

Calculus

Rigor, Concision, Clarity

Jishan Hu, Jian-Shu Li, Wei-Ping Li, Min Yan

Department of Mathematics
The Hong Kong University of Science and Technology

Contents

<i>1</i>	<i>Real Numbers and Functions</i>	<i>1</i>
1.1	Real Number System and Inequalities	1
1.1.1	Real Number System	1
1.1.2	Inequalities	3
1.2	Functions	6
1.3	Coordinate Plane and Graphs of Functions	11
1.4	Summary	18
<i>2</i>	<i>Limit and Continuity</i>	<i>19</i>
2.1	Limits of Sequences	19
2.1.1	Concept of Limits and Properties	19
2.1.2	Rigorous Definition of Limit	26
2.2	Limits of Functions	37
2.2.1	Concept of Limit and Properties	37
2.2.2	Rigorous Definition of Limit	44
2.2.3	Infinite Limit	50
2.3	Continuous Functions	54
2.4	Summary	62
<i>3</i>	<i>Differentiation</i>	<i>65</i>

3.1	<i>Derivative</i>	65
3.2	<i>Computation of Derivative</i>	71
3.3	<i>Differential</i>	80
3.3.1	<i>High Order Approximation</i>	83
3.4	<i>Summary</i>	87
4	<i>Applications of Differentiation</i>	89
4.1	<i>Maximum and Minimum</i>	89
4.2	<i>Mean Value Theorem</i>	95
4.3	<i>L'Hospital's Rule</i>	100
4.4	<i>Increasing and Decreasing</i>	104
4.5	<i>Convexity</i>	112
4.6	<i>Summary</i>	119
5	<i>Integration</i>	123
5.1	<i>Riemann Integration</i>	123
5.2	<i>Properties of Integration</i>	130
5.3	<i>Fundamental Theorem of Calculus</i>	133
5.4	<i>Antiderivative</i>	142
5.5	<i>Summary</i>	151
6	<i>Topics of Integration</i>	153
6.1	<i>Integration of Rational Functions</i>	153
6.2	<i>Numerical Integration</i>	162
6.3	<i>Improper Integral</i>	169
6.3.1	<i>Integral on Unbounded Interval</i>	169
6.3.2	<i>Integral of Unbounded Function</i>	176
6.4	<i>Summary</i>	181
7	<i>Applications of Integration</i>	185
7.1	<i>Applications in Geometry</i>	185
7.1.1	<i>Arc Length</i>	185
7.1.2	<i>Area of Surface of Revolution</i>	192
7.1.3	<i>Volume of Solid of Revolution</i>	196
7.1.4	<i>Cavalieri's Principle</i>	198
7.2	<i>Polar Coordinates</i>	204
7.2.1	<i>Curves in Polar Coordinates</i>	204
7.2.2	<i>Length and Area in Polar Coordinates</i>	209

7.3	<i>Physical Applications</i>	214
7.3.1	<i>Center of Mass</i>	214
7.3.2	<i>Work</i>	219
7.4	<i>Summary</i>	222
8	<i>Infinite Series</i>	225
8.1	<i>Series</i>	225
8.1.1	<i>Sum of Series</i>	225
8.1.2	<i>Comparison Test</i>	228
8.1.3	<i>Integral Comparison Test</i>	232
8.1.4	<i>Absolute and Conditional Convergence</i>	233
8.1.5	<i>Rearrangement</i>	236
8.2	<i>Power Series</i>	239
8.2.1	<i>Taylor Series</i>	239
8.2.2	<i>Power Series</i>	242
8.2.3	<i>Operation of Power Series</i>	244
8.3	<i>Fourier Series</i>	247
8.3.1	<i>Fourier Coefficients</i>	247
8.3.2	<i>Convergence of Fourier Series</i>	250
8.4	<i>Summary</i>	252

1

Real Numbers and Functions

1.1 REAL NUMBER SYSTEM AND INEQUALITIES

In this section, we shall study the real number system, which is the platform for real analysis, and how to solve inequalities.

1.1.1 Real Number System

We start with the set of all integers, $\mathbb{Z}$. It contains all positive integers, negative integers and zero, i.e.,

$$\mathbb{Z} = \{n \mid n = 0, \pm 1, \pm 2, \dots\}.$$

In the set $\mathbb{Z}$, we know that there are a few algebraic operations allowed. These are addition, subtraction and multiplication. Division is generally not allowed in $\mathbb{Z}$, since it will generate fractions. Besides the three algebraic operations, another property on $\mathbb{Z}$ is that any two numbers in $\mathbb{Z}$ can be ordered. In other words, if $a, b \in \mathbb{Z}$ with $a \neq b$, then

$$\text{either } a < b \text{ or } b < a.$$

Because of the restriction of not allowing division, it is convenient to expand the set $\mathbb{Z}$ to the set of all rational numbers:

$$\mathbb{Q} = \{r \mid r = p/q, \text{ where } p, q \in \mathbb{Z} \text{ with } q \neq 0\}.$$

Thus, a rational number is a quotient of two integers, with the denominator not being zero. Obviously, any integer n is also a rational number, since n can be expressed as $\frac{n}{1}$. Hence,

$$\mathbb{Z} \subset \mathbb{Q}.$$

The order relation between two rational numbers also holds in $\mathbb{Q}$.

However, there are many interesting numbers not contained in $\mathbb{Q}$. For instance, the length of the diagonal of the square with side length 1 equals $\sqrt{2}$. This number $\sqrt{2}$ is not a rational number. Another interesting number is π , which is the ratio of the circumference of a circle to the diameter. Therefore the number system $\mathbb{Q}$ is a bit small. We would like to extend $\mathbb{Q}$ to a bigger number system. In fact, there is a bigger number system $\mathbb{R}$, called the the field of real numbers, containing numbers like $\sqrt{2}$ and π . $\mathbb{R}$ has many similar properties as $\mathbb{Q}$. $\mathbb{R}$ has algebraic operations such as addition, subtraction, multiplication and division. Numbers in $\mathbb{R}$ but not rational are called irrational numbers. The set $\mathbb{R}$ is also ordered. We do not discuss in detail the construction of real numbers here, since it is beyond the scope of the book.

Thus, we have three number systems, $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$, satisfying

$$\mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}.$$

Besides the obvious difference that there are irrational numbers which are in $\mathbb{R}$, but not in $\mathbb{Q}$, there is a deeper difference between these two number systems.

In order to understand the difference, we need to introduce the concept of the least upper bound and the greatest lower bound of a subset of $\mathbb{R}$.

Definition 1.1.1 *Given a subset S of $\mathbb{Q}$ or $\mathbb{R}$, we say the set S is bounded above, if there exists a number u of $\mathbb{Q}$ ($\mathbb{R}$ respectively) such that every number in S is no larger than u :*

$$s \leq u, \text{ for any } s \in S.$$

*Such u is called an **upper bound** of S . The smallest upper bound of S , if it exists, is called the **least upper bound**.*

Similarly one can define the concept of **lower bound** and the **greatest lower bound**.

Here is an equivalent way to understand the least upper bound. We say that ℓ is a least upper bound of S in $\mathbb{Q}$ ($\mathbb{R}$ respectively) if ℓ is an upper bound of S and for any $b < \ell$ and $b \in \mathbb{Q}$ ($\mathbb{R}$ respectively), b is not an upper bound of S .

