m+1} x\sin^{n-1} x \\ &\quad + \int(-m\cos^m x\sin^n x + (n-1)\cos^m x(1-\sin^2 x)\sin^{n-2} x)dx \\ &= -\cos^{m+1} x\sin^{n-1} x - (m+n-1)I_{m,n} - (n-1)I_{m,n-2}. \end{aligned}$$

Therefore (the formula can be directly verified for $n=0$)

$$I_{m,n} = -\frac{1}{m+n}\cos^{m+1} x\sin^{n-1} x + \frac{n-1}{m+n}I_{m,n-2}, \quad m+n\neq 0.$$

The formula reduces the power of sine by 2. If we first integrate a copy of $\cos x$, then we get another recursive relation that reduces the power of cosine by 2

$$I_{m,n} = \frac{1}{m+n}\cos^{m-1} x\sin^{n+1} x + \frac{m-1}{m+n}I_{m-2,n}, \quad m+n\neq 0.$$

On the other hand, we can also express $I_{m,n-2}$ and $I_{m-2,n}$ in terms of $I_{m,n}$. After substituting n by $n+2$, we get recursive relations that increase the power by 2

$$\begin{aligned}
I_{m,n} &= \frac{1}{n+1}\cos^{m+1}x\sin^{n+1}x + \frac{m+n+2}{n+1}I_{m,n+2}, \quad n \neq -1, \\
&= -\frac{1}{m+1}\cos^{m+1}x\sin^{n+1}x + \frac{m+n+2}{m+1}I_{m+2,n}, \quad m \neq -1.
\end{aligned}$$

Here is a concrete example of using the recursive relation

$$\begin{aligned}
\int \cos^4 x\sin^6 x dx = I_{4,6} &= -\frac{1}{4+6}\cos^{4+1}x\sin^{6-1}x + \frac{6-1}{4+6}I_{4,6-2} \\
&= -\frac{1}{10}\cos^5 x\sin^5 x + \frac{5}{10}I_{4,4} \\
&= -\frac{1}{10}\cos^5 x\sin^5 x + \frac{5}{10}\left(-\frac{1}{8}\cos^5 x\sin^3 x + \frac{3}{8}I_{4,2}\right) \\
&= -\cos^5 x\left(\frac{1}{10}\sin^5 x + \frac{5}{10\cdot 8}\sin^3 x\right) \\
&\quad + \frac{5\cdot 3}{10\cdot 8}\left(-\frac{1}{6}\cos^5 x\sin x + \frac{1}{6}I_{4,0}\right) \\
&= -\cos^5 x\left(\frac{1}{10}\sin^5 x + \frac{5}{10\cdot 8}\sin^3 x + \frac{5\cdot 3}{10\cdot 8\cdot 6}\sin x\right) \\
&\quad + \frac{5\cdot 3\cdot 1}{10\cdot 8\cdot 6}\left(\frac{1}{4}\cos^3 x\sin x + \frac{3}{4}I_{2,0}\right) \\
&= -\cos^5 x\left(\frac{1}{10}\sin^5 x + \frac{5}{10\cdot 8}\sin^3 x + \frac{5\cdot 3}{10\cdot 8\cdot 6}\sin x\right) \\
&\quad + \frac{5\cdot 3\cdot 1}{10\cdot 8\cdot 6}\frac{1}{4}\cos^3 x\sin x + + \frac{5\cdot 3\cdot 1}{10\cdot 8\cdot 6}\frac{3}{4}\left(\frac{1}{2}\cos x\sin x + \frac{1}{2}I_{0,0}\right) \\
&= -\cos^5 x\left(\frac{1}{10}\sin^5 x + \frac{5}{10\cdot 8}\sin^3 x + \frac{5\cdot 3}{10\cdot 8\cdot 6}\sin x\right) \\
&\quad + \frac{5\cdot 3\cdot 1}{10\cdot 8\cdot 6}\left(\frac{1}{4}\cos^3 x + \frac{3\cdot 1}{4\cdot 2}\cos x\right)\sin x + + \frac{5\cdot 3\cdot 1}{10\cdot 8\cdot 6}\frac{3\cdot 1}{4\cdot 2}x + C.
\end{aligned}$$

Here is another example that requires increasing the power

$$\begin{aligned}
\int \frac{\sin^2 x}{\cos^4 x}dx = I_{-4,2} &= -\frac{1}{-4+1}\cos^{-4+1}x\sin^{2+1}x + \frac{-4+2+2}{-4+1}I_{-4+2,2} \\
&= \frac{1}{3}\frac{\sin^3 x}{\cos^3 x} + C.
\end{aligned}$$

Applying the recursive relation to the definite integral, we have

$$\int_0^{\frac{\pi}{2}} \sin^n x dx = -\frac{1}{n}\cos x\sin^{n-1}x\Big|_0^{\frac{\pi}{2}} + \frac{n-1}{n}\int_0^{\frac{\pi}{2}}\sin^{n-2}x dx = \frac{n-1}{n}\int_0^{\frac{\pi}{2}}\sin^{n-2}x dx.$$

Then we get

$$\int_0^{\frac{\pi}{2}} \sin^n x dx = \frac{n-1}{n}\frac{n-3}{n-2}\cdots\int_0^{\frac{\pi}{2}} \sin^{0 \text{ or } 1} x dx = \begin{cases} \dfrac{(n-1)!!}{n!!}\dfrac{\pi}{2}, & \text{if } n \text{ is even,} \\ \dfrac{(n-1)!!}{n!!}, & \text{if } n \text{ is odd.} \end{cases}$$

Here $\sin^{0 \text{ or } 1} x$ takes power 0 for even n and takes power 1 for odd n. Moreover, we used the *double factorial*

$$n!! = n(n-2)(n-4)\cdots = \begin{cases} 2k(2k-2)(2k-4)\cdots 4\cdot 2, & \text{if } n = 2k, \\ (2k+1)(2k-1)(2k-3)\cdots 3\cdot 1, & \text{if } n = 2k+1. \end{cases}$$

Exercise 3.5.18. Find the recursive relations for $\int x^p \cos ax dx$ and $\int x^p \sin ax dx$. Then compute the integral.

1. $\int x \cos^2 x dx$.
2. $\int x^3 \cos^2 x dx$.
3. $\int x \cos^2 x \sin 2x dx$.
4. $\int x^3 \cos^2 x \sin 2x dx$.
5. $\int_0^{\pi} x^6 \cos x dx$.
6. $\int_0^{\frac{\pi}{2}} x^5 \sin 2x dx$.

Exercise 3.5.19. Find the recursive relations for $\int x^p e^{ax} \cos bx dx$ and $\int x^p e^{ax} \sin bx dx$. Then compute the integral.

1. $\int x^2 e^{-x} \sin 3x dx$.
2. $\int x^2 2^x \cos x dx$.
3. $\int x^3 e^x \cos^2 x dx$.

Exercise 3.5.20. Compute the integral.

1. $\int \sin^6 x dx$.
2. $\int \cos^8 x dx$.
3. $\int \cos^8 x \sin^6 x dx$.
4. $\int \cos^3 x \sin^2 x dx$.
5. $\int \cos^3 x \sin^5 x dx$.
6. $\int \cos^{-2} x \sin^2 x dx$.
7. $\int \dfrac{dx}{\cos^6 x}$.
8. $\int \dfrac{dx}{\sin^2 x \cos^2 x}$.

Exercise 3.5.21. Show that $\int_0^{\frac{\pi}{2}} \sin^{2m} x \cos^{2n} x dx = \dfrac{(2m)!(2n)!}{2^{2m+2n+1} m!n!(m+n)!}$ for natural numbers m, n. Can you find $\int_0^{\frac{\pi}{2}} \sin^m x \cos^n x dx$?

Exercise 3.5.22. Use $(\tan x)' = \sec^2 x = \tan^2 x + 1$ to derive the recursive formula for $\int \sec^m x \tan^n x dx$ similar to Example 3.5.14 and then find the value of $\int_0^{\frac{\pi}{4}} \tan^{2n} x dx$.

Example 3.5.15. Let

$$I_p = \int (ax^2 + bx + c)^p dx, \quad a \neq 0,\ b^2 \neq 4ac.$$

We have

$$\begin{aligned} I_p &= x(ax^2 + bx + c)^p - \int xd(ax^2 + bx + c)^p \\ &= x(ax^2 + bx + c)^p - \int px(2ax + b)(ax^2 + bx + c)^{p-1} dx. \end{aligned}$$

We try to express $px(2ax+b)$ as a combination of (ax^2+bx+c) and $(ax^2+bx+c)'$, up to adding a constant

$$2pax^2 + pbx = A(ax^2 + bx + c) + B(ax^2 + bx + c)' + C.$$

We get $A = 2p$, $B = -\dfrac{pb}{2a}$, $C = \dfrac{p(b^2 - 4ac)}{2a}$. Then

$$\begin{aligned} I_p &= x(ax^2 + bx + c)^p - A \int (ax^2 + bx + c)^p dx \\ &\quad - B \int (ax^2 + bx + c)^{p-1} (ax^2 + bx + c)' dx - C \int (ax^2 + bx + c)^{p-1} dx \\ &= x(ax^2 + bx + c)^p - AI_p - \frac{B}{p}(ax^2 + bx + c)^p - CI_{p-1}. \end{aligned}$$

This gives us the recursive relation

$$I_p = \frac{1}{(2p+1)2a}(2ax + b)(ax^2 + bx + c)^p - \frac{p(b^2 - 4ac)}{(2p+1)2a} I_{p-1}, \quad p \neq -\frac{1}{2}.$$

On the other hand, we may also express I_{p-1} in terms of I_p. After substituting p by $p+1$, we get

$$I_p = \frac{1}{(p+1)(b^2 - 4ac)}(2ax + b)(ax^2 + bx + c)^{p+1} - \frac{(2p+3)2a}{(p+1)(b^2 - 4ac)} I_{p+1}, \quad p \neq -1.$$

For the special case

$$I_p = \int (ax^2 + b)^p dx, \quad a, b \neq 0,$$

the recursive relations become

$$\begin{aligned} I_p &= \frac{1}{2p+1} x(ax^2 + b)^p + \frac{2pb}{2p+1} I_{p-1}, \quad p \neq -\frac{1}{2}; \\ I_p &= -\frac{1}{2(p+1)b} x(ax^2 + b)^{p+1} + \frac{2p+3}{2(p+1)b} I_{p+1}, \quad p \neq -1. \end{aligned}$$

For the special cases of $p = -\frac{1}{2}, -1$, I_p is given by Exercise 3.4.10 (and will be derived in Examples 3.5.30, 3.5.31, 3.5.32)

$$\begin{aligned}
\int \frac{dx}{\sqrt{a^2 - x^2}} &= \arcsin \frac{x}{a} + C, \quad a > 0, \\
\int \frac{dx}{\sqrt{x^2 + a}} &= \log \left| x + \sqrt{x^2 + a} \right| + C, \\
\int \frac{dx}{x^2 - a^2} &= \frac{1}{2a} \log \left| \frac{x - a}{x + a} \right| + C, \\
\int \frac{dx}{x^2 + a^2} &= \frac{1}{a} \arctan \frac{x}{a} + C.
\end{aligned}$$

Then the recursive relations can be used to compute I_p when p is an integer or a half integer. For example, we have

$$\begin{aligned}
\int \sqrt{a^2 - x^2} dx = I_{\frac{1}{2}} &= \frac{1}{2 \cdot \frac{1}{2} + 1} x(a^2 - x^2)^{\frac{1}{2}} + \frac{2 \cdot \frac{1}{2} a^2}{2 \cdot \frac{1}{2} + 1} I_{\frac{1}{2} - 1} \\
&= \frac{1}{2} x \sqrt{a^2 - x^2} + \frac{a^2}{2} \arcsin \frac{x}{a} + C, \\
\int (a^2 - x^2)^{\frac{3}{2}} dx = I_{\frac{3}{2}} &= \frac{1}{2 \cdot \frac{3}{2} + 1} x(a^2 - x^2)^{\frac{3}{2}} + \frac{2 \cdot \frac{3}{2} a^2}{2 \cdot \frac{3}{2} + 1} I_{\frac{3}{2} - 1} \\
&= \frac{1}{4} x(a^2 - x^2)^{\frac{3}{2}} + \frac{3a^2}{4} \left(\frac{1}{2} x \sqrt{a^2 - x^2} + \frac{a^2}{2} \arcsin \frac{x}{a} \right) + C \\
&= -\frac{1}{8} x(2x^2 - 5a^2) \sqrt{a^2 - x^2} + \frac{3a^4}{8} \arcsin \frac{x}{a} + C, \\
\int \frac{dx}{(x^2 + a^2)^2} = I_{-2} &= -\frac{1}{2(-2 + 1)a^2} x(x^2 + a^2)^{-2+1} + \frac{2(-2) + 3}{2(-2 + 1)a^2} I_{-2+1} \\
&= \frac{x}{2a^2(x^2 + a^2)} + \frac{1}{2a^3} \arctan \frac{x}{a} + C.
\end{aligned}$$

Exercise 3.5.23. Compute the integral.

1. $\int x^2 \sqrt{a^2 - x^2} dx.$
2. $\int x \sqrt{a^2 - x^2} dx.$
3. $\int (x+b)^2 \sqrt{a^2 - x^2} dx.$
4. $\int \frac{x^2 dx}{\sqrt{a^2 - x^2}}.$
5. $\int \frac{dx}{(a^2 - x^2)^{\frac{3}{2}}}.$
6. $\int \frac{x dx}{(a^2 - x^2)^{\frac{3}{2}}}.$
7. $\int \frac{dx}{(a^2 - x^2)^{\frac{5}{2}}}.$
8. $\int \frac{dx}{(a^2 - x^2)^3}.$
9. $\int \frac{x dx}{(a^2 - x^2)^3}.$

Exercise 3.5.24. Compute the integral.

1. $\displaystyle\int x^2\sqrt{x^2+a}dx.$
2. $\displaystyle\int x\sqrt{x^2+a}dx.$
3. $\displaystyle\int \frac{x^2dx}{\sqrt{x^2+a}}.$
4. $\displaystyle\int \frac{xdx}{\sqrt{x^2+a}}.$
5. $\displaystyle\int (x^2+a)^{\frac{3}{2}}dx.$
6. $\displaystyle\int \frac{dx}{(x^2+a)^{\frac{3}{2}}}.$
7. $\displaystyle\int \frac{dx}{(x^2+a^2)^3}.$
8. $\displaystyle\int \frac{xdx}{(x^2+a^2)^3}.$
9. $\displaystyle\int \frac{x^2dx}{(x^2+a^2)^3}.$

Exercise 3.5.25. Combine the ideas of Exercise 3.4.11 and Example 3.5.15 to compute the integral.

1. $\displaystyle\int \frac{dx}{(x^2+2x+2)^2}.$
2. $\displaystyle\int \frac{dx}{\sqrt{x^2+2x+2}}.$
3. $\displaystyle\int \frac{dx}{(x^2+2x+2)^{\frac{3}{2}}}.$
4. $\displaystyle\int (x^2+2x+2)^{\frac{3}{2}}dx.$
5. $\displaystyle\int \sqrt{x(1-x)}dx.$
6. $\displaystyle\int \frac{dx}{(x(1-x))^{\frac{3}{2}}}.$

3.5.3 Change of Variable

The chain rule says that, if $\displaystyle\int f(y)dy = F(y)+C$ is the indefinite integral of $f(y)$, and $\phi(x)$ is a differentiable function, then

$$F(\phi(x))' = F'(\phi(x))\phi'(x) = f(\phi(x))\phi'(x).$$

In other words, $F(\phi(x))$ is the antiderivative of $f(\phi(x))\phi'(x)$, or

$$\int f(\phi(x))\phi'(x)dx = F(\phi(x)) + C = \int f(y)dy\bigg|_{y=\phi(x)}.$$

If we use the differential notation $d\phi(x) = \phi'(x)dx$, then the equality becomes

$$\int f(\phi(x))d\phi(x) = \int f(y)dy\bigg|_{y=\phi(x)}.$$

The right side means computing the antiderivative of the function of y first, and then substituting $y=\phi(x)$ into the antiderivative. This is the *change of variable* formula. By Newton-Leibniz formula, we further get the change of variable formula for definite integral

$$\int_a^b f(\phi(x))\phi'(x)dx = \int_a^b f(\phi(x))d\phi(x) = \int_{\phi(a)}^{\phi(b)} f(y)dy.$$

Example 3.5.16. If $\displaystyle\int f(y)dy = F(y)+C$, then by letting $y = ax+b$, we have

$$\int f(ax+b)dx = \frac{1}{a}\int f(ax+b)d(ax+b) = \frac{1}{a}F(ax+b)+C.$$

For example,

$$\begin{aligned}
\int (2x+1)^p dx &=_{y=2x+1} \frac{1}{2}\int y^p dy = \frac{y^{p+1}}{2(p+1)} + C = \frac{(2x+1)^{p+1}}{2(p+1)} + C,\\
\int \sin(3x-2)dx &=_{y=3x-2} \frac{1}{3}\int \sin y dy = -\frac{1}{3}\cos y + C = -\frac{1}{3}\cos(3x-2) + C,\\
\int \frac{dx}{x^2+2x+2} &= \int \frac{dx}{(x+1)^2+1} =_{y=x+1} \int \frac{dx}{y^2+1}\\
&= \arctan y + C = \arctan(x+1) + C.
\end{aligned}$$

Example 3.5.17. The following is a simple change of variable

$$\int xe^{x^2}dx =_{y=x^2} \frac{1}{2}\int e^y dy = \frac{1}{2}e^y + C = \frac{1}{2}e^{x^2} + C.$$

The idea is a "mini-integration" of xdx that can be expressed more clearly by writing

$$\int xe^{x^2}dx = \int e^{x^2}\frac{1}{2}d(x^2) = \frac{1}{2}e^{x^2} + C.$$

After the mini-integration, we view x^2 as the new variable.

The following are more examples following the mini-integration idea

$$\begin{aligned}
\int \frac{dx}{x\log x} &= \int \frac{1}{\log x}\left(\frac{dx}{x}\right) = \int \frac{1}{\log x}d(\log x) = \log|\log x| + C,\\
\int \frac{dx}{x^2+a^2} &= \int \frac{d}{a^2\left(\left(\frac{x}{a}\right)^2+1\right)} = \int \frac{d\left(\frac{x}{a}\right)}{a\left(\left(\frac{x}{a}\right)^2+1\right)} = \frac{1}{a}\arctan\frac{x}{a} + C,\\
\int \frac{xdx}{x^4+a^4} &= \frac{1}{2}\int \frac{d(x^2)}{(x^2)^2+a^4} = \frac{1}{2a^2}\arctan\frac{x^2}{a^2} + C.
\end{aligned}$$

Example 3.5.18. The integral in Example 3.5.1 was computed in Example 3.5.10 again by using the integration by parts. The integral can also be computed by change of variable.

Let $y = x+1$. Then

$$\int x(x+1)^{10}dx = \int (y-1)y^{10}dy = \int (y^{10} - y^{11})dy = \frac{1}{11}y^{11} - \frac{1}{12}y^{12} + C.$$

Substituting $y = x+1$ back, we get

$$\int x(x+1)^{10}dx = \frac{1}{11}(x+1)^{11} - \frac{1}{12}(x+1)^{12} + C.$$

By the same change, we have

$$\begin{aligned}\int \left(\frac{x-1}{x+1}\right)^4 dx &=_{y=x+1} \int \frac{(y-2)^4}{y^4} dy \\ &= \int (1 - 4\cdot 2y^{-1} + 6\cdot 2^2 y^{-2} - 4\cdot 2^3 y^{-3} + 2^4 y^{-4}) dy \\ &= x + 1 - 8\log|x+1| \\ &\quad - 24(x+1)^{-1} + 16(x+1)^{-2} - \frac{16}{3}(x+1)^{-3} + C \\ &= x - \frac{8(9x^2+12x+5)}{3(x+1)^3} - 8\log|x+1| + C.\end{aligned}$$

Note that the second C is the first C plus 1.

Compare the above with the computation of definite integral

$$\begin{aligned}\int_0^1 \left(\frac{x-1}{x+1}\right)^4 dx &=_{y=x+1} \int_1^2 \frac{(y-2)^4}{y^4} dy \\ &= \int_1^2 (1 - 4\cdot 2y^{-1} + 6\cdot 2^2 y^{-2} - 4\cdot 2^3 y^{-3} + 2^4 y^{-4}) dy \\ &= \left(y - 8\log y - 24y^{-1} + 16y^{-2} - \frac{16}{3}y^{-3}\right)_1^2 \\ &= 1 - 8\log 2 - 24\left(\frac{1}{2} - 1\right) + 16\left(\frac{1}{4} - 1\right) - \frac{16}{3}\left(\frac{1}{8} - 1\right) \\ &= \frac{17}{3} - 8\log 2.\end{aligned}$$

Note that the evaluation is done by using the new variable y instead of the old x.

Example 3.5.19. The integrals of inverse trigonometric functions can also be computed by combining integration by parts and change of variable

$$\begin{aligned}\int \arcsin x dx &= x\arcsin x - \int \frac{xdx}{\sqrt{1-x^2}} = x\arcsin x + \frac{1}{2}\int \frac{d(1-x^2)}{\sqrt{1-x^2}} \\ &= x\arcsin x + \sqrt{1-x^2} + C.\end{aligned}$$

Alternatively, we may simply introduce the trigonometric function as the new variable. For example, by $y = \arcsin x$, $x = \sin y$, we have

$$\begin{aligned}\int \arcsin x dx &= \int y d(\sin y) = y\sin y - \int \sin y dy \\ &= y\sin y + \cos y + C = x\arcsin x + \sqrt{1-x^2} + C.\end{aligned}$$

Note that $\cos y = \sqrt{1-x^2}$ is non-negative because $x \in [-1, 1]$ and $y \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. The integration by parts used in both computations are essentially the same.

The idea for integrating $\arcsin x$ can also be used to compute $\int x^m(\arcsin x)^n dx$ and $\int x^m(\arctan x)^n dx$.

Example 3.5.20. To compute $\int \frac{dx}{\sqrt{e^x+a}}$, we introduce

$$y = \sqrt{e^x+a}, \quad y^2 = e^x + a, \quad 2ydy = e^x dx.$$

Then

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

By Examples 3.4.9 and 3.5.2, we have

$$\int \frac{dx}{\sqrt{e^x+a}} = \begin{cases} \frac{1}{\sqrt{a}}\log\left|\frac{y-\sqrt{-a}}{y+\sqrt{-a}}\right| + C = \frac{1}{\sqrt{a}}\log\left|\frac{\sqrt{e^x+a}-\sqrt{a}}{\sqrt{e^x+a}+\sqrt{a}}\right| + C, & \text{if } a > 0, \\ \frac{2}{\sqrt{-a}}\arctan\frac{y}{\sqrt{-a}} + C = \frac{2}{\sqrt{-a}}\arctan\sqrt{-\frac{e^x}{a}-1} + C, & \text{if } a < 0. \end{cases}$$

Example 3.5.21. To compute

$$I = \int_0^\pi \frac{x\sin x}{1+\cos^2 x}dx,$$

we introduce $y = \pi - x$. Then

$$\begin{aligned} I &= -\int_\pi^0 \frac{(\pi-y)\sin y}{1+\cos^2 y}dy = \pi\int_0^\pi \frac{\sin y}{1+\cos^2 y}dy - \int_0^\pi \frac{y\sin y}{1+\cos^2 y}dy \\ &= \pi\int_0^\pi \frac{\sin y}{1+\cos^2 y}dy - I. \end{aligned}$$

Therefore

$$\begin{aligned} I &= \frac{\pi}{2}\int_0^\pi \frac{\sin y}{1+\cos^2 y}dy = -\frac{\pi}{2}\int_{\cos 0}^{\cos\pi} \frac{1}{1+z^2}dz \\ &= \frac{\pi}{2}\int_{-1}^1 \frac{dz}{1+z^2} = \frac{\pi}{2}(\arctan 1 - \arctan(-1)) = \frac{\pi^2}{4}. \end{aligned}$$

Note that the computation of the definite integral makes use of the new variable z only. There is no need to go back to the original variable x.

Exercise 3.5.26. Compute the integral.

1. $\displaystyle\int (x^2+1)(2x-1)^{10}dx.$
2. $\displaystyle\int \log(2x-1)dx.$
3. $\displaystyle\int x\log(2x-1)dx.$
4. $\displaystyle\int \cos(2x-1)dx.$
5. $\displaystyle\int e^{3x}\cos(2x-1)dx.$
6. $\displaystyle\int \sin(2x+1)\cos(2x-1)dx.$

Exercise 3.5.27. Compute the integral.

1. $\displaystyle\int \frac{x}{x^2+1}dx.$
2. $\displaystyle\int \frac{bx+c}{x^2+a^2}dx.$
3. $\displaystyle\int \frac{x}{(x^2+1)^p}dx.$
4. $\displaystyle\int x(x^2+a^2)^p dx.$
5. $\displaystyle\int x^3\sqrt{x^2+1}dx.$
6. $\displaystyle\int \frac{x^3dx}{\sqrt[3]{x^2+a^2}}.$

Exercise 3.5.28. Compute the integral.

1. $\displaystyle\int \sin x\sin(\cos x)dx.$
2. $\displaystyle\int \sqrt{x}\sin(1+x^{\frac{3}{2}})dx.$
3. $\displaystyle\int \cot x dx.$
4. $\displaystyle\int x\tan x^2 dx.$
5. $\displaystyle\int \frac{1}{x}\tan(\log x)dx.$
6. $\displaystyle\int \sin(\log x)dx.$
7. $\displaystyle\int \frac{\sin x dx}{a+\cos^2 x}.$
8. $\displaystyle\int \frac{\sin 2x dx}{a+\cos^2 x}.$
9. $\displaystyle\int \sin 2x\sqrt{a+\cos^2 x}dx.$
10. $\displaystyle\int \frac{\cos x dx}{\sqrt{a+\cos^2 x}}.$
11. $\displaystyle\int \frac{\sin x\cos x dx}{\sqrt{a^2\sin^2 x+b^2\cos^2 x}}.$
12. $\displaystyle\int \frac{\sin x\cos x dx}{\sin^4 x+\cos^4 x}.$

Exercise 3.5.29. Compute the integral.

1. $\displaystyle\int \arccos x dx.$
2. $\displaystyle\int \frac{\arcsin x dx}{\sqrt{1-x^2}}.$
3. $\displaystyle\int \sqrt{1-x^2}\arcsin x dx.$
4. $\displaystyle\int x\sqrt{1-x^2}\arcsin x dx.$
5. $\displaystyle\int \frac{dx}{\sqrt{1-x^2}\arccos x}.$
6. $\displaystyle\int \frac{x^3\arccos x}{\sqrt{1-x^2}}dx.$
7. $\displaystyle\int x^2\arccos x dx.$
8. $\displaystyle\int \frac{\arcsin x}{x^2}dx.$
9. $\displaystyle\int (\arccos x)^2 dx.$
10. $\displaystyle\int x\arctan x dx.$
11. $\displaystyle\int \frac{(\arctan x)^2}{1+x^2}dx.$
12. $\displaystyle\int \frac{dx}{(1+x^2)\arctan x}.$

Exercise 3.5.30. Compute the integral.

1. $\displaystyle\int x^3e^{x^2}dx.$
2. $\displaystyle\int e^{\sqrt{x}}dx.$
3. $\displaystyle\int \frac{e^{-\sqrt{x}}}{\sqrt{x}}dx.$
4. $\displaystyle\int \sqrt{x}e^{\sqrt{x}}dx.$
5. $\displaystyle\int e^{\cos x}\sin x dx.$
6. $\displaystyle\int \frac{e^x dx}{1+e^x}.$

7. $\displaystyle\int \frac{dx}{1+e^x}$.
8. $\displaystyle\int \frac{dx}{e^x+e^{-x}}$.
9. $\displaystyle\int \sqrt{e^x+a}dx$.

Exercise 3.5.31. Find a recursive relation for $\int (e^x+a)^p dx$. Then compute $\int (e^x+a)^{\frac{5}{2}} dx$ and $\displaystyle\int \frac{dx}{(e^x+a)^3}$.

Exercise 3.5.32. Compute the integral.

1. $\displaystyle\int \frac{dx}{x\sqrt{1+\log x}}$.
2. $\displaystyle\int \frac{dx}{x\log x \log(\log x)}$.
3. $\displaystyle\int \frac{1}{x^2-1}\log\frac{x+1}{x-1}dx$.
4. $\displaystyle\int \frac{\log(x+1)-\log x}{x(x+1)}dx$.

Exercise 3.5.33. Compute the integral.

1. $\displaystyle\int \frac{f'(x)}{f(x)^p}dx$.
2. $\displaystyle\int \frac{f'(x)}{1+f(x)^2}dx$.
3. $\displaystyle\int 2^{f(x)}f'(x)dx$.

Exercise 3.5.34. Prove the equalties

1. $\displaystyle\int_0^{\frac{\pi}{2}} f(\sin x)dx = \int_0^{\frac{\pi}{2}} f(\cos x)dx$.
2. $\displaystyle\int_0^{\pi} xf(\sin x)dx = \frac{\pi}{2}\int_0^{\pi} f(\sin x)dx$.

Exercise 3.5.35. Explain why we cannot use the change of variable $y = \dfrac{1}{x}$ to compute the integral $\displaystyle\int_{-1}^{1} \frac{dx}{1+x^2}$.

Exercise 3.5.36. Suppose f is continuous on an open interval containing $[a,b]$. Find the derivative $\displaystyle\frac{d}{dt}\int_a^b f(x+t)dx$.

Exercise 3.5.37. Explain the equalities in Exercise 3.1.11 by change of variable.

Exercise 3.5.38. Prove that $\int_{-a}^{a} f(x)dx = 2\int_0^a f(x)dx$ for even function f. Prove that $\int_{-a}^{a} f(x)dx = 0$ for odd function f.

Example 3.5.22. To compute $\displaystyle\int \frac{dx}{1+\sqrt{x-1}}$, we simply let $y=\sqrt{x-1}$. Then $x = y^2+1$, $dx = 2ydy$, and

$$\begin{aligned}\int \frac{dx}{1+\sqrt{x-1}} &= \int \frac{2ydy}{1+y} = 2\int \left(1-\frac{1}{1+y}\right)dy \\ &= 2y - 2\log(1+y) + C = 2\sqrt{x-1} - 2\log(1+\sqrt{x-1}) + C.\end{aligned}$$

Example 3.5.23. By taking $y = x^6$, we get rid of the square root and cube root at the same time.

$$\begin{aligned}
\int \frac{dx}{\sqrt{x}+\sqrt[3]{x}} &=_{x=y^6} \int \frac{6y^5dy}{y^3+y^2} = 6\int \frac{y^3dy}{y+1} \\
&= 6\int \frac{(y^3+1)-1dy}{y+1} = 6\int \left(y^2-y+1-\frac{1}{1+y}\right)dy \\
&= 2y^3-3y^2+6y-6\log(1+y)+C \\
&= 2\sqrt{x}-3\sqrt[3]{x}+6\sqrt[6]{x}-6\log(1+\sqrt[6]{x})+C.
\end{aligned}$$

Example 3.5.24. To compute $\displaystyle\int \frac{dx}{\sqrt{x+1}+\sqrt{x}+1}$, we introduce

$$y = \sqrt{x+1}+\sqrt{x}.$$

Then

$$\frac{1}{y} = \sqrt{x+1}-\sqrt{x}, \quad y+\frac{1}{y} = 2\sqrt{x+1}, \quad y-\frac{1}{y} = 2\sqrt{x},$$

and

$$x = \frac{1}{4}\left(y-\frac{1}{y}\right)^2, \quad dx = \frac{1}{2}\left(y-\frac{1}{y^3}\right)dy.$$

Therefore

$$\begin{aligned}
\int \frac{dx}{\sqrt{x+1}+\sqrt{x}+1} &= \int \frac{\frac{1}{2}\left(y-\frac{1}{y^3}\right)dy}{y+1} = \frac{1}{2}\int \frac{(y-1)(y+1)(y^2+1)}{(y+1)y^3}dy \\
&= \frac{1}{2}\int \left(1-\frac{1}{y}+\frac{1}{y^2}-\frac{1}{y^3}\right)dy \\
&= \frac{1}{2}\left(y-\log|y|-\frac{1}{y}+\frac{1}{2y^2}\right)+C \\
&= -\frac{1}{2}\log(\sqrt{x+1}+\sqrt{x})+\sqrt{x}+\frac{1}{4}(\sqrt{x+1}+\sqrt{x})^2+C.
\end{aligned}$$

Example 3.5.25. The change of variable in Example 3.5.24 can be used for integrating other functions involving $\sqrt{x+a}$ and $\sqrt{x+b}$. For example, to compute $\displaystyle\int \sqrt{\frac{x+1}{x-1}}dx$, which makes sense for $x > 1$ or $x \le -1$, we introduce $y = \sqrt{x+1}+\sqrt{x-1}$ for $x > 1$. Then

$$\frac{2}{y} = \sqrt{x+1}-\sqrt{x-1}, \quad y+\frac{2}{y} = 2\sqrt{x+1}, \quad y-\frac{2}{y} = 2\sqrt{x-1},$$

and

$$x = \frac{1}{4}\left(y-\frac{2}{y}\right)^2+1, \quad dx = \frac{1}{2}\left(1-\frac{4}{y^4}\right)ydy.$$

Therefore by $x > 1$, we have

$$\begin{aligned}\int\sqrt{\frac{x+1}{x-1}}dx &= \int\frac{y+\frac{2}{y}}{y-\frac{2}{y}}\frac{1}{2}\left(1-\frac{4}{y^4}\right)ydy = \frac{1}{2}\int\left(y+\frac{4}{y}+\frac{4}{y^3}\right)dy\\ &= \frac{1}{4}y^2+2\log|y|-\frac{1}{y^2}+C = 2\log|y|+\frac{1}{4}\left(y+\frac{2}{y}\right)\left(y-\frac{2}{y}\right)+C\\ &= \frac{1}{4}\log(\sqrt{x+1}+\sqrt{x-1})+\sqrt{x^2-1}+C.\end{aligned}$$

For $x \le -1$, we may introduce $y = \sqrt{-x+1}-\sqrt{-x-1}$. Then we have

$$\begin{aligned}\int\sqrt{\frac{x+1}{x-1}}dx &= \frac{1}{2}\int\left(y+\frac{4}{y}+\frac{4}{y^3}\right)dy = 2\log|y|+\frac{1}{4}\left(y+\frac{2}{y}\right)\left(y-\frac{2}{y}\right)+C\\ &= \frac{1}{4}\log(\sqrt{-x-1}-\sqrt{-x+1})-\sqrt{x^2-1}+C.\end{aligned}$$

Example 3.5.26. The integral $\int\sqrt{\frac{x}{1-x}}dx$ is comparable to the integral in Example 3.5.25. Yet the similar change of variable does not work, due to the requirement $0 \le x < 1$. So we introduce

$$y = \sqrt{\frac{x}{1-x}},\quad x = \frac{y^2}{1+y^2},\quad dx = \frac{2y}{(1+y^2)^2}dy,$$

and get

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

The last computation in Example 3.5.15 is used here.

Note that the idea here can also be applied to the integral in Example 3.5.25, by introducing $y = \sqrt{\frac{x+1}{x-1}}$. The advantage of the approach is that we do not need to distinguish $x > 1$ and $x \le 1$.

Exercise 3.5.39. Compute the integrals in Example 3.5.24 and 3.5.25 by using change of variable similar to Example 3.5.26.

Exercise 3.5.40. Compute the integral.

1. $\displaystyle\int (1+\sqrt{x})^p dx.$
2. $\displaystyle\int (1+\sqrt[3]{x})^p dx.$
3. $\displaystyle\int (\sqrt{x}+\sqrt{x+1})^p dx.$
4. $\displaystyle\int \frac{1+\sqrt{x}}{1+\sqrt[3]{x}} dx.$
5. $\displaystyle\int \frac{\sqrt{1+x^2}}{x} dx.$
6. $\displaystyle\int \frac{dx}{x\sqrt{1+x^2}}.$

Example 3.5.27. We can use $(\cos x)' = -\sin x$, $(\sin x)' = \cos x$ and $\cos^2 x + \sin^2 x = 1$ to calculate the antiderivative of $\cos^m x \sin^n x$, in which either m or n is odd.

$$\begin{aligned}
\int \sin^3 x dx &= -\int \sin^2 x d\cos x = \int (\cos^2 x - 1) d\cos x \\
&= \frac{1}{3}\cos^3 x - \cos x + C, \\
\int \cos^4 x \sin^5 x dx &= -\int \cos^4 x \sin^4 x d\cos x = -\int \cos^4 x (1-\cos^2 x)^2 d\cos x \\
&= -\int (\cos^4 x - 2\cos^6 x + \cos^8 x) d\cos x \\
&= -\frac{1}{5}\cos^5 x + \frac{2}{7}\cos^7 x - \frac{1}{9}\cos^9 x + C, \\
\int \tan x dx &= \int \frac{\sin x}{\cos x} dx = -\int \frac{d\cos x}{\cos x} = -\log|\cos x| + C, \\
\int \sec x dx &= \int \frac{dx}{\cos x} = \int \frac{d\sin x}{\cos^2 x} = \int \frac{d\sin x}{1-\sin^2 x} \\
&= \frac{1}{2}\log\frac{1+\sin x}{1-\sin x} + C = \frac{1}{2}\log\frac{(1+\sin x)^2}{1-\sin^2 x} + C = \log\frac{1+\sin x}{|\cos x|} + C \\
&= \log|\sec x + \tan x| + C.
\end{aligned}$$

Example 3.5.28. Similar to Example 3.5.27, we can also use $(\tan x)' = \sec^2 x$, $(\sec x)' = \sec x \tan x$ and $\sec^2 x = 1 + \tan^2 x$ to calculate the antiderivative of $\sec^m x \tan^n x$.

$$\begin{aligned}
\int \sec x \tan^3 x dx &= \int \tan^2 x d\sec x = \int (\sec^2 x - 1) d\sec x = \frac{1}{3}\sec^3 x - \sec x + C, \\
\int \sec^4 x dx &= \int (\tan^2 x + 1) d\tan x = \frac{1}{3}\tan^3 x + \tan x + C, \\
\int \tan^4 x dx &= \int (\sec^2 x - 1)^2 dx = \int (\sec^4 x - 2\sec^2 x + 1) dx \\
&= \left(\frac{1}{3}\tan^3 x + \tan x\right) - 2\tan x + x + C \\
&= \frac{1}{3}\tan^3 x - \tan x + x + C,
\end{aligned}$$

The following is computed in Example 3.5.14 by more complicated method

$$\int \frac{\sin^2 x}{\cos^4 x} dx = \int \sec^2 x \tan^2 x dx = \int \tan^2 x d\tan x = \frac{1}{3}\tan^3 x + C.$$

Example 3.5.29. The method of Example 3.5.28 cannot be directly applied to the antiderivatives of $\sec^n x$ and $\tan^n x$ for odd n. Instead, the idea of Example 3.5.27 can be used.

Using the integration by parts, we have

$$\begin{aligned}\int \sec^3 x dx &= \int \sec x d\tan x = \sec x \tan x - \int \tan x d\sec x \\ &= \sec x \tan x - \int \tan^2 x \sec x dx = \sec x \tan x - \int (\sec^2 x - 1)\sec x dx \\ &= \sec x \tan x - \int \sec^3 x dx + \int \sec x dx.\end{aligned}$$

Then with the help of Example 3.5.27, we get

$$\begin{aligned}\int \sec^3 x dx &= \frac{1}{2}\sec x \tan x + \frac{1}{2}\int \sec x dx \\ &= \frac{1}{2}\sec x \tan x + \frac{1}{2}\log|\sec x + \tan x| + C.\end{aligned}$$

In fact, the integral $\int \sec^3 x dx$ is $I_{-3,0}$ in Example 3.5.14, and the expression above in terms of $\int \sec x dx$ is the expression of $I_{-3,0}$ in terms of $I_{-1,0}$.

Example 3.5.27 also gives

$$\begin{aligned}\int \tan^3 x dx &= \int \tan x(\sec^2 x - 1)dx = \int \tan x d\tan x - \int \tan x dx \\ &= \frac{1}{2}\tan^2 x + \log|\cos x| + C.\end{aligned}$$

Exercise 3.5.41. Compute the integral.

1. $\int \cos^3 x \sin^2 x dx$.
2. $\int \cos^3 x \sin^5 x dx$.
3. $\int \frac{dx}{\cos^6 x}$.
4. $\int \frac{dx}{\sin^2 x \cos^2 x}$.
5. $\int \frac{dx}{\sin x \cos x}$.
6. $\int \cos^{-3} x \sin^5 x dx$.

Exercise 3.5.42. Compute the integral.

1. $\int \csc x dx$.
2. $\int \tan^3 x dx$.
3. $\int \tan^6 x \sec^4 x dx$.
4. $\int \tan^n x \sec^4 x dx$.
5. $\int \cot^6 x \csc^4 x dx$.
6. $\int \tan^2 x \sec x dx$.
7. $\int \tan^3 x \sec x dx$.
8. $\int \tan^5 x \sec^7 x dx$.
9. $\int \tan^3 x \cos^2 x dx$.

10. $\int \cot^5 x \sin^4 x dx$.
11. $\int \csc^4 x \cot^6 x dx$.
12. $\int x \tan x \sec x dx$.

Exercise 3.5.43. Compute the integral.

1. $\int x^3 e^{-x^2} \cos x^2 dx$.
2. $\int x \sin(\log x) dx$.
3. $\int (\sin x)^p \cos^3 x dx$.
4. $\int \frac{\cos 2x dx}{\sin^2 x \cos^2 x}$.
5. $\int \frac{a \sin x + b \cos x}{\sin 2x} dx$.
6. $\int \frac{dx}{a \sin x + b \cos x}$.
7. $\int \frac{1 + \sin x}{1 + \cos x} dx$.
8. $\int \frac{dx}{1 + \cos x}$.
9. $\int \frac{dx}{a + \tan x}$.
10. $\int \frac{A \sin x + B \cos x}{a \sin x + b \cos x} dx$.
11. $\int \frac{dx}{a \sin^2 x + b \cos^2 x}$.
12. $\int \frac{A \sin x + B \cos x}{a \sin^2 x + b \cos^2 x} dx$.
13. $\int \frac{dx}{\sqrt{2} + \sin x + \cos x}$.
14. $\int \frac{\sin x dx}{\sqrt{2} + \sin x + \cos x}$.
15. $\int \frac{x dx}{\cos^2 x}$.

Exercise 3.5.44. Show that there are constants A_n, B_n, C_n, such that

$$\int \frac{dx}{(a \sin x + b \cos x)^n} = \frac{A_n \sin x + B_n \cos x}{(a \sin x + b \cos x)^{n-1}} + C_n \int \frac{dx}{(a \sin x + b \cos x)^{n-2}}.$$

Exercise 3.5.45. For $|a| \neq |b|$, show that there are constants A_n, B_n, C_n, 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 3.5.46. How to calculate $\int \frac{(A \sin x + B \cos x + C) dx}{(a \sin x + b \cos x + c)^n}$?

Exercise 3.5.47. Compute $\int \frac{dx}{\cos(x + a) \cos(x + b)}$ 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 integral.

1. $\int \frac{dx}{\sin(x + a) \cos(x + b)}$.
2. $\int \tan(x + a) \tan(x + b) dx$.
3. $\int \frac{dx}{\sin x - \sin a}$.
4. $\int \frac{dx}{\cos x + \cos a}$.

Example 3.5.30. To integrate a function of $\sqrt{a^2 - x^2}$, with $a > 0$, we may introduce $x = a \sin y$, $dx = a \cos y dy$. Note that the function makes sense only for $|x| \leq a$. Correspondingly, we take $y = \arcsin \frac{x}{a} \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. This implies $\cos y \geq 0$, and $\sqrt{a^2 - x^2} = a \cos y$.

The following example is also computed in Example 3.5.15.

$$\begin{aligned}\int \sqrt{a^2-x^2}dx &= a^2\int \cos^2 y dy = \frac{a^2}{2}\int (1+\cos 2y)dy \\ &= a^2\left(\frac{1}{2}y+\frac{1}{4}\sin 2y\right)+C = \frac{a^2}{2}(y+\sin y\cos y)+C \\ &= \frac{a}{2}\arcsin\frac{x}{a}+\frac{1}{2}x\sqrt{a^2-x^2}+C.\end{aligned}$$

We may also use $x=a\cos y$ instead of $x=a\sin y$.

$$\begin{aligned}\int \frac{dx}{a+\sqrt{a^2-x^2}} &= \int \frac{a\sin y dy}{a+a\sin y} = \int \frac{\sin y(1-\sin y)dy}{(1+\sin y)(1-\sin y)} \\ &= \int \left(\frac{\sin y}{\cos^2 y}-\frac{\sin^2 y}{\cos^2 y}\right)dy = \int (\sec y\tan y-\sec^2 y+1)dy \\ &= \sec y-\tan y+y+C = \frac{a}{x}-\frac{a}{x}\sqrt{1-\frac{x^2}{a^2}}+\arccos\frac{x}{a}+C.\end{aligned}$$

The following is an example of definite integral.

$$\int_0^1 (1-x^2)^p dx = -\int_{\frac{\pi}{2}}^0 (1-\cos^2 y)^p \sin y dy = \int_0^{\frac{\pi}{2}} \sin^{2p+1} y dy.$$

By Example 3.5.15, we know the specific value when $2p+1$ is a natural number.

$$\int_0^1 (1-x^2)^n dx = \frac{(2n)!!}{(2n+1)!!}, \quad \int_0^1 (1-x^2)^{n-\frac{1}{2}} dx = \frac{(2n-1)!!}{(2n)!!}\frac{\pi}{2}.$$

Example 3.5.31. To integrate a function of $\sqrt{x^2+a^2}$, with $a>0$, we may introduce $x=a\tan y$, $dx=a\sec^2 y dy$. We have $y=\arctan\frac{x}{a}\in\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$ and $\sqrt{x^2+a^2}=a\sec y$.