For instance, the numbers 1, 2, 3 are all upper bounds of the set $(0, 1)$ in $\mathbb{R}$. The number 1 is the least upper bound of the set.

The fundamental difference between $\mathbb{Q}$ and $\mathbb{R}$ in the context of analysis is that a bounded set in $\mathbb{R}$ has the least upper bound in $\mathbb{R}$, but a bounded set in $\mathbb{Q}$ may not have the least upper bound in $\mathbb{Q}$. This can be shown by

$$S = \{q \in \mathbb{Q} \mid q^2 \leq 2\} \subset \mathbb{Q},$$

and

$$T = \{r \in \mathbb{R} \mid r^2 \leq 2\} \subset \mathbb{R}.$$

In fact, we have the following general theorem, which will not be proved here.

Theorem 1.1.2 *Let S be a subset of $\mathbb{R}$ bounded above. Then S has the least upper bound.*

We say that the real number system has the **least upper bound property**. In this course, we will use this theorem without proof.

From the previous example, we see that the number system $\mathbb{Q}$ does not have this property.

We use the *coordinate line* (or real line) to represent the real number system. The coordinate line is a line with a distinguished point O , called the *origin*, and an right arrow on the right end indicating the positive direction of the line. A point on the real line is represented by a real number. Points on the left (right resp.) side of the origin represent negative (positive resp.) real numbers. The origin represents the zero 0. These are shown in Fig. 1.1.

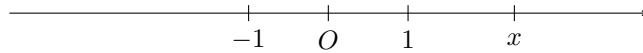


Fig. 1.1 The real line.

1.1.2 Inequalities

Given two real numbers a and b with $a < b$, the set

$$(a, b) = \{x \in \mathbb{R} \mid a < x < b\}$$

is called an *open interval*. The set

$$[a, b] = \{x \in \mathbb{R} \mid a \leq x \leq b\}$$

is called a *closed interval*. There are other kinds of intervals:

$$\begin{aligned} (a, b] &= \{x \in \mathbb{R} \mid a < x \leq b\}, & [a, b) &= \{x \in \mathbb{R} \mid a \leq x < b\}, \\ (a, +\infty) &= \{x \in \mathbb{R} \mid a < x\}, & [a, +\infty) &= \{x \in \mathbb{R} \mid a \leq x\}, \\ (-\infty, b) &= \{x \in \mathbb{R} \mid x < b\}, & (-\infty, b] &= \{x \in \mathbb{R} \mid x \leq b\}. \end{aligned}$$

The whole real line $(-\infty, +\infty) = \mathbb{R}$.

These intervals are special cases of subsets of $\mathbb{R}$. There are operations on subsets: **union**, **intersection** and **complement**. Given two subsets S and T of $\mathbb{R}$, the union $S \cup T$ is defined by

$$S \cup T = \{x \in \mathbb{R} \mid x \in S \text{ or } x \in T\},$$

the intersection $S \cap T$ is defined by

$$S \cap T = \{x \in \mathbb{R} \mid x \in S \text{ and } x \in T\},$$

and the complement S^c is defined by

$$S^c = \{x \in \mathbb{R} \mid x \notin S\}.$$

Here is the intuitive way to explain the words “open” and “closed”. The interval (a, b) is open since all the inequalities are all strict inequalities. The interval $[a, b]$ is closed since all the inequalities are either $\leq$ or $\geq$.

The precise definition of an open set in $\mathbb{R}$ can be stated as follows. A set $S \subset \mathbb{R}$ is **open** if for any point $p \in S$, there exists $(a, b) \subset S$ containing p . A set S is **closed** if the complement of S is open.

There is an subset of $\mathbb{R}$, called the **empty set** $\emptyset$, which does not contain any elements of $\mathbb{R}$. For example, the intersection $(-1, 0) \cap (0, 1) = \emptyset$.

According to the definition, $\mathbb{R}$ is open and $\emptyset$ is closed since $\emptyset$ is the complement of $\mathbb{R}$. The empty set $\emptyset$ is also open according to the definition. In fact, $\mathbb{R}$ and $\emptyset$ are the only two sets that are both closed and open.

Let us recall some properties of inequalities.

1. If $a < b$, then $a + c < b + c$ for any $c \in \mathbb{R}$;
2. If $a < b$ and $c < d$, then $a + c < b + d$;
3. If $a < b$ and $c > 0$, then $ac < bc$.

Example 1.1.1 The inequality $-2x + 3 < 3x + 8$ can be solved by moving terms around to have an equivalent inequality

$$-5x < 5.$$

Multiply both sides by $-1/5$ and switch the direction of the inequality, we get the solution

$$x > -1. \quad \blacksquare$$

Example 1.1.2 To solve the inequality $x^2 - 3x + 2 > 0$, we factorize the polynomial first:

$$x^2 - 3x + 2 = (x - 1)(x - 2).$$

The inequality is equivalent to the fact that $x - 1$ and $x - 2$ have the same signs, i.e.,

$$x - 1 > 0 \quad \text{and} \quad x - 2 > 0; \quad \text{or} \quad x - 1 < 0 \quad \text{and} \quad x - 2 < 0.$$

We obtain

$$x > 2 \quad \text{or} \quad x < 1.$$

A more intuitive way to solve the inequality is as follows.

One should observe that the monomials $x - 1$ and $x - 2$ change sign from negative to positive at 1 and 2, respectively. Since the roots 1 and 2 of the polynomial $x^2 - 3x + 2 = (x - 1)(x - 2)$ divides the real line into three open intervals

$$(-\infty, 1), \quad (1, 2), \quad \text{and} \quad (2, +\infty),$$

we see that this polynomial changes sign when x moves from one interval to an adjacent one, as demonstrated in Fig. 1.2. In other words, the polynomial $(x - 1)(x - 2)$ takes the positive sign in the interval $(-\infty, 1)$, subsequently negative sign in $(1, 2)$ and positive sign in $(2, +\infty)$. Therefore the solution for the original inequality is that x lies in the interval $(-\infty, 1)$ or the interval $(2, +\infty)$. Equivalently, one gets

$$x < 1, \quad \text{or} \quad x > 2. \quad \blacksquare$$

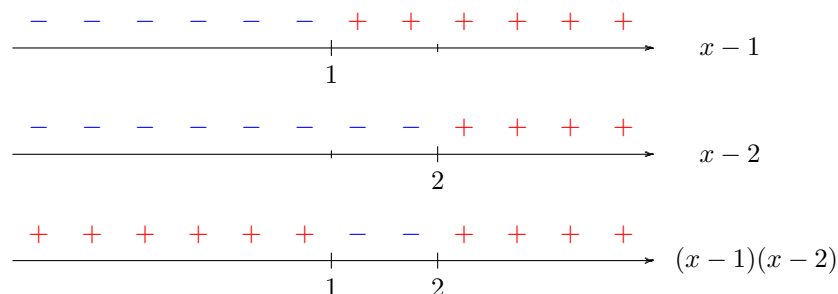


Fig. 1.2 Signs of the polynomials $x-1$, $x-2$, and $(x-1)(x-2)$.