With the help of Example 3.5.27, we have

$$\begin{aligned}\int \frac{dx}{\sqrt{x^2+a^2}} &= \int \frac{a\sec^2 y dy}{a\sec y} = \int \sec y dy \\ &= \log|\sec y+\tan y|+C = \log(\sqrt{x^2+a^2}+x)+C.\end{aligned}$$

With the help of Example 3.5.29, we have

$$\begin{aligned}\int \sqrt{x^2+a^2}x dx &= \int a^2\sec^3 y dy = \frac{a^2}{2}\sec y\tan y+\frac{a^2}{2}\log|\sec y+\tan y|+C \\ &= \frac{1}{2}x\sqrt{x^2+a^2}+\frac{a^2}{2}\log(\sqrt{x^2+a^2}+x)+C.\end{aligned}$$

One may verify that the two integrals satisfy the recursive relation in Example 3.5.15.

Example 3.5.32. To integrate a function of $\sqrt{x^2-a^2}$, with $a>0$, we may introduce $x=a\sec y$, $dx=a\sec y\tan ydy$. We have $y=\text{arcsec}\frac{x}{a}=\arccos\frac{a}{x}\in[0,\pi]$ and $\sqrt{x^2-a^2}=\pm a\tan y$, where the sign depends on whether $x\geq a$ or $x\leq -a$.

With the help of Example 3.5.28, we have

$$\begin{aligned}
\int\frac{dx}{\sqrt{x^2-a^2}}&=\int\frac{a\sec y\tan ydy}{\pm a\tan y}=\int\sec ydy\\
&=\pm\log|\sec y+\tan y|+C=\pm\log\left|x\pm\sqrt{x^2-a^2}\right|+C\\
&=\log\left|x+\sqrt{x^2-a^2}\right|+C,\\
\int\sqrt{x^2-a^2}dx&=\int(\pm a\tan y)a\sec y\tan ydy=\pm a^2\int(\sec^3y-\sec y)dy\\
&=\pm\frac{a^2}{2}\sec y\tan y\mp\frac{a^2}{2}\log|\sec y+\tan y|+C\\
&=\frac{1}{2}x\sqrt{x^2-a^2}-\log\left|x+\sqrt{x^2-a^2}\right|+C.
\end{aligned}$$

Combing with Example 3.5.31, for positive as well as negative a, we have

$$\begin{aligned}
\int\frac{dx}{\sqrt{x^2+a}}&=\log\left|x+\sqrt{x^2+a}\right|+C,\\
\int\sqrt{x^2+a}dx&=\frac{1}{2}x\sqrt{x^2+b}+\frac{a}{2}\log\left|x+\sqrt{x^2+a}\right|+C.
\end{aligned}$$

Example 3.5.33. By completing the square, a quadratic function ax^2+bx+c can be changed to $a(y^2+d)$, where $y=x+\frac{b}{2a}$ and $d=\frac{4ac-b^2}{4a^2}$. For example, if $b^2<4ac$, then

$$\begin{aligned}
\int\frac{dx}{ax^2+bx+c}&=\int\frac{dy}{a(y^2+(\sqrt{d})^2)}=\frac{1}{a\sqrt{d}}\arctan\frac{y}{\sqrt{d}}+C\\
&=\frac{2}{\sqrt{4ac-b^2}}\arctan\frac{2ax+b}{\sqrt{4ac-b^2}}+C.
\end{aligned}$$

Using the recursive relation in Example 3.5.15, we further get

$$\int\frac{dx}{(ax^2+bx+c)^2}=\frac{2ax+b}{(4ac-b^2)(ax^2+bx+c)}-\frac{4a}{(4ac-b^2)^{\frac{3}{2}}}\arctan\frac{2ax+b}{\sqrt{4ac-b^2}}+C.$$

If $b^2\geq 4ac$, then by the similar idea, we may get

$$\begin{aligned}
\int\frac{dx}{ax^2+bx+c}&=\int\frac{dy}{a(y^2-(\sqrt{-d})^2)}=\frac{1}{2a\sqrt{-d}}\log\left|\frac{y-\sqrt{-d}}{y+\sqrt{-d}}\right|+C\\
&=\frac{1}{\sqrt{b^2-4ac}}\arctan\left|\frac{2ax+b-\sqrt{b^2-4ac}}{2ax+b+\sqrt{b^2-4ac}}\right|+C.
\end{aligned}$$

In fact, the quadratic function has two real roots, and it is more direct to calculate the integral by using

$$ax^2+bx+c=a(x-x_1)(x-x_2),\quad x_1,x_2=\frac{-b\pm\sqrt{b^2-4ac}}{2a},$$

and the idea of Examples 3.5.2 and 3.5.3.

Example 3.5.34. We further use the idea of Example 3.5.33 to convert the antiderivatives of functions of $\sqrt{ax^2+bx+c}$ to the computations in Examples 3.5.30, 3.5.31 and 3.5.32. For example, by

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

we let $y=2x-1$ and get

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

Moreover, for $0\le x<1$, we get

$$\begin{aligned}\int\sqrt{\frac{x}{1-x}}dx&=\int\frac{xdx}{\sqrt{x(1-x)}}=\int\frac{\frac{y+1}{2}dy}{\sqrt{1-y^2}}=-\int\frac{d(1-y^2)}{4\sqrt{1-y^2}}+\int\frac{dy}{2\sqrt{1-y^2}}\\&=\frac{1}{4}\sqrt{1-y^2}+\frac{1}{2}\arcsin y+C=\frac{1}{2}\sqrt{x(1-x)}+\frac{1}{2}\arcsin(2x-1)+C.\end{aligned}$$

The reader is left to verify that the result is the same as the one in Example 3.5.26.

Exercise 3.5.48. Compute the integral.

1. $\displaystyle\int\frac{xdx}{\sqrt{1-x^2}}$.
2. $\displaystyle\int\frac{(ax^2+bx+c)dx}{\sqrt{1-x^2}}$.
3. $\displaystyle\int\frac{dx}{x\sqrt{1-x^2}}$.
4. $\displaystyle\int\frac{x^2dx}{(a^2-x^2)^{\frac{3}{2}}}$.
5. $\displaystyle\int\frac{dx}{x(a^2-x^2)^{\frac{3}{2}}}$.
6. $\displaystyle\int(x^2+a^2)^{\frac{3}{2}}dx$.
7. $\displaystyle\int\frac{dx}{x(x^2+a^2)^{\frac{3}{2}}}$.
8. $\displaystyle\int\frac{x^3dx}{(x^2+a^2)^{\frac{3}{2}}}$.
9. $\displaystyle\int(x(x+1))^{\frac{3}{2}}dx$.
10. $\displaystyle\int x^2\sqrt{1-x^2}dx$.
11. $\displaystyle\int\frac{dx}{1-\sqrt{1-x^2}}$.
12. $\displaystyle\int\frac{dx}{\sqrt{1+x^2}+\sqrt{1-x^2}}$.
13. $\displaystyle\int\frac{xdx}{\sqrt{1+x^2}+\sqrt{1-x^2}}$.
14. $\displaystyle\int\frac{dx}{a+\sqrt{x^2+a^2}}$.
15. $\displaystyle\int\frac{xdx}{a+\sqrt{x^2+a^2}}$.

Exercise 3.5.49. Compute the integral.

1. $\displaystyle\int \sin x \log \tan x dx$.
2. $\displaystyle\int \arctan \sqrt{x} dx$.
3. $\displaystyle\int \frac{\log(\sin x)}{\sin^2 x} dx$.
4. $\displaystyle\int e^{\sin x}\left(\cos^2 x + \frac{1}{\cos^2 x}\right) dx$.

Exercise 3.5.50. Derive the formula

$$\int \frac{dx}{\sqrt{ax^2+bx+c}} = \begin{cases} \dfrac{1}{\sqrt{a}} \log\left(\dfrac{2ax+b}{\sqrt{a}} + \sqrt{ax^2+bx+c}\right) + C, & \text{if } a > 0, \\ \dfrac{1}{\sqrt{-a}} \arcsin \dfrac{-2ax-b}{\sqrt{b^2-4ac}} + C, & \text{if } a < 0. \end{cases}$$

Exercise 3.5.51. Use the change of variable $y = x \pm \dfrac{1}{x}$ to compute the integral.

1. $\displaystyle\int \frac{x^2+1}{x^4+1} dx$.
2. $\displaystyle\int \frac{x^2-1}{x^4+1} dx$.
3. $\displaystyle\int \frac{dx}{x^4+1} dx$.
4. $\displaystyle\int_{\frac{1}{2}}^{2} \left(1 + x - \frac{1}{x}\right) e^{x+\frac{1}{x}} dx$.

3.6 Integration of Rational Function

A rational function is the quotient of two polynomials. Examples 3.5.2, 3.5.3, 3.5.18, 3.5.33 are some typical examples of integrating rational functions. In this section, we systematically study how to integrate rational functions and how to convert some integrations into the integration of rational functions.

3.6.1 Rational Function

Example 3.6.1. The idea in Example 3.5.2 can be extended to the integral of rational functions whose denominator is a product of linear functions. For example, to integrate $\dfrac{x^2-2x+3}{x(x+1)(x+2)}$, we postulate

$$\frac{x^2-2x+3}{x(x+1)(x+2)} = \frac{A}{x} + \frac{B}{x+1} + \frac{C}{x+2}.$$

The equality is the same as

$$x^2 - 2x + 3 = A(x+1)(x+2) + Bx(x+2) + Cx(x+1).$$

Taking $x = 0, -1, -2$, we get

$$3 = 2A, \quad 6 = -B, \quad 11 = -2C.$$

Therefore

$$\begin{aligned}\int \frac{x^2-2x+3}{x(x+1)(x+2)}dx &= \int \left(\frac{3}{2x} - \frac{6}{x+1} - \frac{11}{2(x+2)}\right) dx \\ &= \frac{3}{2}\log|x| - 6\log|x+1| - \frac{11}{2}\log|x+2| + C.\end{aligned}$$

Example 3.6.2. If some real root of the denominator has multiplicity, then we need more sophisticated postulation. For example, to integrate $\dfrac{x^2-2x+3}{x(x+1)^3}$, we postulate

$$\frac{x^2-2x+3}{x(x+1)^3} = \frac{A}{x} + \frac{B}{x+1} + \frac{C}{(x+1)^2} + \frac{D}{(x+1)^3}.$$

This is the same as

$$x^2-2x+3 = A(x+1)^3 + Bx(x+1)^2 + Cx(x+1) + Dx.$$

Taking various values, we get

$$\begin{aligned} &x=0\colon 3=A, && x=-1\colon 6=-D, \\ &\text{(coefficient of) } x^3\colon 0=A+B, && \left.\frac{d}{dx}\right|_{x=-1}\colon\ -4=D-C. \end{aligned}$$

Therefore $A=3$, $B=-3$, $C=-2$, $D=-6$, and (C below means the general constant, and is different from the coefficient $C=-2$ above)

$$\begin{aligned}\int \frac{x^2-2x+3}{x(x+1)^3}dx &= \int \left(\frac{3}{x} - \frac{3}{x+1} - \frac{2}{(x+1)^2} - \frac{6}{(x+1)^3}\right) dx \\ &= 3\log\left|\frac{x}{x+1}\right| + \frac{2}{x+1} + \frac{3}{(x+1)^2} + C \\ &= 3\log\left|\frac{x}{x+1}\right| + \frac{2x+5}{(x+1)^2} + C.\end{aligned}$$

Example 3.6.3. In Examples 3.5.2, 3.6.1, 3.6.2, the numerator has lower degree than the denominator. In general, we need to divide polynomials for this to happen. For example, to integrate $\dfrac{x^5}{(x+1)^2(x-1)}$, we first divide x^5 by $(x+1)^2(x-1) = x^3+x^2-x-1$.

$$\begin{array}{r|rrrrrr}
 & & & x^2 & -x & +2 & \\ \hline
x^3+x^2-x-1 & x^5 & & & & & \\
 & -x^5 & -x^4 & +x^3 & +x^2 & & \\ \hline
 & & -x^4 & +x^3 & +x^2 & & \\
 & & x^4 & +x^3 & -x^2 & -x & \\ \hline
 & & & 2x^3 & & -x & \\
 & & & -2x^3 & -2x^2 & +2x & +2 \\ \hline
 & & & & -2x^2 & +x & +2
\end{array}$$

Then

$$\frac{x^5}{(x+1)^2(x-1)} = x^2 - x + 2 + \frac{-2x^2+x+2}{(x+1)^2(x-1)}$$
$$= x^2 - x + 2 - \frac{9}{4(x+1)} + \frac{1}{2(x+1)^2} + \frac{1}{4(x-1)},$$

and

$$\int \frac{x^5}{(x+1)^2(x-1)}dx = \frac{x^3}{3} - \frac{x^2}{2} + 2x - \frac{9}{4}\log|x+1| - \frac{1}{2(x+1)} + \frac{1}{4}\log|x-1| + C.$$

Exercise 3.6.1. Compute the integral.

1. $\int \frac{x^2dx}{1+x}$.
2. $\int \frac{dx}{x^2+x-2}$.
3. $\int \frac{xdx}{x^2+x-2}$.
4. $\int \frac{x^5dx}{x^2+x-2}$.
5. $\int \frac{(2-x)^2dx}{2-x^2}$.
6. $\int \frac{x^4dx}{1-x^2}$.
7. $\int \frac{dx}{x(1+x)(2+x)}$.
8. $\int \frac{dx}{x^2(1+x)}$.
9. $\int \frac{dx}{(x+a)^2(x+b)^2}$.

The examples above illustrate how to integrate rational functions of the form $\frac{b_mx^m + b_{m-1}x^{m-1} + \cdots + b_1x + b_0}{(x-a_1)^{n_1}(x-a_2)^{n_2}\cdots(x-a_k)^{n_k}}$. This means exactly that all the roots of the denominator are real. In general, however, a real polynomial may have complex roots, and a conjugate pair of complex roots corresponds to a real quadratic factor.

Example 3.6.4. To integrate $\frac{1}{x^3-1}$, we note that $x^3 = (x-1)(x^2+x+1)$, where x^2+x+1 has a conjugate pair of complex roots. We postulate

$$\frac{1}{x^3-1} = \frac{A}{x-1} + \frac{Bx+C}{x^2+x+1}.$$

This means $1 = A(x^2+x+1) + (Bx+C)(x-1)$ and gives

$$x=0\colon 1 = A - C; \quad x=1\colon 1 = 3A; \quad x^2\colon 0 = A + B.$$

Therefore $A = \frac{1}{3}$, $B = -\frac{1}{3}$, $C = -\frac{2}{3}$, and

$$\begin{aligned}
\int \frac{dx}{x^3-1} &= \frac{1}{3}\int \frac{dx}{x-1} + \frac{1}{3}\int \frac{-x-2}{x^2+x+1}dx \\
&= \frac{1}{3}\log|x-1| - \frac{1}{6}\int \frac{d(x^2+x+1)}{x^2+x+1} + \frac{1}{2}\int \frac{dx}{x^2+x+1} \\
&= \frac{1}{3}\log|x-1| - \frac{1}{6}\log(x^2+x+1) - \frac{1}{2}\int \frac{dx}{\left(x+\frac{1}{2}\right)^2 + \left(\frac{\sqrt{3}}{2}\right)^2} \\
&= \frac{1}{6}\log\frac{(x-1)^2}{x^2+x+1} - \frac{1}{2}\frac{1}{\frac{\sqrt{3}}{2}}\arctan\frac{x+\frac{1}{2}}{\frac{\sqrt{3}}{2}} + C \\
&= \frac{1}{6}\log\frac{(x-1)^2}{x^2+x+1} - \frac{1}{\sqrt{3}}\arctan\frac{2x+1}{\sqrt{3}} + C.
\end{aligned}$$

Example 3.6.5. To integrate $\frac{x^2}{(x^4-1)^2}$, we postulate

$$\begin{aligned}
\frac{x^2}{(x^4-1)^2} &= \frac{x^2}{(x-1)^2(x+1)^2(x^2+1)^2} \\
&= \frac{A_1}{x-1} + \frac{A_2}{(x-1)^2} + \frac{B_1}{x+1} + \frac{B_2}{(x+1)^2} + \frac{C_1x+D_1}{x^2+1} + \frac{C_2x+D_2}{(x^2+1)^2}.
\end{aligned}$$

Since changing x to $-x$ does not change the left side, we see that $A_1 = -B_1$, $A_2 = B_2$, $C_1 = C_2 = 0$, and the equality becomes

$$\frac{x^2}{(x^4-1)^2} = \frac{2A_1}{x^2-1} + 2A_2\frac{x^2+1}{(x^2-1)^2} + \frac{D_1}{x^2+1} + \frac{D_2}{(x^2+1)^2}.$$

It is then easy to find $A_1 = -\frac{1}{16}$, $A_2 = \frac{1}{16}$, $D_1 = 0$, $D_2 = -\frac{1}{4}$. Therefore with the help of Example 3.5.15,

$$\begin{aligned}
\int \frac{x^2}{(x^4-1)^2}dx &= \int \left(-\frac{1}{16(x-1)} + \frac{1}{16(x-1)^2}\right. \\
&\qquad \left. + \frac{1}{16(x+1)} + \frac{1}{16(x+1)^2} - \frac{1}{4(x^2+1)^2}\right) dx \\
&= \frac{1}{16}\log\left|\frac{x+1}{x-1}\right| - \frac{1}{16(x-1)} \\
&\qquad - \frac{1}{16(x+1)} - \frac{x}{8(x^2+1)} - \frac{1}{8}\arctan x + C \\
&= -\frac{x^3}{4(x^4-1)} + \frac{1}{16}\log\left|\frac{x+1}{x-1}\right| - \frac{1}{8}\arctan x + C.
\end{aligned}$$

Example 3.6.6. To integrate the rational function $\frac{(x^2+1)^4}{(x^3-1)^2}$, we first notice that the degree of the numerator is higher. Therefore we divide $(x^2+1)^4$ by $(x^3-1)^2$.

$$
\begin{array}{r|l}
 & x^2 + 4 \\
\hline
x^6 - 2x^3 + 1 & x^8 + 4x^6 + 6x^4 + 4x^2 + 1 \\
 & -x^8 + 2x^5 - x^2 \\
\hline
 & 4x^6 + 2x^5 + 6x^4 + 3x^2 + 1 \\
 & -4x^6 + 8x^3 - 4 \\
\hline
 & 2x^5 + 6x^4 + 8x^3 + 3x^2 - 3
\end{array}
$$

This means that

$$\frac{(x^2+1)^4}{(x^3-1)^2} = x^2 + 4 + \frac{2x^5+6x^4+8x^3+3x^2-3}{(x^3-1)^2}.$$

Since $x^3 - 1 = (x-1)(x^2+x+1)$, we postulate

$$\frac{2x^5+6x^4+8x^3+3x^2-3}{(x^3-1)^2} = \frac{A_1}{x-1} + \frac{A_2}{(x-1)^2} + \frac{B_1x+C_1}{x^2+x+1} + \frac{B_2x+C_2}{(x^2+x+1)^2}.$$

This can be interpreted as an expression for $2x^5+6x^4+8x^3+3x^2-3$, which gives

$$
\begin{aligned}
x = 0&\colon\ -3 = -A_1 + A_2 + C_1 + C_2, \\
x = 1&\colon\ 16 = 9A_2, \\
x = -1&\colon\ -4 = -2A_1 + A_2 + 4(-B_1+C_1) + 4(-B_2+C_2), \\
x^5&\colon\ 2 = A_1 + B_1, \\
x^4&\colon\ 6 = A_1 + A_2 - B_1 + C_1, \\
\left.\frac{d}{dx}\right|_{x=1}&\colon\ 32 = 9A_1 + 18A_2.
\end{aligned}
$$

Solving the system, we get

$$A_2 = \frac{32}{9}, \quad A_1 = \frac{16}{9}, \quad B_1 = -\frac{14}{9}, \quad C_1 = -\frac{8}{9}, \quad B_2 = 0, \quad C_2 = -\frac{1}{3}.$$

Then

$$\begin{aligned}
\int \frac{(x^2+1)^4}{(x^3-1)^2}dx &= \frac{1}{3}x^3 + 4x + \int \frac{32}{9(x-1)}dx + \int \frac{16}{9(x-1)^2}dx \\
&\quad - \int \frac{14x+8}{9(x^2+x+1)}dx - \int \frac{1}{3(x^2+x+1)^2}dx \\
&= \frac{1}{3}x^3 + 4x + \frac{32}{9}\ln|x-1| - \frac{16}{9(x-1)} \\
&\quad - \int \frac{7d(x^2+x+1)}{9(x^2+x+1)} - \int \frac{dx}{9(x^2+x+1)} - \int \frac{dx}{3(x^2+x+1)^2} \\
&= \frac{1}{3}x^3 + 4x + \frac{32}{9}\log|x-1| - \frac{16}{9(x-1)} - \frac{7}{9}\log(x^2+x+1) \\
&\quad - \int \frac{dx}{9(x^2+x+1)} - \int \frac{dx}{3(x^2+x+1)^2}.
\end{aligned}$$

By

$$x^2+x+1 = \left(x+\frac{1}{2}\right)^2 + \left(\frac{\sqrt{3}}{2}\right)^2, \qquad \frac{x+\frac{1}{2}}{\frac{\sqrt{3}}{2}} = \frac{2x+1}{\sqrt{3}},$$

and Example 3.5.15, we have

$$\int \frac{dx}{x^2+x+1} = \frac{2}{\sqrt{3}}\arctan\frac{2x+1}{\sqrt{3}} + C,$$

and

$$\begin{aligned}
\int \frac{dx}{(x^2+x+1)^2} &= \frac{x+\dfrac{1}{2}}{2\left(\frac{\sqrt{3}}{2}\right)^2(x^2+x+1)} + \frac{1}{2\left(\frac{\sqrt{3}}{2}\right)^3}\arctan\frac{2x+1}{\sqrt{3}} + C \\
&= \frac{2x+1}{3(x^2+x+1)} + \frac{4}{3\sqrt{3}}\arctan\frac{2x+1}{\sqrt{3}} + C.
\end{aligned}$$

Combining everything together, we get

$$\begin{aligned}
\int \frac{(x^2+1)^4}{(x^3-1)^2}dx &= \frac{1}{3}x^3 + 4x - \frac{6x^2+5x+5}{3(x^3-1)} + \frac{13}{3}\log|x-1| - \frac{7}{9}\log|x^3-1| \\
&\quad - \frac{2}{3\sqrt{3}}\arctan\frac{2x+1}{\sqrt{3}} + C.
\end{aligned}$$

In general, the numerator of a rational function is a product of $(x+a)^m$ and $(x^2+bx+c)^n$, where $b^2 < 4c$ so that the factor x^2+bx+c has no real root. After dividing the numerator by the denominator, we can make sure the numerator has lower degree than the denominator. When the numerator has lower degree, the rational function can then be expressed as a sum: For each factor $(x+a)^m$ of the denominator, we have terms

$$\frac{A_1}{x+a}+\frac{A_2}{(x+a)^2}+\cdots+\frac{A_m}{(x+a)^m},$$

and for each factor $(x^2+bx+c)^n$ of the denominator, we have terms

$$\frac{B_1x+C_1}{x^2+bx+c}+\frac{B_2x+C_2}{(x^2+bx+c)^2}+\cdots+\frac{B_nx+C_n}{(x^2+bx+c)^n}.$$

The computation is then reduced to the integration of the terms.

We have

$$\int\frac{A}{(x+a)^m}dx=\begin{cases}-\dfrac{1}{(m-1)(x+a)^{m-1}}+C, & \text{if } m>1,\\ \log|x+a|+C, & \text{if } m=1.\end{cases}$$

The quadratic term can be split into two parts

$$\int\frac{Bx+C}{(x^2+bx+c)^n}dx=\frac{B}{2}\int\frac{d(x^2+bx+c)}{(x^2+bx+c)^n}+\left(C-\frac{B}{2}\right)\int\frac{dx}{(x^2+bx+c)^n}.$$

The first part is easy to compute

$$\int\frac{d(x^2+bx+c)}{(x^2+bx+c)^n}=\begin{cases}-\dfrac{1}{(n-1)(x^2+bx+c)^{n-1}}+C, & \text{if } n>1,\\ \log|x^2+bx+c|+C, & \text{if } n=1.\end{cases}$$

The second part can be computed by the recursive relation in Example 3.5.15

$$\begin{aligned}&\int\frac{dx}{(x^2+bx+c)^n}\\ &=\frac{1}{(4c-b^2)(n-1)}\left(\frac{2x+b}{(x^2+bx+c)^{n-1}}+2(2n-3)\int\frac{dx}{(x^2+bx+c)^{n-1}}\right).\end{aligned}$$

For $n=1,2$, we have

$$\begin{aligned}\int\frac{dx}{x^2+bx+c}&=\frac{2}{\sqrt{4c-b^2}}\arctan\frac{2x+b}{\sqrt{4c-b^2}}+C,\\ \int\frac{dx}{(x^2+bx+c)^2}&=\frac{2x+b}{(4c-b^2)(x^2+bx+c)}+\frac{4}{(4c-b^2)^{\frac{3}{2}}}\arctan\frac{2x+b}{\sqrt{4c-b^2}}+C.\end{aligned}$$

Exercise 3.6.2. Compute the integral.

1. $\displaystyle\int \frac{(1+x)^2 dx}{1+x^2}$.
2. $\displaystyle\int \frac{(2x^2+3)dx}{x^3+x^2-2}$.
3. $\displaystyle\int \frac{dx}{x^3-1}$.
4. $\displaystyle\int \frac{(x+1)^3 dx}{x^3+1}$.
5. $\displaystyle\int \frac{dx}{(x+1)(x^2+1)}$.
6. $\displaystyle\int \frac{dx}{x(x+1)(x^2+x+1)}$.
7. $\displaystyle\int \frac{dx}{(x+1)(x^2+1)(x^3+1)}$.
8. $\displaystyle\int \frac{dx}{(x^2+a^2)(x^2+b^2)}$.
9. $\displaystyle\int \frac{x^3 dx}{(x^2+1)^2}$.
10. $\displaystyle\int \frac{x^2 dx}{(x^2+4x+6)^2}$.
11. $\displaystyle\int \frac{dx}{x^4-1}$.
12. $\displaystyle\int \frac{x^2 dx}{x^4-1}$.
13. $\displaystyle\int \frac{dx}{(x^4-1)^2}$.
14. $\displaystyle\int \frac{dx}{x^4+4}$.
15. $\displaystyle\int \frac{dx}{x^4+x^2+1}$.
16. $\displaystyle\int \frac{dx}{x^6+1}$.
17. $\displaystyle\int \frac{x dx}{(x-1)^2(x^2+2x+2)}$.
18. $\displaystyle\int \frac{(x^4+4x^3+4x^2+4x+4)dx}{x(x+2)(x^2+2x+2)^2}$.

3.6.2 Rational Function of $\sqrt[n]{\dfrac{ax+b}{cx+d}}$

Using suitable changes of variables, some integrals can be changed to integrals of rational functions.

Example 3.6.7. To integrate $\sqrt{\dfrac{x-2}{x-1}}$, we introduce

$$y=\sqrt{\frac{x-2}{x-1}},\quad x=\frac{y^2-2}{y^2-1},\quad dx=\frac{2y}{(y^2-1)^2}dy.$$

Then for $x\ge 2$, we have

$$\begin{aligned}\int\sqrt{\frac{x-2}{x-1}}dx &= \int y\frac{2y}{(y^2-1)^2}dy = \int \frac{1}{2}\left(\frac{1}{y-1}-\frac{1}{y+1}+\frac{1}{(y-1)^2}+\frac{1}{(y+1)^2}\right)\\ &= \frac{1}{2}\log\left|\frac{y-1}{y+1}\right| - \frac{y}{y^2-1}+C\\ &= \frac{1}{2}\log\left|\frac{\sqrt{x-2}-\sqrt{x-1}}{\sqrt{x-2}+\sqrt{x-1}}\right| - \frac{\sqrt{\frac{x-2}{x-1}}}{\frac{x-2}{x-1}-1}+C\\ &= \log(\sqrt{x-1}-\sqrt{x-2})+\sqrt{(x-1)(x-2)}+C.\end{aligned}$$

For $x\le 1$, the answer is $\log(\sqrt{2-x}-\sqrt{1-x})-\sqrt{(2-x)(1-x)}+C$.

The same substituting can be used to compute that, for $x>2$,

$$\begin{aligned}\int\frac{dx}{x\sqrt{x^2-3x+2}}&=\int\frac{dx}{x\sqrt{(x-1)(x-2)}}=\int\sqrt{\frac{x-2}{x-1}}\frac{dx}{x(x-2)}\\&=\int\sqrt{\frac{x-2}{x-1}}\frac{dx}{x(x-2)}=-\int\frac{dy}{y^2-2}\\&=\frac{1}{2\sqrt{2}}\log\left|\frac{y+\sqrt{2}}{y-\sqrt{2}}\right|+C=\frac{1}{2\sqrt{2}}\log\frac{\sqrt{2x-2}+\sqrt{x-2}}{\sqrt{2x-2}-\sqrt{x-2}}+C.\end{aligned}$$

For the case $x<1$, the answer is $-\dfrac{1}{2\sqrt{2}}\log\dfrac{\sqrt{2-2x}+\sqrt{2-x}}{\sqrt{2-2x}-\sqrt{2-x}}+C$.

Example 3.6.8. For $y=\sqrt[3]{\dfrac{x}{x+1}}$, we have $\dfrac{1}{y^3}=1+\dfrac{1}{x}$ and

$$-\frac{3}{y^4}dy=-\frac{1}{x^2}dx=-\left(1-\frac{1}{y^3}\right)^2dx.$$

Therefore by Example 3.6.4, we have

$$\begin{aligned}\int\frac{dx}{\sqrt[3]{x^3+x^2}}&=\int\frac{1}{x}\sqrt[3]{\frac{x}{x+1}}dx=\int\left(1-\frac{1}{y^3}\right)y\frac{-\frac{3}{y^4}dy}{-\left(1-\frac{1}{y^3}\right)^2}=\int\frac{3dy}{y^3-1}\\&=\frac{1}{2}\log\frac{(y-1)^2}{y^2+y+1}-\sqrt{3}\arctan\frac{2y+1}{\sqrt{3}}+C\\&=\frac{1}{2}\log\frac{(y-1)^3}{y^3-1}-\sqrt{3}\arctan\frac{2y+1}{\sqrt{3}}+C\\&=\frac{3}{2}\log(\sqrt[3]{x+1}-\sqrt[3]{x})-\sqrt{3}\arctan\frac{1}{\sqrt{3}}\left(2\sqrt[3]{\frac{x}{x+1}}+1\right)+C.\end{aligned}$$

In general, a function involving $\sqrt[n]{\dfrac{ax+b}{cx+d}}$ can be integrated by introducing

$$y=\sqrt[n]{\frac{ax+b}{cx+d}},\quad x=\frac{dy^n-b}{-cy^n+a},\quad dx=n(ad-bc)\frac{y^{n-1}}{(cy^n-a)^2}dy.$$

Exercise 3.6.3. Compute the integral.

1. $\displaystyle\int\frac{dx}{1+\sqrt{x}}$.
2. $\displaystyle\int\frac{dx}{\sqrt{x}(1+x)}$.
3. $\displaystyle\int\frac{x\sqrt[3]{2+x}}{x+\sqrt[3]{2+x}}dx$.

4. $\displaystyle\int \frac{x^3}{\sqrt[3]{x^2+1}}dx.$
5. $\displaystyle\int \frac{dx}{\sqrt[3]{x}+\sqrt[4]{x}}dx.$
6. $\displaystyle\int \frac{dx}{\sqrt[4]{x^3(a-x)}}.$
7. $\displaystyle\int \frac{\sqrt{x+1}+\sqrt{x-1}}{\sqrt{x+1}-\sqrt{x-1}}dx.$
8. $\displaystyle\int \sqrt{\frac{1-x}{1+x}}dx.$
9. $\displaystyle\int \frac{1}{x^2}\sqrt{\frac{1-x}{1+x}}dx.$
10. $\displaystyle\int \frac{1}{x^2}\sqrt[3]{\frac{1-x}{1+x}}dx.$
11. $\displaystyle\int \frac{dx}{2\sqrt{x}+\sqrt{x+1}+1}.$
12. $\displaystyle\int \frac{dx}{\sqrt{x+a}+\sqrt{x+b}+c}.$
13. $\displaystyle\int \frac{1+\sqrt{x+a}}{1+\sqrt{x+b}}dx.$
14. $\displaystyle\int \sqrt{\frac{x-a}{x-b}}dx.$
15. $\displaystyle\int \sqrt{\frac{x-a}{b-x}}dx.$
16. $\displaystyle\int \sqrt{(x-a)(x-b)}dx.$
17. $\displaystyle\int \sqrt{(x-a)(b-x)}dx.$
18. $\displaystyle\int x\sqrt{(x-a)(b-x)}dx.$
19. $\displaystyle\int x\sqrt{\frac{x-a}{x-b}}dx.$
20. $\displaystyle\int \frac{xdx}{\sqrt{(x-a)(x-b)}}.$

Exercise 3.6.4. Compute the integral.

1. $\displaystyle\int \sqrt{1+e^x}dx.$
2. $\displaystyle\int \frac{dx}{\sqrt{1+e^x}+\sqrt{1-e^x}}.$
3. $\displaystyle\int \frac{dx}{\sqrt{a^x+b}}.$

Exercise 3.6.5. 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 $\int R(x,(ax+b)^r(cx+d)^s)dx$ is changed into the antiderivative of a rational function.

Exercise 3.6.6. Suppose r, s, t are rational numbers. For each of the following cases, find a suitable change of variable, such that $\int x^r(a+bx^s)^t dx$ is changed into the integral 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[1] says that these are the only cases that the antiderivative can be changed to the integral of a rational function.

[1] 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.

3.6.3 Rational Function of $\sin x$ and $\cos x$

A rational function of $\sin x$ and $\cos x$ can be integrated by introducing

$$y = \tan\frac{x}{2}, \quad \sin x = \frac{2y}{1+y^2}, \quad \cos x = \frac{1-y^2}{1+y^2}, \quad dx = \frac{2}{1+y^2}dy.$$

Example 3.6.9. For $a \neq 0$, we have

$$\int \frac{dx}{a+\sin x} = \int \frac{2dy}{\left(a + \frac{2y}{1+y^2}\right)(1+y^2)} = \int \frac{2dy}{ay^2+2y+a}.$$

If $|a| > 1$, then

$$\int \frac{dx}{a+\sin x} = \frac{1}{a\sqrt{1-\frac{1}{a^2}}} \arctan \frac{y+\frac{1}{a}}{\sqrt{1-\frac{1}{a^2}}} + C = \frac{2}{\operatorname{sign}(a)\sqrt{a^2-1}} \arctan \frac{a\tan\frac{x}{2}+1}{\sqrt{a^2-1}} + C.$$

If $|a| < 1$, then

$$\begin{aligned}\int \frac{dx}{a+\sin x} &= \frac{1}{a\sqrt{\frac{1}{a^2}-1}} \log\left|\frac{y+\frac{1}{a}-\sqrt{\frac{1}{a^2}-1}}{y+\frac{1}{a}+\sqrt{\frac{1}{a^2}-1}}\right| + C \\ &= \frac{1}{\operatorname{sign}(a)\sqrt{1-a^2}} \log\left|\frac{a\tan\frac{x}{2}+1-\sqrt{1-a^2}}{a\tan\frac{x}{2}+1+\sqrt{1-a^2}}\right| + C.\end{aligned}$$

If $|a| = 1$, then

$$\int \frac{dx}{a+\sin x} = \frac{-2}{ay+1} + C = \frac{-2}{a\tan\frac{x}{2}+1} + C.$$

The example can be extended to $\displaystyle\int \frac{dx}{a+b\sin x+c\cos x}$. We have

$$b\sin x + c\cos x = \sqrt{b^2+c^2}\sin(x+\theta),$$

where θ is any fixed angle satisfying $\sin\theta = \dfrac{b}{\sqrt{b^2+c^2}}$ and $\cos\theta = \dfrac{c}{\sqrt{b^2+c^2}}$. Then

$$\int \frac{dx}{a+b\sin x+c\cos x} = \int \frac{dy}{a+\sqrt{b^2+c^2}\sin y}, \quad y = x+\theta.$$

Example 3.6.10. Rational functions of $\sin x$ and $\cos x$ can be integrated by a simpler substitution if it has additional property. For example, to integrate the function $\dfrac{\sin x}{\sin x+\cos x}$, we introduce

$$y = \tan x, \quad x = \arctan y, \quad dx = \frac{dy}{y^2+1}.$$

Here we use the tangent of the full angle x, instead of half the angle in Example 3.6.9. Then

$$\int \frac{\sin x dx}{\sin x + \cos x} = \int \frac{\tan x dx}{\tan x + 1} = \int \frac{y \frac{dy}{y^2+1}}{y+1} = \frac{1}{4}\log\frac{y^2+1}{(y+1)^2} + \frac{1}{2}\arctan y + C$$
$$= \frac{1}{2}x - \frac{1}{2}\log|\sin x + \cos x| + C.$$

The key point here is that the integrand is a rational function $R(\sin x, \cos x)$ that satisfies $R(-u,-v) = R(u,v)$. In this case, the integrad can always be written as a rational function of $\tan x$, and the change of variable can be applied.

Example 3.6.11. Note that rational function $\dfrac{\cos x}{\cos x \sin x + \sin^3 x}$ of $\sin x$ and $\cos x$ is odd in the $\sin x$ variable. This is comparable to the function $\cos^m x \sin^n x$ for the case n is odd. We may introduce the same change of variable $y = \cos x$, $dy = -\sin x dx$ like the earlier example and get

$$\int \frac{\cos x dx}{\cos x \sin x + \sin^3 x} = \int \frac{\cos x \sin x dx}{\cos x \sin^2 x + (\sin^2 x)^2} = \int \frac{-y dy}{y(1-y^2) + (1-y^2)^2}$$
$$= \frac{1}{\sqrt{5}}\log\left|\frac{y - \frac{1+\sqrt{5}}{2}}{y - \frac{1-\sqrt{5}}{2}}\right| + \frac{1}{2}\log\left|\frac{y-1}{y+1}\right| + C$$
$$= \frac{1}{\sqrt{5}}\log\left|\frac{2\cos x - 1 - \sqrt{5}}{2\cos x - 1 + \sqrt{5}}\right| + \frac{1}{2}\log\frac{1-\cos x}{1+\cos x} + C.$$

Similarly, a rational function of $\sin x$ and $\cos x$ that is odd in the $\cos x$ variable can be integrated by introducing $x = \sin y$. If $R(-u,-v) = R(u,v)$, then we may introduce $y = \tan x$ to compute $\displaystyle\int R(\sin x, \cos x)dx$.

Exercise 3.6.7. Compute the integral.

1. $\displaystyle\int \frac{1-r^2}{1-2r\cos x + r^2}dx$, $|r| < 1$.
2. $\displaystyle\int \frac{dx}{a - \cos 2x}$.
3. $\displaystyle\int \frac{dx}{a + \tan x}$.
4. $\displaystyle\int \frac{dx}{\cos x + \tan x}$.
5. $\displaystyle\int \frac{dx}{\sin x + \tan x}$.
6. $\displaystyle\int \frac{dx}{2\sin x + \sin 2x}$.
7. $\displaystyle\int \frac{\sin^2 x}{1 + \sin^2 x}dx$.
8. $\displaystyle\int \frac{dx}{\sin(x+a)\ \sin(x+b)}$.
9. $\displaystyle\int \frac{dx}{(1+\cos^2 x)(2+\sin^2 x)}$.
10. $\displaystyle\int \frac{(1+\sin x)dx}{\sin x(1+\cos x)}$.

11. $\displaystyle\int \frac{dx}{(a+\cos x)\sin x}$.

12. $\displaystyle\int \frac{(\sin x+\cos x)dx}{\sin x(\sin x-\cos x)}$.

13. $\displaystyle\int \frac{dx}{(a+\cos^2 x)\sin x}$.

14. $\displaystyle\int \frac{dx}{a^2\sin^2 x+b^2\cos^2 x}$.

15. $\displaystyle\int \frac{1-\tan x}{1+\tan x}dx$.

16. $\displaystyle\int \sqrt{\tan x}dx$.

3.7 Improper Integral

The definition of Riemann integral requires both the function and the interval to be bounded. If either the function or the interval is unbounded, then the integral is *improper*. We may still make sense of an improper integral if it can be viewed as the limit of usual integral of bounded function on bounded interval.

3.7.1 Definition and Property

Example 3.7.1. The function e^{-x} is bounded on the unbounded interval $[0,+\infty)$. To make sense of the improper integral $\displaystyle\int_0^{+\infty} e^{-x}dx$, we consider the integral on any bounded interval

$$\int_0^b e^{-x}dx = 1-e^{-b}.$$

As the bounded interval approaches $[0,+\infty)$, we get

$$\lim_{b\to+\infty}\int_0^b e^{-x}dx = \lim_{b\to+\infty}(1-e^{-b}) = 1.$$

Therefore the improper integral $\displaystyle\int_0^{+\infty} e^{-x}dx$ has value 1. Geometrically, this means that the area of the unbounded region under the graph of the function e^{-x} and over the interval $[0,+\infty)$ is 1.

Example 3.7.2. The function $\log x$ is unbounded on the bounded interval $(0,1]$. Since the integral $\displaystyle\int_0^1 \log x dx$ is improper at 0^+, we consider the integral over $[\epsilon,1]$ for $\epsilon>0$

$$\int_\epsilon^1 \log x dx = (x\log x - x)\Big|_{x=\epsilon}^{x=1} = -1-\epsilon\log\epsilon+\epsilon.$$

Since the right side converges to -1 as $\epsilon\to0^+$, the improper integral converges and has value

$$\int_0^1 \log x dx = \lim_{\epsilon\to0^+}\int_\epsilon^1 \log x dx = -1.$$

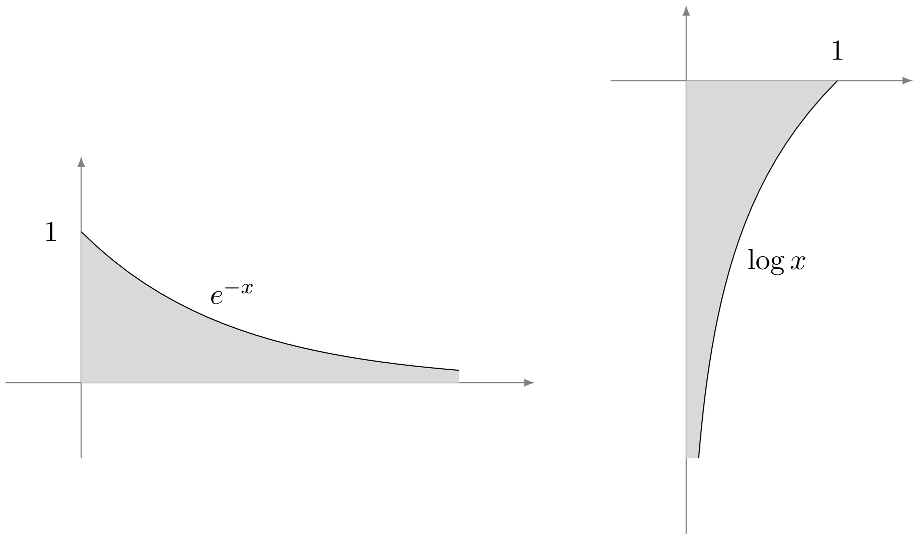


Figure 3.7.1: The unbounded region has area 1.

The unbounded region measured by this improper integral is actually the same as the one in Example 3.7.1, up to a rotation.

Example 3.7.3. Consider the improper integral $\displaystyle\int_a^{+\infty} \frac{dx}{x^p}$, where $a > 0$. We have

$$\int_a^b \frac{dx}{x^p} = \begin{cases} \dfrac{b^{1-p} - a^{1-p}}{1-p}, & \text{if } p \neq 1, \\ \log b - \log a, & \text{if } p = 1. \end{cases}$$

As $b \to +\infty$, we get

$$\int_a^{+\infty} \frac{dx}{x^p} = \begin{cases} \dfrac{a^{1-p}}{p-1}, & \text{if } p > 1, \\ \text{diverge}, & \text{if } p \leq -1. \end{cases}$$

Example 3.7.4. The integral $\displaystyle\int_0^1 \frac{dx}{x^p}$ is improper at 0^+ for $p > 0$. For $\epsilon > 0$, we have

$$\int_\epsilon^1 \frac{dx}{x^p} = \begin{cases} \dfrac{1 - \epsilon^{1-p}}{1-p}, & \text{if } p \neq 1, \\ -\log \epsilon, & \text{if } p = 1. \end{cases}$$

As $\epsilon \to 0^+$, we get

$$\int_0^1 \frac{dx}{x^p} = \begin{cases} \dfrac{1}{1-p}, & \text{if } p < 1, \\ \text{diverge}, & \text{if } p \geq 1. \end{cases}$$

By the same argument, for $a < b$, the improper integrals $\int_a^b (x-a)^p dx$ and $\int_a^b (b-x)^p dx$ converge if and only if $p > -1$.

Example 3.7.5. The integral $\int_{-\infty}^{+\infty} \frac{dx}{x^2+1}$ is improper at $+\infty$ and $-\infty$. The integral on a bounded interval is

$$\int_a^b \frac{dx}{x^2+1} = \arctan b - \arctan a.$$

Then we get

$$\int_{-\infty}^{+\infty} \frac{dx}{x^2+1} = \lim_{\substack{a\to-\infty \\ b\to+\infty}} \int_a^b \frac{dx}{x^2+1} = \lim_{b\to+\infty} \arctan b - \lim_{a\to-\infty} \arctan a = \frac{\pi}{2} - \left(-\frac{\pi}{2}\right) = \pi.$$

Example 3.7.6. Since

$$\int_0^b \cos x dx = \sin b$$

diverges as $b \to +\infty$, the improper integral $\int_0^{+\infty} \cos x dx$ diverges.

In general, an integral may be improper at several places. We may divide the interval into several parts, such that each part contains exactly one improperness. If an integral has one improperness at $+\infty$ or $-\infty$, then we study the limit of the integral on bounded intervals. If an integral has one improperness at a^+ or a^-, then we study the limit of the integral on intervals $[a+\epsilon, b]$ or $[b, a-\epsilon]$.

Example 3.7.7. A naive application of the Newton-Leibniz formula would tell us

$$\int_{-1}^1 \frac{dx}{x} = (\log|x|)\Big|_{x=-1}^{x=1} = \log 1 - \log 1 = 0.$$

However, the computation is wrong since the integrand $\frac{1}{x}$ is not continuous on $[-1,1]$. In fact, the integral $\int_{-1}^1 \frac{dx}{x}$ is improper on both sides of 0, and we need both improper integrals $\int_{-1}^0 \frac{dx}{x}$ and $\int_0^1 \frac{dx}{x}$ to converge and then get

$$\int_{-1}^1 \frac{dx}{x} = \int_{-1}^0 \frac{dx}{x} + \int_0^1 \frac{dx}{x}.$$

Since

$$\int_0^1 \frac{dx}{x} = \lim_{\epsilon\to 0^+} \int_\epsilon^1 \frac{dx}{x} = \lim_{\epsilon\to 0^+} -\log\epsilon = +\infty$$

diverges, the improper integral $\int_0^1 \frac{dx}{x}$ diverges, so that $\int_{-1}^1 \frac{dx}{x}$ also diverges.

Example 3.7.8. To compute the improper integral $\int_{-\infty}^0 xe^x dx$, we start with integration by parts on a bounded interval

$$\int_b^0 xe^x dx = \int_b^0 xde^x = -be^b - \int_b^0 e^x dx = -be^b - 1 + e^b.$$

Taking $b\to -\infty$ on both sides, we get

$$\int_{-\infty}^0 xe^x dx = -1.$$

The example shows that the integration by parts can be extended to improper integrals, simply by taking the limit of the integration by parts formula for the usual proper integrals.

Example 3.7.9. For $a>1$, consider the improper integral $\int_a^{+\infty} \frac{dx}{x(\log x)^p}$. We have

$$\int_a^b \frac{dx}{x(\log x)^p} = \int_a^b \frac{d(\log x)}{(\log x)^p} = \int_{\log a}^{\log b} \frac{dy}{y^p}.$$

Taking $b\to +\infty$ on both sides, we get

$$\int_a^{+\infty} \frac{dx}{x(\log x)^p} = \int_{\log a}^{+\infty} \frac{dy}{y^p}.$$

The equality means that the improper integral on the left converges if and only if the improper integral on the right converges, and the two values are the same. By Example 3.7.3, we see that the improper integral $\int_a^{+\infty} \frac{dx}{x(\log x)^p}$ converges if and only if $p<1$, and

$$\int_a^{+\infty} \frac{dx}{x(\log x)^p} = -\frac{(\log a)^{p+1}}{p+1}, \quad \text{if } p<1.$$

The example shows that the change of variable can also be extended to improper integrals, simply by taking the limit of the change of variable formula for the usual proper intervals.

Exercise 3.7.1. Determine the convergence of improper integrals and evaluate the convergent ones.

1. $\displaystyle\int_0^{+\infty} x^p dx$.
2. $\displaystyle\int_0^1 \frac{dx}{x(-\log x)^p}$.
3. $\displaystyle\int_0^{+\infty} a^x dx$.
4. $\displaystyle\int_{-\infty}^0 a^x dx$.
5. $\displaystyle\int_{-1}^1 \frac{dx}{1-x^2}$.
6. $\displaystyle\int_2^{+\infty} \frac{dx}{1-x^2}$.
7. $\displaystyle\int_{-1}^1 \frac{dx}{\sqrt{1-x^2}}$.
8. $\displaystyle\int_0^{\frac{\pi}{2}} \tan x dx$.
9. $\displaystyle\int_0^{\pi} \sec x dx$.

Exercise 3.7.2. Determine the convergence of improper integrals and evaluate the convergent ones.

1. $\displaystyle\int_1^{+\infty} \frac{dx}{x+1}$.
2. $\displaystyle\int_{-\infty}^{+\infty} \frac{xdx}{x^2+1}$.
3. $\displaystyle\int_1^{+\infty} \frac{dx}{\sqrt[3]{x+1}}$.
4. $\displaystyle\int_0^{+\infty} \frac{x^2dx}{x^3+1}$.
5. $\displaystyle\int_0^{+\infty} \frac{x^2dx}{(x^3+1)^2}$.
6. $\displaystyle\int_0^{+\infty} \frac{dx}{x(x+1)(x+2)}$.
7. $\displaystyle\int_0^{+\infty} \frac{dx}{\sqrt{x}(1+x)}$.
8. $\displaystyle\int_1^9 \frac{dx}{\sqrt[3]{x-9}}$.
9. $\displaystyle\int_0^{+\infty} xe^x dx$.
10. $\displaystyle\int_0^{+\infty} xe^{-x^2} dx$.
11. $\displaystyle\int_0^{+\infty} e^{-\sqrt{x}} dx$.
12. $\displaystyle\int_0^1 x\log x dx$.
13. $\displaystyle\int_1^{+\infty} \frac{\log x}{x^2} dx$.
14. $\displaystyle\int_0^1 \frac{\log x}{\sqrt{x}} dx$.
15. $\displaystyle\int_0^{+\infty} \frac{x\arctan x}{(1+x^2)^2} dx$.
16. $\displaystyle\int_0^{+\infty} e^{-ax}\cos bx dx$.
17. $\displaystyle\int_0^{+\infty} e^{-ax}\sin bx dx$.
18. $\displaystyle\int_0^{+\infty} e^{-x}|\sin x| dx$.

Exercise 3.7.3. Prove that

$$\lim_{n\to\infty}\frac{1}{n}\left(\log\frac{1}{n}+\log\frac{2}{n}+\cdots+\log\frac{n}{n}\right)=\int_0^1 \log x dx.$$

Note that the left side is a "Riemann sum" for the right side. However, since the integral is improper, we cannot directly use the fact that the Riemann sum converges to the integral. Moreover, the limit is the same as $\lim_{n\to\infty}\dfrac{\sqrt[n]{n!}}{n}=e^{-1}$.

3.7.2 Comparison Test

The improper integral is defined by taking limit. Therefore there is always the problem of convergence.

The convergence of the improper integral $\displaystyle\int_a^{+\infty} f(x)dx$ means the convergence of the function $I(b)=\displaystyle\int_a^b f(x)dx$ as $b\to+\infty$. The *Cauchy criterion* for the convergence is that, for any $\epsilon > 0$, there is N, such that

$$b, c > N \implies |I(c) - I(b)| = \left| \int_b^c f(x)dx \right| < \epsilon.$$

The Cauchy criterion for the convergence of other types of improper integrals is similar. For example, if the integral $\int_a^b f(x)dx$ is improper at a^+, then the integral converges if and only if for any $\epsilon > 0$, there is $\delta > 0$, such that

$$c, d \in (a, a+\delta) \implies \left| \int_c^d f(x)dx \right| < \epsilon.$$

The Cauchy criterion shows that the convergence of an improper integral depends only on the behavior of the function near the improper place. Moreover, the Cauchy criterion also implies the following test for convergence.

Theorem 3.7.1 (Comparison Test). *If $|f(x)| \le g(x)$ on (a, b) and the integral $\int_a^b g(x)dx$ converges, then $\int_a^b f(x)dx$ also converges.*

Note that if $|f| \le g$, then $||f|| \le g$. Therefore whenever we use the comparison test, we may always conclude that $\int_a^b |f(x)|dx$ also converges.

We say that $\int_a^b f(x)dx$ *absolutely converges* if $\int_a^b |f(x)|dx$ converges. The comparison test tells us that absolute convergence implies convergence.

Here we justify the comparison test for the integral $\int_a^{+\infty} f(x)dx$ that is improper at $+\infty$. The convergence of $\int_a^{+\infty} g(x)dx$ implies that for any $\epsilon > 0$, there is N, such that

$$c > b > N \implies \int_b^c g(x)dx < \epsilon.$$

The assumption $|f(x)| \le g(x)$ further implies that for $c > b$,

$$\left| \int_b^c f(x)dx \right| \le \int_b^c |f(x)|dx \le \int_b^c g(x)dx.$$

Combining the two implications, we get

$$c > b > N \implies \left| \int_b^c f(x)dx \right| \le \int_b^c g(x)dx < \epsilon.$$

This verifies the Cauchy criterion for the convergence of $\int_a^{+\infty} f(x)dx$.

Example 3.7.10. We know $\dfrac{1}{\sqrt{x^3+1}} < x^{-\frac{3}{2}}$ for $x \geq 1$. Since $\displaystyle\int_1^{+\infty} x^{-\frac{3}{2}} dx$ converges, by the comparison test, we know $\displaystyle\int_1^{+\infty} \frac{dx}{\sqrt{x^3+1}}$ also converges. We note that $\displaystyle\int_0^{+\infty} \frac{dx}{\sqrt{x^3+1}}$ converges too because only the behavior of the function for big x (i.e., near $+\infty$) is involved.

Example 3.7.11. To determine the convergence of $\displaystyle\int_0^{+\infty} e^{-x^2} dx$, we use $0 < e^{-x^2} \leq e^{-x}$ for $x \geq 1$. By Example 3.7.1 and the comparison test, we know $\displaystyle\int_1^{+\infty} e^{-x^2} dx$ converges. Since $\displaystyle\int_0^{1} e^{-x^2} dx$ is a proper integral, we know $\displaystyle\int_0^{+\infty} e^{-x^2} dx$ also converges.

Example 3.7.12. To determine the convergence of $\displaystyle\int_1^{+\infty} \frac{\log x}{x^p} dx$, $p > 0$, we use the comparison $\dfrac{\log x}{x^p} \geq \dfrac{1}{x^p} > 0$ for $x \geq e$. For $p \leq 1$, the divergence of $\displaystyle\int_1^{+\infty} \frac{1}{x^p} dx$ implies the divergence of $\displaystyle\int_1^{+\infty} \frac{\log x}{x^p} dx$.

For $p > 1$, although we also know the convergence of $\displaystyle\int_1^{+\infty} \frac{1}{x^p} dx$, the comparison above cannot be used to conclude the convergence of $\displaystyle\int_1^{+\infty} \frac{\log x}{x^p} dx$. Instead, we choose q satisfying $p > q > 1$. Then by

$$\lim_{x \to +\infty} \frac{\dfrac{\log x}{x^p}}{\dfrac{1}{x^q}} = \lim_{x \to +\infty} \frac{\log x}{x^{p-q}} = 0,$$

we have $\left| \dfrac{\log x}{x^p} \right| \leq \dfrac{1}{x^q}$ for sufficiently large x. By the convergence of $\displaystyle\int_1^{+\infty} \frac{dx}{x^q}$, therefore, we know the converges of $\displaystyle\int_1^{+\infty} \frac{\log x}{x^p} dx$.

We conclude that $\displaystyle\int_1^{+\infty} \frac{\log x}{x^p} dx$ converges if and only if $p > 1$.

Example 3.7.13. The integral $\displaystyle\int_0^{1} \frac{dx}{\sqrt{x(1-x)}}$ is improper at 0^+ and 1^-. By applying the idea of Example 3.7.12 to

$$\lim_{x\to 0^+}\frac{\dfrac{1}{\sqrt{x(1-x)}}}{\dfrac{1}{\sqrt{x}}}=1,\quad \lim_{x\to 1^-}\frac{\dfrac{1}{\sqrt{x(1-x)}}}{\dfrac{1}{\sqrt{1-x}}}=1,$$

and the convergence of $\int_0^1\frac{dx}{\sqrt{x}}$ and $\int_0^1\frac{dx}{\sqrt{1-x}}$, we know that $\int_0^1\frac{dx}{\sqrt{x(1-x)}}$ converges.

Exercise 3.7.4. Compare the integrals $I=\int_a^{+\infty}f(x)dx$ and $J=\int_a^{+\infty}g(x)dx$ that are improper at $+\infty$.

1. Prove that if $\lim_{x\to+\infty}\dfrac{f(x)}{g(x)}$ converges, $g(x)\ge 0$ for sufficiently large x, and J converges, then I also converges.
2. Prove that if $\lim_{x\to+\infty}\dfrac{f(x)}{g(x)}$ converges to a nonzero number, and $g(x)\ge 0$ for sufficiently large x, then I converges if and only if J converges.

Exercise 3.7.5. Suppose $f\ge 0$, prove that the improper integral $\int_a^{+\infty}f(x)dx$ converges if and only if $\int_a^b f(x)dx$, for all $b\in[a,+\infty)$, is bounded. What about the integral $\int_a^b f(x)dx$ that is improper at a^+.

Exercise 3.7.6. Suppose $\int_a^{+\infty}f^2dx$ and $\int_a^{+\infty}g^2dx$ converge. Prove that $\int_a^{+\infty}fgdx$ and $\int_a^{+\infty}(f+g)^2dx$ converge.

Exercise 3.7.7. Determine convergence.

1. $\displaystyle\int_2^{+\infty}\frac{dx}{x^p(\log x)^q}$.
2. $\displaystyle\int_1^{2}\frac{dx}{x^p(\log x)^q}$.
3. $\displaystyle\int_0^{1}\frac{dx}{x^p|\log x|^q}$.
4. $\displaystyle\int_{-\infty}^{1}\frac{dx}{|x|^p|\log x|^q}$.
5. $\displaystyle\int_2^{+\infty}\frac{dx}{x^p+(\log x)^q}$.
6. $\displaystyle\int_0^{+\infty}\frac{x^pdx}{1+x^q}$.
7. $\displaystyle\int_0^{1}\frac{x^pdx}{\sqrt{1-x^q}}$, $q>0$.
8. $\displaystyle\int_0^{1}\frac{dx}{x^p(1-x^q)^r}$, $q>0$.
9. $\displaystyle\int_0^{1}x^pa^xdx$, $a>0$.
10. $\displaystyle\int_1^{+\infty}x^pa^xdx$, $a>0$.
11. $\displaystyle\int_1^{+\infty}x^p\log(1+x^q)dx$.
12. $\displaystyle\int_0^{1}x^p\log(1+x^q)dx$.

Exercise 3.7.8. Determine convergence.

1. $\displaystyle\int_2^{+\infty} \frac{xdx}{\sqrt{x^5-2x^2+1}}$.
2. $\displaystyle\int_0^{1} \frac{xdx}{\sqrt{x^5-2x^2+1}}$.
3. $\displaystyle\int_2^{+\infty} \frac{x\sin xdx}{\sqrt{x^5-2x^2+1}}$.
4. $\displaystyle\int_0^{1} \frac{x\sin xdx}{\sqrt{x^5-2x^2+1}}$.
5. $\displaystyle\int_2^{+\infty} \frac{x\arctan xdx}{\sqrt{x^5-2x^2+1}}$.
6. $\displaystyle\int_0^{1} \frac{x\arctan xdx}{\sqrt{x^5-2x^2+1}}$.

Exercise 3.7.9. Determine convergence.

1. $\displaystyle\int_1^{+\infty} \frac{x+1}{x^3-2x+3}dx$.
2. $\displaystyle\int_0^{10} \frac{dx}{\sqrt{|x^2-4x+3|}}$.
3. $\displaystyle\int_1^{+\infty} \frac{\log x}{\sqrt{x^p+1}}dx$.
4. $\displaystyle\int_0^{1} (1-x)^p|\log x|^q dx$.
5. $\displaystyle\int_1^{2} \frac{dx}{(3x-2-x^2)^p}$.
6. $\displaystyle\int_0^{+\infty} \frac{\log(1+x^2)dx}{1+x^q}$.
7. $\displaystyle\int_0^{1} \frac{dx}{\sqrt{x+\sqrt{x+\sqrt{x}}}}$.
8. $\displaystyle\int_1^{+\infty} \frac{dx}{\sqrt{x+\sqrt{x+\sqrt{x}}}}$.
9. $\displaystyle\int_0^{+\infty} \frac{dx}{\sqrt{1+\sqrt{1+\sqrt{x}}}}$.

Exercise 3.7.10. Determine convergence.

1. $\displaystyle\int_0^{\frac{\pi}{2}} \frac{dx}{\cos^p x}$.
2. $\displaystyle\int_0^{\frac{\pi}{2}} \frac{dx}{x^p\sin^q x}$.
3. $\displaystyle\int_0^{\frac{\pi}{2}} \frac{dx}{\sin^p x\cos^q x}$.
4. $\displaystyle\int_0^{\frac{\pi}{3}} \tan^p xdx$.
5. $\displaystyle\int_0^{\frac{\pi}{3}} \tan^p x\log^q xdx$.
6. $\displaystyle\int_0^{\frac{\pi}{4}} \frac{dx}{|\sin x-\cos x|^p}$.
7. $\displaystyle\int_{\frac{\pi}{4}}^{\frac{\pi}{2}} \frac{dx}{(1-\sin x)^p}$.
8. $\displaystyle\int_0^{\frac{\pi}{2}} (-\log\sin x)^p dx$.
9. $\displaystyle\int_0^{\frac{\pi}{2}} x^p\log\sin xdx$.
10. $\displaystyle\int_0^{\frac{\pi}{2}} |\log\tan x|^p dx$.
11. $\displaystyle\int_0^{+\infty} e^{ax}\cos bxdx$.
12. $\displaystyle\int_0^{+\infty} e^{-\sqrt{x}}\cos bx^2dx$.
13. $\displaystyle\int_1^{+\infty} \frac{x+\cos x}{x^2-\sin x}dx$.
14. $\displaystyle\int_1^{+\infty} \frac{\log(x+\cos x)}{x^2-\sin x}dx$.
15. $\displaystyle\int_1^{+\infty} \frac{\log x+\cos x}{x^2-\sin x}dx$.

Exercise 3.7.11. Find a constant a, such that $\displaystyle\int_0^{+\infty}\left(\frac{1}{\sqrt{x^2+1}}+\frac{a}{x+1}\right)dx$ converges. Moreover, evaluate the integral for this a.

3.7.3 Conditional Convergence

Although the comparison test is very effective, some improper integrals needs to be modified before the comparison test can be applied.

Example 3.7.14. By the comparison test, we know $\int_1^{+\infty} \frac{\sin x}{x^p} dx$ converges for $p > 1$. However, the argument fails for the case $p = 1$. We will show that $\int_1^{+\infty} \frac{\sin x}{x} dx$ still converges. We will also show that, after taking the absolute value, the integral $\int_1^{+\infty} \left| \frac{\sin x}{x} \right| dx$ actually diverges. This means that the comparison test cannot be directly applied to $\int_1^{+\infty} \frac{\sin x}{x} dx$.

Using integration by parts, we have

$$\int_1^b \frac{\sin x}{x} dx = -\int_1^b \frac{1}{x} d\cos x = -\frac{\cos b}{b} + \cos 1 - \int_1^b \frac{\cos x}{x^2} dx.$$

By the comparison test, the improper integral $\int_1^{+\infty} \frac{\cos x}{x^2} dx$ converges. Therefore the right side converges as $b \to +\infty$, and we conclude that $\int_1^{+\infty} \frac{\sin x}{x} dx$ converges.

On the other hand, we have

$$\begin{aligned} \int_1^{n\pi} \left| \frac{\sin x}{x} \right| dx \ge \int_\pi^{n\pi} \left| \frac{\sin x}{x} \right| dx = \sum_{k=2}^n \int_{(k-1)\pi}^{k\pi} \left| \frac{\sin x}{x} \right| dx \ge \sum_{k=2}^n \frac{1}{k\pi} \int_{(k-1)\pi}^{k\pi} |\sin x| dx \\ \ge \sum_{k=2}^n \frac{1}{k\pi} \ge \frac{1}{\pi} \left(\frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} \right). \end{aligned}$$

By Example 1.3.8, the right side diverges to $+\infty$. Therefore $\int_1^{+\infty} \left| \frac{\sin x}{x} \right| dx$ diverges.

Example 3.7.15. By a change of variable, we have

$$\int_0^{+\infty} \sin x^2 dx = \int_0^{+\infty} \sin y d(\sqrt{y}) = \int_0^{+\infty} \frac{\sin y}{2\sqrt{y}} dy.$$

The integral on the right is *proper* at 0^+ and improper at $+\infty$. It converges by an argument similar to Example 3.7.14. Therefore the integral $\int_0^{+\infty} \sin x^2 dx$ converges.

We first used the change of variable, then used the integration by parts, and finally used the comparison test to conclude the convergence of $\int_0^{+\infty} \sin x^2 dx$.

The reader can further use the idea of Example 3.7.14 to show that $\int_0^{+\infty} |\sin x^2| dx$ diverges.

Example 3.7.14 shows that it is possible for an improper integral $\int_a^b f(x)dx$ to converge but the corresponding absolute improper integral $\int_a^b |f(x)|dx$ to diverge. In this case, the integral converges but not absolutely, and we say $\int_a^b f(x)dx$ *conditionally converges.*

The idea in Example 3.7.14 can be elaborated to get the following useful tests.

Theorem 3.7.2 (Dirichlet Test). *Suppose* $\int_a^b f(x)dx$ *is bounded for all* $b \in [a, +\infty)$. *Suppose* $g(x)$ *is monotonic and* $\lim_{x\to+\infty} g(x) = 0$. *Then* $\int_a^{+\infty} f(x)g(x)dx$ *converges.*

Theorem 3.7.3 (Abel Test). *Suppose* $\int_a^{+\infty} f(x)dx$ *converges. Suppose* $g(x)$ *is monotonic and bounded on* $[a, +\infty)$. *Then* $\int_a^{+\infty} f(x)g(x)dx$ *converges.*

The tests basically replaces $\sin x$ and $\dfrac{1}{x}$ in the example by $f(x)$ and $g(x)$. In case $f(x)$ is continuous and $g(x)$ is continuously differentiable, we can justify the tests by repeating the argument in the example. Let $F(x) = \int_a^x f(t)dt$. Then $F(a) = 0$, and

$$\int_a^b f(x)g(x)dx = \int_a^b g(x)dF(x) = g(b)F(b) - \int_a^b F(x)g'(x)dx.$$

Under the assumption of the Dirichlet test, we have $\lim_{b\to+\infty} g(b)F(b) = 0$, and $|F(x)| < M$ for some constant M and all $x \ge a$. Assume the monotonic function $g(x)$ is increasing. Then $g'(x) \ge 0$, and

$$|F(x)g'(x)| \le Mg'(x).$$

Since

$$\int_a^{+\infty} g'(x)dx = \lim_{b\to+\infty} \int_a^b g'(x)dx = \lim_{b\to+\infty} (g(b) - g(a)) = -g(a)$$

converges, by the comparison test, the improper integral

$$\int_a^{+\infty} F(x)g'(x)dx = \lim_{b\to+\infty} \int_a^b F(x)g'(x)dx$$

converges. Therefore $\int_a^{+\infty} f(x)g(x)dx = \lim_{b\to+\infty} \int_a^b f(x)g(x)dx$ also converges. The proof for decreasing $g(x)$ is similar.

Under the assumption of the Abel test, we know both $F(b)$ and $g(b)$ converge as $b \to +\infty$. Therefore $F(x)$ is bounded, and we may apply the comparison test as before. Moreover, the convergence of $\lim_{b\to+\infty} g(b)$ implies the convergence of $\int_a^{+\infty} g'(x)dx$. We conclude again that $\int_a^{+\infty} f(x)g(x)dx$ converges.

Exercise 3.7.12. Determine convergence. Is the convergence absolute or conditional?

1. $\displaystyle\int_0^{+\infty} \frac{\sin x^q}{x^p}dx$.
2. $\displaystyle\int_0^{+\infty} \frac{\cos x^q}{x^p}dx$.
3. $\displaystyle\int_0^{1} \frac{1}{x}\sin\frac{1}{x}dx$.
4. $\displaystyle\int_0^{1} \frac{1}{x^p}\sin\frac{1}{x}dx$.
5. $\displaystyle\int_0^{+\infty} \frac{\cos ax}{1+x^p}dx$.
6. $\displaystyle\int_0^{+\infty} \frac{x^p\sin ax}{1+x^q}dx$.
7. $\displaystyle\int_0^{+\infty} \frac{\sin^2 x}{x}dx$.
8. $\displaystyle\int_0^{+\infty} \frac{\sin^3 x}{x}dx$.
9. $\displaystyle\int_1^{+\infty} \frac{\sin x\arctan x}{x^p}dx$.

Exercise 3.7.13. Construct a function $f(x)$ such that $|f(x)| = 1$ and $\int_0^{+\infty} f(x)dx$ converges.

Finally, we show some examples of using the integration by parts and change of variable to compute improper integrals. We note that the convergence needs to be verified before applying the properties of integration.

Example 3.7.16. In Example 3.7.11, we know the convergence of $\int_0^{+\infty} e^{-x^2}dx$. By the similar idea, especially $\lim_{x\to+\infty}\dfrac{x^pe^{-x^2}}{e^{-x}} = 0$, we know that $\int_0^{+\infty} x^pe^{-x^2}dx$ converges for any $p \ge 0$.

Let

$$I_n = \int_0^{+\infty} x^ne^{-x^2}dx.$$

Then we may apply the integration by parts to get

$$\begin{aligned} I_n &= -\frac{1}{2}\int_0^{+\infty} x^{n-1}de^{-x^2} \\ &= -\frac{1}{2}x^{n-1}e^{-x^2}\Big|_{x=0}^{x=+\infty} + \frac{n-1}{2}\int_0^{+\infty} x^{n-2}e^{-x^2}dx = \frac{n-1}{2}I_{n-2}. \end{aligned}$$

It is known (by using integration of two variable function, for example) that

$$I_0 = \int_0^{+\infty} e^{-x^2}dx = \frac{\sqrt{\pi}}{2}.$$

We can also apply the change of variable to get

$$I_1 = \int_0^{+\infty} xe^{-x^2}dx = \frac{1}{2}\int_0^{+\infty} e^{-x}dx = \frac{1}{2}.$$

Then we can use the recursive relation to compute I_n for all natural number n.

Example 3.7.17. The integral $\int_0^{\frac{\pi}{2}} \log \sin x dx$ is improper at 0. By L'Hospital's rule, we have

$$\lim_{x\to 0^+} \frac{\log \sin x}{\log x} = \lim_{x\to 0^+} \frac{\frac{\cos x}{\sin x}}{\frac{1}{x}} = \lim_{x\to 0^+} \frac{x\cos x}{\sin x} = 1.$$

By the convergence of $\int_0^1 |\log x| dx = -\int_0^1 \log x dx$ in Example 3.7.2 and the comparison test, we see that $\int_0^{\frac{\pi}{2}} \log \sin x dx$ converges.

The value of the improper integral can be computed as follows

$$\begin{aligned}
\int_0^{\frac{\pi}{2}} \log \sin x dx &= \int_0^{\frac{\pi}{4}} \log \sin x dx + \int_{\frac{\pi}{4}}^{\frac{\pi}{2}} \log \sin x dx \\
&= \int_0^{\frac{\pi}{4}} \log \sin x dx - \int_{\frac{\pi}{4}}^{0} \log \cos x dx \\
&= \int_0^{\frac{\pi}{4}} (\log \sin x + \log \cos x) dx = \int_0^{\frac{\pi}{4}} \log \left(\frac{1}{2}\sin 2x\right) dx \\
&= \int_0^{\frac{\pi}{4}} \log \sin 2x dx - \frac{\pi}{4}\log 2 = \frac{1}{2}\int_0^{\frac{\pi}{2}} \log \sin x dx - \frac{\pi}{4}\log 2.
\end{aligned}$$

Note that all the deductions are legitimate because all the improper integrals involved converge. Therefore we conclude that

$$\int_0^{\frac{\pi}{2}} \log \sin x dx = -\frac{\pi}{2}\log 2.$$

Exercise 3.7.14. Compute improper integral.

1. $\int_0^1 (\log x)^n dx$.
2. $\int_0^{+\infty} x^n e^{-x} dx$.
3. $\int_0^{+\infty} \frac{dx}{(1+x^2)^n}$.
4. $\int_0^1 \frac{x^n dx}{\sqrt{1-x^2}}$.
5. $\int_0^{\frac{\pi}{2}} \log \cos x dx$.

Exercise 3.7.15. The *Gamma function* is

$$\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt.$$

1. Show that the function is defined for $x > 0$.

2. Show the other formulae for the Gamma function

$$\Gamma(x) = 2\int_0^\infty t^{2x-1}e^{-t^2}dt = a^x\int_0^\infty t^{x-1}e^{-at}dt.$$

3. Show that $\Gamma(x+1) = x\Gamma(x)$ and $\Gamma(n) = (n-1)!$.

3.8 Application to Geometry

3.8.1 Length of Curve

Curves in a Euclidean space are often presented by parametrization. For example, the unit circle centered at the origin of $\mathbb{R}^2$ may be parametrized by the angle

$$x = \cos\theta, \quad y = \sin\theta, \quad 0 \le \theta \le 2\pi.$$

The *helix* in $\mathbb{R}^3$

$$x = r\cos\theta, \quad y = r\sin\theta, \quad z = \frac{h}{2\pi}\theta, \quad 0 \le \theta \le 2\pi,$$

moves along the circle of radius r from the viewpoint of the (x, y)-coordinates, and moves up in the z-direction in constant speed, such that each round moves up by height h.

In general, a *parametrized curve* in $\mathbb{R}^2$ is given by

$$x = x(t), \quad y = y(t), \quad a \le t \le b.$$

The *initial point* of the curve is $(x(a), y(a))$, and the *end point* is $(x(b), y(b))$. To compute the length of the curve between the two points, we consider the length $s(t)$ from the initial point $(x(a), y(a))$ to the point $(x(t), y(t))$. We find the change $s'(t)$ and then integrate the change to get $s(t)$. The length of the whole curve is $s(b)$.

Similar to the argument for the area of the region $G_{[a,b]}(f)$, we need to be careful about the sign. In the subsequent discussion, we pretend everything is positive (which at least gives you the right derivative), and further argument about the negative case is omitted. Moreover, we restrict the argument to the case $x(t)$ and $y(t)$ are nice. In fact, we will assume the two functions are continuously differentiable. In general, we may break the curve into finitely many continuously differentiable pieces and add the lengths of the pieces together.

As the parameter t is changed by Δt, the change $\Delta s = s(t + \Delta t) - s(t)$ of the length is the length of the curve segment from $(x(t), y(t))$ to $(x(t + \Delta t), y(t + \Delta t))$. The curve segment is approximated by the straight line connecting the two points. Therefore the length of the curve is approximated by the length of the straight line

$$\Delta s \approx \sqrt{(x(t+\Delta t) - x(t))^2 + (y(t+\Delta t) - y(t))^2} = \sqrt{(\Delta x)^2 + (\Delta y)^2}.$$

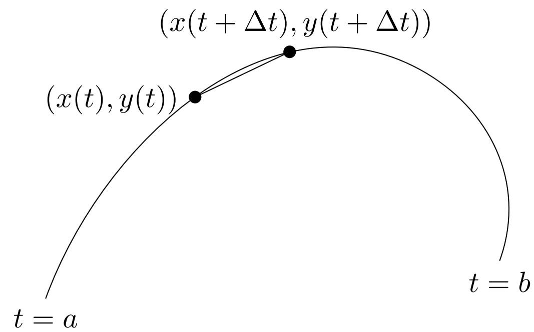


Figure 3.8.1: Length of curve.

Dividing the change Δt of parameter, we get

$$\frac{\Delta s}{\Delta t} \approx \sqrt{\left(\frac{\Delta x}{\Delta t}\right)^2 + \left(\frac{\Delta y}{\Delta t}\right)^2}.$$

The approximation gets more refined as $\Delta t \to 0$. By taking the limit as $\Delta t \to 0$, the approximation becomes an equality

$$s'(t) = \lim_{\Delta t\to 0} \frac{\Delta s}{\Delta t} = \lim_{\Delta t\to 0} \sqrt{\left(\frac{\Delta x}{\Delta t}\right)^2 + \left(\frac{\Delta y}{\Delta t}\right)^2} = \sqrt{x'(t)^2 + y'(t)^2}.$$

Therefore the length function $s(t)$ is the antiderivative of $\sqrt{x'(t)^2 + y'(t)^2}$, or

$$ds = \sqrt{x'(t)^2 + y'(t)^2}dt,$$

and we have

$$\text{length of curve} = s(b) = \int_a^b \sqrt{x'(t)^2 + y'(t)^2}dt.$$

Example 3.8.1. The length of the unit circle is

$$\int_0^{2\pi} \sqrt{(-\sin\theta)^2 + (\cos\theta)^2}d\theta = \int_0^{2\pi} d\theta = 2\pi.$$

More generally, an ellipse $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$ can be parametrized as

$$x = a\cos\theta, \quad y = b\sin\theta, \quad 0 \le \theta \le 2\pi.$$

The length of the ellipse is the so called *elliptic integral*

$$\int_0^{2\pi} \sqrt{(-a\sin\theta)^2 + (b\cos\theta)^2}d\theta = a\int_0^{2\pi} \sqrt{1 + K\cos^2\theta}d\theta, \quad K = \frac{b^2}{a^2} - 1.$$

The integral cannot be computed as an elementary expression if $a \ne b$.

Example 3.8.2. The graph of a function $f(x)$ on $[a,b]$ is a curve

$$x = t, \quad y = f(t), \quad t \in [a,b].$$

The length of the graph is

$$\int_a^b \sqrt{1+f'(x)^2}dx,$$

where the variable t is substituted to the more familiar x.

For example, the parabola $y = x^2$ is cut by the diagonal $y = x$. With the help of Example 3.5.31, the finite segment corresponding to $x \in [0,1]$ has length

$$\begin{aligned}\int_0^1 \sqrt{1+(2x)^2}dx &= \frac{1}{2}\int_0^2 \sqrt{1+x^2}dx \\ &= \frac{1}{4}\left(x\sqrt{1+x^2} + \log(\sqrt{1+x^2}+x)\right)_0^2 = \frac{1}{2}\sqrt{5} + \frac{1}{4}\log(\sqrt{5}+2).\end{aligned}$$

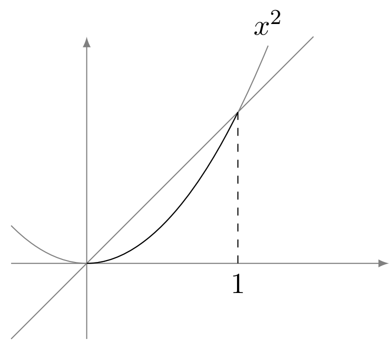


Figure 3.8.2: Parabola x^2 cut by the diagonal.

Example 3.8.3. The *astroid* $x^{\frac{2}{3}} + y^{\frac{2}{3}} = 1$ can be parametrized as

$$x = \cos^3 t, \quad y = \sin^3 t, \quad t \in [0, 2\pi].$$

Note that the range $[0,2\pi]$ for t corresponds to moving around the astroid exactly once. Therefore the perimeter is

$$\int_0^{2\pi} \sqrt{(-3\cos^2 t \sin t)^2 + (3a\sin^2 t\cos t)^2}dt = \int_0^{2\pi} 3|\sin t\cos t|dt = 6.$$

Example 3.8.4. The argument about the length of curves also applies to curves in $\mathbb{R}^3$ and leads to

$$\text{length of curve} = \int_a^b \sqrt{x'(t)^2 + y'(t)^2 + z'(t)^2}dt.$$

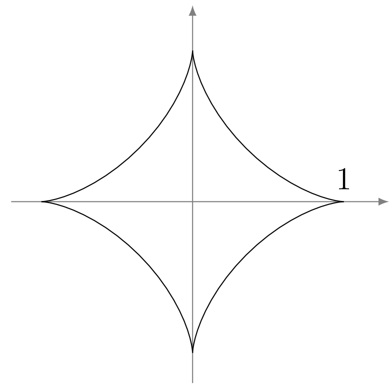


Figure 3.8.3: Astroid.

For example, the length of one round of the helix is

$$\int_0^{2\pi}\sqrt{(-r\sin\theta)^2+(r\cos\theta)^2+\left(\frac{h}{2\pi}\right)^2}d\theta=\int_0^{2\pi}\sqrt{r^2+\left(\frac{h}{2\pi}\right)^2}d\theta$$
$$=2\pi\sqrt{r^2+\left(\frac{h}{2\pi}\right)^2}=\sqrt{(2\pi r)^2+h^2}.$$

The result has a simple geometrical interpretation: By cutting along a vertical line, the cylinder can be "flattened" into a plane. Then the helix becomes the hypotenuse of a right triangle with horizontal length $2\pi r$ and vertical length h.

Example 3.8.5. When a circle rolls along a straight line, the track of one point on the circle is the *cycloid*. Let r be the radius of the circle, and assume the point is at the bottom at the beginning. After rotating angle t, the center of the circle is at (rt,r), and the point is at $(rt,r)+r(-\cos(t-\frac{\pi}{2}),\sin(t-\frac{\pi}{2}))$. Therefore the cycloid is parameterized by

$$x=rt-r\sin t,\quad y=r-r\cos t.$$

As the circle makes one complete rotation, we get one period of the cycloid, corresponding to $t\in[0,2\pi]$. The length of this one period is

$$\int_0^{2\pi}\sqrt{(r-r\cos t)^2+(r\sin t)^2}dt=r\int_0^{2\pi}\sqrt{2(1-\cos t)}dt=r\int_0^{2\pi}2\left|\sin\frac{t}{2}\right|dt=8r.$$

Exercise 3.8.1. Compute length.

1. $y^2=2x$, $x\in[0,a]$.
2. $x^2=2py$, $x\in[0,a]$.
3. $y=e^x$, $x\in[0,a]$.
4. $y=\log x$, $x\in[1,a]$.

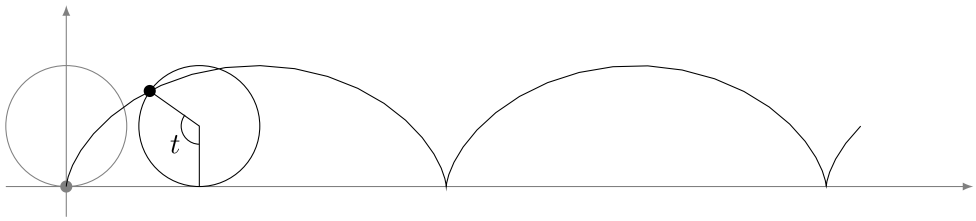


Figure 3.8.4: Cycloid.

5. $y = \log(4 - x^2)$, $x \in [-1, 1]$.
6. $y = \log\cos x$, $x \in \left[0, \dfrac{\pi}{4}\right]$.
7. $y = \log\sec x$, $x \in \left[0, \dfrac{\pi}{4}\right]$.
8. $y = \dfrac{e^x + e^{-x}}{2}$, $x \in [-a, a]$.
9. $y^2 = x^3$, $x \in [0, a]$.
10. $y^4 = x^3$, $x \in [0, a]$.
11. $y = \log\dfrac{e^x + 1}{e^x - 1}$, $x \in [1, 2]$.
12. $y = \displaystyle\int_0^x \sqrt{t^3 - 1}dt$, $x \in [1, 4]$.

Exercise 3.8.2. Compute length.

1. $\sqrt{x} + \sqrt{y} = 1$, $x, y \geq 0$.
2. $x = t$, $y = \log t$, $t \in [1, 2]$.
3. $x = e^t - t$, $y = e^t + t$, $t \in [0, 1]$.
4. $x = e^t \cos t$, $y = e^t \sin t$, $t \in [0, \pi]$.
5. $x = \cos^2 t$, $y = \cos t \sin t$, $t \in [0, \pi]$.
6. $x = 3\cos t - \cos 3t$, $y = 3\sin t - \sin 3t$, $t \in [0, \pi]$.
7. $x = \cos t + \log\tan\dfrac{t}{2}$, $y = \sin t$, $t \in \left[\dfrac{\pi}{4}, \dfrac{\pi}{2}\right]$.
8. $x = \cos t + t\sin t$, $y = \sin t - t\cos t$, $z = t^2$, $t \in [0, 2\pi]$.
9. $x = a^2 e^t$, $y = b^2 e^{-t}$, $z = \sqrt{2}abt$, $t \in [0, 1]$.

Exercise 3.8.3. Compute the length of *Cornu's spiral*

$$x = \int_0^t \cos\frac{\pi u^2}{2}du, \quad y = \int_0^t \sin\frac{\pi u^2}{2}du.$$

Exercise 3.8.4. Think of the rolling circle that produces the cycloid as a disk. What is the track of a point on the disk that is not necessarily on the circle (i.e., the boundary of the disk)? Find the formula for computing the length of this track.

Exercise 3.8.5. Suppose a line is wrapped around a circle. When the line is unwrapped from the circle, the track of one point on the line is the *involute* of the circle. Let r be the radius of the circle and let t be the unwrapped angle.

1. Find the parameterized formula for the involute.

2. Find the length of the involute as the line is unwrapped by half of the circle.

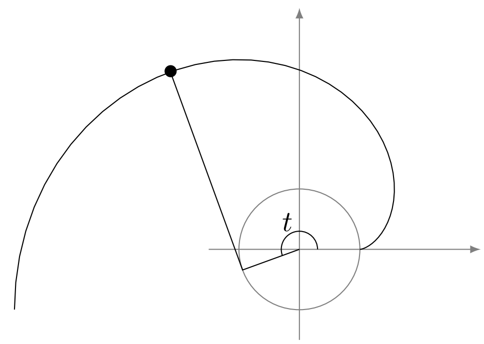


Figure 3.8.5: Involute of circle.

3.8.2 Area of Region

Being defined as area, the integration is naturally adapted to the computation of area. We start with the area of region bounded by two functions.

Example 3.8.6. The curve $y = x^2$ and the straight line $y = x$ enclose a region over $0 \le x \le 1$. The area of the region is the area below x subtracting the area below x^2, which is

$$\int_0^1 x dx - \int_0^1 x^2 dx = \int_0^1 (x - x^2) dx = \frac{1}{6}.$$

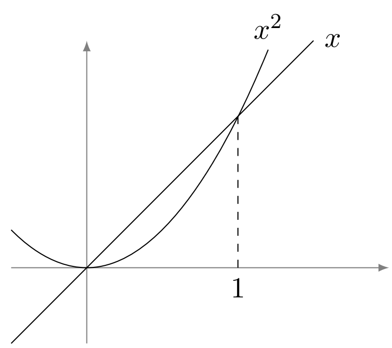


Figure 3.8.6: Region between x and x^2.

Example 3.8.7. To compute the area of the region bounded by $y = x^2 - 2x$ and $y = x$. We denote the (positive) areas of the four indicated regions by A_1, A_2, A_3, A_4. Then

$$\int_0^2 xdx = A_1, \quad \int_0^2 (x^2 - 2x)dx = -A_2, \quad \int_2^3 xdx = A_3 + A_4, \quad \int_2^3 (x^2 - 2x)dx = A_4.$$

The area we are interested in is

$$\begin{aligned} A_1 + A_2 + A_3 &= A_1 - (-A_2) + (A_3 + A_4) - A_4 \\ &= \int_0^2 xdx - \int_0^2 (x^2 - 2x)dx + \int_2^3 xdx - \int_2^3 (x^2 - 2x)dx \\ &= \int_0^3 xdx - \int_0^3 (x^2 - 2x)dx = \int_0^3 [x - (x^2 - 2x)]dx = \frac{9}{2}. \end{aligned}$$

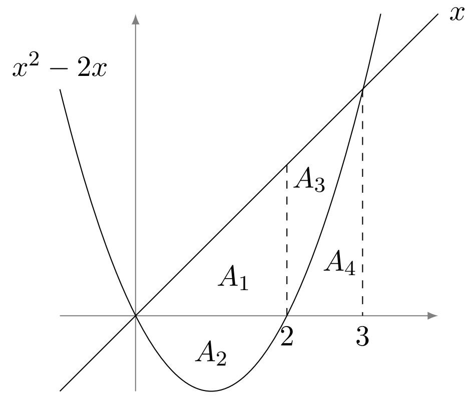


Figure 3.8.7: Region between x and $x^2 - 2x$.