Example 1.1.3 The inequality $\frac{x-2}{x-1} < 0$ can be solved by using the same idea as above to get $1 < x < 2$. ■

Example 1.1.4 To solve the inequality $x^3 - 3x^2 > -2x$, we first change it to an equivalent inequality

$$x(x-1)(x-2) > 0.$$

The roots of the polynomial are 0, 1 and 2. They divide the real line into intervals

$$(-\infty, 0), \quad (0, 1), \quad (1, 2) \quad \text{and} \quad (2, +\infty).$$

The polynomial $x(x-1)(x-2)$ takes negative sign when x is very negative. Hence $x(x-1)(x-2)$ is negative in $(-\infty, 0)$, positive in $(0, 1)$, negative in $(1, 2)$, and positive in $(2, +\infty)$, as shown in Fig. 1.3. The solution for the inequality is that x is in $(0, 1)$ or $(2, +\infty)$. Equivalently

$$0 < x < 1 \quad \text{and} \quad x > 2. \quad \blacksquare$$



Fig. 1.3 Signs of the polynomial $x(x-1)(x-2)$.

Exercises

1.1.1 Show $\sqrt{2}$ is an irrational number.

1.1.2 Show $\sqrt{3}$ is an irrational number.

1.1.3 Solve the following inequalities:

1. $x^2 - 3x - 4 < 0$;

2. $|x^2 - 2x| < 3$;

3. $|x^2 - 2x| > 1$;

4. $x^3 + 4x > 4x^2$;

5. $(x-1)(x-2)(x-3)(x-4) > 0$;

6. $(x-1)(x-2)^2(x-3)^3(x-4)^4 < 0$;

7. $\frac{1}{x-1} + \frac{2}{2x+1} > 1$;

8. $\left(\frac{1}{x-1} - \frac{1}{x+1}\right)^2 > 1$.

1.2 FUNCTIONS

Given a subset D of $\mathbb{R}$, a function is a rule that assigns a real number $f(x)$ to any $x \in D$. For example, the function $f(x) = x^2$ is the rule that assigns the square of x for any x in $D = \mathbb{R}$. The subset D is called the **domain** of the function. The letter x is called the **independent variable**. Thus, the domain of a function is the set of all allowable values of the independent variable of the function. We often write $y = f(x)$ for the function $f(x)$. The letter y is called the **dependent variable**. The set of possible values of dependent variable is called the **codomain**. The **range** of a function is the set of all actual values the function may return corresponding to the domain. In other words, the range is the set $f(D)$. Obviously, it is always a subset of the codomain.

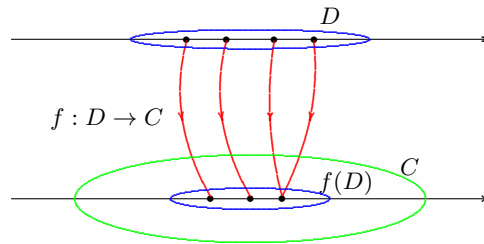


Fig. 1.4 An illustration of a function $f : D \rightarrow C$.

Often we use the notation $f : D \rightarrow C$, where D is the domain, and C is the codomain. This notation implicitly suggests that $f(D) \subset C$.

In many cases, a function has an explicit expression such as $f(x) = x^2$. The function $f(x) = x^2$ can be defined on the whole $\mathbb{R}$ or any subset of $\mathbb{R}$. However there are many functions whose domains are smaller than the whole $\mathbb{R}$. For example, the largest possible domain of $f(x) = \sqrt{x}$ is the interval $[0, +\infty)$. Sometimes, the domain of a function may not be given explicitly. If a function has an explicit expression, we may find the largest possible domain of $f(x)$ in many cases. For example, the largest possible domain of the function $f(x) = \frac{1}{x}$ is the real line with the origin deleted.

Example 1.2.1 The function $f(x) = \sqrt{x^2 - 3x + 2}$ is meaningful when

$$x^2 - 3x + 2 \geq 0.$$

From Example 1.1.2, we get

$$x \leq 1, \quad \text{or} \quad x \geq 2.$$

Thus the domain is the union of two intervals $(-\infty, 1] \cup [2, +\infty)$. ■

Here are examples of functions we will encounter frequently in this book.

(i) Polynomial functions

$$f(x) = a_0x^n + a_1x^{n-1} + \dots + a_{n-1}x + a_n.$$

The largest possible domain for a polynomial function is the real line.

(ii) Rational functions

$$F(x) = \frac{f(x)}{g(x)},$$

where both $f(x)$ and $g(x)$ are polynomials. The largest possible domain for $F(x)$ is the real line with roots of $g(x)$ deleted.

(iii) Algebraic functions are functions obtained from polynomials via addition, multiplication, division and taking roots. For instance, $\sqrt{x+1} + x$ is an algebraic function.

(iv) Transcendental functions stand for those which are not algebraic functions. For example, e^x , $\ln x$, $\sin x$, $\cos x$, $\tan x$ and $\cot x$ are all transcendental functions.

(v) There are also functions expressed in pieces such as

$$f(x) = \begin{cases} 1, & \text{if } x < 0, \\ -1, & \text{if } x > 0, \end{cases} \quad \text{and} \quad g(x) = \begin{cases} -x^2, & \text{if } x < -1, \\ x, & \text{if } -1 \leq x \leq 1, \\ x^3, & \text{if } x > 1. \end{cases}$$

Here are means to generate more varieties of functions.

Given a function f with domain A and a function g with the domain B , there is a function h defined by

$$h(x) = f(x) + g(x).$$

The domain of h is the intersection $A \cap B$. We usually denote the function h by $f + g$ and write

$$(f + g)(x) = f(x) + g(x).$$

Similarly, we can define a new function $f \cdot g$ by

$$(f \cdot g)(x) = f(x) \cdot g(x)$$

with the domain $A \cap B$; a new function f/g by

$$(f/g)(x) = f(x)/g(x)$$

with the domain

$$A \cap \{x \in B \mid g(x) \neq 0\};$$

and a new function cf by

$$(cf)(x) = cf(x)$$

with the domain A for a constant c . If the range of f is contained in the domain B of g , we can define the **composition** h of g and f by $h(x) = g(f(x))$. Usually we use $g \circ f$ to denote the composition h .

Example 1.2.2 Let us look at some more complicated functions. Suppose $f(x) = e^x$, $g(x) = \sin x$, and $h(x) = x^2$. The composition function $f \circ g$ is

$$f \circ g(x) = f(g(x)) = e^{\sin x}.$$

The composition function $g \circ f$ is

$$g \circ f(x) = \sin e^x.$$

We see that $f \circ g \neq g \circ f$. The composition function $h \circ (f \circ g)$ is

$$h \circ (f \circ g)(x) = h(f \circ g(x)) = (e^{\sin x})^2 = e^{2 \sin x}.$$

The composition function $(f \circ g) \circ h$ is

$$(f \circ g) \circ h(x) = f \circ g(h(x)) = e^{\sin x^2}. \quad \blacksquare$$

Inverse Functions

A function $f : D \rightarrow C$, with the domain D and the codomain C , is called **one-to-one**, or **injective**, if

$$f(x_1) \neq f(x_2) \quad \text{whenever} \quad x_1 \neq x_2.$$

The term “one-to-one” is in contrast to “many-to-one”.

A function $f : D \rightarrow C$ is called from D **onto** C , or **surjective**, if for every $y \in C$, there exists at least one $x \in D$ such that $f(x) = y$. In fact, f is surjective iff (if and only if) C is the range of f , that is, $C = f(D)$.

If $f : D \rightarrow C$ is injective as well as surjective, then it is called **bijective**.