The examples suggest that, if $f(x) \geq g(x)$ on $[a, b]$, then the area of the region between f and g over $[a, b]$ is

$$\int_a^b f(x)dx - \int_a^b g(x)dx = \int_a^b (f(x) - g(x))dx.$$

In general, we can divide $[a, b]$ into some intervals, such that on each interval, one of the following happens: $f(x) \geq 0 \geq g(x)$, $f(x) \geq g(x) \geq 0$, $0 \geq f(x) \geq g(x)$. Then an argument similar to Example 3.8.7 shows that the total area is indeed given by the formula above.

Example 3.8.8. The functions $\sin x$ and $\cos x$ intersect at many places and enclose many regions. One such region is over the interval $\left[\frac{\pi}{4}, \frac{5\pi}{4}\right]$, on which we have $\sin x \geq \cos x$. The area of the region is

$$\int_{\frac{\pi}{4}}^{\frac{5\pi}{4}} (\sin x - \cos x)dx = 2\sqrt{2}.$$

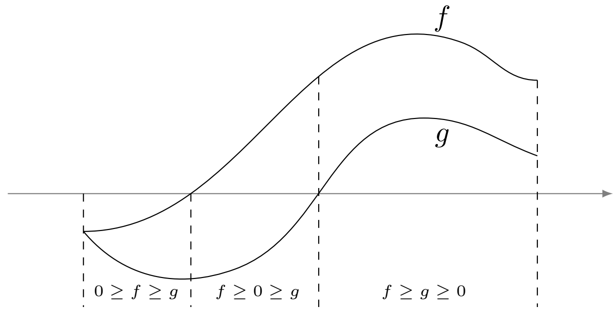


Figure 3.8.8: The area is $\displaystyle\int_a^b (f(x) - g(x))dx$.

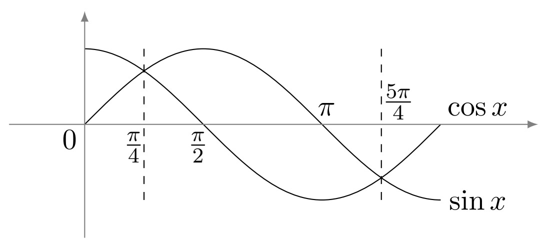


Figure 3.8.9: Region between $\sin x$ and $\cos x$.

Example 3.8.9. The region between the parabola $y^2 - x = 1$ and the straight line $x + y = 1$ is between the functions

$$f(x) = \begin{cases} \sqrt{x+1}, & \text{if } -1 \leq x \leq 0, \\ 1 - x, & \text{if } 0 \leq x \leq 3, \end{cases} \qquad g(x) = -\sqrt{x+1}.$$

The area is

$$\int_{-1}^3 f(x)dx - \int_{-1}^3 g(x)dx = \int_{-1}^0 \sqrt{x+1}dx + \int_0^3 (1-x)dx - \int_{-1}^3 (-\sqrt{x+1})dx = \frac{9}{2}.$$

Note that the region is obtained by rotating the region in Example 3.8.7. Naturally the results are the same. The previous example actually suggests another way of computing the area, by exchanging the roles of x and y.

Example 3.8.10. Consider $f(x) = x^5 + 2x^2 - 2x - 3$ and $g(x) = x^5 - x^3 + x^2 - 3$. We have

$$f(x) - g(x) = x^3 + x^2 - 2x = x(x-1)(x+2).$$

Therefore the two functions intersect at $x = -2, 0, 1$ and enclose two regions. The first region is over $[-2, 0]$, on which $f(x) \geq g(x)$. The second region is over $[0, 1]$,

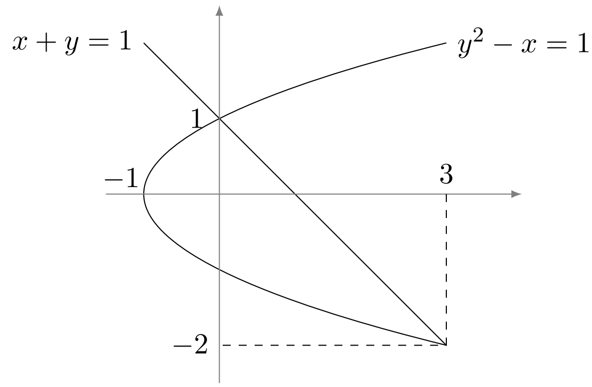


Figure 3.8.10: Region between a parabola and a straight line.

on which $f(x) \leq g(x)$. The areas of the regions are

$$\text{Area over } [-2, 0] = \int_{-2}^{0} (f(x) - g(x))dx = \frac{8}{3},$$
$$\text{Area over } [0, 1] = \int_{0}^{1} (f(x) - g(x))dx = \frac{5}{12}.$$

Note that we may change the functions to $f(x) = x^5 + 2x^2 - 2x - 3 + e^{x^2}$ and $g(x) = x^5 - x^3 + x^2 - 3 + e^{x^2}$ and get the same result. Although it is hard (actually impossible) to compute the exact values of $\int_a^b f(x)dx$ and $\int_a^b g(x)dx$. Yet we can still compute the area.

Exercise 3.8.6. Compute area of the region with the given bounds.

1. $y = \sqrt{x}$, y-axis, $y = 1$.
2. $y = e^x$, $y = x$, on $[0, 1]$.
3. $y = \log x$, $y = x$, $y = 0$, $y = 1$.
4. $y = x^2$, $y = 2x - x^2$.
5. $y = \sin x$, $y = \cos x$, on $[0, \frac{\pi}{4}]$.
6. $y = e^x$, $y = x^2 - 1$, on $[-1, 1]$.
7. $y = \log x$, $y^2 = x + 2$, $y = -1$, $y = 1$.
8. $x = y^2 - 4y$, $x = 2y - y^2$.
9. $y = 2x - x^2$, $x + y = 0$.
10. $y = x$, $y = x + \sin^2 x$, on $[0, \pi]$.

Exercise 3.8.7. Explain that, if $0 \geq f \geq g$ on $[a, b]$, then the area of the region between the graphs of f and g over $[a, b]$ is $\int_a^b (f(x) - g(x))dx$.

Exercise 3.8.8. Explain that, the area of the region between the graphs of f and g over $[a, b]$ is $\int_a^b |f(x) - g(x)|dx$, even when we may have $f > g$ some place and $f < g$ some other place.

In practise, a region is often enclosed by a closed boundary curve (or several closed curves if the region has holes). It is often more convenient to describe curves by their parameterisations. For example, the unit disk is enclosed by the unit circle, which can be conveniently parameterised as $x = \cos t$, $y = \sin t$, $t \in [0, 2\pi]$.

A parameterisation of a curve can be considered as a movement along the curve, and therefore imposes a direction on the curve. We will always make the standard assumption that, as we move along a parameterised boundary curve, the region is always on the left of curve. Figure 3.8.2 illustrates the meaning of the assumption. For a region without holes, this means that the curve has *counterclockwise* direction. The unit circle parameterisation above is such an example. If the region has holes, then the "inside boundary components" should have *clockwise* direction.

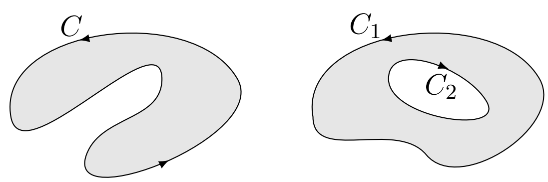


Figure 3.8.11: The region is always on the left of the boundary curve.

Consider a simple region in Figure 3.8.2, such that the boundary curve can be divided into the graphs of two functions $y = y_1(x)$ and $y = y_2(x)$ for $x \in [\alpha, \beta]$. Suppose the boundary curve has parameterisation $\phi(t) = (x(t), y(t))$, $t \in [a, b]$, such that y_1 and y_2 correspond respectively to $t \in [a, c]$ and $t \in [c, b]$. The direction of the parameterisation satisfies our assumption.

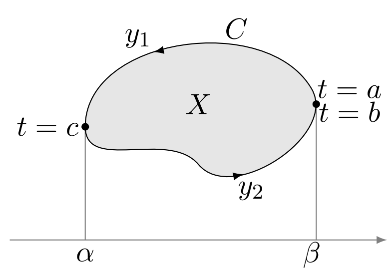


Figure 3.8.12: Calculate the area by integrating along the boundary curve.

The area of the region is $\int_\alpha^\beta (y_1(x) - y_2(x))dx$. We may use the parameterisation of the boundary curve as the change of variable to get the following formula for the area

$$\begin{aligned}\int_\alpha^\beta (y_1(x)-y_2(x))dx &= \int_\alpha^\beta y_1(x)dx - \int_\alpha^\beta y_2(x)dx \\ &= \int_c^a y(t)dx(t) - \int_c^b y(t)dx(t) \\ &= -\int_a^b y(t)dx(t) = -\int_C ydx.\end{aligned}$$

We note that the negative sign comes from our assumption of the direction. The upper part y_1 is supposed to positively contribute $\int_\alpha^\beta y_1(x)dx$ to the area of X. However, the direction of y_1 is leftward, opposite to the x-direction (the direction of dx). This introduces a negative sign. Similarly, the lower part y_2 should negatively contribute to the area, and yet has the rightward direction, the same as the x-direction. This also introduces a negative sign.

We may further use the integration by parts to get another formula for the area

$$-\int_a^b y(t)dx(t) = -y(b)x(b) + y(a)x(a) + \int_a^b x(t)dy(t) = \int_C xdy.$$

Here we have $x(a)=x(b)$ and $y(a)=y(b)$ because the boundary curve is closed. The positive sign on the right can be explained as follows. The area is supposed to be the contribution from the right boundary part subtracting the contribution from the left boundary part. From the picture, we see that the direction of the right part is upward, the same as the y-direction (the direction of dy), and the direction of the left part is downward, opposite to the y-direction.

Example 3.8.11. The boundary circle of the unit disk is parameterised by $x(t)=\cos t$, $y=\sin t$, $t\in[0,2\pi]$. Since the parameterisation satisfies our assumption, we may use it to calculate the area of the unit disk

$$-\int_0^{2\pi} y(t)dx(t) = \int_0^{2\pi} \sin^2 t dt = \pi.$$

Example 3.8.12. Consider the region enclosed by the Archimedean spiral $x=t\cos t$, $y=t\sin t$, $t\in[0,\pi]$, and the x-axis. The boundary of the region consists of the spiral and the interval $[-\pi,0]$ on the x-axis. After checking that the direction of the boundary satisfies the assumption, we get the area

$$-\int_0^\pi (t\sin t)(t\cos t)'dt - \int_{-\pi}^0 0dx = -\int_0^\pi (t\sin t\cos t - t^2\sin^2 t)dt = \frac{1}{6}\pi^3.$$

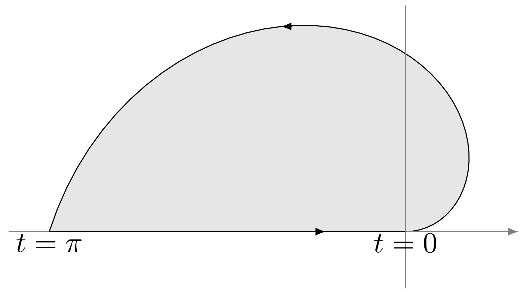


Figure 3.8.13: Archimedean spiral.

Exercise 3.8.9. Explain that the area of the region on the left of Figure 3.8.2 may be calculated by $-\int_C ydx$. You may need to break the boundary into four parts y_1, y_2, y_3, y_4. Moreover, show that the area may also be calculated by $\int_C xdy$.

Exercise 3.8.10. Explain that the area of the region on the right of Figure 3.8.2 may be calculated by $-\int_{C_1} ydx - \int_{C_2} ydx$ and $\int_{C_1} xdy + \int_{C_2} xdy$.

Exercise 3.8.11. Explain that, if the direction of the boundary curve C is opposite to our assumption, then the area is $\int_C ydx$.

Exercise 3.8.12. Compute the areas of the regions enclosed by the curves.

1. Ellipse $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$.
2. Astroid $x^{\frac{2}{3}} + y^{\frac{2}{3}} = 1$.
3. $\sqrt{|x|} + \sqrt{|y|} = 1$.
4. Hyperbola $x^2 - y^2 = 1$ and $x = a$ $(a > 1)$.
5. Sprial $x = e^t \cos t$, $y = e^t \sin t$, $t \in [0, \pi]$, and the x-axis.
6. One period of the cycloid in Example 3.8.5 and the x-axis.

3.8.3 Surface of Revolution

If we revolve a curve on the plane with respect to a straight line, we get a surface. For example, the sphere is obtained by revolving a circle around any straight line passing through the center of the circle, and the torus is obtained by revolving a circle around any straight line not intersecting the circle.

Let $(x(t), y(t))$, $t \in [a, b]$, be a parametrized curve in the upper half of the (x, y)-plane (this means $y(t) \geq 0$). To find the area of the surface obtained by revolving the curve around the x-axis, we let $A(t)$ be the area of the surface obtained by revolving the $[a,t]$ segment of the curve around the x-axis. Again the subsequent argument ignores the sign.

As the parameter is changed by Δt, the change $\Delta A = A(t+\Delta t) - A(t)$ of the area is the area of surface obtained by revolving the curve segment from $(x(t), y(t))$ to $(x(t+\Delta t), y(t+\Delta t)) = (x,y) + (\Delta x, \Delta y)$. Since the curve segment is approximated by the straight line connecting the two points, the area ΔA is approximated by the area of the revolution of the straight line.

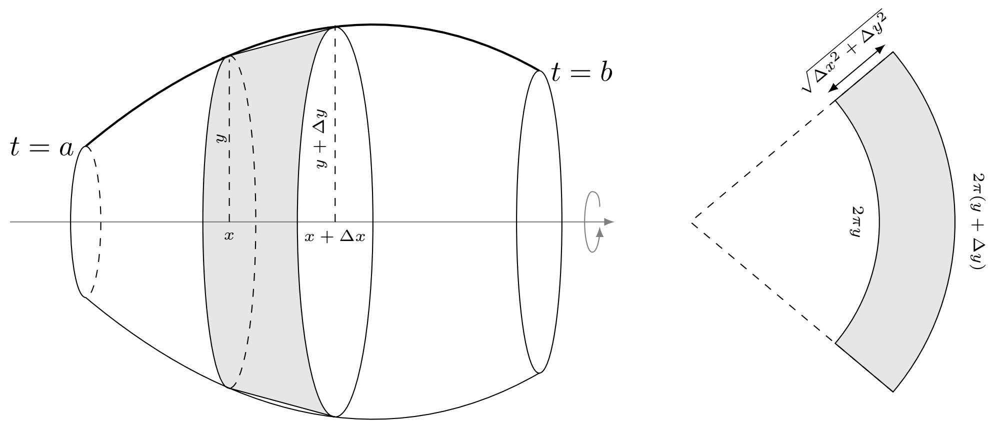


Figure 3.8.14: Area of surface of revolution.

The revolution of the straight line can be expanded to lie on the plane. It is part of the annulus of thickness $\sqrt{\Delta x^2 + \Delta y^2}$. Moreover, the inner arc has length $2\pi y(t)$ and the outer arc has length $2\pi y(t+\Delta t)$. Therefore the area of the partial annulus gives the approximation

$$\Delta A \approx \frac{1}{2}(2\pi y(t) + 2\pi y(t+\Delta t))\sqrt{\Delta x^2 + \Delta y^2} = \pi(y(t) + y(t+\Delta t))\sqrt{\Delta x^2 + \Delta y^2}.$$

Dividing the change Δt of the parameter, we get

$$\frac{\Delta A}{\Delta t} \approx \pi(y(t) + y(t+\Delta t))\sqrt{\left(\frac{\Delta x}{\Delta t}\right)^2 + \left(\frac{\Delta y}{\Delta t}\right)^2}.$$

The approximation gets more refined as $\Delta t \to 0$. By taking the limit as $\Delta t \to 0$, the approximation becomes an equality

$$\begin{aligned} A'(t) = \lim_{\Delta t \to 0} \frac{\Delta A}{\Delta t} &= \lim_{\Delta t \to 0} \pi(y(t) + y(t+\Delta t))\sqrt{\left(\frac{\Delta x}{\Delta t}\right)^2 + \left(\frac{\Delta y}{\Delta t}\right)^2} \\ &= 2\pi y(t)\sqrt{x'(t)^2 + y'(t)^2}. \end{aligned}$$

This leads to

$$\text{area of surface of revolution} = A(b) = 2\pi \int_a^b y(t)\sqrt{x'(t)^2 + y'(t)^2}dt.$$

We note that $ds = \sqrt{x'(t)^2 + y'(t)^2}dt$ is used for computing the length of curve, and we can write

$$\text{area of surface of revolution} = 2\pi \int_a^b y(t)ds.$$

Here y is really the distance from the curve to the axis of revolution.

Example 3.8.13. The 2-dimensional sphere of radius r is obtained by revolving the half circle

$$x = r\cos\theta, \quad y = r\sin\theta, \quad \theta \in [0, \pi]$$

around the x-axis. Since the length of circular arc is given by $ds = rd\theta$, the area of the sphere is

$$2\pi \int_0^\pi (r\sin\theta)rd\theta = 4\pi r^2.$$

Example 3.8.14. The torus is obtained by revolving a circle on the upper half plane around the x-axis. Let the radius of the circle be a and let the center of the circle be $(0, b)$. Then $a < b$ and the circle may be parametrized as

$$x = a\cos\theta, \quad y = a\sin\theta + b, \quad \theta \in [0, 2\pi].$$

The length is given by $ds = ad\theta$, so that the area of the torus is

$$2\pi \int_0^{2\pi} (a\sin\theta + b)ad\theta = 4\pi^2 ab.$$

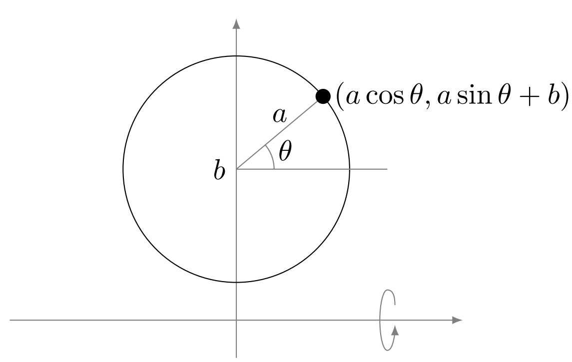


Figure 3.8.15: Torus.

Example 3.8.15. Take the segment $y = x^2$, $x \in [0,1]$ of the parabola in Example 3.8.2. If we revolve the parabola around the x-axis, then $ds = \sqrt{1+4x^2}dx$, the distance from the curve to the axis of rotation (i.e., the x-axis) is x^2. Therefore the area of the surface of revolution is

$$2\pi \int_0^1 x^2\sqrt{1+4x^2}dx$$

With the help of Example 3.5.15 and the computation in Example 3.8.2, we have

$$\begin{aligned}\int_0^1 x^2\sqrt{1+4x^2}dx &= \frac{1}{8}\int_0^2 x^2\sqrt{1+x^2}dx = \frac{1}{8}\int_0^2 \left((1+x^2)^{\frac{3}{2}} - (1+x^2)^{\frac{1}{2}}\right)dx \\ &= \frac{1}{8}\left(\frac{1}{2\cdot\frac{3}{2}+1}x(1+x^2)^{\frac{3}{2}}\Big|_0^2 + \left(\frac{2\cdot\frac{3}{2}}{2\cdot\frac{3}{2}+1} - 1\right)\int_0^2 (1+x^2)^{\frac{1}{2}}dx\right) \\ &= \frac{1}{8}\left(\frac{2}{5}5^{\frac{3}{2}} - \frac{2}{5}\left(\sqrt{5} + \frac{1}{2}\log(2+\sqrt{5})\right)\right) = \frac{1}{\sqrt{5}} - \frac{1}{40}\log(2+\sqrt{5}).\end{aligned}$$

So the area is $\dfrac{2\pi}{\sqrt{5}} - \dfrac{\pi}{20}\log(2+\sqrt{5})$.

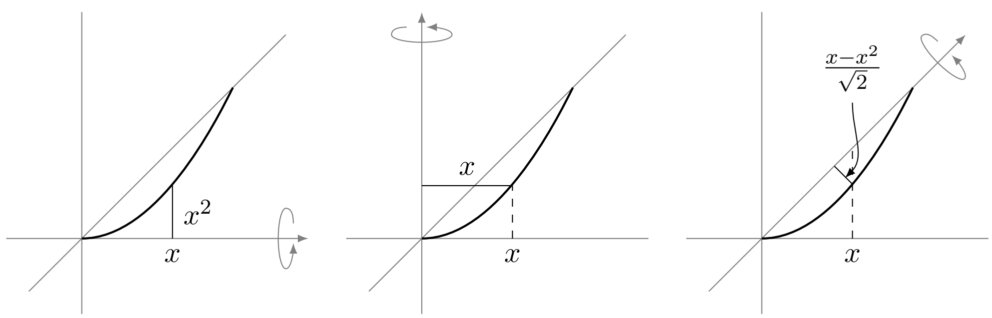


Figure 3.8.16: Revolving a parabola segment around different axes.

If we revolve around the y-axis, then we get a paraboloid. We still have $ds = \sqrt{1+4x^2}dx$, but the distance from the curve to the axis of rotation (i.e., the y-axis) is now x. Therefore the area of the paraboloid is

$$2\pi \int_0^1 x\sqrt{1+4x^2}dx = 2\pi\, \frac{2}{3\cdot 8}(1+4x^2)^{\frac{3}{2}}\Big|_0^1 = \frac{\pi}{6}(5\sqrt{5}-1).$$

Finally, if we revolve around the diagonal $y = x$, then the distance from the curve to the axis of rotation is $\dfrac{x-x^2}{\sqrt{2}}$, and the area is

$$\begin{aligned}2\pi\int_0^1 \frac{x-x^2}{\sqrt{2}}\sqrt{1+4x^2}dxdx &= \frac{1}{\sqrt{2}}\left(\frac{\pi}{6}(5\sqrt{5}-1) - \frac{2\pi}{\sqrt{5}} + \frac{\pi}{20}\log(2+\sqrt{5})\right) \\ &= \sqrt{2}\pi\left(\frac{13}{60}\sqrt{5} - \frac{1}{12} - \frac{1}{40}\log(2+\sqrt{5})\right).\end{aligned}$$

Example 3.8.15 shows how to adapt the formula for the area of the surface of revolution to the more general case of any parametrized curve $(x(t), y(t))$ with respect to a straight line $\alpha x + \beta y + \gamma = 0$. Assume the curve is on the "positive side" of the straight line

$$\alpha x(t) + \beta y(t) + \gamma \geq 0, \quad \text{for all } t \in [a, b].$$

Then the distance $y(t)$ should be replaced by $\dfrac{\alpha x(t) + \beta y(t) + \gamma}{\sqrt{\alpha^2 + \beta^2}}$. We still have $ds = \sqrt{x'(t)^2 + y'(t)^2}dt$. Therefore we get the general formula

$$\text{area of surface of revolution} = 2\pi \int_a^b \frac{(\alpha x(t) + \beta y(t) + \gamma)\sqrt{x'(t)^2 + y'(t)^2}}{\sqrt{\alpha^2 + \beta^2}} dt.$$

Exercise 3.8.13. Find the formula for the area of the surface of revolution of the graph of a function $y = f(x)$ around the x-axis. What about revolving around the y-axis? What about revolving around the line $x = a$?

Exercise 3.8.14. Find the area of the surface of revolution.

1. $y = x^3$, $x \in [0, 2]$, around x-axis.
2. $x^2 = 2py$, $x \in [0, 1]$, around y-axis.
3. $y = e^x$, $x \in [0, 1]$, around x-axis.
4. $y = e^x$, $x \in [0, 1]$, around y-axis.
5. $y = e^x$, $x \in [0, 1]$, around $y = 1$.
6. $y = \tan x$, $x \in [0, \frac{\pi}{4}]$, around x-axis.
7. $y^2 = \dfrac{e^x + e^{-x}}{2}$, $x \in [-a, a]$, around x-axis.
8. $y^2 = x^3$, $x \in [0, 1]$, around x-axis.
9. Ellipse $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$, around x-axis.
10. Astroid $x^{\frac{2}{3}} + y^{\frac{2}{3}} = 1$, around x-axis.

Exercise 3.8.15. Find the area of the surface obtained by revolving one period of the cycloid in Example 3.8.5 around the x-axis.

Exercise 3.8.16. Find the area of the surface obtained by revolving the involute of the circle in Example 3.8.5 around the x-axis.

3.8.4 Solid of Revolution

If we revolve a region in the plane with respect to a straight line, we get a solid. For example, the ball is obtained by revolving a disk around any straight line passing through the center of the disk, and the solid torus is obtained by revolving a disk around any straight line not intersecting the disk.

Consider the region $G_{[a,b]}(f)$ for a function $f(x) \geq 0$ over the interval $[a, b]$. To find the volume of the solid obtained by revolving the region around the x-axis, we let $V(x)$ be the volume of the part of solid obtained by revolving $G_{[a,x]}(f)$ around the x-axis. Then the change $\Delta V = V(x+\Delta x) - V(x)$ is the volume of the solid obtained by revolving $G_{[x,x+\Delta x]}(f)$.

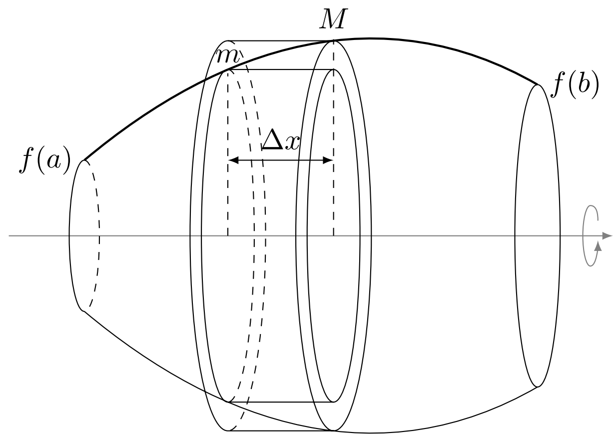


Figure 3.8.17: Volume of solid of revolution.

Let $m = \min_{[x,x+\Delta x]} f$ and $M = \max_{[x,x+\Delta x]} f$. Then $G_{[x,x+\Delta x]}(f)$ is sandwiched between the rectangles $[x, x+\Delta x] \times [0, m]$ and $[x, x+\Delta x] \times [0, M]$. Therefore the revolution of $G_{[x,x+\Delta x]}(f)$ is sandwiched between the revolutions of the two rectangles. The revolutions of rectangles are cylinders and have volumes $\pi m^2 \Delta x$ and $\pi M^2 \Delta x$. Therefore we get

$$\pi m^2 \Delta x \le \Delta V \le \pi M^2 \Delta x.$$

This implies

$$\pi m^2 \le \frac{\Delta V}{\Delta x} \le \pi M^2.$$

If f is continuous, then $\lim_{\Delta x \to 0} m = \lim_{\Delta x \to 0} M = f(x)$. By the sandwich rule, we get

$$V'(x) = \lim_{\Delta x \to 0} \frac{\Delta V}{\Delta x} = \pi f(x)^2.$$

This leads to

$$\text{volume of solid of revolution} = V(b) = \pi \int_a^b f(x)^2 dx.$$

Example 3.8.16. The 3-dimensional ball of radius r is obtained by revolving the half disk around the x-axis. The half disk is the region between $\sqrt{r^2 - x^2}$ and the x-axis over $[-r, r]$. Therefore the volume of the ball is

$$\pi \int_{-1}^{1} (\sqrt{r^2 - x^2})^2 dx = \frac{4}{3}\pi r^3.$$

Example 3.8.17. The solid torus is obtained by revolving a disk in the upper half plane around the x-axis. Let the radius of the disk be a and let the center of the disk be $(0, b)$. Then $a < b$ and the disk is the region between $y_1(x) = b + \sqrt{a^2 - x^2}$ and $y_2(x) = b - \sqrt{a^2 - x^2}$ over the interval $[-a, a]$. The torus is the solid obtained by revolving $G_{[-a,a]}(y_1)$ subtracting the solid obtained by revolving $G_{[-a,a]}(y_2)$. Therefore the volume of the torus is the volume of the first solid subtracting the second

$$\begin{aligned}\pi \int_{-a}^{a} y_1(x)^2 dx - \pi \int_{-a}^{a} y_2(x)^2 dx &= \pi \int_{-a}^{a} (y_1(x)^2 - y_2(x)^2) dx \\ &= \pi \int_{-a}^{a} ((b + \sqrt{a^2 - x^2})^2 - (b - \sqrt{a^2 - x^2})^2) dx \\ &= \pi \int_{-a}^{a} 4b\sqrt{a^2 - x^2} dx \\ &= 4\pi b \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} a^2 \cos^2 t dt = 2\pi^2 a^2 b.\end{aligned}$$

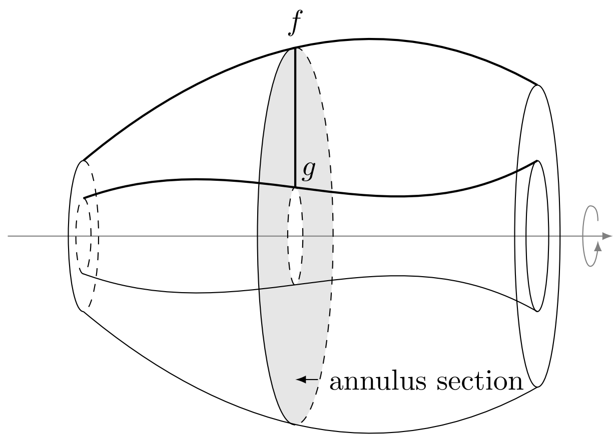


Figure 3.8.18: Solid of revolution of the region between two functions.

Example 3.8.17 shows that, if $f \geq g \geq 0$ on $[a, b]$, then the volume of the solid of revolution obtained by revolving the region between f and g over $[a, b]$ around the x-axis is

$$\pi \int_{a}^{b} (f(x)^2 - g(x)^2) dx.$$

Now we extend the discussion before Example 3.8.11 about calculating the area of a plan region by the integrating along the boundary curve. Suppose the region X in Figure 3.8.2 lies in the upper half plane. Then similar to the earlier discussion, the volume of the solid obtained by rotating X around the x-axis is

$$\begin{aligned}\pi\int_{\alpha}^{\beta}(y_1(x)^2-y_2(x)^2)dx &= \pi\int_{\alpha}^{\beta}y_1(x)^2dx-\pi\int_{\alpha}^{\beta}y_2(x)^2dx\\ &= \pi\int_{c}^{a}y(t)^2dx(t)-\pi\int_{c}^{b}y(t)^2dx(t)\\ &= -\pi\int_{a}^{b}y(t)^2dx(t)=-\pi\int_{C}y^2dx.\end{aligned}$$

So all the earlier discussion about the area can be applied to the volume of the solid of revolution.

Example 3.8.18. The volume of the 3-dimensional ball of radius r is obtained by revolving the half disk around the x-axis. The boundary of the half disk consists of the half circle $x=\cos t$, $y=\cos t$, $t\in[0,\pi]$, and the interval $[-1,1]$ on the x-axis. Moreover, the parameterisation of the boundary curve satisfies our assumption. Therefore the volume of the ball is

$$\pi\int_{0}^{\pi}(r\sin t)^2d(r\cos t)+\pi\int_{-1}^{1}0^2dx=\pi r^3\int_{0}^{\pi}(1-\cos^2 t)d(\cos t)=\frac{4}{3}\pi r^3.$$

Example 3.8.19. We use the parameterisation $x=a\cos t$, $y=b+a\sin t$, $t\in[0,2\pi]$ of the circle to calculate the volume of the torus in Example 3.8.17

$$\begin{aligned}-\pi\int_{0}^{2\pi}(b+a\sin t)^2d(a\cos t) &= \pi a\int_{0}^{2\pi}(b^2\sin t+2ab\sin^2 t+a^2\sin^2 t)dt\\ &= 2\pi^2a^2b.\end{aligned}$$

Example 3.8.20. Consider the region enclosed by the Archimedean spiral and the x-axis in Example 3.8.20. The volume of the solid obtained by revolving the region around the x-axis is

$$-\int_{0}^{\pi}(t\sin t)^2d(t\cos t)=-\int_{0}^{\pi}(t^2\sin^2 t\cos t-t^3\sin^3 t)dt=\frac{2}{3}\pi^3-4\pi.$$

Exercise 3.8.17. Find volume of the solid obtained by revolving the region in Exercise 3.8.12 around the x-axis.

Exercise 3.8.18. Use integration by parts to explain that the volume of the solid of revolution can also be calculated by $2\pi\int_C xydy$. Then use this formula to calculate the volumes of the ball, the torus, and the solids obtained by revolving the regions in Exercise 3.8.12 around the x-axis.

The formula will be the "shell method" in Example 3.8.27.

Exercise 3.8.19. Explain that, if the direction of the boundary curve C is opposite to our assumption, then the volume of the solid of revolution around the x-axis is $\pi\int_C y^2dx$.

Exercise 3.8.20. For the solid obtained by revolving a region in the lower half plane around the x-axis, how should the formula $-\pi\int_C y^2dx$ for the volume be modified?

Next we consider the general case of revolving a region X around a straight line $L\colon \alpha x+\beta y+\gamma=0$. We assume X is on the "positive side" of L in the sense that the parameterisation $(x(t),y(t))$ of the boundary curve C of X satisfies

$$\alpha x(t)+\beta y(t)+\gamma\ge 0,\ \text{ for all } t\in[a,b].$$

Moreover, we still assume that the direction of C satisfies our assumption. Then in the formula $-\pi\int_C y^2dx$, y should be understood as the distance $\dfrac{\alpha x(t)+\beta y(t)+\gamma}{\sqrt{\alpha^2+\beta^2}}$ from C to L, and dx should be understood as the progression

$$\frac{\beta dx-\alpha dy}{\sqrt{\alpha^2+\beta^2}}=\frac{\beta x'(t)-\alpha y'(t)}{\sqrt{\alpha^2+\beta^2}}dt$$

along the direction $\dfrac{(\beta,-\alpha)}{\sqrt{\alpha^2+\beta^2}}$ of L. Therefore the volume of the solid of revolution is

$$-\pi\int_a^b\frac{(\alpha x(t)+\beta y(t)+\gamma)^2(\beta x'(t)-\alpha y'(t))}{(\sqrt{\alpha^2+\beta^2})^3}dt$$

We note that the negative sign is due to the mismatch (See Figure 3.8.4) of the direction of the boundary curve and the direction of the progression along L. In general, we may determine the sign by comparing the direction of the parameter and the direction of progression.

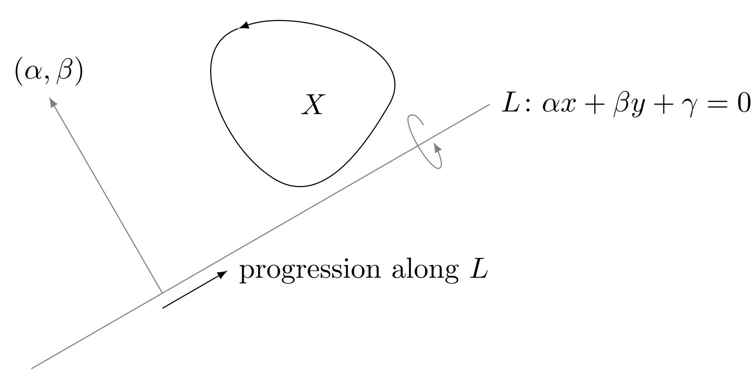


Figure 3.8.19: Revolving a region X around a line.

For the special case that X is above the horizontal line $y = b$ ($(\alpha, \beta) = (0, 1)$), the volume of the solid of revolution around the line is $-\pi \int_C (y - b)^2 dx$. If X is on the *right* of the y-axis (i.e., the line $x = 0$, with $(\alpha, \beta) = (1, 0)$), then the volume of the solid of revolution around the y-axis is

$$-\pi \int_a^b x(t)^2(-y'(t))dt = \pi \int_C x^2 dy.$$

The negative sign in front of y' comes from the fact that the progression for the line $x = 0$ goes *downwards*, the opposite of the y-direction. If X is on the *right* of the vertical line $x = a$, then the volume of the solid of revolution around the vertical line is $\pi \int_C (x - a)^2 dy$.

Example 3.8.21. Take the segment $y = x^2$, $x \in [0, 1]$, of the parabola in Example 3.8.2. If we revolve the region X between the parabola and the x-axis around the x-axis, then the volume of the solid is

$$\pi \int_{x=0}^{x=1} y^2 dx = \pi \int_0^1 (x^2)^2 dx = \frac{1}{5}\pi.$$

If we revolve X around the y-axis, then the volume of the solid is

$$\pi \int_{x=0}^{x=1} x^2 dy = \pi \int_{y=0}^{y=1} y dy = \frac{1}{2}\pi.$$

Let Y be the region between the parabola and the vertical line $x = 1$. If we revolve Y around the vertical line $x = 1$, then the volume of the solid is

$$\pi \int_{x=0}^{x=1} (1 - x)^2 dy = \pi \int_0^1 (1 - x)^2 d(x^2) = \frac{7}{6}\pi.$$

If we revolve Y around the y-axis instead, then the volume of the solid is

$$\pi \int_{x=0}^{x=1} (1^2 - x^2) dy = \pi \int_{y=0}^{y=1} (1 - y) dy = \frac{1}{2}\pi.$$

Let Z be the region between the parabola $y = x^2$ and the diagonal $y = x$. If we revolve Z around the x-axis, then the volume of the solid is

$$\pi \int_{x=0}^{x=1} (x^2 - (x^2)^2) dx = \frac{2}{15}\pi.$$

If we revolve Z around the line $x = -1$, then the volume of the solid is

$$\begin{aligned} \pi \int_{x=0}^{x=1} ((x + 1)^2 d(x^2) - (x + 1)^2 dx) &= \pi \int_0^1 (x + 1)^2 (2x - 1) dx \\ &= \pi \int_1^2 z^2 (2z - 3) dz = \frac{1}{2}\pi. \end{aligned}$$

If we revolve Z around the diagonal $y = x$, then to make sure the region is on the positive side of the diagonal, we should write the diagonal as $x - y = 0$, with $(\alpha, \beta) = (1, -1)$. The distance between the parabola and the diagonal is $\dfrac{x - x^2}{\sqrt{2}}$. The progression of the parabola in the direction $\dfrac{(\beta, -\alpha)}{\sqrt{\alpha^2 + \beta^2}} = \dfrac{(-1, -1)}{\sqrt{2}}$ of the line is

$$\frac{-dx - dy}{\sqrt{2}} = \frac{-dx - d(x^2)}{\sqrt{2}} = \frac{-(1 + 2x)}{\sqrt{2}} dx.$$

Therefore the volume of the solid is

$$-\pi \int_{x=0}^{x=1} \left(\frac{x - x^2}{\sqrt{2}} \right)^2 \frac{-(1 + 2x)}{\sqrt{2}} dx = \frac{1}{30\sqrt{2}} \pi.$$

If we revolve Z around the line $y = x - 1$, we should write the line as $-x + y + 1 = 0$ for Z to be on the positive side, with $(\alpha, \beta) = (-1, 1)$. The distance from the diagonal and the parabola to the line are $\dfrac{1}{\sqrt{2}}$ and $\dfrac{1}{\sqrt{2}} - \dfrac{x - x^2}{\sqrt{2}}$. The progressions of the diagonal and the parabola in the direction $\dfrac{(\beta, -\alpha)}{\sqrt{\alpha^2 + \beta^2}} = \dfrac{(1, 1)}{\sqrt{2}}$ of the line are

$$\frac{dx + dx}{\sqrt{2}} = \sqrt{2} dx \text{ and } \frac{dx + d(x^2)}{\sqrt{2}} = \frac{1 + 2x}{\sqrt{2}} dx.$$

Therefore the volume of the solid is

$$\pi \int_{x=0}^{x=1} \left(\left(\frac{1}{\sqrt{2}} \right)^2 \sqrt{2} dx - \left(\frac{1}{\sqrt{2}} - \frac{x - x^2}{\sqrt{2}} \right)^2 \frac{1 + 2x}{\sqrt{2}} dx \right) = \frac{3}{10\sqrt{2}} \pi.$$

Exercise 3.8.21. Let $A \le f(x) \le B$ for $x \in [a, b]$. Find the formula for the volume of the solid of revolution of the region between the graph of function f and $y = A$ around the line $y = C$, where $C \notin (A, B)$.

Exercise 3.8.22. Find the formula for the volume of the solid obtained by revolving a region X for which the parameterised boundary has the right direction.

1. X is on the left of the y-axis, around the y axis.

2. X is on the left of $x = a$, around $x = a$.

3. X is below $y = b$, around $y = b$.

4. X is on the negative side of $x + y = 0$, around $x + y = 0$.

5. X is on the negative side of $x + y = 1$, around $x + y = 1$.

Exercise 3.8.23. Find the volume of the solid obtained by revolving the region between the curve and the axis of revolution.

1. $y = x^3$, $x \in [0, 2]$, around x-axis.
2. $x^2 = 2py$, $x \in [0, 1]$, around y-axis.
3. $y = e^x$, $x \in [0, 1]$, around x-axis.
4. $y = e^x$, $x \in [0, 1]$, around y-axis.
5. $y = e^x$, $x \in [0, 1]$, around $x = 1$.
6. $y = \tan x$, $x \in \left[0, \frac{\pi}{4}\right]$, around x-axis.
7. $y^2 = \dfrac{e^x + e^{-x}}{2}$, $x \in [-a, a]$, around x-axis.
8. $y^2 = x^3$, $x \in [0, 1]$, around x-axis.
9. Ellipse $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} \le 1$, around x-axis.
10. Astroid $x^{\frac{2}{3}} + y^{\frac{2}{3}} \le 1$, around x-axis.

Exercise 3.8.24. Find the volume of the solid of revolution.

1. Region bounded by $y = x$ and $y = x^2$, around x-axis.
2. Region bounded by $y = x$ and $y = x^2$, around $x = 2$.
3. Region bounded by $y = x$ and $y = x^2$, around $x = y$.
4. Region bounded by $y = x$ and $y = x^2$, around $x + y = 0$.
5. Region bounded by $y^2 = x + 1$ and $x + y = 1$, around $x + y = 1$.
6. Region bounded by $y^2 = x + 1$ and $x + y = 1$, around $x = 3$.
7. Region bounded by $y^2 = x + 1$ and $x + y = 1$, around $y = 1$.
8. Region bounded by $y = \log x$, $y = 0$, $y = 1$, and y-axis, around y-axis.
9. Region bounded by $y = \cos x$ and $y = \sin x$, around $y = 1$.
10. Triangle with vertices $(0, 0), (1, 2), (2, 1)$, around x-axis.
11. Ellipse $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} \le 1$, around $y = b$.
12. Ellipse $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} \le 1$, around $bx + ay = 2ab$.
13. Region bounded by $y = \dfrac{1}{1 + |x|}$ and the x-axis, around x-axis.
14. Region bounded by $y = e^{-|x|}$ and the x-axis, around x-axis.

3.8.5 Cavalieri's Principle

The formulae for the area of surface of revolution and the volume of solid of revolution follow from a more general principle.

In general, an n-dimensional solid X has n-dimensional size. For $n = 1$, X is a curve and the size is the length. For $n = 2$, X is a region in $\mathbb{R}^2$ or more generally a surface, and the size is the area. For $n = 3$, X is typically a region in $\mathbb{R}^3$ but can also be a "3-dimensional surface" such as the 3-dimensional sphere in $\mathbb{R}^4$, and the size is the volume.

To find the size of an n-dimensional solid X, we may decompose X into sections X_t of one lower dimension (i.e., X_t has dimension $n-1$). For 2-dimensional X, this means that X is decomposed into a one parameter family of curves. For 3-dimensional X, this means that X is decomposed into a one parameter family of surfaces. The decomposition is *equidistant* if the distance between two nearby pieces does not depend on the location where the distance is measured. In this case, we have the distance function $s(t)$, such that the distance between the sections X_t and $X_{t+\Delta t}$ is $\Delta s = s(t+\Delta t) - s(t)$. If X spans from distance $s = a$ to distance $s = b$, then

$$\text{size of } X = \int_{t=a}^{t=b} \text{size}(X_t)ds = \int_a^b \text{size}(X_t)s'(t)dt.$$

A consequence of the formula is the following *principle of Cavalieri*: If two solids X and Y have equidistant decompositions X_t and Y_t, such that X_t and Y_t have the same size, and the distance between X_t and $X_{t'}$ is the same as the distance between Y_t and $Y_{t'}$, then X and Y have the same size.