Let $f : D \rightarrow R$ be a bijective function.¹ We can define a new function g whose domain is R as follows. For each $r \in R$, there exists a unique d in D such that $f(d) = r$. We define a new function g so that $g(r) = d$. The function g has the property that

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

and

¹When we write $f : D \rightarrow R$, we implicitly mean that the codomain R is the range $f(D)$ (R for “range”). If we do not intend to emphasize this distinction, we will use the letter C in general. One can always write the codomain of any function to be $\mathbb{R} = (-\infty, +\infty)$, since it contains all real numbers. When the domain of a function is not shown explicitly, it is usually the set where the function is defined. For example, for $f(x) = \sqrt{x-1}$, the domain is $x \geq 1$.

$$g \circ f(x) = x, \quad \text{for any } x \in D.$$

The function g is called the **inverse function** of f , sometimes denoted by f^{-1} .

Example 1.2.3 Consider $f(x) = \frac{ax+b}{cx+d}$, with $ad \neq bc$.²

Let us first assume that $c \neq 0$. It is easy to see that the domain of the function f is $\mathbb{R} \setminus \{-d/c\}$. However, it is not so obvious what the range of f is.

The inverse function of f can be found as follows. First, by writing $y = \frac{ax+b}{cx+d}$ and solving for x to have

$$x = -\frac{dy-b}{cy-a}.$$

We switch the symbols of the dependent and independent variables to have the inverse function

$$g(x) = -\frac{dx-b}{cx-a},$$

whose domain is $\mathbb{R} \setminus \{a/c\}$.

This suggests that the function $f(x) = \frac{ax+b}{cx+d}$ is bijective from $\mathbb{R} \setminus \{-d/c\}$ to $\mathbb{R} \setminus \{a/c\}$. In fact, we can show that it is both injective and surjective.

1. Injective. If $x_1 \neq x_2$, we have

$$ad(x_2 - x_1) \neq bc(x_2 - x_1),$$

which implies

$$(ax_1 + b)(cx_2 + d) \neq (ax_2 + b)(cx_1 + d).$$

Since $cx_1 + d \neq 0$ and $cx_2 + d \neq 0$, we have

$$\frac{ax_1 + b}{cx_1 + d} \neq \frac{ax_2 + b}{cx_2 + d}.$$

That is, the function f is injective.

2. Surjective. For any $y \in \mathbb{R} \setminus \{a/c\}$, take $x = -\frac{dy-b}{cy-a}$. Then for such x , we have $y = \frac{ax+b}{cx+d}$.

This means that the function f is surjective.

Now let us examine the compositions of f and g . Elementary calculations give

$$f \circ g(x) = \frac{a\left(-\frac{dx-b}{cx-a}\right) + b}{c\left(-\frac{dx-b}{cx-a}\right) + d} = \frac{(-a)(dx-b) + b(cx-a)}{(-c)(dx-b) + d(cx-a)} = x, \quad \text{for any } x \in \mathbb{R} \setminus \{a/c\};$$

²If $ad = bc$, then the function is a constant.

and

$$g \circ f(x) = -\frac{d\left(\frac{ax+b}{cx+d}\right) - b}{c\left(\frac{ax+b}{cx+d}\right) - a} = -\frac{d(ax+b) - b(cx+d)}{c(ax+b) - a(cx+d)} = x, \quad \text{for any } x \in \mathbb{R} \setminus \{-d/c\}.$$

If $c = 0$, then $a, d \neq 0$, since $ad \neq bc$. In this case, it is easy to show that f is a linear function and bijective from $\mathbb{R}$ to $\mathbb{R}$. ■

Example 1.2.4 The domain of the function $f(x) = e^x$ is $D = \mathbb{R}$ and the range is $R = (0, +\infty)$, while the codomain C can be either $[0, +\infty)$ or $\mathbb{R}$. It is bijective from $\mathbb{R}$ to $(0, +\infty)$. The function $g(x) = \ln x : (0, +\infty) \rightarrow \mathbb{R}$ is the inverse of f , since the following two identities hold:

$$f \circ g(x) = e^{\ln x} = x, \quad \text{for any } x \in R = (0, +\infty)$$

and

$$g \circ f(x) = \ln(e^x) = x, \quad \text{for any } x \in D = (-\infty, +\infty).$$

The function $f(x) = x^2 : D = \mathbb{R} \rightarrow [0, +\infty)$ has no inverse function since it is not one-to-one. For instance, $f(-1) = 1 = f(1)$. However the function $h(x) = x^2 : [0, +\infty) \rightarrow [0, +\infty)$ has an inverse $h^{-1}(x) = \sqrt{x}$. ■

The trigonometric functions $\sin x$, $\cos x$, $\tan x$ and $\cot x$ are periodic functions. If we only consider the sine function $\sin x$ over the interval $(-\pi/2, \pi/2)$, then it is one-to-one and hence has an inverse, denoted by $\arcsin x$ defined over $(-1, 1)$. Similarly, the cosine function $\cos x$ defined over $(0, \pi)$ has an inverse function $\arccos x$ over $(-1, 1)$. The function $\tan x$ defined over $(-\pi/2, \pi/2)$ has an inverse $\arctan x$ defined over the whole real line. The function $\cot x$ defined over $(0, \pi)$ has an inverse $\operatorname{arccot} x$ defined over the whole real line.

We say a function is **increasing** if

$$f(x_1) > f(x_2) \quad \text{when } x_1 > x_2;$$

is **decreasing** if

$$f(x_1) < f(x_2) \quad \text{when } x_1 > x_2.$$

Both are called monotonic functions. It is easy to see that any monotonic function is one-to-one and hence has inverse function from the domains to its range.

Example 1.2.5 The function $f(x) = x^3$ is increasing. The function $g(x) = \sin x$ is increasing in the interval $(-\pi/2, \pi/2)$. The function $h(x) = \cos x$ is decreasing over the interval $(0, \pi)$. ■

Exercises

1.2.1 Determine the domains and ranges of the following functions:

$$1. f(x) = \frac{x^2}{1-x};$$

$$2. f(x) = \frac{x}{2-3x};$$

$$3. f(x) = \sqrt{x-x^3};$$

$$4. f(x) = \sqrt{\frac{1+x}{1-x}};$$

$$5. f(x) = \ln(x^2 - 1);$$

$$6. f(x) = \ln(x+1) + \ln(x-1);$$

$$7. f(x) = \sqrt{\sin x};$$

$$8. f(x) = \ln(\sin x).$$

1.2.2 Show that each of the functions is increasing in the specified interval:

$$1. f(x) = x^2 \quad (0 \leq x < +\infty);$$

$$3. f(x) = \tan x \quad \left(-\frac{\pi}{2} < x < \frac{\pi}{2}\right);$$

$$2. f(x) = \sin x \quad \left(-\frac{\pi}{2} \leq x \leq \frac{\pi}{2}\right);$$

$$4. f(x) = 2x + \sin x \quad (-\infty < x < +\infty).$$

1.2.3 Show that each of the functions is decreasing in the specified interval:

$$1. f(x) = x^2 \quad (-\infty < x \leq 0);$$

$$3. f(x) = \cot x \quad (0 < x < \pi);$$

$$2. f(x) = \cos x \quad (0 \leq x \leq \pi);$$

$$4. f(x) = (x-1)(x-2) \quad (-\infty < x \leq 3/2).$$

1.2.4 Determine the region where the function is monotonic:

$$1. f(x) = x^3;$$

$$3. f(x) = \frac{1+x}{1-x};$$

$$2. f(x) = ax^2 + bx + c;$$

$$4. f(x) = x^2 - 4.$$

1.2.5 Find the inverse function:

$$1. f(x) = 2x + 1;$$

$$4. f(x) = \frac{1+x}{1-x} \quad (x \neq 1);$$

$$2. f(x) = x^2 \quad (-\infty < x \leq 0);$$

$$5. f(x) = \sqrt{1-x^2} \quad (-1 \leq x \leq 0);$$

$$3. f(x) = x^2 \quad (0 \leq x < +\infty);$$

$$6. f(x) = \sqrt{1-x^2} \quad (0 \leq x \leq 1).$$

1.3 COORDINATE PLANE AND GRAPHS OF FUNCTIONS

Draw a horizontal line with a right arrow on the right end and a vertical line with an upward arrow on the top end. The two perpendicular lines intersect at one point named the **origin**. The two lines are real lines. The horizontal line is called the ***x*-axis**. The vertical line is called the ***y*-axis**. Each point P on the plane corresponds to a pair (a, b) as follows. Project the point vertically towards

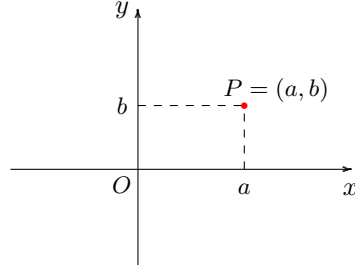
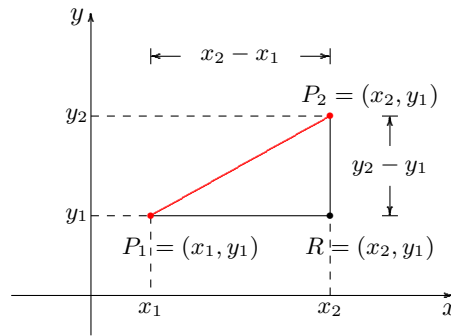


Fig. 1.5 The coordinate plane.

the x -axis to get a point a on the x -axis. Similarly project the point horizontally towards the y -axis to get a point b on the y -axis. The points a and b are called **x -coordinate** and **y -coordinate** respectively. This apparatus is called the **coordinate plane**.

Fig. 1.6 The distance from $P_1 = (x_1, y_1)$ to $P_2 = (x_2, y_2)$.

Given two points $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$. Without loss of generality, assume $x_1 < x_2$ and $y_1 < y_2$. Let $R = (x_2, y_1)$. We get a right triangle P_1P_2R . The edge P_1R has length equal to $x_2 - x_1$. The edge RP_2 has length equal to $y_2 - y_1$. The distance from P_1 to P_2 equals the length of the edge P_1P_2 . Therefore we have the distance formula

$$\text{distance from } P_1 \text{ to } P_2 = \sqrt{(y_2 - y_1)^2 + (x_2 - x_1)^2}.$$

Given a function $y = f(x)$. For each x , we get a point $(x, f(x))$ on the coordinate plane. When x moves in the domain D of f , the point $(x, f(x))$ moves on the plane. The trace of $(x, f(x))$ for $x \in D$ is called the **graph** of f .

Lines and Linear Functions

Example 1.3.1 Consider the linear function $f(x) = ax + b$. We know that the graph of f is a straight line, shown in Fig. 1.7.

When $a = 0 = b$, the graph of the function f is the x -axis, a line.

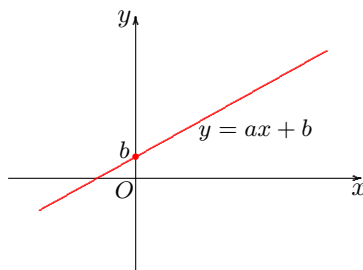
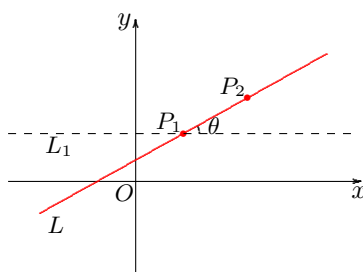
Fig. 1.7 The graph of the straight line $y = ax + b$.

Fig. 1.8 A line is determined by the coordinates of two points.

In the general case, as in Fig. 1.8, given any two distinct points $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ on the graph of f , we have $y_1 = ax_1 + b$ and $y_2 = ax_2 + b$.

$$y_2 - y_1 = (ax_2 + b) - (ax_1 + b) = a(x_2 - x_1), \quad a = \frac{y_2 - y_1}{x_2 - x_1}.$$

Therefore a is determined by (x_1, y_1) and (x_2, y_2) . Subsequently b is also determined by these two points: $b = y_1 - ax_1$. ■

The coefficient a in $f(x) = ax + b$ is called the **slope** of the line and b is the **y -intercept** of the line since the intersection of the L with the y -axis is $(0, b)$.

The slope of a line is a very important concept. In the example above, as shown in Fig. 1.8, let θ be the angle between the line L_1 and the line L . It is easy to see that

$$\tan \theta = \frac{y_2 - y_1}{x_2 - x_1} = a.$$

Given two non-vertical lines L_1 and L_2 . Let a_i be the slope of L_i for $i = 1, 2$. Suppose L_1 and L_2 are perpendicular, as in Fig. 1.9. Let θ_1 be the angle between the x -axis and L_1 and θ_2 be the angle between the x -axis and L_2 . Then

$$\theta_2 - \theta_1 = \pm\pi/2,$$

$$\cos(\theta_2 - \theta_1) = 0,$$

$$\cos \theta_2 \cos \theta_1 + \sin \theta_2 \sin \theta_1 = 0.$$

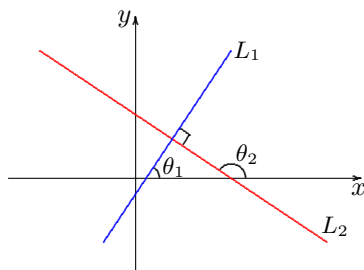


Fig. 1.9 Two perpendicular straight lines.

Therefore we get

$$a_1 \cdot a_2 = \tan \theta_1 \cdot \tan \theta_2 = -1.$$

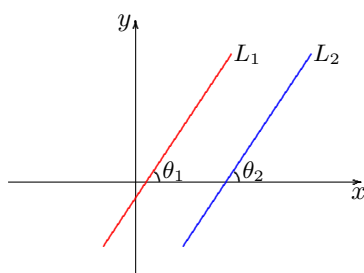
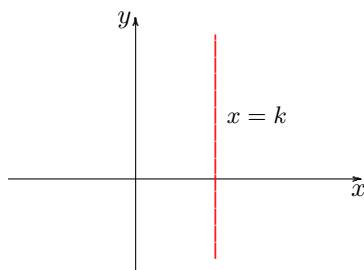


Fig. 1.10 Two parallel straight lines.

Given two non-vertical lines L_1 and L_2 . If they are parallel, as in Fig. 1.10, then the angle between L_1 and the x -axis is equal to the angle between L_2 and the x -axis. Therefore the slopes of two lines are the same:

$$a_1 = a_2.$$

The vertical line can be describe by the equation $x = k$, as shown in Fig. 1.11.