Example 3.8.22. Let X be a region inside the plane. We decompose X by intersecting with vertical lines $X_x = X \cap x \times \mathbb{R}$. The decomposition is equidistant, with the x-coordinate as the distance. Thus the area of X is $\displaystyle\int_a^b \text{length}(X_x)dx$. In the special case X is the region between $f(x)$ and $g(x)$, where $f(x) \geq g(x)$, the section X_x is the interval $[g(x), f(x)]$ and has length $f(x) - g(x)$. Then we recover the formula $\displaystyle\int_a^b (f(x) - g(x))dx$ in Section 3.8.2.

Example 3.8.23. Let X be a region inside $\mathbb{R}^3$. We decompose X by intersecting with vertical planes $X_x = X \cap x \times \mathbb{R}^2$. The decomposition is equidistant, with the x-coordinate as the distance. Thus the volume of X is $\displaystyle\int_a^b \text{area}(X_x)dx$. In the special case X is obtained by rotating the region between $f(x)$ and $g(x)$, where $f(x) \geq g(x) \geq 0$, around the x-axis, the section X_x is the annulus with outer radius $f(x)$ and inner radius $g(x)$. The section has area $\pi(f(x)^2 - g(x)^2)$, and we get the

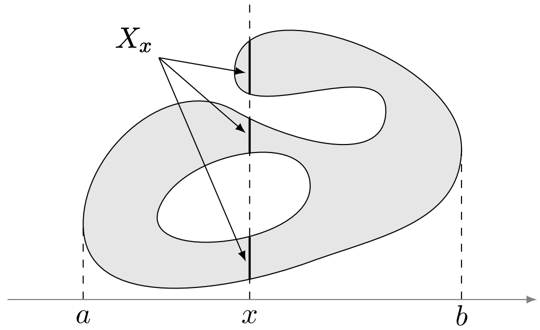


Figure 3.8.20: Area of a region in $\mathbb{R}^2$.

formula $\pi \int_a^b (f(x)^2 - g(x)^2)dx$ for the solid of revolution.

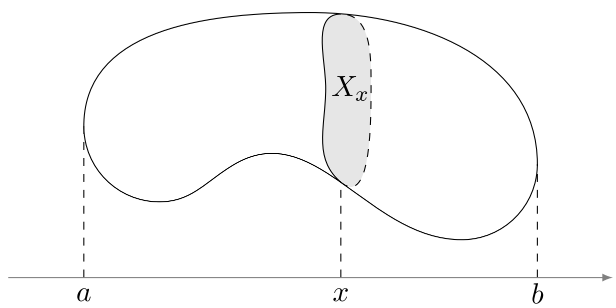


Figure 3.8.21: Volume of a solid in $\mathbb{R}^3$.

Example 3.8.24. Let R be a region in the plane. Let P be a point not in the plane. Connecting P to all points in R by straight lines produces the pyramid X with base R and apex P.

We may put R on the (x, y)-plane in $\mathbb{R}^3$ and assume that $P = (0, 0, h)$ lies in the positive z-axis, where h is the distance from P to the plane. Let A be the area of R. We decompose the pyramid by the horizontal planes, so that z is the distance. The section X_z is similar to R, so that the area of X_z is proportional to the square of its distance $h - z$ to P. We find the area of X_z to be $\left(\frac{h-z}{h}\right)^2 A$, and the volume of the pyramid is

$$\int_0^h \left(\frac{h-z}{h}\right)^2 A dx = \frac{1}{3}hA.$$

Example 3.8.25. Let X be the intersection of two round solid cylinders of radius 1 in

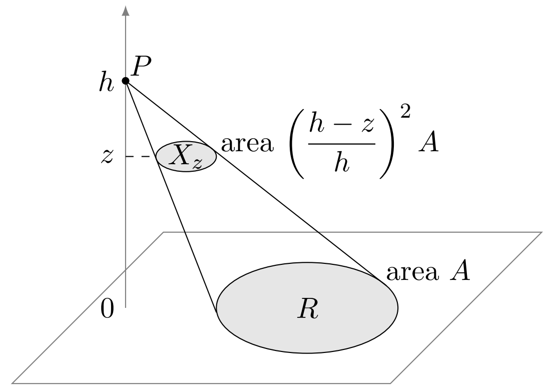


Figure 3.8.22: Pyramid.

orthogonal position. We put the two cylinders in $\mathbb{R}^3$, by assuming the two cylinders to be $x^2+y^2 \le 1$ and $x^2+z^2 \le 1$. Then we decompose the solid by intersecting with the planes perpendicular to the x-axis. The section X_x is a square of side length $2\sqrt{1-x^2}$ and therefore has area $4(1-x^2)$. The volume of the intersection solid X is

$$\int_{-1}^{1} 4(1-x^2)dx = \frac{16}{3}.$$

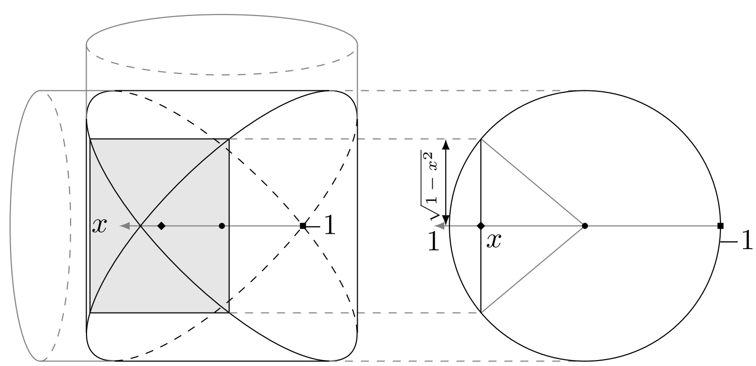


Figure 3.8.23: Orthogonal intersection of two cylinders.

Exercise 3.8.25. Explain the formula in Section 3.8.3 for the area of surface of revolution by using suitable equidistant decomposition.

Exercise 3.8.26. Explain that if a solid is stretched by a factor A in the x-direction, by B in the y-direction, and by C in the z-direction, then the volume of the solid is multiplied by the factor ABC.

Exercise 3.8.27. Find the volume of solid.

1. Ellipsoid $\dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}+\dfrac{z^2}{c^2}\le 1$.

2. Solid bounded by $\dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}-\dfrac{z^2}{c^2}=1$ and $z=\pm c$.

3. Intersection of the sphere $x^2+y^2+z^2\le 1$ and the cylinder $x^2+y^2\le x$.

4. Solid bounded by $x+y+z^2=1$ and inside the first quadrant.

Exercise 3.8.28. Find the volume of solid.

1. A solid with a disk as the base, and the parallel sections perpendicular with the base are equilateral triangles.

2. A solid with a disk as the base, and the parallel sections perpendicular with the base are squares.

3. Cylinder cut by two planes, one is perpendicular to the cylinder and the other form angle α with the cylinder. The two planes do not intersect inside the cylinder.

4. Cylinder cut by two planes forming respective angles α and β with the cylinder. The two planes do not intersect inside the cylinder.

5. A wedge cut out of a cylinder, by two planes forming respective angles α and β with the cylinder, such that the intersection of two planes is a diameter of the cylinder.

So far we used parallel lines and planes to construct the decomposition. We may also use equidistant curves and surfaces to construct the decomposition.

Example 3.8.26. We may decompose a region X in the plane by concentric circles. The decomposition is equidistant, with the radius r of the circles as the distance. The section X_r consists of the points in X of distance r from the origin and is typically an arc from angle $\phi(r)$ to angle $\psi(r)$. The length of the arc X_r is $(\psi(r)-\phi(r))r$, so that the area of X is $\displaystyle\int_a^b(\psi(r)-\phi(r))rdr$.

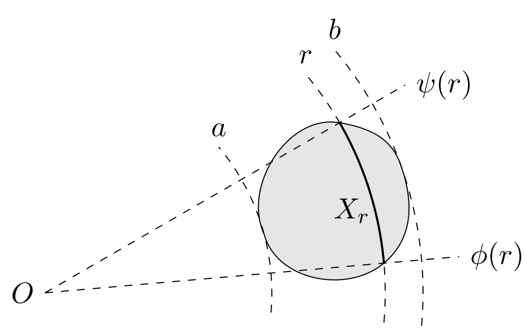


Figure 3.8.24: Equidistant decomposition by concentric circles.

For example, for the disk centered at $(1,0)$ and of radius 1, we have $\phi = -\arccos\frac{r}{2}$ and $\psi = \arccos\frac{r}{2}$, $r \in [0,2]$. The area of the disk of radius 1 is (taking $t = \arccos\frac{r}{2}$, $r = 2\cos t$)

$$\begin{aligned}\int_0^2 2r\arccos\frac{r}{2}dr &= \int_{\frac{\pi}{2}}^0 2t(2\cos t)d(2\cos t) = 8\int_0^{\frac{\pi}{2}} t\cos t\sin tdt \\ &= 4\int_0^{\frac{\pi}{2}} t\sin 2tdt = \int_0^{\pi} u\sin udu = \pi.\end{aligned}$$

Example 3.8.27. Let X be a region in the right plane (i.e., the right side of y-axis). Let Y be the solid obtained by revolving X around the y-axis. We may use the cylinders centered at the y-axis to decompose Y. The decomposition is equidistant, with x as the distance. Let X_x be the intersection of X with the vertical line $x \times \mathbb{R}$. Then the section Y_x is the cylinder obtained by revolving X_x around the y-axis. The area of the section is $2\pi x(\text{length of } X_x)$. Therefore the volume of the solid of revolution is $2\pi\int_a^b x(\text{length of } X_x)dx$. In particular, if X is the region between functions $f(x)$ and $g(x)$, where $f(x) \geq g(x)$ on $[a,b]$, then the volume is $2\pi\int_a^b x(f(x)-g(x))dx$.

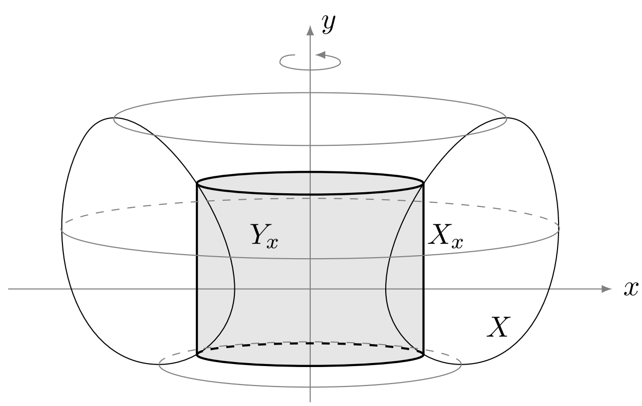


Figure 3.8.25: Equidistant decomposition by concentric cylinders.

For example, consider the solid torus in Example 3.8.17. The disk is the region between $x = \sqrt{a^2-(y-b)^2}$ and $x = -\sqrt{a^2-(y-b)^2}$, for $y \in [b-a, b+a]$. If we use the formula above (note that x and y are exchanged), we get the volume of the solid torus

$$\begin{aligned}2\pi\int_{b-a}^{b+a} y(2\sqrt{a^2-(y-b)^2})dy &= 4\pi\int_{-a}^{a}(t+b)\sqrt{a^2-t^2})dt \\ &= 8\pi\int_0^a b\sqrt{a^2-t^2}dt = 4\pi^2a^2b.\end{aligned}$$

Exercise 3.8.29. Compute the volumes of the solids of revolution in Example 3.8.21 by using the formula in Example 3.8.27.

Exercise 3.8.30. Compute the volumes of the solids of revolution in Exercise 3.8.23 by using the formula in Example 3.8.27.

Exercise 3.8.31. Compute the volumes of the solids of revolution in Exercise 3.8.24 by using the formula in Example 3.8.27.

Exercise 3.8.32. After Example 3.8.21, we presented the formula for computing the volume of a solid obtained by revolving a region in $\mathbb{R}^2$ bounded by a parameterized curve. Can you derive the similar formula by using the idea from Example 3.8.27?

Exercise 3.8.33. In Section 3.8.4 and Example 3.8.27, we have two ways of computing the volume of a solid of revolution. For the following simple case, explain that the two ways give the same result: Let $f(x)$ be and invertible non-negative function on $[0,a]$, such that $f(a)=0$ and both $f(x)$ and $f^{-1}(y)$ are continuously differentiable. The solid is obtained by revolving the region between the graph of f and the two axis.

Example 3.8.28. Finally, we compute the size of high dimensional objects. Let α_n be the volume of the n-dimensional sphere S^n of radius 1. Then

$$\alpha_0 = 2, \quad \alpha_1 = 2\pi, \quad \alpha_2 = 4\pi.$$

Moreover, the n-dimensional sphere of radius r has volume $\alpha_n r^n$.

To compute α_n, we decompose S^n by intersecting with "horizontal hyperplanes". The hyperplanes are indexed by the angle t. The section at angle t is the $(n-1)$-dimensional sphere S^{n-1} of radius $\cos t$, and form an equidistant decomposition. In fact, the angle t can be used to measure the distance between the sections. Since the section at t has volume $\alpha_{n-1}\cos^{n-1} t$ and the range of t is $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, we conclude that

$$\alpha_n = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \alpha_{n-1}\cos^{n-1} t dt = 2\alpha_{n-1} I_{n-1},$$

where $I_{n-1} = \int_0^{\frac{\pi}{2}} \cos^{n-1} t dt = \int_0^{\frac{\pi}{2}} \sin^{n-1} t dt$ has been computed in Example 3.5.14

$$I_{2k} = \frac{(2k)!}{2^{2k+1}(k!)^2}\pi, \quad I_{2k+1} = \frac{2^{2k}(k!)^2}{(2k+1)!}.$$

Thus

$$\alpha_n = 2\alpha_{n-1}I_{n-1} = 4\alpha_{n-2}I_{n-1}I_{n-2} = 4\alpha_{n-2}\frac{\pi}{2(n-1)} = \frac{2\pi}{n-1}\alpha_{n-2}.$$

By the values of α_1 and α_2, we conclude that

$$\alpha_n = \begin{cases} \dfrac{(2\pi)^{\frac{n}{2}}}{2\cdot 4\cdots(n-2)}, & \text{if } n \text{ is even,} \\ \dfrac{2(2\pi)^{\frac{n-1}{2}}}{1\cdot 3\cdots(n-2)}, & \text{if } n \text{ is odd.} \end{cases}$$

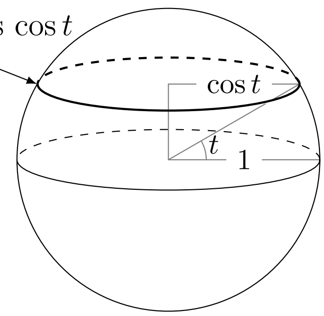


Figure 3.8.26: Decomposing n-dimensional sphere of radius 1.

Exercise 3.8.34. Let β_n be the volume of the ball B^n of radius 1.

1. Similar to Example 3.8.28, use the intersection with horizontal hyperplanes to derive the relation between β_n and β_{n-1}. Then use the special values β_1 and β_2, and Example 3.5.14 to compute β_n.

2. Use the decomposition of B^n by concentric $(n-1)$-dimensional spheres to derive the relation between β_n and α_{n-1}. Then use Example 3.8.28 to find β_n.

The two methods should give the same result.

Exercise 3.8.35. Suppose R is a region in $\mathbb{R}^{n-1}$ with volume. Suppose P is a point in $\mathbb{R}^n$ of distance h from $\mathbb{R}^{n-1}$. By connecting P to all points of R by straight lines, we get a pyramid X with base R and apex P. Find the relation between the volumes of X and R.

3.9 Polar Coordinate

The polar coordinate locates a point on the plane by its distance r to the origin and the angle θ indicating the direction from the viewpoint of origin. It is roughly related to the cartesian coordinates (x, y) by

$$x = r\cos\theta, \quad y = r\sin\theta; \quad r = \sqrt{x^2+y^2}, \quad \theta = \arctan\frac{y}{x}.$$

We say "roughly" because the relation between (x, y) and (r, θ) is not a one-to-one correspondence. For example, the last formula literally restricts θ, as the value of inverse tangent function, to be within $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. In fact, the angle for a point in the plane is unique only up to adding an integer multiple of 2π, and is more precisely determined by

$$(\cos\theta, \sin\theta) = \frac{(x,y)}{\sqrt{x^2+y^2}}.$$

Another way to say this is that θ is unique if we restrict to $[0, 2\pi)$ (or $[-\pi, \pi)$, etc.).

For the convenience of presenting polar equations, we also allow r to be negative, by specifying that $(-r, \theta)$ and $(r, \theta+\pi)$ represent the same point. In other words, $(-r,\theta)$ and (r,θ) are symmetric with respect to the origin. The cost of such extension is more ambiguity in the polar coordinates of a point because all the following represent the same point

$$(r,\theta),\ (-r,\theta\pm\pi),\ (r,\theta\pm 2\pi),\ (-r,\theta\pm 3\pi),\ \dots.$$

3.9.1 Curves in Polar Coordinate

Example 3.9.1. The equation $r = c$ is the circle of radius $|c|$ centered at the origin. The equation $\theta = c$ is a straight line passing through the origin.

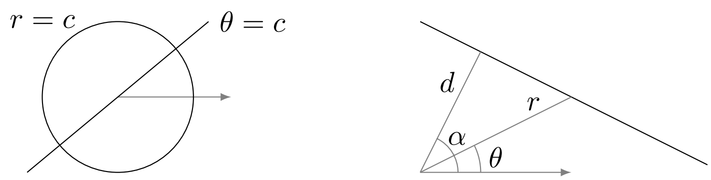


Figure 3.9.1: $r = c$, $\theta = c$, and polar equation for general straight line.

The equation for a general straight line is

$$r = \frac{d}{\cos(\alpha - \theta)}.$$

Moreover, $r = a\cos\theta$ is the circle of diameter a passing through the origin.

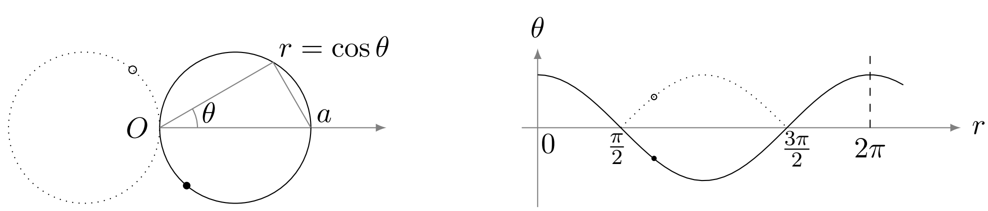


Figure 3.9.2: Circle $r = a\cos\theta$.

Exercise 3.9.1. Find the cartesian equation.

1. $r = 2$.
2. $r = -2$.
3. $r = \sin\theta$.
4. $r\sin\theta = 1$.
5. $r = \tan\theta\sec\theta$.
6. $r = \cos\theta + \sin\theta$.

Exercise 3.9.2. Find the polar equation.

1. $x = 1$.
2. $y = -1$.
3. $x + y = 1$.
4. $x = y^2$.
5. $x^2 + y^2 = x$.
6. $xy = 1$.

Exercise 3.9.3. What is the polar equation of the curve obtained by flipping $r = f(\theta)$ with respect to the origin? Then use your conclusion to find the curve $r = -\cos\theta$.

Exercise 3.9.4. What is the relation between the curves $r = f(\theta)$ and $r = -f(\theta + \pi)$?

Exercise 3.9.5. What is the polar equation of the curve obtained by rotating $r = f(\theta)$ by angle α? Then use your conclusion to answer the following.

1. What is the curve $r = \sin\theta$?
2. Find the polar equation for a general circle passing through the origin.

Exercise 3.9.6. Find the polar equation of a general circle.

Example 3.9.2. The *Archimedean spiral* is $r = \theta$. Note that $r < 0$ when $\theta < 0$, so that a flipping with respect to the origin is needed when we draw the part of the spiral corresponding to $\theta < 0$. The symmetry with respect to the y-axis is due to the fact that if (r, θ) satisfies $r = \theta$, then $(-r, -\theta)$ also satisfies $r = \theta$.

The *Fermat's spiral* is $r^2 = \theta$. The symmetry with respect to the origin is due to the fact that if (r, θ) satisfies $r^2 = \theta$, then $(-r, \theta)$ also satisfies $r^2 = \theta$.

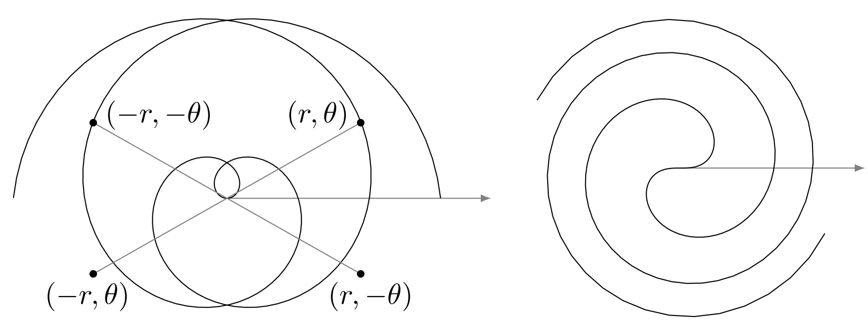


Figure 3.9.3: Spirals $r = \theta$ and $r^2 = \theta$.

Example 3.9.3. The curve $r = 1 + \cos\theta$ is a *cardioid*. Its clockwise rotation by 90° is another cardioid $r = 1 + \sin\theta$. More generally, the curve $r = a + \cos\theta$ is a *limaçon*.

The curve intersects itself when $|a| < 1$ and does not intersect itself when $|a| > 1$. The symmetry with respect to the x-axis is due to the fact that if (r, θ) satisfies $r = a + \cos\theta$, then $(r, -\theta)$ also satisfies the equation.

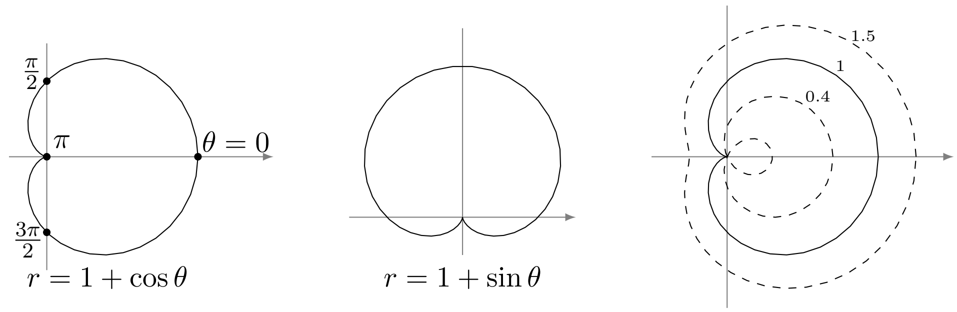


Figure 3.9.4: Cardioids and limaçons $r = a + \cos\theta$, $a = 0.4, 1, 1.5$.

The cardioid originates from the following geometrical construction. Consider a circle C of diameter 1 rolling outside of a circle A of equal diameter 1. This is the same as the circle rolling inside a big circle B of diameter 3. The track traced by a point on C is the cardioid. Note that the origin O of the polar coordinate should be a point on A, not the center of A.

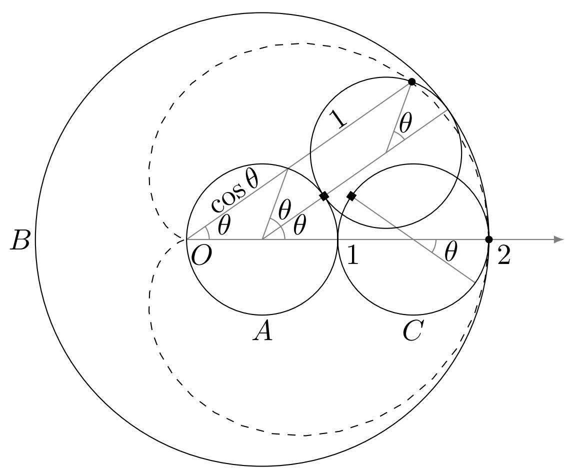


Figure 3.9.5: Origin of the cardioid.

If we imagine the rolling circle C as part of a rolling disk D, and we fix a point in D of distance d from the center of C. Then the track tranced by the point is the limaçon $r = 1 + 2d\cos\theta$, with the origin of the polar coordinate being a point of distance d form the center of A.

Example 3.9.4. The curve $r = \cos 2\theta$ is the *four-leaved rose*, and $r = \cos 3\theta$ is the *three-leaved rose*. The circle $r = \cos\theta$ can be considered as the one-leaved rose.

In general, the curve $r = \cos n\theta$ can be described as follows. For θ in the arc $I = \left[-\frac{\pi}{2n}, \frac{\pi}{2n}\right]$, the value of r goes from 0 to 1 and then back to 0, so that the corresponding curve is one leaf occupying $\frac{\pi}{n}$ angle of the whole circle. This is the leaf in $\left[-\frac{\pi}{4}, \frac{\pi}{4}\right]$ for $n = 2$ and in $\left[-\frac{\pi}{6}, \frac{\pi}{6}\right]$ for $n = 3$. For θ in the second arc $I + \frac{\pi}{n}$, we need to rotate this first leaf by angle $\frac{\pi}{n}$ and then flipping with respect to the origin (because r becomes negative), which gives a leaf occupying $I + \frac{\pi}{n} + \pi$. This is the leaf in $\left[\frac{5\pi}{4}, \frac{7\pi}{4}\right]$ for $n = 2$ and in $\left[\frac{7\pi}{6}, \frac{9\pi}{6}\right]$ for $n = 3$. For θ in the third arc $I + \frac{2\pi}{n}$, we get the leaf obtained by rotating the first leaf by angle $\frac{2\pi}{n}$ (no flipping needed now because r becomes non-negative again), which gives a leaf occupying $I + \frac{2\pi}{n}$. This is the leaf in $\left[\frac{3\pi}{4}, \frac{5\pi}{4}\right]$ for $n = 2$ and in $\left[\frac{3\pi}{6}, \frac{5\pi}{6}\right]$ for $n = 3$. Keep going, we see two distinct patterns depending on the parity of n.

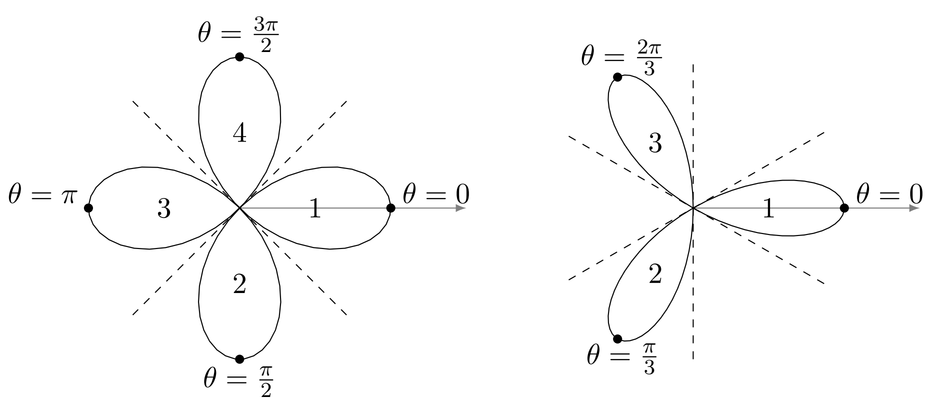


Figure 3.9.6: Four-leaved rose $r = \cos 2\theta$ and three-leaved rose $r = \cos 3\theta$.

More generally, we may consider $r = \cos p\theta$. Again we get first leaf occupying $I = \left[-\frac{\pi}{2p}, \frac{\pi}{2p}\right]$, the second leaf occupying $I + \frac{\pi}{p} + \pi$, the third leaf occupying $I + \frac{2\pi}{p}$, etc. The pattern could be very complicated, depending on whether p is rational or irrational, and in case p is rational, the parity of the numerator and denominator of p.

Finally, $r = \sin p\theta$ is obtained by rotating $r = \cos 2\theta$ by $\frac{\pi}{2p}$. We also get many leaved roses by other rotations.

Exercise 3.9.7. Describe the curve.

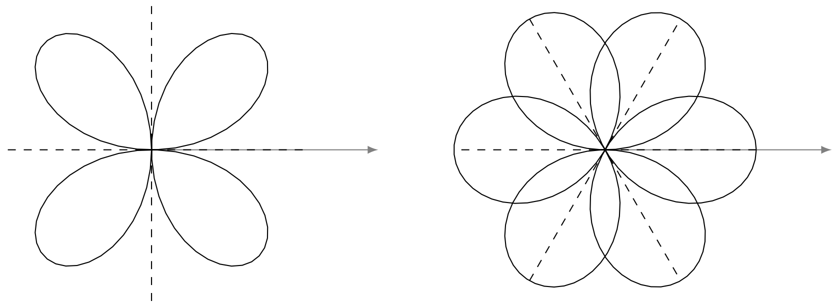


Figure 3.9.7: Many leaved roses $r = \sin 2\theta$ and $r = \sin \frac{3}{2}\theta$.

1. $r = -\theta$.
2. $r = \theta + \pi$.
3. $r = 2\theta$.
4. $r^2 = -\theta$.
5. $r^2 = 4\theta$.
6. $r^2 = \theta + \pi$.
7. $e^r = \theta$.
8. $r\theta = 1$.
9. $r = 2 + \cos\theta$.
10. $r = 2 + 3\cos\theta$.
11. $r = 2 - \cos\theta$.
12. $r = \cos\theta + \sin\theta$.
13. $r = 1 + \cos\theta + \sin\theta$.
14. $r = \cos 4\theta$.
15. $r = 2\sin 5\theta$.
16. $r = -3\sin 6\theta$.
17. $r = \sin 2\theta - \cos 2\theta$.
18. $r = \sin 2\theta + 2\cos 2\theta$.
19. $r = \cos\dfrac{4}{3}\theta$.
20. $r = \sin\dfrac{5}{3}\theta$.
21. $r = \cos\dfrac{1}{3}\theta$.
22. $r = \cos\dfrac{2}{3}\theta$.
23. $r = \cos\dfrac{2}{3}\theta + \sin\dfrac{2}{3}\theta$.
24. $r^2 = \sin 2\theta$.
25. $r^2 = -\cos 4\theta$.
26. $r = 1 + 2\cos\dfrac{1}{2}\theta$.
27. $r = 2 + \cos\dfrac{1}{2}\theta$.

3.9.2 Geometry in Polar Coordinate

The curve $r = f(\theta)$ for $\theta \in [\alpha, \beta]$ is the parameterized curve

$$x = f(\theta)\cos\theta, \quad y = f(\theta)\sin\theta, \quad \theta \in [\alpha, \beta]$$

in the cartesian coordinate. The length of the curve is

$$\int_\alpha^\beta \sqrt{(f(\theta)\cos\theta)'^2 + (f(\theta)\sin\theta)'^2}\,d\theta = \int_\alpha^\beta \sqrt{f^2 + f'^2}\,d\theta.$$

For the area in terms of polar coordinate, assume $f \ge 0$ and consider the region $X_{[\alpha,\beta]}(f)$ bounded by $r = f(\theta)$, $\theta \in [\alpha, \beta]$, and the rays $\theta = \alpha$ and $\theta = \beta$. Using the idea of Section 3.1.1, let $A(\theta)$ be the area of the region $X_{[\alpha,\theta]}(f)$. Then the change $A(\theta + h) - A(\theta)$ is the area of $X_{[\theta,\theta+h]}(f)$. Since $X_{[\theta,\theta+h]}(f)$ is sandwiched between fans of angle between θ, $\theta + h$ and radii $m = \min_{[\theta,\theta+h]} f$, $M = \max_{[\theta,\theta+h]} f$, we get

$$\frac{1}{2}m^2h \le A(\theta + h) - A(\theta) \le \frac{1}{2}M^2h.$$

Here the left and right sides are the known areas of the fans. The inequality is the same as

$$\frac{1}{2}m^2 \le \frac{A(\theta+h)-A(\theta)}{h} \le \frac{1}{2}M^2.$$

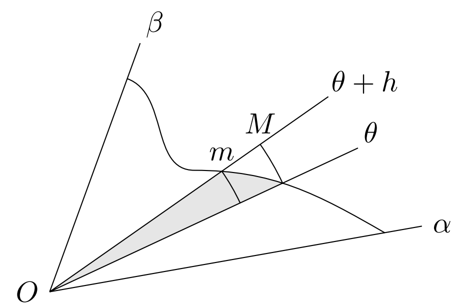


Figure 3.9.8: Estimate the change of area.

If f is continuous, then $\lim_{h\to 0} m = \lim_{h\to 0} M = f(\theta)$. By the sandwich rule, we get

$$A'(\theta) = \frac{1}{2}f(\theta)^2.$$

Therefore the area of $X_{[\alpha,\beta]}(f)$ is

$$A(\beta) = \frac{1}{2}\int_\alpha^\beta f(\theta)^2 d\theta.$$

Example 3.9.5. The cardioid $r = 1+\sin\theta$ has length

$$\begin{aligned}\int_0^{2\pi} \sqrt{(1+\sin\theta)^2+(1+\sin\theta)'^2}d\theta &= \int_0^{2\pi}\sqrt{2(1+\sin\theta)}d\theta \\ &= \int_{\frac{1}{4}\pi}^{-\frac{3}{4}\pi}\sqrt{2(1+\cos 2t)}d\left(\frac{\pi}{2}-2t\right) \\ &= 4\int_{-\frac{3}{4}\pi}^{\frac{1}{4}\pi}|\cos t|dt = 4\int_0^\pi \cos t dt = 8.\end{aligned}$$

The region enclosed by the cardioid has area

$$\frac{1}{2}\int_0^{2\pi}(1+\sin\theta)^2 d\theta = \frac{1}{2}\int_0^{2\pi}(1+2\sin\theta+\sin^2\theta)d\theta = \frac{3}{2}\pi.$$

Example 3.9.6. Let $p > \dfrac{1}{2}$. Then one leaf of the rose $r = \cos p\theta$ is from the angle

$-\frac{\pi}{2p}$ to the angle $\frac{\pi}{2p}$. The length of the leaf is

$$\begin{aligned}\int_{-\frac{\pi}{2p}}^{\frac{\pi}{2p}} \sqrt{(\cos p\theta)^2 + (\cos p\theta)'^2} d\theta &= \int_{-\frac{\pi}{2p}}^{\frac{\pi}{2p}} \sqrt{\cos^2 p\theta + p^2 \sin^2 p\theta} d\theta \\ &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sqrt{p^{-2}\cos^2 t + \sin^2 t} dt \\ &= \frac{1}{2}\int_0^{2\pi} \sqrt{1 + (p^{-2} - 1)\cos^2 t} dt.\end{aligned}$$

This is the elliptic integral in Example 3.8.1. Moreover, the area of the leaf is

$$\frac{1}{2}\int_{-\frac{\pi}{2p}}^{\frac{\pi}{2p}} (\cos p\theta)^2 d\theta = \frac{\pi}{4p}.$$

Example 3.9.7. The cardioid $r = 1 + \cos\theta$ and the circle $r = 3\cos\theta$ intersect at $\theta = \pm\frac{\pi}{3}$. The area of the region outside the cardioid and inside the circle is

$$\frac{1}{2}\int_{-\frac{\pi}{3}}^{\frac{\pi}{3}} ((3\cos\theta)^2 - (1 + \cos\theta)^2) d\theta = \pi.$$

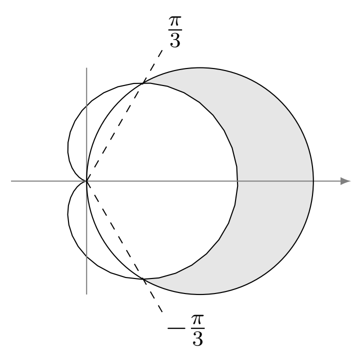


Figure 3.9.9: Ourside cardioid $r = 1 + \cos\theta$ and inside circle $r = 3\cos\theta$.

Example 3.9.8. We try to find the volume of the solid of revolution obtained by revolving the region between the two leaves of the limaçon $r = a + \cos\theta$, $0 < a < 1$, around the x-axis. In the cartesian coordinate, the curve is parameterized by

$$x = (a + \cos\theta)\cos\theta, \quad y = (a + \cos\theta)\sin\theta, \quad \theta \in [0, \pi].$$

Let $\theta = \alpha$ at the origin O. Then the volume we are looking for is the volume of the solid of revolution from $\theta = 0$ to $\theta = \alpha$, subtracting the volume of the solid of revolution from $\theta=\alpha$ to $\theta=\pi$. As θ goes from 0 to α, we are moving opposite to the direction of the x-axis. Therefore the first volume is $-\pi\int_{\theta=0}^{\theta=\alpha} y^2dx$. As θ goes from α to π, we are moving in the direction of the x-axis. Therefore the second volume is $\pi\int_{\theta=\alpha}^{\theta=\pi} y^2dx$. We conclude that the volume we are looking for is

$$
\begin{aligned}
-\pi\int_{\theta=0}^{\theta=\alpha} y^2dx-\pi\int_{\theta=\alpha}^{\theta=\pi} y^2dx &= -\pi\int_{\theta=0}^{\theta=\pi} y^2dx \\
&= -\pi\int_0^{\pi}(a+\cos\theta)^2\sin^2\theta d[(a+\cos\theta)\cos\theta] \\
&= -\pi\int_0^{\pi}(a+\cos\theta)^2(1-\cos^2\theta)(a+2\cos\theta)d(\cos\theta) \\
&= -\pi\int_1^{-1}(a+t)^2(1-t^2)(a+2t)dt=\frac{4}{3}\pi a(a^2+1).
\end{aligned}
$$

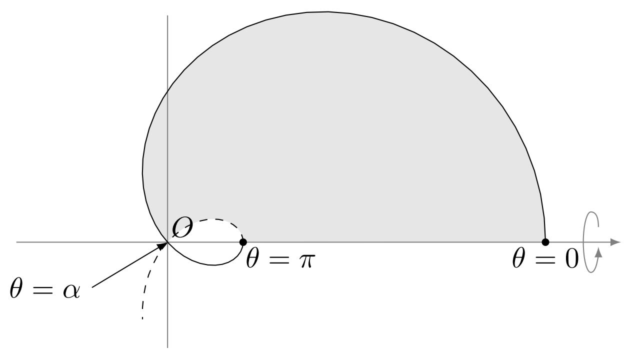


Figure 3.9.10: Revolving the region between two leaves of a limaçon.

Exercise 3.9.8. What is the length of lemiçon? What is the area of the region enclosed by lemiçon? Note that for $|c|>1$, we have two parts of the lemiçon and two regions.

Exercise 3.9.9. Find length of the part of the cardioid $r=1+\cos\theta$ in the first quadrant. Moreover, find the area of the region enclosed by this part and the two axes.

Exercise 3.9.10. Find the area of the region enclosed by *strophoid* $r=2\cos\theta-\sec\theta$.

Exercise 3.9.11. Find length.

1. $r=\theta$, $\theta\in[0,\pi]$.
2. $r=\theta^2$, $\theta\in[0,\pi]$.
3. $r=e^{\theta}$, $\theta\in[0,2\pi]$.

Exercise 3.9.12. Find area.

1. Bounded by $r=\theta$, $\theta\in[0,\pi]$ and the x-axis.

2. Outside $r = 1$ and inside $r = 2\cos\theta$.
3. Inside $r = 1$ and outside $r = 2\cos\theta$.
4. Inside both $r = 1$ and $r = 1 + \cos\theta$.
5. Outside $r = 3\sin\theta$ and inside $r = 2 - \sin\theta$.
6. Inside both $r = \cos 2\theta$ and $r = \sin 2\theta$.
7. Inside both $r = 1 + c\cos\theta$ and $r = 1 + c\sin\theta$, $|c| < 1$.
8. Inside both $r = 1 + c\cos\theta$ and $r = 1 - c\cos\theta$, $|c| < 1$.
9. Inside both $r^2 = \cos 2\theta$ and $r^2 = \sin 2\theta$.
10. Outside $r = 1$ and inside $r = 2\cos 3\theta$.
11. Between the two loops of $r = 1 + 2\cos 3\theta$.

3.10 Application to Physics

3.10.1 Work and Pressure

Integration is also widely used to compute physical quantities. If under a constant force F, an object moves by a distance d in the direction of the force, then the *work* done by the force is Fd. In general, however, the force may vary. For simplicity, assume the object moves along the x-axis, from $x = a$ to $x = b$, and a horizontal force $F(x)$ is applied when the object is at location x. Then the work done by the force when the objects moves a little bit from x to $x + \Delta x$ is approximately $\Delta W \approx F(x)\Delta x$.

Similar to the earlier argument, let $W(x)$ be the work done by the force when the object moves from a to x. Since the work is *additive*, we have $\Delta W = W(x + \Delta x) - W(x)$. The approximation $\dfrac{\Delta W}{\Delta x} \approx F(x)$ becomes more accurate as $\Delta x \to 0$, and we get an equality after taking the limit

$$W'(x) = \lim_{\Delta x \to 0} \frac{W(x + \Delta x) - W(x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta W}{\Delta x} = F(x).$$

This implies that the work done for the whole trip from a to b is

$$W(b) = \int_a^b F(x)dx.$$

Example 3.10.1. Suppose one end of spring is fixed and the other end is attached to an object. In the natural position, when the spring is neither stretched nor compressed, no force is exercised on the object. When the position of the object deviate from the natural position by x, however, Hooke's law says that the spring exercises a force $F(x) = -kx$ on the object. Here k is the *spring constant*, and the negative sign indicates that the direction of the force is opposite to the direction of the deviation.

If the object starts at distance a from its natural position, then the work done by the spring in pulling the object to its natural position is

$$\int_0^a kx dx = \frac{k}{2}a^2.$$

Here we use the positive sign because the direction of movement is the same as the direction of the force.

The argument about the work done by a force is quite typical. In general, if a quantity is additive, then the quantity can be decomposed into small pieces. The estimation of each small piece tells us the change of the quantity. The whole quantity is then the integration of the change.

In the subsequent examples, we will only analyze a small piece of an additive quantity. We will omit the limit part of the argument and directly write down the corresponding integration.

Example 3.10.2. We want to find the work it takes to pump a bucket of liquid out of the top of the bucket.

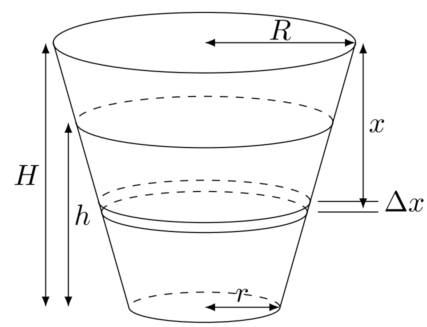


Figure 3.10.1: Bucket of liquid.

Suppose the bucket has base diameter r, top diameter R, and height H. Suppose the liquid has density ρ and depth h. We decompose liquid into horizontal sections. At distance x from the top, the section is a disk of radius $r(x)$ satisfying

$$\frac{r(x) - r}{R - r} = \frac{H - x}{H}.$$

The liquid of thickness Δx and at distance x from the top has (approximate) weight $g\rho\pi r(x)^2\Delta x$ (g is the gravitational constant). The work it takes to lift this piece of liquid to the top of bucket is $\Delta W \approx (g\rho\pi r(x)^2\Delta x)x = \pi g\rho x r(x)^2\Delta x$. Since the liquid spans from $x = H - h$ to $x = H$, the total work needed is

$$\begin{aligned} W &= \pi g\rho \int_{H-h}^{H} x r(x)^2 dx = \frac{\pi g\rho}{H^2}\int_{H-h}^{H} x\left[(R-r)(H-x)+rH\right]^2 dx \\ &= \pi g\rho H^2R^2\left(a^2b + \frac{1}{2}a(2a-3b)b^2 + \frac{1}{3}(1-a)(1-3a)b^3 - \frac{1}{4}(1-a)^2b^4\right), \end{aligned}$$

where $a = \dfrac{r}{R}$ and $b = \dfrac{h}{H}$.

Example 3.10.3. We want to find the force exercised by water on a dam.

Let ρ be the density of water. At the depth x, the pressure of water is ρx per unit area. Now suppose the dam is a vertical trapezoid with base length l, top length L, and height H. We decompose dam into horizontal sections. At distance x from the top, the section is a strip of height Δx and length $l(x)$ satisfying

$$\frac{l(x) - l}{L - l} = \frac{H - x}{H}.$$

The force exercised on the strip is $\Delta F \approx (\rho x)l(x)\Delta x$. Since the water spans from $x = 0$ to $x = H$, the total force

$$F = \rho\int_0^H x l(x)dx = \rho\int_0^H x\left(L - \frac{L-l}{H}x\right)dx = \frac{1}{6}\rho H^2(L + 2l).$$

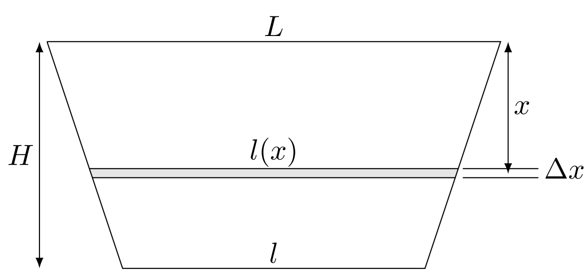


Figure 3.10.2: Hydraulic dam.

Exercise 3.10.1. A spring has natural length a. If the force F is needed to stretch the spring to length b, how muck work is needed to stretch the spring from the natural length to the length b?

Exercise 3.10.2. A ball of radius R is full of liquid of density ρ. Due to the gravity, the liquid leaks out of a hole at the bottom of the ball. How much work is done by the gravity in draining all the liquid?

Exercise 3.10.3. A circular disk of radius r is fully submerged in liquid of density ρ, such that the center of the disk is at depth h. What is the force exercised by the liquid on one side of the plate? Note that the plate may be inclined at some angle.

Exercise 3.10.4. A ball of radius r is fully submerged in liquid of density ρ, such that the center of the disk is at depth h. What is the force exercised by the liquid on the ball?

Exercise 3.10.5. A cable of mass m and length l has a mass M tied to the lower end. How much word is done in using the cable to lift the mass M to the top end of the cable?

Exercise 3.10.6. Newton's law of gravitation says that two bodies with masses m and M attract each other with a force $F = \dfrac{gmM}{d^2}$, where d is the distance between the bodies. Suppose the radius of the earth is R and the mass is M. How much work is needed to launch a satellite of mass m vertically to a circular orbit of height H? What is the minimal initial velocity needed for the satellite to escape the earth's gravity?

3.10.2 Center of Mass

Consider n masses $m_1, m_2, \dots, m_n$ distributed at the locations $x_1, x_2, \dots, x_n$ along a straight line. The *center of mass* is

$$\bar{x} = \frac{m_1x_2 + m_2x_2 + \cdots + m_nx_n}{m_1 + m_2 + \cdots + m_n}.$$

The center has the physical meaning that the total moment of the system with respect to $\bar{x}$ is zero, or the system is balanced with respect to $\bar{x}$.

Now suppose we have masses distributed throughout an interval $[a, b]$, with the density $\rho(x)$ at location x. We partition the interval into small pieces

$$P\colon a = x_0 < x_1 < x_2 < \cdots < x_n = b.$$

Then the system is decomposed into n pieces. The i-th piece can be approximately considered as a mass $m_i = \rho(x_i^*)(x_i - x_{i-1})$ located at x_i^*, for some $x_i^* \in [x_{i-1}, x_i]$. The whole system is approximated by the system of n pieces, and has approximate center of mass

$$\bar{x}_P = \frac{\sum_{i=1}^n \rho(x_i^*)(x_i - x_{i-1})x_i^*}{\sum_{i=1}^n \rho(x_i^*)(x_i - x_{i-1})}.$$

The denominator is the Riemann sum of the function $x\rho(x)$ and the numerator is the Riemann sum of the function $\rho(x)$ (see the beginning of Section 3.3.1). Therefore as the partition gets more and more refined, the limit becomes the center of mass

$$\bar{x} = \frac{\displaystyle\int_a^b x\rho(x)dx}{\displaystyle\int_a^b \rho(x)dx}.$$

The center of mass can be extended to higher dimensions, simply by considering each coordinate separately. For example, the system of n masses $m_1, m_2, \dots, m_n$ at $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ in the plane has the center of mass $(\bar{x}, \bar{y})$ given by

$$\bar{x} = \frac{\sum m_i x_i}{\sum m_i}, \quad \bar{y} = \frac{\sum m_i y_i}{\sum m_i}.$$

Now consider masses distributed along a curve $(x(t), y(t))$, $t \in [a, b]$, with the density $\rho(t)$ at location t. Take a partition P of $[a, b]$. The curve is approximated by straight line segments connecting $(x(t_{i-1}), y(t_{i-1}))$ to $(x(t_i), y(t_i))$. The i-th straight line segment has length $\Delta s_i = \sqrt{(x(t_i) - x(t_{i-1}))^2 + (y(t_i) - y(t_{i-1}))^2}$ and can be approximately considered as a mass $m_i = \rho(t_i^*)\Delta s_i$ located at $(x(t_i^*), y(t_i^*))$, for some $t_i^* \in [t_{i-1}, t_i]$. The whole system is approximated by the system of n pieces, and has approximate center of mass

$$\bar{x}_P = \frac{\sum_{i=1}^n (\rho(t_i^*)\Delta s_i) x(t_i^*)}{\sum_{i=1}^n \rho(t_i^*)\Delta s_i}, \quad \bar{y}_P = \frac{\sum_{i=1}^n (\rho(t_i^*)\Delta s_i) y(t_i^*)}{\sum_{i=1}^n \rho(t_i^*)\Delta s_i}.$$

As the partition gets more and more refined, the limit becomes the center of mass

$$\bar{x} = \frac{\displaystyle\int_a^b x(t)\rho(t)ds}{\displaystyle\int_a^b \rho(t)ds}, \quad \bar{y} = \frac{\displaystyle\int_a^b y(t)\rho(t)ds}{\displaystyle\int_a^b \rho(t)ds}, \quad ds = \sqrt{x'(t)^2 + y'(t)^2}dt.$$

Example 3.10.4. For constant density $\rho(x) = \rho$ distributed on the interval, the center of mass is the middle point

$$\bar{x} = \frac{\displaystyle\int_a^b x\rho dx}{\displaystyle\int_a^b \rho dx} = \frac{\rho\frac{1}{2}(b^2 - a^2)}{\rho(b - a)} = \frac{a + b}{2}.$$

If the density is $\rho(x) = \lambda + \mu x$, which is linearly increasing, then the center of mass is

$$\bar{x} = \frac{\displaystyle\int_a^b x(\lambda + \mu x)dx}{\displaystyle\int_a^b (\lambda + \mu x)dx} = \frac{3\lambda(a + b) + 2\mu(a^2 + ab + b^2)}{3(2\lambda + \mu(a + b))}.$$

Example 3.10.5. Consider the semi-circular curve of radius r and constant density ρ. We have

$$x = r\cos\theta, \quad y = r\sin\theta, \quad ds = rd\theta, \quad 0 \le \theta \le \pi,$$

and the center of mass is

$$\bar{x} = \frac{\displaystyle\int_0^\pi (r\cos\theta)\rho r d\theta}{\displaystyle\int_0^\pi \rho r d\theta} = 0, \quad \bar{y} = \frac{\displaystyle\int_0^\pi (r\sin\theta)\rho r d\theta}{\displaystyle\int_0^\pi \rho r d\theta} = \frac{2r}{\pi}.$$

Exercise 3.10.7. Find the center of mass of the parabola $y = x^2$, $x \in [0,2]$, of constant density.

Exercise 3.10.8. Find the center of mass of a triangle of constant density and with vertices at $(-1,0)$, $(0,\sqrt{15})$ and $(7,0)$.

Exercise 3.10.9. Let $m_{[a,b]}$ and $\bar{x}_{[a,b]}$ be the mass and the center of mass of a distribution of masses on $[a,b]$ with the density $\rho(x)$. Let $[a,b] = [a,c] \cup [c,b]$ and similarly introduce $m_{[a,c]}$, $m_{[c,b]}$, $\bar{x}_{[a,c]}$, $\bar{x}_{[c,b]}$. Show that the center of mass has the distribution property

$$m_{[a,b]} = m_{[a,c]} + m_{[c,b]}, \quad \bar{x}_{[a,b]} = \frac{m_{[a,c]}\bar{x}_{[a,c]} + m_{[c,b]}\bar{x}_{[c,b]}}{m_{[a,c]} + m_{[c,b]}}.$$

Does the property extend to curves in $\mathbb{R}^2$?

Chapter 4

Series

4.1 Series of Numbers

A *series* is an infinite sum

$$\sum_{n=1}^{\infty} a_n = a_1 + a_2 + a_3 + \cdots .$$

The following are some examples.

$$\begin{aligned}
\sum_{n=0}^{\infty} r^n &= 1 + r + r^2 + r^3 + \cdots + r^n + \cdots , \\
\sum_{n=1}^{\infty} n &= 1 + 2 + 3 + \cdots + n + \cdots , \\
\sum_{n=1}^{\infty} \frac{1}{n(n+1)} &= \frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \frac{1}{3 \cdot 4} + \cdots + \frac{1}{n(n+1)} + \cdots , \\
\sum_{n=1}^{\infty} \frac{1}{n^p} &= 1 + \frac{1}{2^p} + \frac{1}{3^p} + \cdots + \frac{1}{n^p} + \cdots , \\
\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} &= 1 - \frac{1}{2} + \frac{1}{3} - \cdots + \frac{(-1)^{n+1}}{n} + \cdots , \\
\sum_{n=0}^{\infty} \frac{1}{n!} &= 1 + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} + \cdots .
\end{aligned}$$

Like sequences, series do not have to start at $n = 1$. For example, it is more convenient for the *geometric series* $\sum_{n=0}^{\infty} r^n$ to start at $n = 0$.

4.1.1 Sum of Series

Definition 4.1.1. The *partial sum* of a series $\sum_{n=1}^{\infty} a_n$ is

$$s_n = \sum_{i=1}^{n} a_i = a_1 + a_2 + \cdots + a_n.$$

If the partial sum converges, then the series *converges* and has *sum* (or *value*)

$$\sum_{n=1}^{\infty} a_n = \lim_{n\to\infty} s_n.$$

If the partial sum diverges, then the series *diverges*.

If finitely many terms in a series are modified, added or dropped, then the new partial sum s_n' and the original partial sum s_n are related by $s_n' = s_{n+n_0} + C$ for some constants n_0 and C. This implies that the convergence of series is not affected, although the sum may be affected.

The arithmetic properties of the sequence limit implies

$$\sum (a_n + b_n) = \sum a_n + \sum b_n, \quad \sum c a_n = c \sum a_n.$$

However, there is no formula for $\sum a_n b_n$ or $\sum \dfrac{a_n}{b_n}$.

Example 4.1.1. Let $s_n = 1 + r + r^2 + \cdots + r^n$ be the partial sum of geometric series $\sum_{n=0}^{\infty} r^n$. Then

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

Therefore $s_n = \dfrac{1-r^{n+1}}{1-r}$ and

$$\sum_{n=0}^{\infty} r^n = \begin{cases} \dfrac{1}{1-r}, & \text{if } |r| < 1, \\ \text{diverges}, & \text{if } |r| \geq 1. \end{cases}$$

Example 4.1.2. The computation in Example 1.3.1 gives the partial sum of $\sum \dfrac{1}{n(n+1)}$.

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

Therefore

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

Example 4.1.3. Example 2.7.5 shows that the partial sum of $\sum \frac{1}{n!}$ satisfies $|s_n - e| = |R_n(1)| \le \frac{e}{(n+1)!}$, which implies $\sum_{n=0}^{\infty} \frac{1}{n!} = e$. Exercise 1.3.18 gives an alternative argument. Of course the argument, which uses the Lagrange form of the remainder (Theorem 2.7.1), can be extended to the series $\sum \frac{x^n}{n!}$. The partial sum satisfies

$$|s_n - e^x| = |R_n(x)| = \frac{e^c}{(n+1)!}|x|^{n+1} \le \frac{e^{|x|}}{(n+1)!}|x|^{n+1}, \quad |c| < |x|.$$

Since for fixed x, the right side converges to 0 as $n \to \infty$, we conclude that $\sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x$.

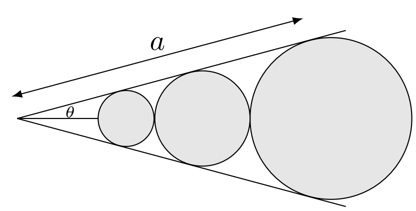
Exercise 4.1.1. Suppose the partial sum $s_n = \frac{n}{2n+1}$. Find the series $\sum a_n$ and its sum.

Exercise 4.1.2. Decimal expressions for rational numbers have repeating patterns. For example, we have

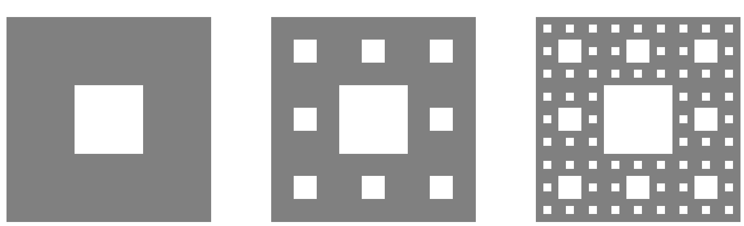
$$\begin{aligned} 1.2\overline{34} = 1.2343434\cdots &= 1.2 + \frac{34}{1000} + \frac{34}{100000} + \frac{34}{10000000} + \cdots \\ &= 1.2 + \frac{34}{1000}\sum_{n=0}^{\infty}\frac{1}{100^n} = 1.2 + \frac{34}{1000}\frac{1}{1-\frac{1}{100}} = \frac{611}{495}. \end{aligned}$$

1. Find rational expressions for $1.\overline{23}$, $1.\overline{230}$, $1.\overline{023}$.
2. Final the decimal based series representing the rational numbers $\frac{5}{12}$, $\frac{43}{35}$.

Exercise 4.1.3. What is the total area of infinitely many disks?



Exercise 4.1.4. The *Sierpinski carpet* is obtained from the unit square by successively deleting "one third squares". Find the area of the carpet.



Exercise 4.1.5. Two lines L and L' form an angle θ at P. A boy starts on L at distance a from P and walk to L' along shortest path. After reaching L', he walks back to L along shortest path. Then he walks to L' again along shortest path, and keeps walking back and forth. What is the total length of his trip?

Exercise 4.1.6. Find the area between curves $y = x^n$ and $y = x^{n+1}$ and use this to conclude that $\sum_{n=1}^{\infty} \dfrac{1}{n(n+1)} = 1$.

Exercise 4.1.7. Compute the partial sum and the sum of series.

1. $\sum_{n=1}^{\infty} nr^n$.
2. $\sum \dfrac{2^n + (-1)^n 3^{n-1}}{5^{n+1}}$.
3. $\sum_{n=0}^{\infty} \dfrac{1}{(a+nd)(a+(n+1)d)}$.
4. $\sum_{n=2}^{\infty} \dfrac{1}{n(n+1)(n+2)}$.
5. $\sum_{n=2}^{\infty} \log\left(1 - \dfrac{1}{n^2}\right)$.
6. $\sum_{n=1}^{\infty} \dfrac{n}{(n+1)!}$.

Exercise 4.1.8. Suppose $x_n > 0$. Compute $\sum_{n=1}^{\infty} \dfrac{x_n}{(1+x_1)(1+x_2)\cdots(1+x_n)}$.

Exercise 4.1.9. The *Fibonacci sequence* $1, 1, 2, 3, 5, \dots$ is defined recursively by $a_0 = a_1 = 1$, $a_n = a_{n-1} + a_{n-2}$. Prove the following

$$\frac{1}{a_{n-1}a_{n+1}} = \frac{1}{a_{n-1}a_n} - \frac{1}{a_n a_{n+1}}, \quad \sum_{n=2}^{\infty} \frac{1}{a_{n-1}a_{n+1}} = 1, \quad \sum_{n=2}^{\infty} \frac{a_n}{a_{n-1}a_{n+1}} = 2.$$

Exercise 4.1.10. Use the Lagrange form of the remainder to show that the Taylor series for $\dfrac{1}{1-x}$ converges for $|x| < 1$.

Exercise 4.1.11. Use the Lagrange form of the remainder to show that the Taylor series for $\cos x$ and $\sin x$ converge for any x

$$\sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!} = \cos x, \quad \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!} = \sin x.$$

4.1.2 Convergence of Series

Theorem 4.1.2. *If $\sum a_n$ converges, then $\lim_{n\to\infty} a_n = 0$.*

This is a consequence of

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} (s_n - s_{n-1}) = \lim_{n\to\infty} s_n - \lim_{n\to\infty} s_{n-1} = s - s = 0.$$

By the theorem, the series $\sum 1$, $\sum n$, $\sum(-1)^n$, $\sum \frac{n}{n+1}$ diverge. By Example 1.1.20, the series $\sum \sin na$ converges if and only if a is an integer multiple of π.

If $a_n \geq 0$ for sufficiently large n, then the partial sum sequence is increasing for large n, and Theorem 1.3.2 becomes the following.

Theorem 4.1.3. *If $a_n \geq 0$, then $\sum a_n$ converges if and only if the partial sums are bounded.*

Example 4.1.4. The terms in the series $\sum = \frac{1}{n^p}$ are positive. Therefore the convergence is equivalent to the boundedness of the partial sum. For $p \geq 2$, we have $\frac{1}{n^p} < \frac{1}{(n-1)n}$ and the following bound from Examples 1.3.1 and 4.1.2

$$\frac{1}{1^p} + \frac{1}{2^p} + \cdots + \frac{1}{n^p} \leq 1 + \frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \cdots + \frac{1}{(n-1)n} = 2 - \frac{1}{n} < 2.$$

By Theorem 4.1.3, therefore, the series converges for $p \geq 2$.

For $p = 1$, we used Cauchy criterion in Example 1.3.8 to show that the *harmonic series* $\sum \frac{1}{n}$ diverges.

Example 4.1.5. The even partial sum of the series $\sum \frac{(-1)^{n+1}}{n}$ is the partial sum of the series

$$\sum \left(\frac{1}{2n-1} - \frac{1}{2n} \right) = \left(1 - \frac{1}{2}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \cdots .$$

The terms of the series above are positive, and the partial sum has upper bound

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

By Theorem 4.1.3, $\sum \left(\frac{1}{2n-1} - \frac{1}{2n} \right)$ converges. This means that the even partial sum s_{2n} of $\sum \frac{(-1)^{n+1}}{n}$ converges. By $s_{2n+1} = s_{2n} + \frac{1}{2n+1}$, the odd partial sum s_{2n+1} converges to the same limit. Therefore $\sum \frac{(-1)^{n+1}}{n}$ converges.

Exercise 4.1.12. Show the divergence of $\sum \frac{1}{\sqrt[n]{a}}$ and $\sum \frac{n}{2n-1}$.

Exercise 4.1.13. Determine convergence.

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

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

3. $1+\frac{1}{2^2}+\frac{1}{3^3}+\frac{1}{4^4}+\cdots$.

4. $1+\frac{1}{\sqrt{1\cdot 2}}+\frac{1}{\sqrt{3\cdot 4}}+\frac{1}{\sqrt{5\cdot 6}}+\cdots$.

Exercise 4.1.14. Use Theorem 4.1.3 to argue about the convergence of $\sum r^n$ for $0\le r<1$ and $\sum \frac{1}{n!}$.

Exercise 4.1.15. What is wrong with the following calculation?

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

We used the Cauchy criterion (Theorem 1.3.3) for the divergence of harmonic series $\sum \frac{1}{n}$. In general, applying the Cauchy criterion to the partial sum shows that $\sum a_n$ converges if and only if for any $\epsilon \ge 0$, there is N, such that (since $|s_m - s_n|$ is symmetric in m and n, we may always assume $n>m$)

$$n>m>N \implies |s_m - s_n| = |a_{m+1}+a_{m+2}+\cdots+a_n| < \epsilon.$$

We may further modify the criterion by taking $m+1$ to be m.

Theorem 4.1.4 (Cauchy Criterion). *A series $\sum a_n$ converges if and only if for any $\epsilon>0$, there is N, such that*

$$n\ge m>N \implies |a_m + a_{m+1}+\cdots+a_n| < \epsilon.$$

Theorem 4.1.2 is a special case of the Cauchy criterion by taking $m=n$.

4.2 Comparison Test

Series $\sum a_n$ are very much like improper integrals $\int_a^{+\infty} f(x)dx$. The two can be compared in two aspects. First the convergence of the two can be compared, through the integral test. Second all the convergence theorems for $\int_a^{+\infty} f(x)dx$, such as the comparison test, Dirichlet test and Abel test, have parallels for the convergence of series.

4.2.1 Integral Test

Theorem 4.2.1 (Integral Test). *Suppose $f(x)$ is a decreasing function on $[1,+\infty)$ satisfying $\lim_{x\to+\infty} f(x)=0$. Then*

$$f(1)+f(2)+\cdots+f(n)=\int_1^n f(x)dx+\gamma+\epsilon_n,$$

for a constant $0\le\gamma\le f(1)$ and a decreasing sequence ϵ_n converging to 0. In particular, the series $\sum f(n)$ converges if and only if the improper integral $\displaystyle\int_a^{+\infty} f(x)dx$ converges.

Let

$$x_n=f(1)+f(2)+\cdots+f(n)-\int_1^n f(x)dx.$$

By f decreasing, we get

$$\begin{aligned}x_n-x_{n-1}&=f(n)-\int_{n-1}^n f(x)dx\le 0,\\ x_n-f(n)&=f(1)+\cdots+f(n-1)-\int_1^n f(x)dx\ge 0\\ &=\sum_{k=1}^{n-1}\left(f(k)-\int_k^{k+1} f(x)dx\right)\ge 0.\end{aligned}$$

The first inequality implies x_n is decreasing, and the second inequality implies $x_n\ge f(n)\ge 0$. Therefore $\lim x_n=\gamma$ converges, and the theorem follows. We have $0\le\gamma\le x_1=f(1)$.

Example 4.2.1. For $p>0$, the function $\dfrac{1}{x^p}$ is decreasing and converges to 0 as $x\to+\infty$. By Theorem 4.2.1, therefore, the series $\sum\dfrac{1}{n^p}$ converges if and only if the improper integral $\displaystyle\int_1^{+\infty}\frac{dx}{x^p}$ converges. By Example 3.7.3, this happens if and only if $p>1$.

Although the harmonic series $\sum\dfrac{1}{n}$ diverges, Theorem 4.2.1 estimates the partial sum

$$1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}=\log n+\gamma+\epsilon_n,$$

where ϵ_n decreases and converges to 0, and

$$\gamma=0.577215664901532860606512090082\cdots$$

is the *Euler-Mascheroni constant.*

Example 4.2.2. For $p > 0$ and $x > e$, the integral test can be applied to the function $\dfrac{1}{x(\log x)^p}$. We conclude that $\sum \dfrac{1}{n(\log n)^p}$ converges if and only if the improper integral $\displaystyle\int_a^{+\infty} \frac{dx}{x(\log x)^p}$ converges. By Example 3.7.9, this means $p > 1$.

Example 4.2.3. We will show that $\sum_{n=1}^{\infty} \dfrac{1}{n^2} = \dfrac{\pi^2}{6}$ in Example 4.5.13. In fact, for even k, $\sum_{n=1}^{\infty} \dfrac{1}{n^k}$ can be calculated as a rational multiple of π^k. However, very little is known about the sum for odd k. Still, we may use the idea of Theorem 4.2.1 to estimate the remainder

$$\int_{n+1}^{+\infty} f(x)dx = \sum_{k=n+1}^{\infty} \int_k^{k+1} f(x)dx \le \sum_{k=n+1}^{\infty} f(k) \le \sum_{k=n}^{\infty} \int_k^{k+1} f(x)dx = \int_n^{+\infty} f(x)dx.$$

For example, the 10-th partial sum of $\sum_{n=1}^{\infty} \dfrac{1}{n^3}$ is

$$s_{10} = \frac{1}{1^3} + \frac{1}{2^3} + \cdots + \frac{1}{10^3} = 1.197532\cdots.$$

By

$$\int_{10}^{\infty} \frac{dx}{x^3} = \frac{1}{2(10)^2} = 0.005, \quad \int_{11}^{\infty} \frac{dx}{x^3} = \frac{1}{2(11)^2} = 0.004132\cdots,$$

we get

$$\begin{aligned} 1.201664\cdots &= 1.197532\cdots + 0.004132\cdots \\ &\le \sum_{n=1}^{\infty} \frac{1}{n^3} \le 1.197532\cdots + 0.005 = 1.202532\cdots. \end{aligned}$$

If we want to get the approximate value of $\sum_{n=1}^{\infty} \dfrac{1}{n^3}$ up to the 6-th digit, then we my try to find n satisfying

$$\int_n^{n+1} \frac{dx}{x^3} = \frac{2n+1}{2n^2(n+1)^2} < \frac{1}{n^3} < 0.000001.$$

So we may take $n = 100$ and get

$$\sum_{n=1}^{100} \frac{1}{n^3} + \int_{101}^{\infty} \frac{dx}{x^3} \le \sum_{n=1}^{\infty} \frac{1}{n^3} \le \sum_{n=1}^{100} \frac{1}{n^3} + \int_{100}^{\infty} \frac{dx}{x^3}.$$

Exercise 4.2.1. Determine the convergence of $\sum \dfrac{1}{n(\log n)(\log(\log n))^p}$.

Exercise 4.2.2. Find suitable function $f(n)$, such that the sequence $1+\frac{1}{\sqrt{2}}+\cdots+\frac{1}{\sqrt{n}}-f(n)$ converges to a limit γ. Then express the sum of the series $\sum_{n=0}^{\infty}(-1)^n\frac{1}{\sqrt{n}}$ in terms of γ.

Exercise 4.2.3. Estimate $\sum_{n=1}^{\infty}\frac{1}{n^{\frac{3}{2}}}$ to within 0.01.

4.2.2 Comparison Test

Theorem 4.2.2 (Comparison Test). *Suppose $|a_n| \le b_n$ for sufficiently large n. If $\sum b_n$ converges, then $\sum a_n$ also converges.*

The test is completely parallel to the similar test (Theorem 3.7.1) for the convergence of improper integrals, and can be proved similarly by using the Cauchy criterion (Theorem 4.1.4).

For the special case $b_n = |a_n|$, the test says that if $\sum |a_n|$ converges, then $\sum a_n$ converges. In other words, absolute convergence implies convergence. We note that the conclusion of the comparison test is always absolute convergence.

Example 4.2.4. Consider the series $\sum \frac{\log n}{n^p}$. If $p \le 1$, then $\frac{\log n}{n^p} \ge \frac{1}{n}$. By the comparison test, the divergence of $\sum \frac{1}{n}$ implies the divergence of $\sum \frac{\log n}{n^p}$.

If $p > 1$, then choose q satisfying $p > q > 1$. We have

$$\frac{\log n}{n^p} = \frac{\log n}{n^{p-q}}\frac{1}{n^q} < \frac{1}{n^q} \text{ for large } n.$$

Here the inequality is due to the fact that $p - q > 0$ implies $\lim \frac{\log n}{n^{p-q}} = 0$. By Example 4.2.1, $\sum \frac{1}{n^q}$ converges. Then by the comparison test, we conclude that $\sum \frac{\log n}{n^p}$ converges.

The key idea of the example above is to compare $a_n = \frac{\log n}{n^p}$ with $b_n = \frac{1}{n^q}$ by using the limit of their quotient. By

$$\lim \frac{a_n}{b_n} = \lim \frac{\log n}{n^{p-q}} = 0,$$

we get $\frac{a_n}{b_n} < 1$ for sufficiently large n. Since both a_n and b_n are positive, we may apply the comparison test to conclude that the convergence of $\sum b_n$ implies the convergence of $\sum a_n$.

In general, if $a_n, b_n > 0$ and $\lim \frac{a_n}{b_n} = l$ converges, then by the comparison test, the convergence of $\sum b_n$ implies the convergence of $\sum a_n$. Moreover, if $l \neq 0$, then we also have $\lim \frac{b_n}{a_n} = \frac{1}{l}$, and we conclude that $\sum a_n$ converges if and only if $\sum b_n$ converges.

Example 4.2.5. For $\sum \frac{n + \sin n}{n^3 + n + 2}$, we make the following comparison

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

By the convergence of $\sum \frac{1}{n^2}$, we get the convergence of $\sum \frac{n + \sin n}{n^3 + n + 2}$.

Similarly, by the comparison

$$\lim_{n\to\infty} \frac{\dfrac{2^n + n^2}{\sqrt{5^{n-1} - n^4 3^n}}}{\left(\dfrac{2}{\sqrt{5}}\right)^n} = \sqrt{5},$$

and the convergence of $\sum \left(\frac{2}{\sqrt{5}}\right)^n$, the series $\sum \frac{2^n + n^2}{\sqrt{5^{n-1} - n^4 3^n}}$ converges.

Example 4.2.6. By Example 2.5.14, we know $\left(1 + \frac{1}{x}\right)^x - e = -\frac{e}{2x} + o\left(\frac{1}{x}\right)$. This implies that for sufficiently large n, $\left(1 + \frac{1}{n}\right)^n - e$ is negative and comparable to $\frac{1}{n}$. Since the harmonic series $\sum \frac{1}{n}$ diverges, we conclude that $\sum \left[\left(1 + \frac{1}{n}\right)^n - e\right]$ diverges.

Example 4.2.7. By Example 3.7.14, we know that $\int_1^{+\infty} \frac{|\sin x|}{x^p} dx$ converges if and only if $p > 1$. By a change of variable, we also know that, for $a \neq 0$, $\int_1^{+\infty} \frac{|\sin ax|}{x^p} dx$ converges if and only if $p > 1$. However, we cannot use the integral test (Theorem 4.2.1) to get the similar conclusion for $\sum \frac{|\sin na|}{n^p}$. The problem is that $\frac{|\sin ax|}{x^p}$ is not a decreasing function.

By $\frac{|\sin na|}{n^p} \leq \frac{1}{n^p}$ and the comparison test, we know $\sum \frac{|\sin na|}{n^p}$ converges for $p > 1$. The series also converges if a is a multiple of π, because all the terms are 0. It remains to consider the case $p \leq 1$ and a is not a multiple of π.

First assume $0 < a \le \frac{\pi}{2}$. For any natural number k, the interval $\left[k\pi + \frac{\pi}{4}, k\pi + \frac{3\pi}{4}\right]$ has length $\frac{\pi}{2}$ and therefore must contain $n_k a$ for some natural number n_k. Then $|\sin n_k a| \ge \frac{1}{\sqrt{2}}$, and for $p \le 1$,

$$\sum_{n=1}^{\infty} \frac{|\sin na|}{n^p} \ge \sum_{n=1}^{\infty} \frac{|\sin na|}{n} \ge \sum_{k=1}^{\infty} \frac{|\sin n_k a|}{n_k} \ge \frac{1}{\sqrt{2}} \sum_{k=1}^{\infty} \frac{1}{n_k}.$$

By $n_k \le k\pi + \frac{3\pi}{4}$, we get $\frac{1}{n_k} \ge \frac{4}{4k+3}\frac{a}{\pi} > \frac{a}{4k}$. Then by $\sum \frac{1}{k} = +\infty$, we get $\sum \frac{1}{n_k} = +\infty$ and $\sum \frac{|\sin na|}{n^p} = +\infty$.

In general, if a is not an integer multiple of π, then there is b, such that $0 < b \le \frac{\pi}{2}$ 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 $\sum \frac{|\sin na|}{n^p}$ diverges for $p \le 1$.

Exercise 4.2.4. Show that if $a_n > 0$ and $\sum a_n$ converges, then $\sum a_n^2$ converges. Moreover, show that the converse is not true.

Exercise 4.2.5. Show that if $\sum a_n^2$, then $\sum \frac{a_n}{n}$ converges.

Exercise 4.2.6. Show that if $\sum a_n^2$ and $\sum b_n^2$ converge, then $\sum a_n b_n$ and $\sum (a_n + b_n)^2$ converge.

Exercise 4.2.7. Determine the convergence.

1. $\sum \frac{\sqrt{4n^5 + 5n^4}}{3n^2 - 2n^3}$.
2. $\sum \frac{3n^2 - 2n^3}{\sqrt{4n^5 + 5n^4}}$.
3. $\sum \frac{3n^2 + (-1)^n 2n^3}{4n^5 + 5n^4}$.

Exercise 4.2.8. Determine the convergence, $p, q, r, s > 0$.

1. $\sum \frac{1}{n^p + (\log n)^q}$.
2. $\sum \frac{1}{n^p (\log n)^q}$.
3. $\sum \frac{n^r + (\log n)^s}{n^p + (\log n)^q}$.
4. $\sum \frac{n^r (\log n)^s}{n^p + (\log n)^q}$.
5. $\sum \frac{n^r + (\log n)^s}{n^p (\log n)^q}$.
6. $\sum \frac{1}{n^p (\log n)^q (\log(\log n))^r}$.

Exercise 4.2.9. Determine the convergence, $b, d, p, q > 0$.

1. $\sum \frac{1}{(a + nb)^p}$.
2. $\sum \frac{(c + nd)^q}{(a + nb)^p}$.
3. $\sum \frac{1}{(a + nb)^p (c + nd)^q}$.

4. $\sum \dfrac{(\log(c+nd))^q}{(a+nb)^p}$.
5. $\sum \dfrac{1}{(a+nb)^p(\log(c+nd))^q}$.
6. $\sum \dfrac{(\log(c+nd))^q}{(a+nb)^p}$.

Exercise 4.2.10. Determine the convergence, $p, q > 0$.

1. $\sum((n^p+a)^r - (n^p+b)^r)$.
2. $\sum \left[\left(\dfrac{n^p+a}{n^p+b}\right)^q - 1\right]$.

Exercise 4.2.11. Determine the convergence.

1. $\sum \dfrac{1}{n\sqrt{n}}$.
2. $\sum \dfrac{1}{n^{1+\frac{1}{n}}}$.
3. $\sum \dfrac{1}{n^{1+\frac{1}{\log n}}}$.
4. $\sum \dfrac{1}{(\log n)^n}$.
5. $\sum \dfrac{n^2}{(\log n)^n}$.
6. $\sum \dfrac{1}{\sqrt[n]{\log n}}$.
7. $\sum \dfrac{(\log n)^n}{n^n}$.
8. $\sum \dfrac{n^{\log n}}{(\log n)^n}$.

Exercise 4.2.12. Determine the convergence, $p, q > 0$.

1. $\sum \sin\dfrac{1}{n}$.
2. $\sum \dfrac{1}{n^p}\sin\dfrac{1}{n^q}$.
3. $\sum \dfrac{n^2 - n\sin n}{n^3+\cos n}$.
4. $\sum \dfrac{n^2 - n\sin n}{n^3+\cos n}\sin\dfrac{1}{n}$.
5. $\sum \left(\cos\dfrac{1}{n^p} - 1\right)$.
6. $\sum \cos\dfrac{1}{n^p}\sin\dfrac{1}{n^q}$.

Exercise 4.2.13. Determine the convergence.

1. $\sum \dfrac{1}{5^n-1}$.
2. $\sum \dfrac{3^{n+1}}{5^{n-1}-n^2 2^n}$.
3. $\sum \dfrac{5^{n-1}-n^2 2^n}{3^{n+1}}$.

Exercise 4.2.14. Determine the convergence, $a, b > 0$.

1. $\sum \sqrt{a^n+b^n}$.
2. $\sum \dfrac{1}{\sqrt{a^n+b^n}}$.
3. $\sum \dfrac{1}{a^n+b^n}$.
4. $\sum (a^n+b^n)^p$.
5. $\sum \dfrac{n^2}{na^n+b^n}$.
6. $\sum \dfrac{1}{\sqrt[n]{a^n+b^n}}$.

Exercise 4.2.15. Determine the convergence.

1. $\sum x^{n^2}$.
2. $\sum nx^{n^2}$.
3. $\sum x^{\sqrt{n}}$.
4. $\sum nx^{\sqrt{n}}$.
5. $\sum n^2x^{n^2}$.
6. $\sum n^px^{n^q}$.

Exercise 4.2.16. Determine the convergence.

1. $\sum a^{n^p}$.
2. $\sum (a^{\frac{1}{n}} - 1)$.
3. $\sum \left(e^{\frac{1}{n}} - 1 - \dfrac{1}{n}\right)$.
4. $\sum \left(n^{\frac{1}{n^p}} - 1\right)$.
5. $\sum \left(1 + \dfrac{a\log n}{n}\right)^n$.
6. $\sum \left(\dfrac{an+b}{cn+d}\right)^n$.
7. $\sum n^3\left(\dfrac{a+(-1)^n}{b+(-1)^n}\right)^n$.

8. $\sum\left(\frac{1}{\sqrt{n}}-\sqrt{\log\frac{n+1}{n}}\right)$.

9. $\sum\left(1-\frac{1}{n}\right)^{n^2}$.

10. $\sum\left(1+\frac{1}{n}\right)^{2n-n^2}$.

11. $\sum\frac{n^2}{\left(a+\frac{1}{n}\right)^n}$.

12. $\sum\frac{n^{2n}}{(n+a)^{n+b}(n+b)^{n+a}}$.

13. $\sum\left(2\sqrt[n]{a}-\sqrt[n]{b}-\sqrt[n]{c}\right)$.

14. $\sum\left(\cos\frac{a}{n}\right)^{n^2}$.

15. $\sum\left(-\log\cos\frac{1}{n}\right)^p$.

16. $\sum\log\left(n^p\sin\frac{a}{n^q}\right)$.

17. $\sum\frac{|\cos na|}{n^p}$.

Exercise 4.2.17. Determine the convergence.

1. $\sum\int_n^{n+1}e^{-\sqrt{x}}\sin x dx$.

2. $\sum\int_n^{n+1}\frac{\sin x}{x^p}dx$.

3. $\sum\int_{\frac{1}{n+1}}^{\frac{1}{n}}\frac{\log x}{x^p}dx$.

4. $\sum\int_0^{\frac{1}{n}}\frac{x^p}{1+x^2}dx$.

5. $\sum\int_0^{\frac{1}{n}}|\sin x|^p dx$.

6. $\sum\int_0^1\sin x^n dx$.

Exercise 4.2.18. Suppose a_n is a bounded sequence. Show that $\sum\frac{1}{n}(a_n-a_{n+1})$ converges.

Exercise 4.2.19. The decimal representations of positive real numbers are actually the sum of series. For example,

$$\pi=3.1415926\cdots=3+0.1+0.04+0.001+0.0005+0.00009+0.000002+0.0000006+\cdots.$$

Explain why the expression always converges.

4.2.3 Special Comparison Test

We compare a series $\sum a_n$ with the geometric series $\sum r^n$, which we know converges if and only if $|r|<1$. If $|a_n|\le r^n$ for some $r<1$, then the comparison test implies that $\sum a_n$ converges. We note that the condition $|a_n|\le r^n$ for some $r<1$ is the same as $\sqrt[n]{|a_n|}\le r<1$.

Theorem 4.2.3 (Root Test). *Suppose $|a_n|\le r^n$ for some $r<1$ and sufficiently large n. Then $\sum a_n$ converges.*

Example 4.2.8. To determine the convergence of $\sum(n^5+2n+3)x^n$, we note that $\lim_{n\to\infty}\sqrt[n]{|(n^5+2n+3)x^n|}=|x|$. If $|x|<1$, then we can pick r satisfying $|x|<r<1$. By $\lim_{n\to\infty}\sqrt[n]{|(n^5+2n+3)x^n|}<r$ and the order rule, we get $\sqrt[n]{|(n^5+2n+3)x^n|}<r$ for sufficiently large n. Then by the root test, we conclude that $\sum(n^5+2n+3)x^n$ converges for $|x|<1$.

If $|x|\ge 1$, then the term $(n^5+2n+3)x^n$ of the series does not converge to 0. By Theorem 4.1.2, the series diverges for $|x|\ge 1$.

The example suggests that, in practice, it is often more convenient to use the limit version of the root test. Suppose $\lim_{n\to\infty} \sqrt[n]{|a_n|} < 1$. Then fix r satisfying $\lim_{n\to\infty} \sqrt[n]{|a_n|} < r < 1$. By the order rule, we have $\sqrt[n]{|a_n|} < r$ for sufficiently large n. Then the root test shows that $\sum a_n$ converges. On the other hand, if $\lim_{n\to\infty} \sqrt[n]{|a_n|} > 1$, then we have $\sqrt[n]{|a_n|} > 1$ for sufficiently large n. This implies $|a_n| > 1$, and $\sum a_n$ diverges by Theorem 4.1.2.

Exercise 4.2.20. Determine the convergence, $a, b > 0$.

1. $\sum \dfrac{(\log n)^n}{n^2}$.
2. $\sum \dfrac{n^p}{(\log n)^n}$.
3. $\sum \dfrac{1}{a^n + b^n}$.
4. $\sum (a^n + b^n)^p$.
5. $\sum n^p x^n$.
6. $\sum n^p x^{n^q}$.
7. $\sum \left(\dfrac{an+b}{cn+d}\right)^n$.
8. $\sum \left(1 + \dfrac{a}{n}\right)^{n^2}$.
9. $\sum \left(1 + \dfrac{a}{n}\right)^{-n^2}$.
10. $\sum \left(1 + \dfrac{a}{n}\right)^{2n-n^2}$.
11. $\sum \dfrac{n^p}{\left(a + \dfrac{b}{n}\right)^n}$.
12. $\sum n^3 \left(\dfrac{a + (-1)^n}{b + (-1)^n}\right)^n$.

Next we turn to another way of comparing series. Theorem 2.3.3 compares two functions by comparing their derivatives (i.e., the changes of functions). Similarly, we may compare two sequences a_n and b_n by either comparing the differences $a_{n+1} - a_n$ and $b_{n+1} - b_n$, or the ratios $\dfrac{a_{n+1}}{a_n}$ and $\dfrac{b_{n+1}}{b_n}$. The comparison of ratio is especially suitable for the comparison of series.

Suppose $a_n, b_n > 0$, and $\dfrac{a_{n+1}}{a_n} \le \dfrac{b_{n+1}}{b_n}$ for $n \ge N$. Then for $c = \dfrac{a_N}{b_N}$, the two sequences a_n and cb_n are equal at $n = N$, and $\dfrac{a_{n+1}}{a_n} \le \dfrac{cb_{n+1}}{cb_n}$ implies that the second sequence has bigger change than the first one, at least for $n \ge N$. This should imply $a_n \le cb_n$ for $n \ge N$. The following is the rigorous argument

$$a_n = a_N \frac{a_{N+1}}{a_N} \frac{a_{N+2}}{a_{N+1}} \cdots \frac{a_n}{a_{n-1}} \le cb_N \frac{b_{N+1}}{b_N} \frac{b_{N+2}}{b_{N+1}} \cdots \frac{b_n}{b_{n-1}} = cb_n.$$

By the comparison test, if $\sum b_n$ converges, then $\sum a_n$ converges.

Theorem 4.2.4 (Ratio Test). *Suppose* $\left|\dfrac{a_{n+1}}{a_n}\right| \le \dfrac{b_{n+1}}{b_n}$ *for sufficiently large* n. *If* $\sum b_n$ *converges, then* $\sum a_n$ *converges.*

We note that the assumption implies that the terms b_n have the same sign for sufficiently large n. By changing all b_n to $-b_n$ if necessary, we may assume that $b_n > 0$ for sufficiently large n.

Example 4.2.9. The series $\sum \dfrac{(2n)!}{(n!)^2}x^n$ satisfies

$$\lim_{n\to\infty}\left|\frac{a_n}{a_{n-1}}\right| = \lim_{n\to\infty}\left|\frac{\dfrac{(2n)!}{(n!)^2}x^n}{\dfrac{(2n-2)!}{((n-1)!)^2}x^{n-1}}\right| = \lim_{n\to\infty}\frac{2n(2n-1)}{n^2}|x| = 4|x|.$$

If $4|x| < 1$, then we fix r satisfying $4|x| < r < 1$. By the order rule, we have

$$\left|\frac{a_n}{a_{n-1}}\right| < r = \frac{r^n}{r^{n-1}} \text{ for large } n.$$

By comparing with the power series $\sum b_n = \sum r^n$, Theorem 4.2.4 implies that $\sum a_n$ converges. If $4|x| > 1$, then we get

$$\left|\frac{a_n}{a_{n-1}}\right| > 1 \text{ for large } n.$$

Therefore $|a_n|$ is increasing and does not converge to 0. By Theorem 4.1.2, $\sum a_n$ diverges.

We conclude that $\sum \dfrac{(2n)!}{(n!)^2}x^n$ converges for $|x| < \dfrac{1}{4}$ and diverges for $|x| > \dfrac{1}{4}$. For $x = \dfrac{1}{4}$, we cannot compare with the geometric series $\sum r^n$. Instead, we may try to compare with $\sum \dfrac{1}{n^p}$. The terms $a_n = \dfrac{(2n)!}{(n!)^2 4^n} > 0$, and

$$\frac{a_n}{a_{n-1}} = \frac{2n(2n-1)}{4n^2} = 1 - \frac{1}{2n}, \quad \frac{\dfrac{1}{n^p}}{\dfrac{1}{(n-1)^p}} = 1 - \frac{p}{n} + o\left(\frac{1}{n}\right).$$

So we expect a_n to be comparable to $\dfrac{1}{n^{\frac{1}{2}}}$. Since $\sum \dfrac{1}{n^{\frac{1}{2}}}$ diverges, we expect $\sum a_n$ diverges. For a rigorous argument, we wish to show that

$$\frac{a_n}{a_{n-1}} \ge \frac{\dfrac{1}{n^p}}{\dfrac{1}{(n-1)^p}} \text{ for some } p \le 1 \text{ and large } n.$$

Of course this holds for $p = 1 > \dfrac{1}{2}$. Therefore by the ratio test, we conclude that $\sum \dfrac{(2n)!}{(n!)^2 4^n}$ diverges.

We will show in Example 4.3.4 that, for $r = -\dfrac{1}{4}$, the series $\sum (-1)^n \dfrac{(2n)!}{(n!)^2 4^n}$ converges.