Fig. 1.11 Verticle line $x = k$.

The coordinate plane provides a platform to describe geometric objects such as lines and curves by equations $F(x, y) = 0$ of two variables x and y . For example, the graph of a function $f(x)$ corresponds to the equation $F(x, y) = y - f(x) = 0$, as in Fig. 1.12.

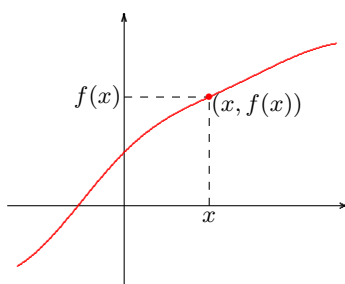


Fig. 1.12 The graph of a function $F(x, y) = y - f(x) = 0$.

Circles, Ellipses, Hyperbolas and Parabolas

The **circle** of radius r centered at the point (a, b) is represented by the equation

$$(x - a)^2 + (y - b)^2 = r^2$$

from the distance formula, shown in Fig. 1.13.

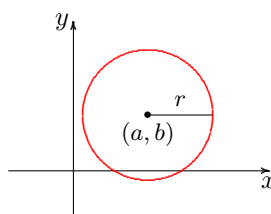


Fig. 1.13 The circle $(x - a)^2 + (y - b)^2 = r^2$.

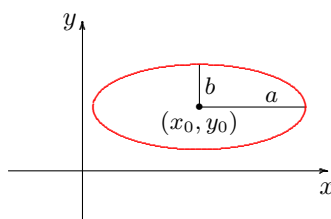


Fig. 1.14 The ellipse $\frac{(x - x_0)^2}{a^2} + \frac{(y - y_0)^2}{b^2} = 1$.

The equation

$$\frac{(x - x_0)^2}{a^2} + \frac{(y - y_0)^2}{b^2} = 1$$

describes a curve called an **ellipse**, shown in Fig. 1.14. The point (x_0, y_0) is called the **center** of the ellipse.

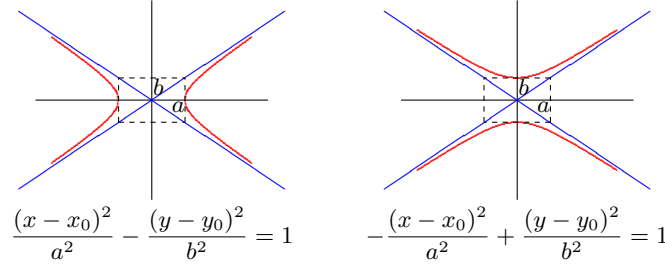


Fig. 1.15 Hyperbolas. Both are centered at (x_0, y_0) .

The equation

$$\frac{(x - x_0)^2}{a^2} - \frac{(y - y_0)^2}{b^2} = 1$$

describes a curve called a **hyperbola**. The equation

$$-\frac{(x - x_0)^2}{a^2} + \frac{(y - y_0)^2}{b^2} = 1$$

is another hyperbola. Both are shown in Fig. 1.15.

The graph of the function

$$f(x) = ax^2$$

is called a **parabola**. The curve passes the point $(0, 0)$ that is called **vertex** of the parabola. It opens upwards or downwards depending upon the sign of a . The equation $ay^2 - x = 0$ provides another kind of parabola opening rightwards if $a > 0$ and leftwards if $a < 0$. These are demonstrated in Fig. 1.16.

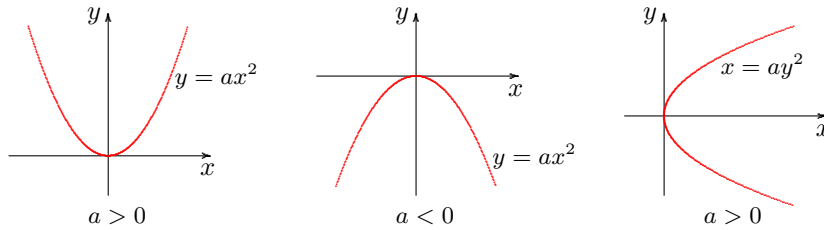


Fig. 1.16 Parabolas.

Let C be the curve corresponding to the equation $F(x, y) = 0$. The curve $F(x - a, y - b) = 0$ is obtained by moving the curve C horizontally by a and vertically by b , equivalently, moving the coordinate plane horizontally by $-a$ and vertically by $-b$, shown in Fig. 1.17. It is called a **translation** of $F(x, y) = 0$.

In Fig. 1.18, the curve $F(-x, y) = 0$ is obtained by reflecting the curve C with respect to y -axis. The curve $F(x, -y) = 0$ is obtained by reflecting the curve C with respect to x -axis.

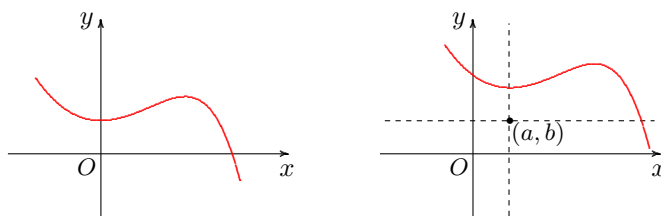


Fig. 1.17 Translation of a curve: from $F(x, y) = 0$ to $F(x - a, y - b) = 0$.

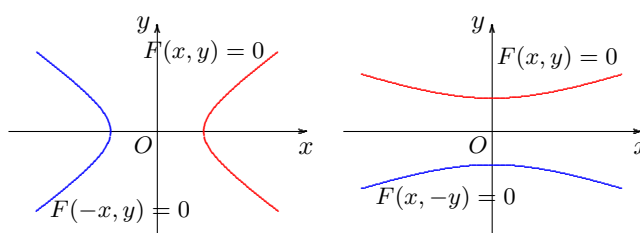


Fig. 1.18 Reflections of a curve $F(x, y) = 0$ with respect to both axes.

Example 1.3.2 To sketch the graph of $f(x) = 2x^2 + 4x + 4$, we rewrite it as

$$y = 2(x + 1)^2 + 2,$$

or

$$y - 2 = 2(x - (-1))^2.$$

The graph of the function f is the parabola $y = 2x^2$ with the vertex moved to $(-1, 2)$, as shown in Fig. 1.19. ■

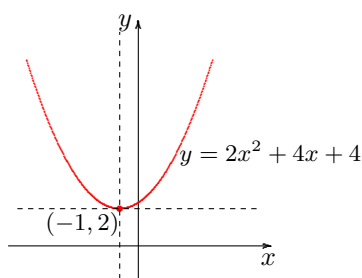


Fig. 1.19 Translation of $y = 2x^2$.

Exercises

1.3.1 Graph the linear function $f(x) = ax$ for $a = 0, 1/2, 1, 2$ and -1 .

1.3.2 Graph the linear function $f(x) = x + b$ for $b = 0, 1, 2$ and -1 .

1.3.3 Graph the linear functions:

(a) $f(x) = 2x + 1$;

(c) $f(x) = x - 2$;

(b) $f(x) = 2 - x$;

(d) $f(x) = -\frac{x}{2} - 1$.

1.3.4 Graph the following equations:

1. $x^2 + y^2 - 4x + 2y + 3 = 0$;

6. $x^2 + 2x + y = 0$;

2. $4x^2 + 4y^2 + 8x - 4y + 3 = 0$;

7. $x^2 - 4x + y + 5 = 0$;

3. $2x^2 + y^2 - 4x + 2y + 2 = 0$;

8. $4y^2 + 4x - 4y + 3 = 0$;

4. $3x^2 + y^2 + 3x + y = 0$;

9. $2x^2 - y^2 - 4x + 2y + 2 = 0$;

5. $x^2 + 2x - y = 0$;

10. $3x^2 - y^2 + 3x + y = 0$.

1.4 SUMMARY

Definitions

- A function $f : D \rightarrow C$, with the domain D and the codomain C , is called **one-to-one**, or **injective**, if

$$f(x_1) \neq f(x_2) \quad \text{whenever} \quad x_1 \neq x_2.$$

A function $f : D \rightarrow C$ is called from D **onto** C , or **surjective**, if for every $y \in C$, there exists at least one $x \in D$ such that $f(x) = y$.

A function $f : D \rightarrow C$ is called **bijective**, if it is injective as well as surjective.

- A function f is **increasing** if $f(x_1) > f(x_2)$ when $x_1 > x_2$; is **decreasing** if $f(x_1) < f(x_2)$ when $x_1 > x_2$.
- Given a subset $S \subset \mathbb{R}$, the set S is bounded above, if there exists a number $u \in \mathbb{R}$ such that every number in S is no larger than u : $s \leq u$, for any $s \in S$. Such u is called an **upper bound** of S . The smallest upper bound of S , if it exists, is called the **least upper bound**.

Theorems

- **Least Upper Bound Property** Let S be a subset of $\mathbb{R}$ bounded above. Then S has the least upper bound.

2

Limit and Continuity

2.1 LIMITS OF SEQUENCES

Limit is one of the most basic concepts in calculus. We start with limit of sequences and later, we will study limit of functions.

2.1.1 Concept of Limits and Properties

We introduce the concept of sequences and their limits, and discuss properties of limits. We will concentrate on the ideas, leaving the rigorous treatment to the later subsection.

A sequence $\{x_n\}$ is

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

We call x_n the n -th term of the sequence. In this course, we will only consider sequence of numbers, which means that x_n are real numbers. Here are some examples

$$\begin{aligned} a_n = n: & \quad 1, 2, 3, \dots, n, \dots; \\ b_n = 2: & \quad 2, 2, 2, \dots, 2, \dots; \\ c_n = \frac{1}{n}: & \quad 1, \frac{1}{2}, \dots, \frac{1}{n}, \dots; \\ d_n = (-1)^n: & \quad 1, -1, 1, \dots, (-1)^n, \dots; \\ e_n = \sin n: & \quad \sin 1, \sin 2, \sin 3, \dots, \sin n, \dots \end{aligned}$$

Note that the index n in a sequence does not have to start from 1. For example, the sequence $\{d_n\}$ actually starts at $n = 0$ (or any even integer).

Let us compare these sequences. When n gets larger, a_n gets larger, b_n is constant, c_n gets closer to zero, d_n and e_n does not approach anything. The sequences $\{b_n\}$ and $\{c_n\}$ sharing the property of approaching a finite number L ($L = 2$ for b_n and $L = 0$ for c_n) when n goes to infinity. We say the sequence converges to the limit L . In contrast, we say the sequences $\{a_n\}$, $\{d_n\}$ and $\{e_n\}$ diverge since they do not approach any finite number when n goes to infinity.

Definition 2.1.1 (Non-rigorous) A sequence $\{a_n\}$ converges to a finite number L if a_n approaches L when n goes to infinity. We write

$$\lim_{n \rightarrow \infty} a_n = L,$$

and call the sequence $\{a_n\}$ **convergent**. We call a sequence **divergent** if it does not approach any finite number when n goes to infinity.

Since the limit describes the behavior when n gets very large, the modification of finitely many terms in a sequence does not change the convergence and the limit value.

Intuitively, we know that if a is close to 3 and b is close to 5, then $a + b$ is close to $3 + 5 = 8$. The intuition leads to the following properties.

Proposition 2.1.2 (Arithmetic Rule) Suppose $\{a_n\}$ and $\{b_n\}$ converge. Then $\{a_n + b_n\}$, $\{ca_n\}$, $\{a_nb_n\}$ converge (c denotes a constant) and

$$\lim_{n \rightarrow \infty} (a_n + b_n) = \lim_{n \rightarrow \infty} a_n + \lim_{n \rightarrow \infty} b_n, \quad \lim_{n \rightarrow \infty} ca_n = c \lim_{n \rightarrow \infty} a_n, \quad \lim_{n \rightarrow \infty} a_nb_n = \lim_{n \rightarrow \infty} a_n \cdot \lim_{n \rightarrow \infty} b_n.$$

Moreover, if $\lim_{n \rightarrow \infty} b_n \neq 0$, then $\left\{\frac{a_n}{b_n}\right\}$ also converges, and

$$\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = \frac{\lim_{n \rightarrow \infty} a_n}{\lim_{n \rightarrow \infty} b_n}.$$

Proposition 2.1.3 If $\{a_n\}$ converges, then $\{|a_n|\}$ also converges and $\lim_{n \rightarrow \infty} |a_n| = \left|\lim_{n \rightarrow \infty} a_n\right|$. On the other hand, if $\lim_{n \rightarrow \infty} |a_n| = 0$, then $\lim_{n \rightarrow \infty} a_n = 0$.

Example 2.1.1 Based on the limits $\lim_{n \rightarrow \infty} c = c$, $\lim_{n \rightarrow \infty} \frac{1}{n} = 0$ and the arithmetic rule, the limit of the sequence $\left\{\frac{2n^2 + n}{n^2 - n + 1}\right\}$ may be computed as follows

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

Example 2.1.2 We have $\lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} = 0$ and $\frac{1}{\sqrt{n}} = \left|\frac{(-1)^n}{\sqrt{n}}\right|$. By Proposition 2.1.3, we get

$$\lim_{n \rightarrow \infty} \frac{(-1)^n}{\sqrt{n}} = 0. \quad \blacksquare$$

The following property is very useful for deriving more sophisticated limits.

Proposition 2.1.4 (Sandwich Rule) *If $a_n \leq b_n \leq c_n$ and $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} c_n = L$, then $\lim_{n \rightarrow \infty} b_n = L$.*

The rule reflects the intuition that if a and c are close to 5, then anything between a and c should also be close to 5.