There are several generalisations we can make from the example. First, if we apply the ratio test to the case $\sum b_n = \sum r^n$ is the geometric series, we find that

$$\left|\frac{a_{n+1}}{a_n}\right| \le r < 1 \text{ for large } n \implies \sum a_n \text{ converges.}$$

The limit version of this specialised ratio test is

$$\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| < 1 \implies \sum a_n \text{ converges.}$$

On the other hand, the example also shows that

$$\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| > 1 \implies \sum a_n \text{ diverges.}$$

Second, when the comparison with the geometric series does not work, we may compare with $\sum \frac{1}{n^p}$. Suppose

$$\left|\frac{a_{n+1}}{a_n}\right| \le 1 - \frac{p}{n} \text{ for some } p > 1 \text{ and large } n. \tag{4.2.1}$$

We find q satisfying $p > q > 1$. Then the property above implies

$$\left|\frac{a_{n+1}}{a_n}\right| \le 1 - \frac{q}{n} + o\left(\frac{1}{n}\right) = \frac{\dfrac{1}{n^q}}{\dfrac{1}{(n-1)^q}} \text{ for large } n.$$

By applying the ratio test to $b_n = \frac{1}{n^p}$, we conclude that $\sum a_n$ converges. The use of criterion (4.2.1) for the convergence of series is the *Raabe test*.

The Raabe test also has the limit version. We note that (4.2.1) is equivalent to

$$n\left(1 - \left|\frac{a_{n+1}}{a_n}\right|\right) \ge p > 1 \text{ for large } n.$$

This will be satisfied if we can verify

$$\lim_{n\to\infty} n\left(1 - \left|\frac{a_{n+1}}{a_n}\right|\right) > 1.$$

Exercise 4.2.21. Determine convergence.

1. $\dfrac{4}{2}+\dfrac{4\cdot 7}{2\cdot 6}+\dfrac{4\cdot 7\cdot 10}{2\cdot 6\cdot 10}+\cdots$.
2. $\dfrac{2}{4}+\dfrac{2\cdot 6}{4\cdot 7}+\dfrac{2\cdot 6\cdot 10}{4\cdot 7\cdot 10}+\cdots$.
3. $\dfrac{2}{4}+\dfrac{2\cdot 5}{4\cdot 7}+\dfrac{2\cdot 5\cdot 8}{4\cdot 7\cdot 10}+\cdots$.
4. $\dfrac{2}{4\cdot 7}+\dfrac{2\cdot 5}{4\cdot 7\cdot 10}+\dfrac{2\cdot 5\cdot 8}{4\cdot 7\cdot 10\cdot 13}+\cdots$.

Exercise 4.2.22. Determine convergence.

1. $\sum\dfrac{a(a+1)\cdots(a+n)}{b(b+1)\cdots(b+n)}$.
2. $\sum\dfrac{a(a+1^p)\cdots(a+n^p)}{b(b+1^p)\cdots(b+n^p)}$.
3. $\sum\dfrac{a^p(a+c)^p\cdots(a+nc)^p}{b^q(b+d)^q\cdots(b+nd)^q}\dfrac{1}{n^r}$.
4. $\sum\dfrac{(a+c)(a+2c)^2\cdots(a+nc)^n}{(b+d)(b+2d)^2\cdots(b+nd)^n}$.
5. $\sum\dfrac{(a_1+b_11+c_11^2)\cdots(a_1+b_1n+c_1n^2)}{(a_2+b_21+c_21^2)\cdots(a_2+b_2n+c_2n^2)}$.
6. $\sum\dfrac{a(a+1)\cdots(a+n)}{b(b+1)\cdots(b+n)}\dfrac{c(c+1)\cdots(c+n)}{d(d+1)\cdots(d+n)}$.

Exercise 4.2.23. Determine convergence. There might be come special values of r for which you cannot yet make conclusion.

1. $\sum\dfrac{(n!)^2}{(2n)!}r^n$.
2. $\sum\dfrac{(3n)!}{(n!)^3}r^n$.
3. $\sum\dfrac{n!(2n)!}{(3n)!}r^n$.
4. $\sum\dfrac{n^n}{n!}r^n$.
5. $\sum\dfrac{n!}{n^n}r^n$.
6. $\sum\dfrac{n!}{(n+1)^n}r^n$.
7. $\sum\dfrac{n^{n+1}}{(n+1)!}r^n$.
8. $\sum\dfrac{(2n)!}{n^{2n}}r^n$.

Exercise 4.2.24. Prove the divergent part of the Raabe test.

1. If $\dfrac{a_{n+1}}{a_n}\ge 1-\dfrac{1}{n}$ for sufficiently large n, then $\sum a_n$ diverges.
2. If $a_n>0$ and $\lim_{n\to\infty} n\left(1-\left|\dfrac{a_{n+1}}{a_n}\right|\right)<1$, then $\sum a_n$ diverges.

4.3 Conditional Convergence

Like improper integrals, the comparison test implies that a series can have three mutually exclusive possibilities:

- *Absolute Convergence*: $\sum|a_n|$ converges ($\implies$ $\sum a_n$ converges).
- *Conditional Convergence*: $\sum|a_n|$ diverges and $\sum a_n$ converges.
- *Divergence*: $\sum a_n$ diverges ($\implies$ $\sum|a_n|$ diverges).

4.3.1 Test for Conditional Convergence

The series $\sum\dfrac{(-1)^{n+1}}{n}$ in Example 4.1.5 is a typical conditionally convergent series. Its absolute value series is the harmonic series $\sum\dfrac{1}{n}$, which we know diverges. We cannot apply the comparison test to the whole series because the conclusion of comparison test is always absolute convergence. In fact, we applied Theorem 4.1.3 to the even partial sum of the series in Example 4.1.5. The following is an elaboration of the idea of Example 4.1.5.

Proposition 4.3.1 (Leibniz Test). *If a_n is decreasing for sufficiently large n and $\lim_{n\to\infty} a_n = 0$, then $\sum(-1)^n a_n$ converges.*

The series

$$\sum_{n=0}^{\infty}(-1)^n a_n = a_0 - a_1 + a_2 - a_3 + \cdots$$

is called an *alternating series*. If a_n is decreasing (for all n), then the odd partial sum

$$\begin{aligned} s_{2n+1} &= (a_0 - a_1) + (a_2 - a_3) + (a_4 - a_5) + \cdots + (a_{2n} - a_{2n+1}) \\ &= a_0 - (a_1 - a_2) - (a_3 - a_4) - \cdots - (a_{2n-1} - a_{2n}) - a_{2n+1} \end{aligned}$$

is increasing and has upper bound a_0. Therefore $\lim s_{2n+1}$ converges. By $s_{2n} = s_{2n+1} - a_{2n+1}$ and $\lim a_{2n+1} = 0$, we have $\lim s_{2n} = \lim s_{2n+1}$ and therefore the whole partial sum sequence converges.

Example 4.3.1. By the Leibniz test, the series

$$\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^p} = 1 - \frac{1}{2^p} + \frac{1}{3^p} - \cdots$$

converges for $p > 0$. By Example 4.2.1, the series absolutely converges for $p > 1$, conditionally converges for $0 < p \le 1$, and diverges for $p \le 0$.

Example 4.3.2. Consider the series $\sum n^a b^n$. By $\lim_{n\to\infty}\sqrt[n]{|n^a b^n|} = |b|$ and the root test, $\sum n^a b^n$ absolutely converges for $|b| < 1$ and diverges for $|b| > 1$.

If $b = 1$, then the series is $\sum n^a$, which converges if and only if $a < -1$. If $b = -1$, then the series is $\sum(-1)^n n^a$, which by Example 4.3.1 converges if and only if $a < 0$.

In conclusion, the series $\sum n^a b^n$ absolutely converges for either $|b| < 1$, or $a < -1$ and $|b| = 1$, conditionally converges for $-1 \le a < 0$ and $b = -1$, and diverges otherwise.

Example 4.3.3. Consider the alternating series $\sum(-1)^n\dfrac{n^2+a}{n^3+b}$. The corresponding absolute value series is comparable to the harmonic series and therefore diverges. If we can show that $f(x) = \dfrac{x^2+a}{x^3+b}$ is decreasing, therefore, then the Leibniz test implies the conditional convergence. By

$$f'(x) = \frac{-x^4 - 3ax^2 + 2xb}{(x^3+b)^2},$$

the function indeed decreases for sufficiently large x.

Example 4.3.4. In Example 4.2.9, we determined the convergence of $\sum \frac{(2n)!}{(n!)^2}x^n$ for all x except $x = -\frac{1}{4}$. For $x = -\frac{1}{4}$, the series is alternating, and $|a_n|$ is decreasing by $\left|\frac{a_n}{a_{n-1}}\right| = 1 - \frac{1}{2n} < 1$. If we can show that a_n converges to 0, then we can apply the Leibniz test.

We compare with the ratio of $|a_n|$ with the ratio of $\frac{1}{n^p}$

$$\left|\frac{a_n}{a_{n-1}}\right| = 1 - \frac{1}{2n} \le \frac{\frac{1}{n^p}}{\frac{1}{(n-1)^p}} = 1 - \frac{p}{n} + o\left(\frac{1}{n}\right).$$

This happens if we pick $p = 0.4$ and n is sufficiently large. The comparison of the ratio implies $|a_n| < \frac{c}{n^{0.4}}$ for a constant a and sufficiently large n. This further implies that $\lim |a_n| = 0$. By the Leibniz test, we conclude that $\sum(-1)^n \frac{(2n)!}{(n!)^2 4^n}$ converges.

Combined with Examples 4.2.9, we conclude that $\sum \frac{(2n)!}{(n!)^2}x^n$ absolutely converges for $|x| < \frac{1}{4}$, conditionally converges for $x = -\frac{1}{4}$, and diverges otherwise.

Exercise 4.3.1. Suppose $a_n > 0$ and $\frac{a_n}{a_{n-1}} = 1 - \frac{p}{n} + o\left(\frac{1}{n}\right)$ for some $p > 0$. Prove that $\lim a_n = 0$ and $\sum(-1)^n a_n$ converges.

Exercise 4.3.2. Determine absolute or conditional convergence.

1. $\frac{4}{2} - \frac{4\cdot 7}{2\cdot 6} + \frac{4\cdot 7\cdot 10}{2\cdot 6\cdot 10} - \cdots$.
2. $\frac{2}{4} - \frac{2\cdot 6}{4\cdot 7} + \frac{2\cdot 6\cdot 10}{4\cdot 7\cdot 10} - \cdots$.
3. $\frac{2}{4} - \frac{2\cdot 5}{4\cdot 7} + \frac{2\cdot 5\cdot 8}{4\cdot 7\cdot 10} - \cdots$.
4. $\frac{2}{4\cdot 7} - \frac{2\cdot 5}{4\cdot 7\cdot 10} + \frac{2\cdot 5\cdot 8}{4\cdot 7\cdot 10\cdot 13} - \cdots$.

Exercise 4.3.3. Determine absolute or conditional convergence.

1. $\sum(-1)^n\dfrac{a(a+1^r)\cdots(a+n^r)}{b(b+1^r)\cdots(b+n^r)}$.
2. $\sum(-1)^n\dfrac{(a+1)(a+2)^2\cdots(a+n)^n}{(b+1)(b+2)^2\cdots(b+n)^n}$.

Exercise 4.3.4. Determine absolute or conditional convergence.

1. $\sum\dfrac{(-1)^n n^2}{n^3+n+2}$.
2. $\sum\dfrac{n^2+\sin n}{(-1)^n n^3+n+2}$.
3. $\sum\dfrac{(-1)^n}{(a+nb)^p}$.
4. $\sum\dfrac{r^n}{n^p}$.
5. $\sum\dfrac{r^n}{n^p(\log n)^q}$.
6. $\sum(-1)^{\frac{n(n-1)}{2}}\dfrac{1}{n^p}$.
7. $\sum(-1)^n n^q a^{n^p}$.
8. $\sum\dfrac{(-1)^n}{n^{p+\frac{q}{\log n}}}$.
9. $\sum\dfrac{(-1)^n n^{n+p}}{(an^2+bn+c)^{\frac{n}{2}+q}}$.

Exercise 4.3.5. Determine the absolute or conditional convergence for the undecided cases in Exercise 4.2.23.

Like the convergence of improper integrals, we also have the analogues of the Dirichlet and Abel tests.

Proposition 4.3.2 (Dirichlet Test). *Suppose the partial sum of $\sum a_n$ is bounded. Suppose b_n is monotonic and $\lim_{n\to\infty} b_n = 0$. Then $\sum a_nb_n$ converges.*

Proposition 4.3.3 (Abel Test). *Suppose $\sum a_n$ converges. Suppose b_n is monotonic and bounded. Then $\sum a_nb_n$ converges.*

Example 4.3.5. In Example 4.2.7, we showed that $\sum\dfrac{|\sin na|}{n^p}$ diverges when a is not an integer multiple of π. Then Example 3.7.14 suggests that the series should converge conditionally.

By the Dirichlet test, if we can show that the partial sum

$$s_n = \sin a + \sin 2a + \cdots + \sin na$$

is bounded, then the series converges. By

$$\begin{aligned}
2s_n\sin\frac{a}{2} &= \left(\cos\left(a-\frac{a}{2}\right)-\cos\left(a+\frac{a}{2}\right)\right)+\left(\cos\left(2a-\frac{a}{2}\right)-\cos\left(2a+\frac{a}{2}\right)\right)\\
&\quad+\cdots+\left(\cos\left(na-\frac{a}{2}\right)-\cos\left(na+\frac{a}{2}\right)\right)\\
&=\cos\frac{a}{2}-\cos\left(na+\frac{a}{2}\right),
\end{aligned}$$

we get

$$|s_n|\le\frac{1}{\left|\sin\dfrac{a}{2}\right|}.$$

The right side is a bound for the partial sums in case a is not a multiple of π.

Exercise 4.3.6. Derive the Leibniz test and the Abel test from the Dirichlet test.

Exercise 4.3.7. Prove that if $\sum \frac{a_n}{n^p}$ converges, then $\sum \frac{a_n}{n^q}$ converges for any $q > p$.

Exercise 4.3.8. Determine the absolute and conditional convergence.

1. $\sum \dfrac{\cos na}{(n+b)^p}$.
2. $\sum (-1)^n \dfrac{\cos na}{n+b}$.
3. $\sum \dfrac{\sin na}{n^p(\log n)^q}$.
4. $\sum (-1)^n \dfrac{\sin^2 na}{n^p}$.
5. $\sum \dfrac{\sin^3 na}{n^p(\log n)^q}$.
6. $\sum (-1)^{\frac{n(n-1)}{2}} \dfrac{\sin na}{n^p}$.

Exercise 4.3.9. Determine absolute or conditional convergence.

1. $\sum \dfrac{1}{n+(-1)^n n^2}$.
2. $\sum \dfrac{(-1)^n}{(\sqrt{n}+(-1)^n)^p}$.
3. $\sum \dfrac{(-1)^n}{(n+(-1)^n)^p}$.
4. $\sum \dfrac{(-1)^{\frac{n(n-1)}{2}}}{\sqrt{n}+(-1)^n}$.
5. $\sum \dfrac{(-1)^n}{n^p+(-1)^n}$.

Exercise 4.3.10. Determine the convergence.

1. $\sum \sin \sqrt{n^2+a}\pi$.
2. $\sum (-1)^n \left(1 - \dfrac{a \log n}{n}\right)^n$.
3. $\sum \left(\log \dfrac{an+b}{cn+d}\right)^n$.

Exercise 4.3.11. Let $[x]$ be the biggest integer $\le x$. Determine the convergence of $\sum \dfrac{(-1)^{[\sqrt{n}]}}{n^p}$ and $\sum \dfrac{(-1)^{[\log n]}}{n^p}$.

4.3.2 Absolute v.s. Conditional

The distinction between absolute and conditional convergence has implications on how we can manipulate series. For example, we have $a+b+c+d = c+b+a+d$. However, we need to be more careful in rearranging orders in an infinite sum.

Theorem 4.3.4. *The sum of an absolutely convergent series does not depend on the order. On the other hand, given any conditionally convergent series and any number s, it is possible to rearrange the order so that the sum of the rearranged series is s.*

Example 4.3.6. We know from Example 4.3.1 that the series $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}$ converges conditionally. The partial sum can be estimated from the partial sum of the harmonic series in Example 4.2.1

$$\begin{aligned}
&1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\cdots+\frac{1}{2n-1}-\frac{1}{2n}\\
&=\left(1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\cdots+\frac{1}{2n-1}+\frac{1}{2n}\right)-2\left(\frac{1}{2}+\frac{1}{4}+\frac{1}{6}+\cdots+\frac{1}{2n}\right)\\
&=(\log 2n+\gamma+\epsilon_{2n})-(\log n+\gamma+\epsilon_n)=\log 2+(\epsilon_{2n}-\epsilon_n).
\end{aligned}$$

This implies

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

If the terms are rearranged, so that one positive term is followed by two negative terms, then the partial sum is

$$\begin{aligned}
&1-\frac{1}{2}-\frac{1}{4}+\frac{1}{3}-\frac{1}{6}-\frac{1}{8}+\cdots+\frac{1}{2n-1}-\frac{1}{4n-2}-\frac{1}{4n}\\
&=\left(1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\cdots+\frac{1}{2n-1}+\frac{1}{2n}\right)\\
&\quad-\left(\frac{1}{2}+\frac{1}{4}+\frac{1}{6}+\cdots+\frac{1}{2n}\right)-\left(\frac{1}{2}+\frac{1}{4}+\frac{1}{6}+\cdots+\frac{1}{4n}\right)\\
&=(\log 2n+\gamma+\epsilon_{2n})-\frac{1}{2}(\log n+\gamma+\epsilon_n)-\frac{1}{2}(\log 2n+\gamma+\epsilon_{2n})\\
&=\frac{1}{2}\log 2+\frac{1}{2}(\epsilon_{2n}-\epsilon_n).
\end{aligned}$$

If two positive terms are followed by one negative term, then the partial sum is

$$\begin{aligned}
&1+\frac{1}{3}-\frac{1}{2}+\frac{1}{5}+\frac{1}{7}-\frac{1}{4}+\cdots+\frac{1}{4n-1}+\frac{1}{4n-3}-\frac{1}{2n}\\
&=\left(1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\cdots+\frac{1}{4n}\right)\\
&\quad-\left(\frac{1}{2}+\frac{1}{4}+\frac{1}{6}+\cdots+\frac{1}{4n}\right)-\left(\frac{1}{2}+\frac{1}{4}+\frac{1}{6}+\cdots+\frac{1}{2n}\right)\\
&=(\log 4n+\gamma+\epsilon_{4n})-\frac{1}{2}(\log 2n+\gamma+\epsilon_{2n})-\frac{1}{2}(\log n+\gamma+\epsilon_n)\\
&=\frac{3}{2}\log 2+\frac{1}{2}(\epsilon_{4n}-\epsilon_{2n}-\epsilon_n).
\end{aligned}$$

We get

$$\begin{aligned}
1-\frac{1}{2}-\frac{1}{4}+\frac{1}{3}-\frac{1}{6}-\frac{1}{8}+\cdots&=\frac{1}{2}\log 2,\\
1+\frac{1}{3}-\frac{1}{2}+\frac{1}{5}+\frac{1}{7}-\frac{1}{4}+\cdots&=\frac{3}{2}\log 2.
\end{aligned}$$

Exercise 4.3.12. Rearrange the series $1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\cdots$ so that p positive terms are followed by q negative terms and the pattern repeated. Show that the sum of new series is $\log 2+\frac{1}{2}\log\frac{p}{q}$.

Exercise 4.3.13. Show that $1-\frac{1}{\sqrt{2}}+\frac{1}{\sqrt{3}}-\frac{1}{\sqrt{4}}+\cdots$ converges, but $1-\frac{1}{\sqrt{2}}-\frac{1}{\sqrt{4}}+\frac{1}{\sqrt{3}}-\frac{1}{\sqrt{6}}-\frac{1}{\sqrt{8}}+\frac{1}{\sqrt{5}}-\frac{1}{\sqrt{10}}-\frac{1}{\sqrt{12}}+\cdots$ diverges.

Another distinction between absolute and conditional convergence is reflected on the product of two series.

Theorem 4.3.5. *Suppose* $\sum_{n=1}^{\infty} a_n$ *and* $\sum_{n=1}^{\infty} b_n$ *converge absolutely. Then* $\sum_{i,j=1}^{\infty} a_i b_j$ *also converge absolutely, and*

$$\sum_{i,j=1}^{\infty} a_i b_j = \left(\sum_{n=1}^{\infty} a_n\right)\left(\sum_{n=1}^{\infty} b_n\right).$$

Note that the infinite sum $\sum_{i,j=1}^{\infty} a_i b_j$ is a "double series" with two indices i and j. There are many ways of arranging this series into a single series. For example, the following is the "diagonal arrangement"

$$\begin{aligned}\sum (ab)_k &= a_1b_1 + a_1b_2 + a_2b_1 + \cdots \\ &\quad + a_1b_{n-1} + a_2b_{n-2} + \cdots + a_{n-1}b_1 + \cdots,\end{aligned}$$

and the following is the "square arrangement"

$$\begin{aligned}\sum (ab)_k &= a_1b_1 + a_1b_2 + a_2b_2 + a_2b_1 + \cdots \\ &\quad + a_1b_n + a_2b_n + \cdots + a_nb_{n-1} + a_nb_n + a_nb_{n-1} + \cdots + a_nb_1 + \cdots.\end{aligned}$$

Under the condition of the theorem, the series is supposed to converge absolutely. Then by Theorem 4.3.4, all arrangements give the same sum.

Example 4.3.7. We know from Example 4.1.4 that $\sum_{n=0}^{\infty}\frac{x^n}{n!}$ absolutely converges to e^x. By Theorem 4.3.5, we have

$$\begin{aligned}e^xe^y &= \left(\sum_{n=0}^{\infty}\frac{x^n}{n!}\right)\left(\sum_{n=0}^{\infty}\frac{y^n}{n!}\right) = \sum_{i,j=0}^{\infty}\frac{x^iy^j}{i!j!} = \sum_{n=0}^{\infty}\sum_{i+j=n}\frac{x^iy^j}{i!j!} \\ &= \sum_{n=0}^{\infty}\frac{1}{n!}\sum_{i=0}^{n}\frac{n!}{i!(n-i)!}x^iy^j = \sum_{n=0}^{\infty}\frac{1}{n!}(x+y)^n = e^{x+y}.\end{aligned}$$

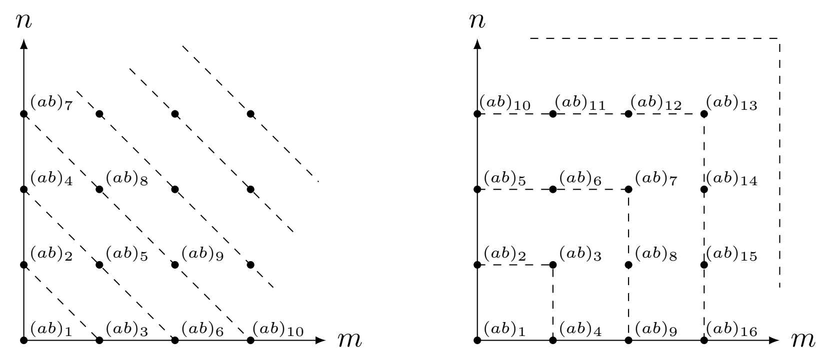


Figure 4.3.1: Diagonal and square arrangements.

In the second to the last equality, we used the binomial expansion.

Exercise 4.3.14. If you take the product of a geometric series with itself, what conclusion can you make?

Exercise 4.3.15. Suppose $\sum \frac{(-1)^n}{\sqrt{n}} = l$. Show that the square arrangement of the product of the series with itself converges to l^2. What is the sum of the diagonal arrangement?

4.4 Power Series

A *power series* at x_0 is

$$\sum_{n=0}^{\infty} a_n (x - x_0)^n.$$

By a simple change of variable, it is sufficient to consider power series at 0

$$\sum_{n=0}^{\infty} a_n x^n.$$

4.4.1 Convergence of Taylor Series

If $f(x)$ has derivatives of arbitrary order at x_0, then the high order approximations of the function gives us the *Taylor series*

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n = f(x_0)+f'(x_0)(x-x_0)+\frac{f''(x_0)}{2!}(x-x_0)^2+\cdots+\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n+\cdots .$$

The partial sum of the series is the n-th order Taylor expansion $T_n(x)$ of $f(x)$ at x_0.

In Example 4.1.3, we used the Lagrange form of the remainder $R_n(x) = f(x) - T_n(x)$ (Theorem 2.7.1) to show that the Taylor series of e^x converges to e^x. Exercise 4.1.11 further showed that the Taylor series of $\sin x$ and $\cos x$ converge to the trigonometric functions.

Example 4.4.1. Consider the Taylor series $\sum_{n=0}^{\infty} \dfrac{(-1)^{n+1}}{n} x^n$ of $\log(1+x)$. We have

$$\frac{d^n}{dx^n}\log(1+x) = \frac{(-1)^{n-1}(n-1)!}{(1+x)^n},$$

and the Lagrange form of the remainder gives

$$|R_n(x)| = \frac{1}{(n+1)!}\frac{n!}{|1+c|^{n+1}}|x|^{n+1} = \frac{|x|^{n+1}}{(n+1)|1+c|^{n+1}},$$

where c lies between 0 and x. If $-\dfrac{1}{2} \le x \le 1$, then $|x| \le |1+c|$, and we get $|R_n(x)| < \dfrac{1}{n+1}$. Therefore $\lim_{n\to\infty} R_n(x) = 0$, and the Taylor series converges to $\log(1+x)$.

The Taylor series is the harmonic series at $x = -1$ and therefore diverges. For $|x| > 1$, the terms of the Taylor series diverges to ∞, and therefore the Taylor series also diverges. The remaining case is $-1 < x < -\dfrac{1}{2}$.

In Exercise 4.4.2, a new form of the remainder is used to show that the Taylor series actually converges to $\log(1+x)$ for all $-1 < x \le 1$.

Example 4.4.2. The Taylor series of $(1-x)^{-1}$ is the geometric series $\sum_{n=0}^{\infty} x^n$. Example 4.1.1 shows that the Taylor series converges to the function for $|x| < 1$ and diverges for $|x| \ge 1$.

The Taylor series of $(1+x)^p$ is $\sum_{n=0}^{\infty} \dfrac{p(p-1)\cdots(p-n+1)}{n!} x^n$. The Lagrange form of the remainder gives

$$\begin{aligned}
|R_n(x)| &= \frac{|p(p-1)\cdots(p-n)|}{(n+1)!}|1+c|^{p-n-1}|x|^n \\
&= \frac{|p(p-1)\cdots(p-n)|}{(n+1)!}\frac{|1+c|^{p-1}|x|^n}{|1+c|^n},
\end{aligned}$$

where c lies between 0 and x. For $-\dfrac{1}{2} < x \le 1$, we have $|x| \le |1+c|$ and $|1+c|^{p-1}$ is bounded. Moreover, by Exercise 4.3.1, for $p > -1$, we have

$$\lim_{n\to\infty} \frac{|p(p-1)\cdots(p-n)|}{(n+1)!} = 0.$$

Then we conclude that $\lim_{n\to\infty} R_n(x) = 0$ for $-\dfrac{1}{2} < x \le 1$ and $p > -1$, and the Taylor series converges to $(1+x)^p$.

The Taylor series diverges for $|x| > 1$. Using a new form of the remainder, Exercise 4.4.2 shows that the Taylor series actually converges to $(1+x)^p$ for $|x| < 1$ and any p.

Example 4.4.3. In Example 2.5.18, we showed that the function

$$f(x) = \begin{cases} e^{-\frac{1}{|x|}}, & \text{if } x \ne 0, \\ 0, & \text{if } x = 0, \end{cases}$$

has all the high derivatives equal to 0. Therefore the Taylor series of the function is $\sum 0 = 0$, although the function is not 0.

A smooth function is *analytic* if it is always equal to its Taylor series. As pointed out after Example 2.5.18, the analytic property means that the function can be measured by polynomials.

Exercise 4.4.1. Use the Lagrange form of the remainder to show that *Cauchy form* of the remainder is

$$R_n(x) = \frac{f^{(n+1)}(c)}{n!}(x-c)^n(x-x_0),$$

where c lies between x_0 and x. Use this to show that the Taylor series of $\log(1+x)$ and $(1+x)^p$ at $x_0 = 0$ converge to the respective functions for any $|x| < 1$.

Exercise 4.4.2. The *Cauchy form* of the remainder is

$$R_n(x) = \frac{f^{(n+1)}(c)}{n!}(x-c)^n(x-x_0),$$

where c lies between x_0 and x. Use this to show that the Taylor series of $\log(1+x)$ and $(1+x)^p$ at $x_0 = 0$ converge to the respective functions for any $|x| < 1$.

4.4.2 Radius of Convergence

We regard a power series $\sum a_n x^n$ as a function with variable x. The domain of the function consists of those x, such that the series converges.

We may use the root test to find the domain. Suppose $\lim_{n\to\infty} \sqrt[n]{|a_n|}$ converges. Then

$$\lim_{n\to\infty} \sqrt[n]{|a_n x^n|} = |x| \lim_{n\to\infty} \sqrt[n]{|a_n|}.$$

Let

$$R = \frac{1}{\lim_{n\to\infty} \sqrt[n]{|a_n|}}.$$

By the limit version of the root test (see the discussion after Theorem 4.2.3), the power series converges for $|x| < R$ and diverges for $|x| > R$.

Example 4.4.4. For the geometric series $\sum x^n = 1+x+x^2+\cdots$, we have $\lim_{n\to\infty} \sqrt[n]{|1|} = 1$. Therefore the geometric series converges for $|x| < 1$ and diverges for $|x| > 1$. Moreover, the series also diverges for $|x| = 1$.

For the Taylor series $\sum_{n=0}^{\infty} \dfrac{(-1)^{n+1}}{n} x^n$ of $\log(1+x)$, we have $\lim_{n\to\infty} \sqrt[n]{\left|\dfrac{(-1)^{n+1}}{n}\right|} = \lim_{n\to\infty} \dfrac{1}{\sqrt[n]{n}} = 1$. Therefore the series converges for $|x| < 1$ and diverges for $|x| > 1$. We also know that the series converges at $x = 1$ and diverges at $x = -1$.

Example 4.4.5. The series $\sum x^{2n}$ has $a_n = 1$ for even n and $a_n = 0$ for odd n. The sequence $\sqrt[n]{|a_n|}$ diverges because it has two limits: $\lim_{\text{even}} \sqrt[n]{|a_n|} = 1$ and $\lim_{\text{odd}} \sqrt[n]{|a_n|} = 0$.

On the other hand, as a function, we have $\sum x^{2n} = f(x^2)$, where $f(x) = \sum x^n$ is the geometric series. Since the domain of f is $|x| < 1$, the domain of $f(x^2)$ is $|x^2| < 1$, which is equivalent to $|x| < 1$. Therefore $\sum x^{2n}$ converges for $|x| < 1$ and diverges for $|x| > 1$.

Theorem 4.4.1. *For any power series $\sum a_n x^n$, there is $R \ge 0$, such that the series absolutely converges for $|x| < R$ and diverges for $|x| > R$.*

Using the root test, we have proved the theorem in case $\lim_{n\to\infty} \sqrt[n]{|a_n|}$ converges. Example 4.4.11 suggests that the theorem is true in general. The theorem is a consequence of the fact that, if $\sum a_n r^n$ converges and $|x| < |r|$, then $\sum a_n x^n$ absolutely converges. Specifically, the convergence of $\sum a_n r^n$ implies $\lim_{n\to\infty} a_n r^n = 0$. This further implies that $|a_n r^n| < 1$ for sufficiently big n. Then for any fixed x satisfying $|x| < |r|$, we have

$$|a_n x^n| = |a_n r^n| \cdot \left|\frac{x}{r}\right|^n \le \left|\frac{x}{r}\right|^n.$$

Since $\left|\dfrac{x}{r}\right| < 1$ implies the convergence of $\sum \left|\dfrac{x}{r}\right|^n$, by the comparison test, the series $\sum a_n x^n$ absolutely converges.

The number R is the *radius of convergence* of power series. If $R = 0$, then the power series converges only for $x = 0$. If $R = +\infty$, then the power series converges for all x.

The same radius of convergence applies to $\sum a_n (x - x_0)^n$. The power series converges on $(x_0 - R, x_0 + R)$, and diverges on $(-\infty, x_0 - R)$ and on $(x_0 + R, +\infty)$.

We had the formula for radius in case $\sqrt[n]{|a_n|}$ converges. In general, the sequence may have many possible limits (for various subsequences). Let the *upper limit* $\overline{\lim} \sqrt[n]{|a_n|}$ be the maximum of all the possible limits. Then the radius is

$$R = \frac{1}{\overline{\lim}_{n\to\infty} \sqrt[n]{|a_n|}}.$$

One can verify the formula for Example 4.4.11.

Example 4.4.6. For any p, we have $\lim_{n\to\infty}\sqrt[n]{n^p}=1$. Therefore the radius of convergence for the power series $\sum n^p x^n$ is 1. The example already appeared in Example 4.3.2.

Example 4.4.7. By $\lim_{n\to\infty}\sqrt[n]{2^n+3^n}=3$, the series $\sum(2^n+3^n)x^n$ converges for $|x|<\frac{1}{3}$ and diverges for $|x|>\frac{1}{3}$. The series also diverges for $|x|=\frac{1}{3}$ because the terms do not converge to 0.

We note that the radius of convergence is $\frac{1}{2}$ for $\sum 2^n x^n$ and $\frac{1}{3}$ for $\sum 3^n x^n$. The radius for the sum of the two series is the smaller one.

Example 4.4.8. By $\lim_{n\to\infty}\sqrt[n]{n^n}=+\infty$, the series $\sum n^n x^n$ diverges for all $x\neq 0$.

By $\lim_{n\to\infty}\sqrt[n]{\dfrac{1}{n^n}}=0$, the series $\sum\dfrac{(-1)^n}{n^n}x^n$ converges for all x.

Example 4.4.9. In Example 4.2.9, we use the ratio test to show that the radius of convergence of $\sum\dfrac{(2n)!}{(n!)^2}x^n$ is $\dfrac{1}{4}$. The idea can be used to show that the radius of convergence for $\sum a_n x^n$ is $R=\lim_{n\to\infty}\left|\dfrac{a_n}{a_{n+1}}\right|$, provided that the limit converges.

For example, by

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

the radius of convergence for $\sum n^p x^n$ is 1. Moreover, the radius of convergence for the Taylor series of e^x is

$$\lim_{n\to\infty}\frac{\dfrac{1}{n!}}{\dfrac{1}{(n+1)!}}=\lim_{n\to\infty}(n+1)=+\infty.$$

Example 4.4.10. The *Bessel function* of order 0 is

$$J_0(x)=\sum_{n=0}^{\infty}\frac{(-1)^n x^{2n}}{2^{2n}(n!)^2}.$$

The radius of convergence is the square root of the radius of convergence of the series

$$\sum_{n=0}^{\infty}\frac{(-1)^n x^n}{2^{2n}(n!)^2}.$$

By

$$\lim_{n\to\infty}\left|\frac{\dfrac{(-1)^n}{2^{2n}(n!)^2}}{\dfrac{(-1)^{n+1}}{2^{2n+2}((n+1)!)^2}}\right| = \lim_{n\to\infty} 4(n+1)^2 = +\infty,$$

the later series converges for all x. Therefore the Bessel function is defined for all x.

Note that we cannot calculate $\lim_{n\to\infty}\left|\dfrac{a_n}{a_{n+1}}\right|$ directly for the Bessel function. In fact, the limit diverges.

Exercise 4.4.3. Suppose a_n can be divided into two subsequences $a_{n'}$ and $a_{n''}$. Suppose $\lim_{n'\to\infty}\sqrt[n']{|a_{n'}|} = l'$ and $\lim_{n''\to\infty}\sqrt[n'']{|a_{n''}|} = l''$ converge. Prove that $\dfrac{1}{\max\{l', l''\}}$ is the radius of convergence for $\sum a_n x^n$.

Exercise 4.4.4. Suppose $\lim_{n\to\infty}\left|\dfrac{a_n}{a_{n+1}}\right| = R$ converges. Prove that R is the radius of convergence for $\sum a_n x^n$.

Exercise 4.4.5. Determine the radius of convergence.

1. $\sum(-1)^n\dfrac{x^n}{n^p}$.
2. $\sum n^p(x-1)^n$.
3. $\sum n^p(2x-1)^n$.
4. $\sum n^p(2x+3)^n$.
5. $\sum\left(\dfrac{a^n}{n}+\dfrac{b^n}{n^2}\right)x^n$.
6. $\sum\dfrac{x^n}{a^n+b^n}$.
7. $\sum n^{\sqrt{n}}x^n$.
8. $\sum n!x^n$.
9. $\sum\dfrac{(-1)^{n+1}}{\sqrt{n!}}x^n$.
10. $\sum\dfrac{(n!)^2}{(2n)!}x^n$.
11. $\sum\dfrac{(3n)!}{n!(2n)!}x^n$.
12. $\sum a^{n^2}x^n$.
13. $\sum a^{n^2}x^{n^2}$.
14. $\sum 2^n x^{n^2-1}$.
15. $\sum(2+(-1)^n)^n x^n$.
16. $\sum\dfrac{(2+(-1)^n)^n}{\log n}x^n$.

Exercise 4.4.6. Find the radius of convergence.

1. $\sum\left(\dfrac{n+1}{n}\right)^n x^n$.
2. $\sum\left(\dfrac{n+1}{n}\right)^{n^2} x^n$.
3. $\sum\left(\dfrac{n+a}{n+b}\right)^{n^2} x^{n+2}$.
4. $\sum(-1)^n\left(\dfrac{n+1}{n}\right)^{n^2} x^n$.
5. $\sum(-1)^n\left(\dfrac{n+1}{n}\right)^{n^2} x^{n^2}$.
6. $\sum\left(\dfrac{an+b}{cn+d}\right)^n x^{n-2}$.

Exercise 4.4.7. Find the domain of the *Bessel function* of order 1

$$J_1(x) = \sum_{n=0}^{\infty}\frac{(-1)^n x^{2n+1}}{n!(n+1)!2^{2n+1}}.$$

Exercise 4.4.8. Find the domain of the *Airy function*

$$A(x) = 1 + \frac{x^3}{2\cdot 3} + \frac{x^6}{2\cdot 3\cdot 5\cdot 6} + \frac{x^9}{2\cdot 3\cdot 5\cdot 6\cdot 8\cdot 9} + \cdots .$$

Exercise 4.4.9. Suppose the radii of convergence for $\sum a_n x^n$ and $\sum b_n x^n$ are R and R'. What can you say about the radii of convergence for the following power series?

$$\sum (a_n + b_n)x^n, \quad \sum (a_n - b_n)x^n, \quad \sum (-1)^n a_n x^n, \quad \sum a_n(2x-1)^n,$$

$$\sum a_n x^{2n}, \quad \sum a_n x^{n+2}, \quad \sum a_{2n} x^n, \quad \sum a_{n+2} x^n,$$
$$\sum a_n x^{n^2}, \quad \sum a_{n^2} x^n, \quad \sum a_{2n} x^{2n}, \quad \sum a_{n^2} x^{n^2}.$$

4.4.3 Function Defined by Power Series

Examples 2.5.18 and 4.4.3 suggest that if a function is the sum of a power series, then the function is particularly nice. In fact, the function should be nicer than functions with derivatives of any order.

Because power series converge absolutely within the radius of convergence, by Theorem 4.3.5, we can multiply two power series together within the common radius of convergence.

Theorem 4.4.2. *Suppose $f(x) = \sum a_n x^n$ and $g(x) = \sum b_n x^n$ have radii of convergence R and R'. Then for $|x| < \min\{R, R'\}$, we have*

$$f(x)g(x) = \sum c_n x^n, \quad c_n = a_0 b_n + a_1 b_{n-1} + \cdots + a_n b_0.$$

The product should be the sum of $a_i b_j x^{i+j}$. We get the power series $\sum c_n x^n$ by gathering all the terms with power x^n.

The power series can also be differentiated or integrated term by term within the radius of convergence.

Theorem 4.4.3. *Suppose $f(x) = \sum_{n=0}^{\infty} a_n x^n$ for $|x| < R$. Then*

$$f'(x) = \sum_{n=1}^{\infty} (a_n x^n)' = a_1 + 2a_2 x + 3a_3 x^2 + \cdots + n a_n x^{n-1} + \cdots ,$$

and

$$\int_0^x f(t)dt = \sum_{n=0}^{\infty} \int_0^x a_n t^n dt = a_0 x + \frac{a_1}{2}x^2 + \frac{a_2}{3}x^3 + \cdots + \frac{a_n}{n+1}x^{n+1} + \cdots$$

for $|x| < R$.

Example 4.4.11. Taking the derivative of $\dfrac{1}{1-x} = \sum_{n=0}^{\infty} x^n$, we get

$$\frac{1}{(1-x)^2} = 1 + 2x + 3x^2 + 4x^3 + \cdots + nx^{n-1} + \cdots,$$
$$\frac{2}{(1-x)^3} = 2\cdot 1 + 3\cdot 2x + 4\cdot 3x^2 + 5\cdot 4x^3 + \cdots + n(n-1)x^{n-2} + \cdots.$$

Therefore

$$\begin{aligned} 1^2x + 2^2x^2 + \cdots + n^2x^n + \cdots &= \sum_{n=1}^{\infty} n^2x^n = x\sum_{n=1}^{\infty} nx^{n-1} + x^2\sum_{n=2}^{\infty} n(n-1)x^{n-2} \\ &= x\frac{1}{(1-x)^2} + x^2\frac{2}{(1-x)^3} = \frac{x(1+x)}{(1-x)^3}. \end{aligned}$$

If we integrate instead, then we get

$$\log(1-x) = -\int_0^x \frac{dx}{1-x} = -x - \frac{x^2}{2} - \frac{x^3}{3} - \cdots - \frac{x^n}{n!} - \cdots, \quad \text{for } |x| < 1.$$

Substituting $-x$ for x, we get the Taylor series of $\log(1+x)$

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots + (-1)^{n+1}\frac{x^n}{n!} + \cdots, \quad \text{for } |x| < 1.$$

Note that in Example 4.4.1, by estimating the remainder, we were able to prove the equality rigorously only for $-\dfrac{1}{2} < x < 1$. Here by using term wise integration, we get the equality for all x within the radius of convergence.

Example 4.4.12. By integrating $\dfrac{1}{1+x^2} = \sum_{n=0}^{\infty}(-1)^n x^{2n}$, we get the Taylor series of $\arctan x$

$$\arctan x = \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1}x^{2n+1} = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots + \frac{(-1)^n}{2n+1}x^{2n+1} + \cdots \quad \text{for } |x| < 1.$$

Exercise 4.4.10. Use the product of power series to verify the identity $\sin 2x = 2\sin x\cos x$.

Exercise 4.4.11. Find Taylor series and determine the radius of convergence.

1. $\dfrac{1}{(x-1)(x-2)}$, at 0.
2. $\sqrt{x}$, at $x = 2$.
3. $\sin x^2$, at 0.
4. $\sin^2 x$, at 0.
5. $\sin x$, at $\dfrac{\pi}{2}$.
6. $\sin 2x$, at $\dfrac{\pi}{2}$.
7. $\arcsin x$, at 0.
8. $\arctan x$, at 0.
9. $\displaystyle\int_0^x \frac{\sin t}{t}dt$, at 0.

Exercise 4.4.12. Given the Taylor series $\sum_{n=0}^{\infty} a_n x^n$ of $f(x)$, find the Taylor series of $\dfrac{f(x)}{1+x}$.

Exercise 4.4.13. Show that the function $f(x) = \sum_{n=0}^{\infty} \dfrac{x^n}{(n!)^2}$ satisfies $xf'' + f' - f = 0$.

Exercise 4.4.14. Show that the Airy function in Exercise 4.4.8 satisfies $f'' - xf = 0$.

Exercise 4.4.15. Show that the Bessel functions in Example 4.4.10 and Exercise 4.4.7 satisfy

$$xJ_0'' + J_0' + xJ_0 = 0, \quad x^2J_1'' + xJ_1' + (x^2-1)J_1 = 0.$$

A power series may or may not converge at the radius of convergence (i.e., at $\pm R$). If it converges, then the following gives the value of the sum.

Theorem 4.4.4. *Suppose* $\sum_{n=0}^{\infty} a_n x^n$ *converges for* $|x| < R$ *and* $x = R$. *Then* $\sum_{n=0}^{\infty} a_n R^n = \lim_{x\to R^-} \sum_{n=0}^{\infty} a_n x^n$.

The theorem says that, if $f(x) = \sum_{n=0}^{\infty} a_n x^n$ is also defined at R, then $f(x)$ is left continuous at R. We also have the similar statement at the other end $-R$.

Example 4.4.13. By Examples 4.4.11 and 4.3.1, we know $\log(1+x) = \sum_{n=1}^{n} \dfrac{(-1)^{n+1}}{n} x^n$ converges for $|x| < 1$, and the series converges at $x = 1$. Since $\log(1+x)$ is continuous at $x = 1$, by Theorem 4.4.4, we get

$$\sum_{n=1}^{n} \frac{(-1)^{n+1}}{n} = \lim_{x\to 1^-} \sum_{n=1}^{n} \frac{(-1)^{n+1}}{n} x^n = \lim_{x\to 1^-} \log(1+x) = \log(1+1) = \log 2.$$

We computed the sum in Example 4.3.6 by another way.

Exercise 4.4.16. Find the sum. Discuss what happens at the radius of convergence.

1. $\sum_{n=1}^{\infty} n^2 x^n$.
2. $\sum_{n=1}^{\infty} n^3 x^n$.
3. $\sum_{n=2}^{\infty} \dfrac{(x-1)^n}{n(n-1)}$.
4. $\sum_{n=1}^{\infty} \dfrac{x^n}{n(n+1)(n+2)}$.
5. $\sum_{n=0}^{\infty} \dfrac{x^{2n+1}}{2n+1}$.
6. $\sum_{n=0}^{\infty} \dfrac{x^n}{2n+1}$.

Exercise 4.4.17. Find the sum. Discuss what happens at the radius of convergence.

1. $\sum_{n=1}^{\infty} (-1)^n \dfrac{x^{2n+1}}{n!}$.
2. $\sum_{n=1}^{\infty} (-1)^n \dfrac{x^{2n}}{(2n+1)!}$.
3. $\sum_{n=1}^{\infty} \dfrac{x^n}{2^n(2n-1)!}$.

Exercise 4.4.18. 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. $\displaystyle\int_0^x \frac{\sin t}{t} dt$.
3. $\arctan x$.
4. $\displaystyle\int_0^x e^{-t^2} dt$.
5. $\log(x + \sqrt{1+x^2})$.
6. $\displaystyle\int_0^x \frac{\log(1-t)}{t} dt$.

4.5 Fourier Series

If $f(x+p) = f(x)$ for all x and a constant p, then we say $f(x)$ is a *periodic function* of *period* p. For example, the functions $\sin x$ and $\cos x$ have period 2π, and $\tan x$ has period π.

A periodic function of period p is also a periodic function of period kp for any integer k. For example, $\cos nx$ and $\sin nx$ have the period $p = \dfrac{2\pi}{n}$ as well as the period $np = 2\pi$.

If $f(x)$ has period p, then $f(x+a)$ still has period p, and $f(ax)$ is periodic with period $\dfrac{p}{a}$.

A combination of periodic functions of the same period p is still periodic of period p. For example, $\sin x + \cos x$, $\sin 3x \cos^2 x$, $\sqrt{2\sin^2 x + \cos^4 x}$ are periodic of period 2π.

The Taylor series approximates a function near a point by linear combinations of power functions. Similarly, we wish to approximate a periodic function by linear combinations of simple periodic functions such as sine and cosine. Specifically, we wish a periodic function $f(x)$ of period 2π to be approximated as

$$\begin{aligned} f(x) &\sim a_0 + \sum_{n=1}^{\infty}(a_n \cos nx + b_n \sin nx) \\ &= a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + \cdots . \end{aligned}$$

Note that there is no b_0 because $a_0 = a_0 \cos 0x + b_0 \sin 0x$. Moreover, like the Taylor series, we use $\sim$ instead of $=$ to indicate that the equality is yet to be established.

The approximation of a periodic function by (linear combinations of) trigonometric functions is not measured by the values at single points, but rather the overall approximation in terms of the integral of the difference function. This means that we can only expect that the sum of the trigonometric series to be equal to the function "almost everywhere".

4.5.1 Fourier Coefficient

Let $f(x)$ be a periodic function of period 2π. Our first problem is to find the coefficients a_n and b_n in the trigonometric series. By Exercise 3.1.6, the trigonometric functions are "orthogonal" in the sense that

$$\int_0^{2\pi} \cos mx \sin nx dx = 0;$$
$$\int_0^{2\pi} \cos mx \cos nx dx = \begin{cases} 0, & \text{if } m \neq n, \\ \pi, & \text{if } m = n \neq 0, \\ 2\pi, & \text{if } m = n = 0; \end{cases}$$
$$\int_0^{2\pi} \sin mx \sin nx dx = \begin{cases} 0, & \text{if } m \neq n \text{ or } m = n = 0, \\ \pi, & \text{if } m = n \neq 0. \end{cases}$$

We expect the sum of the trigonometric series to be equal to $f(x)$ as far as integrations are concerned. We also assume that the integration of infinite series can be calculated term by term. Then we get

$$\begin{aligned} \int_0^{2\pi} f(x) \cos nx dx &= a_0 \int_0^{2\pi} \cos nx dx + \sum_{k=1}^{\infty} a_m \int_0^{2\pi} \cos mx \cos nx dx \\ &\quad + \sum_{k=1}^{\infty} b_m \int_0^{2\pi} \sin mx \cos nx dx = \begin{cases} \pi a_n, & \text{if } n \neq 0, \\ 2\pi a_0, & \text{if } n = 0; \end{cases} \\ \int_0^{2\pi} f(x) \sin nx dx &= a_0 \int_0^{2\pi} \sin nx dx + \sum_{m=1}^{\infty} a_m \int_0^{2\pi} \cos mx \sin nx dx \\ &\quad + \sum_{m=1}^{\infty} b_m \int_0^{2\pi} \sin mx \sin nx dx = \pi b_n. \end{aligned}$$

Definition 4.5.1. The *Fourier series* of a periodic function $f(x)$ of period 2π is

$$f(x) \sim a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx),$$

with the *Fourier coefficients*

$$\begin{aligned} a_0 &= \frac{1}{2\pi} \int_0^{2\pi} f(x) dx, \\ a_n &= \frac{1}{\pi} \int_0^{2\pi} f(x) \cos nx dx, \quad n \neq 0, \\ b_n &= \frac{1}{\pi} \int_0^{2\pi} f(x) \sin nx dx, \quad n \neq 0. \end{aligned}$$

Example 4.5.1. Similar to Example 2.5.1, by

$$\begin{aligned}\sin^4 x &= \frac{1}{4}(1-\cos 2x)^2 = \frac{1}{4}\left(1 - 2\cos 2x + \frac{1}{2}(1-\cos 4x)\right)\\ &= \frac{3}{8} - \frac{1}{2}\cos 2x + \frac{1}{8}\cos 4x,\end{aligned}$$

the right side is the Fourier series of $\sin^4 x$. The coefficients give

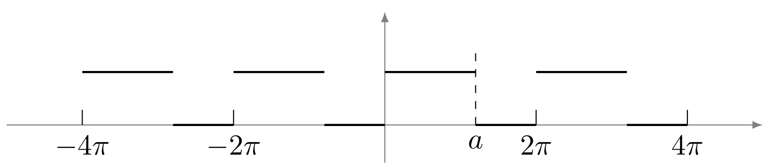
$$\int_0^{2\pi} \sin^4 x dx = \frac{3\pi}{4}, \quad \int_0^{2\pi} \sin^4 x \cos 2x dx = -\frac{\pi}{2}, \quad \int_0^{2\pi} \sin^4 x \cos 4x dx = \frac{\pi}{8}.$$

Example 4.5.2. A periodic function is determined by its value on one interval of period length. For example, if $f(x)$ is a periodic function of period 2π and satisfies

$$f(x) = \begin{cases} 1, & \text{if } 0 \le x < a, \\ 0, & \text{if } a \le x < 2\pi, \end{cases}$$

then

$$f(x) = \begin{cases} 0, & \text{if } 2k\pi \le x < 2k\pi + a, \\ 1, & \text{if } 2k\pi + a \le x < 2(k+1)\pi. \end{cases}$$



The Fourier coefficients are

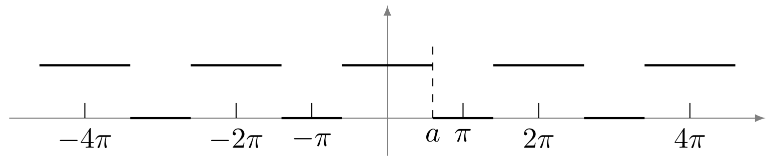
$$\begin{aligned}a_0 &= \frac{1}{2\pi}\int_0^a 1dx = \frac{a}{2\pi},\\ a_n &= \frac{1}{\pi}\int_0^a \cos nx dx = \frac{\sin na}{n\pi},\\ b_n &= \frac{1}{\pi}\int_0^a \sin nx dx = \frac{1-\cos na}{n\pi}.\end{aligned}$$

and the Fourier series is

$$f(x) \sim \frac{a}{2\pi} + \sum_{n=1}^{\infty} \frac{1}{n\pi}(\sin na \cos nx + (1-\cos na)\sin nx).$$

Example 4.5.3. Let $f(x)$ be the even periodic function of period 2π satisfying

$$f(x) = \begin{cases} 1, & \text{if } |x| \le a, \\ 0, & \text{if } a < |x| \le \pi. \end{cases}$$



We have

$$a_0 = \frac{1}{2\pi}\int_0^{2\pi} f(x)dx = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)dx = \frac{1}{\pi}\int_0^{\pi} f(x)dx = \frac{1}{\pi}\int_0^{a} 1dx = \frac{a}{\pi},$$
$$a_n = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos nxdx = \frac{2}{\pi}\int_0^{\pi} f(x)\cos nxdx = \frac{2}{\pi}\int_0^{a} \cos nxdx = \frac{2a}{n\pi}.$$

The calculation used the fact that $\int_0^p f(x)dx = \int_a^{a+p} f(x)dx$ for any periodic function of period p. We may also calculate b_n and find $b_n = 0$. In fact, for even function, we expect that all the odd terms $b_n \sin nx$ to vanish.

Exercise 4.5.1. Suppose $f(x)$ is an even periodic function of period 2π. Prove that

$$a_0 = \frac{1}{\pi}\int_0^{\pi} f(x)dx, \quad a_n = \frac{2}{\pi}\int_0^{\pi} f(x)\cos nxdx, \quad b_n = 0.$$

Exercise 4.5.2. Suppose $f(x)$ is an odd periodic function of period 2π. Prove that

$$a_0 = a_n = 0, \quad b_n = \frac{2}{\pi}\int_0^{\pi} f(x)\sin nxdx.$$

Exercise 4.5.3. Extend $f(x)$ on $\left(0, \frac{\pi}{2}\right)$ to a periodic function of period 2π, such that its Fourier series is of the form $\sum_{n=1}^{\infty} a_n \cos(2n-1)x$? How about $\sum_{n=1}^{\infty} b_n \sin(2n-1)x$?

Exercise 4.5.4. Given the Fourier series of $f(x)$ and $g(x)$, what is the Fourier series of $af(x) + bg(x)$? Use the idea and Example 4.5.2 to find the Fourier series of the periodic function $f(x)$ of period 2π satisfying

$$f(x) = \begin{cases} 1, & \text{if } a \le x < b, \\ 0, & \text{if } 0 \le x < a \text{ or } b \le x < 2\pi. \end{cases}$$

Exercise 4.5.5. Suppose $f(x)$ is a periodic function of period 2π. What is the relation between the Fourier series of $f(x)$ and $f(x+a)$? Use the idea and Example 4.5.3 to derive Example 4.5.2.

A periodic function $f(x)$ of period p may be converted to a periodic function $f\left(\frac{p}{2\pi}x\right)$ of period 2π. Then the Fourier series of $f\left(\frac{p}{2\pi}x\right)$ gives the Fourier series of $f(x)$. Alternatively, the basic periodic functions $\cos nx$ and $\sin nx$ of period 2π

give the basic periodic functions $\cos\dfrac{2n\pi}{p}x$ and $\sin\dfrac{2n\pi}{p}x$ of period p, and we expect the Fourier series of $f(x)$ to be

$$f(x) \sim a_0 + \sum_{n=1}^{\infty}\left(a_n \cos\frac{2n\pi}{p}x + b_n \sin\frac{2n\pi}{p}x\right).$$

By an argument similar to the case of period 2π, we get the Fourier coefficients

$$\begin{aligned}
a_0 &= \frac{1}{p}\int_0^p f(x)dx,\\
a_n &= \frac{2}{p}\int_0^p f(x)\cos\frac{2n\pi}{p}xdx, \quad n \neq 0,\\
b_n &= \frac{2}{p}\int_0^p f(x)\sin\frac{2n\pi}{p}xdx, \quad n \neq 0.
\end{aligned}$$

Exercise 4.5.6. Suppose $f(x)$ is a periodic function of period p.

1. Write down the Fourier series for $f\left(\dfrac{p}{2\pi}x\right)$, together with the formulae for its coefficients.

2. Convert the first part to statements about the original $f(x)$.

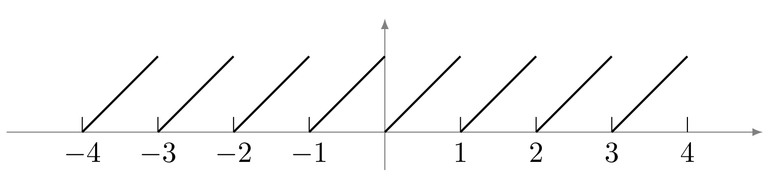
Exercise 4.5.7. Use Example 4.5.2 to derive the Fourier series of the periodic function of period 1 satisfying

$$f(x) = \begin{cases} 1, & \text{if } 0 \le x < a,\\ 0, & \text{if } a \le x < 1.\end{cases}$$

Exercise 4.5.8. Derive the formulae for the Fourier coefficients of periodic even or odd functions of period p, similar to Exercises 4.5.1 and 4.5.2.

Example 4.5.4. The function x on $(0,1)$ extends to a periodic function of period 1

$$f(x) = x - k, \quad k < x < k+1.$$



Note that we do not care about the value at the integer points because it does not affect the Fourier coefficients, which are

$$\begin{aligned}
a_0 &= \frac{1}{1}\int_0^1 x dx = \frac{1}{2}, \\
a_n &= \frac{2}{1}\int_0^1 x\cos 2n\pi x dx = \frac{1}{n\pi}\int_0^1 x d\sin 2n\pi x = -\frac{1}{n\pi}\int_0^1 \sin 2n\pi x dx = 0, \\
b_n &= \frac{2}{1}\int_0^1 x\sin 2n\pi x dx = -\frac{1}{n\pi}\int_0^1 x d\cos 2n\pi x \\
&= -\frac{1}{n\pi}\left(1 - \int_0^1 \cos 2n\pi x dx\right) = -\frac{1}{n\pi}.
\end{aligned}$$

We note that the reason for $a_n = 0$ for $n \neq 0$ is that $f(x) - \frac{1}{2}$ is an odd function.

The Fourier series is

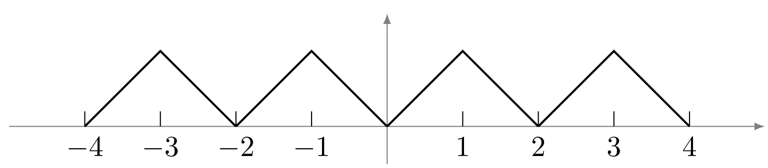
$$x \sim \frac{1}{2} - \frac{1}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}\sin 2n\pi x, \quad x \in (0,1).$$

We indicate $x \in (0,1)$ because the function equals x only on the interval. The function is $x-1$ instead of x on the interval $(1,2)$.

Example 4.5.5. The function x on $(0,1)$ extends to an even periodic function of period 2

$$f(x) = |x - 2k|, \quad 2k-1 < x < 2k+1.$$

This is also the extension of the function $|x|$ on $(-1,1)$ to a periodic function of period 2.



Using Exercises 4.5.1 and 4.5.8, we get $b_n = 0$ and

$$\begin{aligned}
a_0 &= \frac{2}{2}\int_0^1 f(x)dx = \int_0^1 x dx = \frac{1}{2}, \\
a_n &= \frac{2}{1}\int_0^1 f(x)\cos n\pi x dx = 2\int_0^1 x\cos n\pi x dx = \begin{cases} -\dfrac{4}{n^2\pi^2}, & \text{if } n \text{ is odd}, \\ 0, & \text{if } n \text{ is even}. \end{cases}
\end{aligned}$$

Replacing n by $2n+1$, the Fourier series is

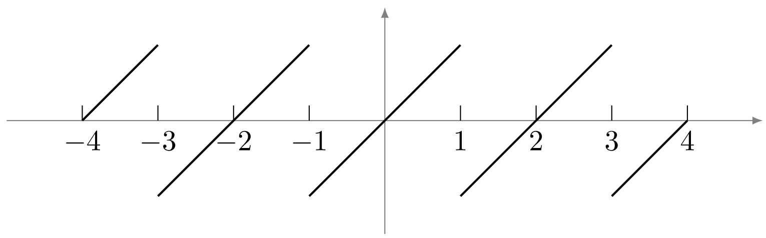
$$|x| \sim \frac{1}{2} - \frac{4}{\pi^2}\sum_{n=0}^{\infty}\frac{1}{(2n+1)^2}\cos(2n+1)\pi x, \quad x \in (-1,1).$$

We may also extend the function to an odd periodic function of period 2. This is also the extension of the function x on $(-1,1)$ to a periodic function of period 2. The Fourier coefficients $a_n = 0$ for the odd function, and

$$b_n = 2\int_0^1 x \sin n\pi x dx = \frac{(-1)^{n+1}2}{n\pi}.$$

The Fourier series is

$$x \sim \sum_{n=1}^{\infty} \frac{(-1)^{n+1}2}{n\pi} \sin n\pi x, \quad x \in (-1,1).$$

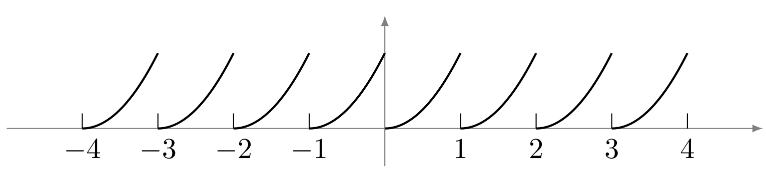


Example 4.5.6. Let $f(x)$ be the periodic function of period 1 extending the function x^2 on $(0,1)$. Then

$$\begin{aligned} a_0 &= \int_0^1 x^2 dx = \frac{1}{3}, \\ a_n &= 2\int_0^1 x^2 \cos 2n\pi x dx = \frac{1}{n^2\pi^2}, \\ b_n &= 2\int_0^1 x^2 \sin 2n\pi x dx = -\frac{1}{n\pi}. \end{aligned}$$

The Fourier series is

$$x^2 \sim \frac{1}{3} + \sum_{n=1}^{\infty} \left(\frac{1}{n^2\pi^2} \cos 2n\pi x - \frac{1}{n\pi} \sin 2n\pi x \right), \quad x \in (0,1).$$



Exercise 4.5.9. The periodic function is given on one interval of period length. Find the Fourier series.

1. $\sin^2 x$ on $(-\pi,\pi)$.
2. $\sin x$ on $(0,\pi)$.
3. $|\sin x|$ on $(0,2\pi)$.
4. $\sin x$ on $(0,p)$.
5. $\cos x$ on $(0,p)$.
6. $|x|$ on $(-p,p)$.
7. x on (a,b).
8. $x\sin x$ on $(-\pi,\pi)$.
9. e^x on $(0,1)$.

Exercise 4.5.10. Use the Fourier coefficient to calculate the integral.

1. $\int_0^{\pi} \sin^2 x\cos 2x dx$.
2. $\int_0^{2\pi} \sin^6 x dx$.
3. $\int_0^{2\pi} \sin x\cos 2x\sin 3x dx$.
4. $\int_0^{2\pi} \sin x\cos 2x\cos 3x dx$.
5. $\int_0^{2\pi} \sin^3 x\sin 3x dx$.
6. $\int_0^{2\pi} \sin^3 x\cos 3x dx$.

Exercise 4.5.11. Write the formula for the Fourier coefficients of the even periodic (of period $2p$) extension of a function $f(x)$ on $(0,p)$. What about the odd extension?

Exercise 4.5.12. Extend the function on $(0,p)$ to even and odd functions of period $2p$ and compute the Fourier series.

1. x^2 on $(0,1)$.
2. $\sin x$ on $(0,\pi)$.
3. $\cos x$ on $(0,p)$.

Exercise 4.5.13. Given the Fourier series of functions $f(x)$ and $g(x)$ of period p. Find the Fourier series of the following periodic functions.

1. $f(ax)$.
2. $f(x)\cos\dfrac{2\pi x}{p}$.
3. $f(x)\sin\dfrac{2\pi x}{p}$.
4. $\dfrac{1}{2h}\int_{x-h}^{x+h} f(t)dt$.
5. $\dfrac{1}{h}\int_{x}^{x+h} f(t)dt$.
6. $\int_0^{p} f(t)g(x-t)dt$.

Exercise 4.5.14. Use Examples 4.5.4 and 4.5.5 to find the Fourier series.

1. x on $(0,p)$.
2. $-x$ on $(-p,0)$.
3. $|x|$ on $(-p,p)$.
4. 0 on $(-1,0)$ and x on $(0,1)$.
5. ax on $(-1,0)$ and bx on $(0,1)$.
6. ax on $(-p,0)$ and bx on $(0,p)$.

4.5.2 Complex Form of Fourier Series

The sine and cosine functions are related to the exponential function via the use of complex numbers

$$e^{ix}=\cos x+i\sin x,\quad \cos x=\frac{e^{ix}+e^{-ix}}{2},\quad \sin x=\frac{e^{ix}-e^{-ix}}{2i},\quad i=\sqrt{-1}.$$

Correspondingly, the Fourier series may be rewritten

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{inx}, \quad c_0 = a_0, \quad c_n = \frac{a_n - ib_n}{2}, \quad c_{-n} = \frac{a_n + ib_n}{2},$$

where complex Fourier coefficients are

$$c_n = \frac{1}{2\pi}\int_0^{2\pi} f(x)e^{-inx}dx.$$

In this course, $f(x)$ is a real valued function, and a_n, b_n are real numbers. Therefore $c_{-n}e^{i(-n)x}$ is the complex conjugation of $c_n e^{inx}$, and the usual Fourier series is given by

$$a_0 = c_0, \quad a_n \cos nx + b_n \sin nx = 2\mathrm{Re}(c_n e^{inx}).$$

Example 4.5.7. For the function in Example 4.5.2, the complex Fourier coefficient is

$$\begin{aligned} c_0 &= \frac{1}{2\pi}\int_0^a dx = \frac{a}{2\pi}, \\ c_n &= \frac{1}{2\pi}\int_0^a e^{-inx}dx = \frac{1}{-2in\pi}e^{-inx}|_0^a = \frac{i}{2n\pi}(e^{-ina} - 1). \end{aligned}$$

The complex Fourier series is

$$f(x) \sim \frac{a}{2\pi} + \sum_{n\neq 0} \frac{i}{2n\pi}(e^{-ina} - 1)e^{inx}, \quad x \in (0, 2\pi).$$

By

$$\begin{aligned} 2\mathrm{Re}\left(\frac{i}{2n\pi}(e^{-ina} - 1)e^{inx}\right) &= \mathrm{Re}\left(\frac{i}{n\pi}((\cos na - 1) - i\sin na)(\cos nx + i\sin nx)\right) \\ &= \frac{1}{n\pi}(\sin na\cos nx - (\cos na - 1)\sin nx), \end{aligned}$$

we recover the Fourier series in terms of trigonometric functions in Example 4.5.2.

Example 4.5.8. Consider the periodic function of period 2π given by e^x on $(0, 2\pi)$. The complex Fourier coefficient is (recall that $e^{2ik\pi} = 1$)

$$c_n = \frac{1}{2\pi}\int_0^{2\pi} e^x e^{-inx}dx = \frac{1}{2\pi(1-in)}e^{(1-in)x}|_0^{2\pi} = \frac{e^{2\pi} - 1}{2n\pi(1-in)}.$$

The complex Fourier series is

$$e^x \sim \frac{e^{2\pi} - 1}{2\pi}\sum_{n=-\infty}^{\infty} \frac{1}{1-in}e^{inx}, \quad x \in (0, 2\pi).$$

By

$$\mathrm{Re}\left(\frac{1}{1-in}e^{inx}\right) = \mathrm{Re}\left(\frac{1+in}{1+n^2}(\cos nx + i\sin nx)\right) = \frac{\cos nx - n\sin nx}{1+n^2},$$

we get the Fourier series in terms of trigonometric functions

$$e^x \sim \frac{e^{2\pi}-1}{\pi}\left(\frac{1}{2}+\sum_{n=1}^{\infty}\frac{\cos nx - n\sin nx}{1+n^2}\right), \quad x\in(0,2\pi).$$

By changing x to $2\pi x$, we get the Fourier series for the periodic function of period 1 given by $e^{2\pi x}$ on $(0,1)$

$$e^{2\pi x} \sim \frac{e^{2\pi}-1}{\pi}\left(\frac{1}{2}+\sum_{n=1}^{\infty}\frac{\cos 2n\pi x - n\sin 2n\pi x}{1+n^2}\right), \quad x\in(0,1).$$

Example 4.5.9. Consider the function $f(x) = \dfrac{a\sin x}{1-2a\cos x + a^2}$, with $|a|<1$. We rewrite the function and take the Taylor expansion in terms of $e^{inx} = z^n$, $z = e^{ix}$,

$$\begin{aligned}
\frac{a\sin x}{1-2a\cos x+a^2} &= \frac{a\dfrac{e^{ix}-e^{-ix}}{2i}}{1-2a\dfrac{e^{ix}+e^{-ix}}{2}+a^2}\\
&= \frac{a}{2i}\frac{e^{ix}-e^{-ix}}{(1-ae^{ix})(1-ae^{-ix})} = \frac{1}{2i}\left(\frac{1}{1-ae^{ix}}-\frac{1}{1-ae^{-ix}}\right)\\
&= \frac{1}{2i}\left(\sum_{n=0}^{\infty}(ae^{ix})^n - \sum_{n=0}^{\infty}(ae^{-ix})^n\right) = \frac{1}{2i}\sum_{n\neq 0}a^{|n|}e^{inx}\\
&= \frac{1}{2i}\sum_{n=1}^{\infty}a^n(e^{inx}-e^{-inx}) = \sum_{n=1}^{\infty}a^n\sin nx.
\end{aligned}$$

Note that we have geometric series because $|ae^{ix}| = |ae^{-ix}| = |a| < 1$. Moreover, the Fourier series is actually equal to the function. The example also shows the connection between the Fourier series and the power series.

Exercise 4.5.15. Find the complex form of the Fourier series of a^x on $(0,2\pi)$. Then by taking $a = e^{\frac{1}{2\pi}}$, derive the Fourier series of the function e^x on $(0,1)$.

Exercise 4.5.16. Find the complex form of the Fourier series of $e^{\frac{ix}{2}}$ on $(0,2\pi)$. Then derive the Fourier series of $\cos x$ and $\sin x$ on $(0,\pi)$ by taking the real and imaginary parts.

Exercise 4.5.17. What is the complex form of the Fourier series for a periodic function of period p?

Exercise 4.5.18. Find complex form of Fourier series.

1. $\sin^2 x$ on $(-\pi, \pi)$.
2. e^{ix} on $(0, p)$.
3. $|x|$ on $(-p, p)$.
4. x^2 on $(0, 1)$.
5. $x \sin x$ on $(-\pi, \pi)$.
6. $x \cos x$ on $(-\pi, \pi)$.

Exercise 4.5.19. Find Fourier series, where $|a| < 1$.

1. $\dfrac{1 - a\cos x}{1 - 2a\cos x + a^2}$.
2. $\dfrac{1 - a^2}{1 - 2a\cos x + a^2}$.
3. $\log(1 - 2a\cos x + a^2)$.

4.5.3 Derivative and Integration of Fourier Series

The derivative of a periodic function is still periodic. We may use $f(p) = f(0)$ (which is the periodic property) and the integration by parts to compute the Fourier coefficients A_n, B_n of $f'(x)$

$$\begin{aligned}
A_0 &= \frac{1}{p}\int_0^p f'(x)dx = \frac{1}{p}(f(p) - f(0)) = 0, \\
A_n &= \frac{2}{p}\int_0^p f'(x)\cos\frac{2n\pi}{p}xdx = \frac{2}{p}\int_0^p \cos\frac{2n\pi}{p}xdf(x) \\
&= \frac{2}{p}\left(f(p) - f(0) + \int_0^p f(x)\frac{2n\pi}{p}\sin\frac{2n\pi}{p}xdx\right) = \frac{2n\pi}{p}b_n, \\
B_n &= \frac{2}{p}\int_0^p f'(x)\sin\frac{2n\pi}{p}xdx = \frac{2}{p}\int_0^p \sin\frac{2n\pi}{p}xdf(x) \\
&= -\frac{2}{p}\int_0^p f(x)\frac{2n\pi}{p}\cos\frac{2n\pi}{p}xdx = -\frac{2n\pi}{p}a_n.
\end{aligned}$$

This shows that we may differentiate the Fourier series term by term.

Proposition 4.5.2. *Suppose $f(x)$ is a periodic function of period p that is continuous and piecewise continuously differentiable. If the Fourier coefficients of $f(x)$ are a_n, b_n, then the Fourier series of $f'(x)$ is*

$$f'(x) \sim \frac{2\pi}{p}\sum_{n=1}^{\infty} n\left(b_n\cos\frac{2n\pi}{p}x - a_n\sin\frac{2n\pi}{p}x\right).$$

The integration $F(x) = \displaystyle\int_0^x f(t)dt$ of a periodic function of period p is still periodic if $a_0 = \dfrac{1}{p}\displaystyle\int_0^p f(x)dx = 0$. Then for $n \neq 0$, the other Fourier coefficients of

$F(x)$ are

$$\begin{aligned}
A_n &= \frac{2}{p}\int_0^p F(x)\cos\frac{2n\pi}{p}xdx = \frac{1}{n\pi}\int_0^p F(x)d\sin\frac{2n\pi}{p}x \\
&= -\frac{1}{n\pi}\int_0^p f(x)\sin\frac{2n\pi}{p}xdx = -\frac{p}{2n\pi}b_n, \\
B_n &= \frac{2}{p}\int_0^p F(x)\sin\frac{2n\pi}{p}xdx = -\frac{1}{n\pi}\int_0^p F(x)d\cos\frac{2n\pi}{p}x \\
&= -\frac{1}{n\pi}\left(\int_0^p f(x)dx - \int_0^p f(x)\cos\frac{2n\pi}{p}xdx\right) = \frac{p}{2n\pi}a_n.
\end{aligned}$$

The 0-th coefficient is

$$A_0 = \frac{1}{p}\int_0^p F(x)dx = \frac{1}{p}\left(xF(x)|_{x=0}^{x=p} - \int_0^p xf(x)dx\right) = -\frac{1}{p}\int_0^p xf(x)dx.$$

This shows that, in case $a_0 = 0$, we may almost integrate the Fourier series term by term.

Proposition 4.5.3. *Suppose $f(x)$ is a periodic function of period p. If the Fourier coefficients of $f(x)$ are a_n, b_n and $a_0 = 0$, then the Fourier series of $F(x) = \int_0^x f(t)dt$ is*

$$\int_0^x f(t)dt \sim A_0 + \frac{p}{2\pi}\sum_{n=1}^{\infty}\frac{1}{n}\left(-b_n\cos\frac{2n\pi}{p}x + a_n\sin\frac{2n\pi}{p}x\right).$$

Example 4.5.10. In Example 4.5.4, we found the Fourier series

$$x \sim \frac{1}{2} - \frac{1}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}\sin 2n\pi x, \quad x \in (0,1).$$

We wish to integrate to get the Fourier series of x^2 on $(0,1)$. However, to satisfy the condition of Proposition 4.5.3, we should consider the Fourier series of $x - \frac{1}{2}$, which has vanishing 0-th coefficient. Then $F(x) = \int_0^x \left(t - \frac{1}{2}\right)dt = \frac{1}{2}x^2 - \frac{1}{2}x$ has Fourier series

$$\frac{1}{2}x^2 - \frac{1}{2}x \sim -\frac{1}{12} + \frac{1}{2\pi^2}\sum_{n=1}^{\infty}\frac{1}{n^2}\cos 2n\pi x, \quad x \in (0,1).$$

Here the 0-th coefficient is

$$A_0 = -\int_0^1 x\left(x - \frac{1}{2}\right)dx = -\frac{1}{12}.$$

Then the Fourier series of $x^2 = 2\left(\frac{1}{2}x^2 - \frac{1}{2}x\right) + x$ on $(0,1)$ is

$$\begin{aligned} x^2 &\sim 2\left(-\frac{1}{12} + \frac{1}{2\pi^2}\sum_{n=1}^{\infty}\frac{1}{n^2}\cos 2n\right) + \left(\frac{1}{2} - \frac{1}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}\sin 2n\pi x\right) \\ &= \frac{1}{3} + \sum_{n=1}^{\infty}\left(\frac{1}{n^2\pi^2}\cos 2n\pi x - \frac{1}{n\pi}\sin 2n\pi x\right), \quad x \in (0,1). \end{aligned}$$

We obtained this Fourier series in Example 4.5.6 by direct computation.

Exercise 4.5.20. Derive the Fourier series of x^3 on $(0,1)$ from the Fourier series of x^2.

Exercise 4.5.21. Derive the Fourier series of $|x|$ on $(-1,1)$ from the Fourier series of its derivative.

Exercise 4.5.22. Suppose $f(x)$ is continuously differentiable on $[0,p]$, with perhaps different $f(0^+)$ and $f(p^-)$. Then we can extend both $f(x)$ and $f'(x)$ to periodic functions of period p. If the Fourier coefficients of the extended $f(x)$ are a_n, b_n, prove that

$$f'(x) \sim \frac{f(p^-) - f(0^+)}{p} + \frac{2}{p}\sum_{n=1}^{\infty}\left((f(p^-) - f(0^+) + n\pi b_n)\cos\frac{2n\pi}{p}x + n\pi a_n \sin\frac{2n\pi}{p}x\right).$$

4.5.4 Sum of Fourier Series

Like the Taylor series, the Fourier series may not always converge to the function. The following is one good case when the Fourier series converges.

Theorem 4.5.4. *Suppose $f(x)$ is a periodic function. If $f(x)$ has one sided limits $f(x_0^-)$ and $f(x_0^+)$ at x_0, and there is M, such that*

$$x < x_0 \text{ and close to } x_0 \implies |f(x) - f(x_0^-)| \le M|x - x_0|,$$

and

$$x > x_0 \text{ and close to } x_0 \implies |f(x) - f(x_0^+)| \le M|x - x_0|.$$

Then the Fourier series of $f(x)$ converges to $\dfrac{f(x_0^+) + f(x_0^-)}{2}$ at $x = x_0$.

The condition means that the value of $f(x)$ lies in two "corners" on the two sides of x_0.

Example 4.5.11. In Example 4.5.4, we find the Fourier series of x on $(0,1)$. The function satisfies the condition of Theorem 4.5.4 at $x_0 = 0$. We conclude that

$$\frac{0+1}{2} = \frac{f(0^-) + f(0^+)}{2} = \frac{1}{2} - \frac{1}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}\sin 2n\pi 0.$$

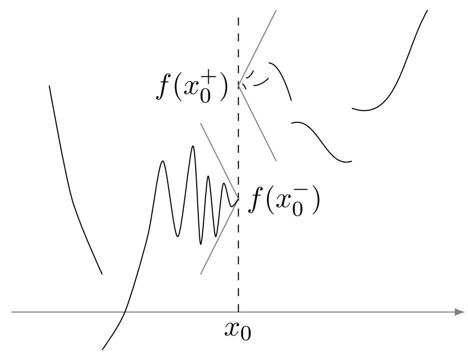


Figure 4.5.1: Condition for the convergence of Fourier series.

This is trivially true. We get similar trivial equalities at $x_0 = 1$ and $x_0 = \dfrac{1}{2}$.

If we take $x_0 = \dfrac{1}{3}$, then

$$\begin{aligned}\frac{1}{3} &= \frac{1}{2} - \frac{1}{\pi}\sum_{k=1}^{\infty}\left(\frac{1}{3k+1}\sin\frac{2(3k+1)\pi}{3} + \frac{1}{3k+2}\sin\frac{2(3k+2)\pi}{3} + \frac{1}{3k+3}\sin\frac{2(3k+3)\pi}{3}\right)\\ &= \frac{1}{2} - \frac{1}{\pi}\sum_{k=1}^{\infty}\left(\frac{1}{3k+1}\frac{\sqrt{3}}{2} - \frac{1}{3k+2}\frac{\sqrt{3}}{2}\right).\end{aligned}$$

This means that

$$1 - \frac{1}{2} + \frac{1}{4} - \frac{1}{5} + \frac{1}{7} - \frac{1}{8} + \frac{1}{10} - \frac{1}{11} + \frac{1}{13} - \frac{1}{14} + \cdots = \frac{\pi}{3\sqrt{3}}.$$

Similarly, by evaluating the Fourier series at $x_0 = \dfrac{1}{4}$ and $x_0 = \dfrac{1}{8}$, we get

$$\begin{aligned}1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \frac{1}{13} - \frac{1}{15} + \frac{1}{17} - \frac{1}{19} + \cdots &= \frac{\pi}{4},\\ 1 + \frac{1}{3} - \frac{1}{5} - \frac{1}{7} + \frac{1}{9} + \frac{1}{11} - \frac{1}{13} - \frac{1}{15} + \frac{1}{17} + \frac{1}{19} - \cdots &= \frac{\pi}{2\sqrt{2}}.\end{aligned}$$

Example 4.5.12. The periodic function of period 2 given by $|x|$ on $(-1, 1)$ satisfies the condition of Theorem 4.5.4 everywhere. Evaluating the Fourier series in Example 4.5.4 at $x = 0$, we get

$$0 = \frac{1}{2} - \frac{4}{\pi^2}\sum_{n=0}^{\infty}\frac{1}{(2n+1)^2}.$$

This means that

$$\sum_{n=0}^{\infty}\frac{1}{(2n+1)^2} = 1 + \frac{1}{3^2} + \frac{1}{5^2} + \frac{1}{7^2} + \cdots = \frac{\pi^2}{8}.$$

If the sum also includes the even terms, then we have

$$\sum_{n=1}^{\infty}\frac{1}{n^2} = \sum_{n=0}^{\infty}\frac{1}{(2n+1)^2} + \sum_{n=1}^{\infty}\frac{1}{(2n)^2} = \sum_{n=0}^{\infty}\frac{1}{(2n+1)^2} + \frac{1}{4}\sum_{n=1}^{\infty}\frac{1}{n^2}.$$

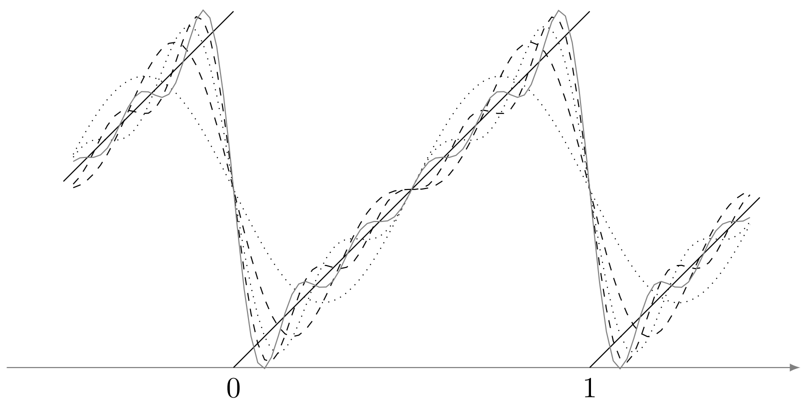


Figure 4.5.2: Partial sums of the Fourier series for x on $(0,1)$.

Therefore

$$\sum_{n=1}^{\infty}\frac{1}{n^2}=1+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{4^2}+\cdots=\frac{4}{3}\sum_{n=0}^{\infty}\frac{1}{(2n+1)^2}=\frac{\pi^2}{6}.$$

If we evaluate at $x=\frac{1}{4}$, then we get

$$\frac{1}{4}=\frac{1}{2}-\frac{4}{\sqrt{2}\pi^2}\left(\frac{1}{1^2}+\frac{1}{3^2}-\frac{1}{5^2}-\frac{1}{7^2}+\frac{1}{9^2}+\frac{1}{11^2}-\frac{1}{13^2}-\frac{1}{15^2}+\cdots\right).$$

Combined with the sum above, we get

$$\frac{1}{1^2}+\frac{1}{3^2}+\frac{1}{9^2}+\frac{1}{11^2}+\frac{1}{17^2}+\frac{1}{19^2}+\cdots=\left(\frac{1}{12}+\frac{1}{16\sqrt{2}}\right)\pi^2,$$
$$\frac{1}{5^2}+\frac{1}{7^2}+\frac{1}{13^2}+\frac{1}{15^2}+\frac{1}{21^2}+\frac{1}{23^2}+\cdots=\left(\frac{1}{12}-\frac{1}{16\sqrt{2}}\right)\pi^2.$$

Exercise 4.5.23. Use the Fourier series of x^2 on $(0,1)$ to compute $\sum_{n=1}^{\infty}\frac{1}{n^2}$.

Exercise 4.5.24. Use the Fourier series of x^2 on $(-1,1)$ to get the Fourier series of x^3 and x^4 on $(-1,1)$. Then evaluate the Fourier series of x^4 to get $\sum_{n=1}^{\infty}\frac{1}{n^4}$.

4.5.5 Parseval's Identity

In Example 4.5.1, we had the Fourier series

$$\sin^4 x=\frac{3}{8}-\frac{1}{2}\cos 2x+\frac{1}{8}\cos 4x.$$

Then

$$\int_0^{2\pi}\sin^8 x dx=\int_0^{2\pi}\left(\frac{3}{8}-\frac{1}{2}\cos 2x+\frac{1}{8}\cos 4x\right)^2 dx.$$

We may expand the square on the right. We get square terms such as $\frac{1}{2^2}\cos^2 2x$ and cross terms such as $2\frac{1}{2}\cdot\frac{1}{8}\cos 2x\cos 4x$. The square term has nonzero integral

$$\int_0^{2\pi}\cos^2 2xdx=\pi.$$

The cross terms has vanishing integral

$$\int_0^{2\pi}\cos 2x\cos 4xdx=0.$$

Therefore

$$\int_0^{2\pi}\sin^8 xdx=\left(\frac{3}{8}\right)^2 2\pi+\left(\frac{1}{2}\right)^2\pi+\left(\frac{1}{8}\right)^2\pi=\frac{35}{64}\pi.$$

The idea (which is essentially Pythagorean theorem) leads to the following formula.

Theorem 4.5.5 (Parseval's Identity). *Suppose $f(x)$ is a periodic function of period p. Then its Fourier coefficients satisfy*

$$2a_0^2+\sum_{n=1}^{\infty}(a_n^2+b_n^2)=\frac{2}{p}\int_0^p|f(x)|^2dx.$$

The identity means that the Fourier series, considered as a conversion between periodic functions and sequences of numbers, preserves the "Euclidean length". The complex form of Parseval's identity is

$$\sum_{n=-\infty}^{\infty}|c_n|^2=\frac{1}{2\pi}\int_0^{2\pi}|f(x)|^2dx.$$

Example 4.5.13. Applying Parseval's identity to the Fourier series of x on $(0,1)$, we get

$$2+\sum_{n=1}^{\infty}\frac{1}{n^2\pi^2}=2\int_0^1 x^2dx=\frac{2}{3}.$$

This is the same as

$$\frac{1}{1^2}+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{4^2}+\cdots=\frac{\pi^2}{6}.$$

Applying the identity to the Fourier series of x^2 on $(0,1)$, we get

$$\frac{2}{9}+\sum_{n=1}^{\infty}\frac{1}{n^4\pi^4}+\sum_{n=1}^{\infty}\frac{1}{n^2\pi^2}=2\int_0^1 x^4dx=\frac{2}{5}.$$

Using $\sum_{n=1}^{\infty}\frac{1}{n^2}=\frac{\pi^2}{6}$, we get

$$\frac{1}{1^4}+\frac{1}{2^4}+\frac{1}{3^4}+\frac{1}{4^4}+\cdots=\frac{\pi^4}{90}.$$

Example 4.5.14. Applying Parseval's identity to the complex form of the Fourier series of e^x on $(0, 2\pi)$ in Example 4.5.8, we get

$$\frac{(e^{2\pi}-1)^2}{4\pi^2}\sum_{n=-\infty}^{\infty}\frac{1}{1+n^2}=\frac{1}{2\pi}\int_0^{2\pi}e^{2x}dx=\frac{e^{4\pi}-1}{4\pi}.$$

The equality leads to

$$1+2\sum_{n=1}^{\infty}\frac{1}{1+n^2}=\frac{e^{2\pi}+1}{e^{2\pi}-1}\pi,$$

or

$$\sum_{n=1}^{\infty}\frac{1}{1+n^2}=\frac{1}{2}+\frac{1}{5}+\frac{1}{10}+\frac{1}{17}+\frac{1}{26}+\cdots=\frac{\pi(e^{2\pi}+1)}{2(e^{2\pi}-1)}-\frac{1}{2}.$$

Exercise 4.5.25. Apply Parseval's identity to the Fourier series in Example 4.5.2 to find $\sum_{n=1}^{\infty}\frac{\sin^2 na}{n^2}$ and $\sum_{n=1}^{\infty}\frac{\cos^2 na}{n^2}$.

Exercise 4.5.26. Apply Parseval's identiy to the Fourier series of x^3 and x^4 on $(0, 1)$ to find $\sum_{n=1}^{\infty}\frac{1}{n^6}$ and $\sum_{n=1}^{\infty}\frac{1}{n^8}$.

Exercise 4.5.27. Find $\sum_{n=1}^{\infty}\frac{1}{1+n^2}$ by evaluating the Fourier series in Example 4.5.8.