Example 2.1.3 By

$$-\frac{1}{n} \leq \frac{\sin n}{n} \leq \frac{1}{n}, \quad \lim_{n \rightarrow \infty} \frac{1}{n} = 0,$$

and the sandwich rule, we get $\lim_{n \rightarrow \infty} \frac{\sin n}{n} = 0$. ■

Example 2.1.4 To compute the limit $\lim_{n \rightarrow \infty} (\sqrt{n+2} - \sqrt{n})$, we note that

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

By the sandwich rule and $\lim_{n \rightarrow \infty} \frac{2}{\sqrt{n}} = 2 \lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} = 0$, we get $\lim_{n \rightarrow \infty} (\sqrt{n+2} - \sqrt{n}) = 0$. ■

Example 2.1.5 Let $a > 0$. We show that

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

In the case $a \geq 1$, we let $\alpha_n = \sqrt[n]{a} - 1$. Then

$$a = (1 + \alpha_n)^n = 1 + n\alpha_n + \frac{n(n-1)}{2}\alpha_n^2 + \cdots + \alpha_n^n \geq 1 + n\alpha_n,$$

and we get

$$0 \leq \alpha_n \leq \frac{a-1}{n}.$$

By the sandwich rule and $\lim_{n \rightarrow \infty} \frac{a-1}{n} = (a-1) \lim_{n \rightarrow \infty} \frac{1}{n} = 0$, we get $\lim_{n \rightarrow \infty} \alpha_n = 0$ and

$$\lim_{n \rightarrow \infty} \sqrt[n]{a} = \lim_{n \rightarrow \infty} \alpha_n + 1 = 1.$$

If $0 < a \leq 1$, we have $\frac{1}{a} \geq 1$. By the arithmetic rule,

$$\lim_{n \rightarrow \infty} \sqrt[n]{a} = \lim_{n \rightarrow \infty} \frac{1}{\sqrt[n]{\frac{1}{a}}} = \frac{1}{\lim_{n \rightarrow \infty} \sqrt[n]{\frac{1}{a}}} = 1. \quad \blacksquare$$

Example 2.1.6 To compute the limit $\lim_{n \rightarrow \infty} \left(1 - \frac{1}{\sqrt[n]{3}}\right) \sin n$, we note that

$$-1 + \frac{1}{\sqrt[n]{3}} \leq \left(1 - \frac{1}{\sqrt[n]{3}}\right) \sin n \leq 1 - \frac{1}{\sqrt[n]{3}}.$$

By $\lim_{n \rightarrow \infty} \left(1 - \frac{1}{\sqrt[n]{3}}\right) = 1 - \frac{1}{\lim_{n \rightarrow \infty} \sqrt[n]{3}} = 0$ and the sandwich rule, we get $\lim_{n \rightarrow \infty} \left(1 - \frac{1}{\sqrt[n]{3}}\right) \sin n = 0$. ■

Example 2.1.7 To compute the limit $\lim_{n \rightarrow \infty} \sqrt[n]{\frac{2n+1}{n-1}}$, we note that $1 < \frac{2n+1}{n-1} < 6$ for $n > 1$. Then

$$1 < \sqrt[n]{\frac{2n+1}{n-1}} < \sqrt[n]{6}$$

and the sandwich rule tell us $\lim_{n \rightarrow \infty} \sqrt[n]{\frac{2n+1}{n-1}} = 1$. ■

Example 2.1.8 We show the limit $\lim_{n \rightarrow \infty} \frac{a^n}{n!} = 0$ for the special case $a = 5$.

For $n > 5$, we have

$$0 < \frac{5^n}{n!} = \frac{5 \cdot 5 \cdot 5 \cdot 5 \cdot 5 \cdot 5 \cdot 5 \cdots 5}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdots n} \leq \frac{5 \cdot 5 \cdot 5 \cdot 5 \cdot 5 \cdot 5}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6} \cdot \frac{5}{n} = \frac{5^5}{24n}.$$

By $\lim_{n \rightarrow \infty} \frac{5^5}{24n} = \frac{5^5}{24} \lim_{n \rightarrow \infty} \frac{1}{n} = 0$ and the sandwich rule, we get $\lim_{n \rightarrow \infty} \frac{5^n}{n!} = 0$. ■

Example 2.1.9 Let $|a| < 1$. We show the limit $\lim_{n \rightarrow \infty} a^n = 0$.

For $0 < a < 1$, write $a = \frac{1}{1+b}$. Then $b > 0$ and

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

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

If $-1 < a < 0$, we have $-|a|^n \leq a^n \leq |a|^n$. By $\lim_{n \rightarrow \infty} |a|^n = 0$ and the sandwich rule, we also get $\lim_{n \rightarrow \infty} a^n = 0$. ■

A sequence $\{a_n\}$ is bounded above if there is a number B (called an upper bound), such that $a_n \leq B$ for all n . The sequence is bounded below if there is a number B (called a lower bound), such that $a_n \geq B$ for all n . A sequence is bounded if it has both upper and lower bounds.

Proposition 2.1.5 Any convergent sequence is bounded.

The property reflects the intuition that if a is close to π , then a is between 3 and 4.

Example 2.1.10 The sequences $\{n\}$, $\left\{\frac{n^2 + (-1)^n}{n+1}\right\}$ diverge because they are not bounded above.

On the other hand, the sequence $1, -1, 1, -1, \dots$ is bounded and diverges. Therefore the converse of Proposition 2.1.5 does not hold. ■

The following property reflects the intuition that larger number should be closer to larger limit.

Proposition 2.1.6 (Order Rule) Suppose $\{a_n\}$ and $\{b_n\}$ converge.

1. If $a_n \geq b_n$, then $\lim_{n \rightarrow \infty} a_n \geq \lim_{n \rightarrow \infty} b_n$.
2. If $\lim_{n \rightarrow \infty} a_n > \lim_{n \rightarrow \infty} b_n$, then $a_n > b_n$ for sufficiently large n .

On the other hand, the condition $a_n > b_n$ does not necessarily imply $\lim_{n \rightarrow \infty} a_n > \lim_{n \rightarrow \infty} b_n$. This can be demonstrated by $a_n = \frac{1}{n}$ and $b_n = \frac{1}{n^2}$.

Example 2.1.11 By $\lim_{n \rightarrow \infty} \frac{2n^2 + n}{n^2 - n + 1} = 2$, we know $1 < \frac{2n^2 + n}{n^2 - n + 1} < 3$ for sufficiently large n . In fact, the inequalities hold for all n .

By $\lim_{n \rightarrow \infty} \sqrt[n]{\frac{2n+1}{n-1}} = 1$, we know $\lim_{n \rightarrow \infty} \sqrt[n]{\frac{2n+1}{n-1}} < 2$ for sufficiently large n . In fact, the inequality holds for $n > 2$. ■

Monotonic Sequences

A sequence $\{a_n\}$ is **increasing** if

$$a_1 \leq a_2 \leq a_3 \leq \dots \leq a_n \leq a_{n+1} \leq \dots$$

It is **decreasing** if

$$a_1 \geq a_2 \geq a_3 \geq \dots \geq a_n \geq a_{n+1} \geq \dots$$

Both are **monotonic** sequences.

The sequences $\left\{\frac{1}{n}\right\}$, $\left\{\frac{1}{2^n}\right\}$, $\{\sqrt[n]{2}\}$ are decreasing. The sequences $\left\{-\frac{1}{n}\right\}$, $\{n\}$ are increasing.

Proposition 2.1.7 Any bounded monotonic sequence converges.

An increasing sequence $\{a_n\}$ is bounded if it has upper bound, because the first term of the sequence is already the lower bound. Similar remark can be made for a decreasing sequence.

Example 2.1.12 The sequence

$$\sqrt{2}, \sqrt{2 + \sqrt{2}}, \sqrt{2 + \sqrt{2 + \sqrt{2}}}, \dots$$

is inductively given by

$$a_1 = \sqrt{2}, \quad a_{n+1} = \sqrt{2 + a_n}.$$

We claim that the sequence $\{a_n\}$ is increasing and bounded above by 2.

We have $a_1 = \sqrt{2} < 2$ and $a_2 = \sqrt{2 + \sqrt{2}} > a_1$. Now assume $a_k < 2$ and $a_{k+1} > a_k$. Then

$$a_{k+1} = \sqrt{2 + a_k} < \sqrt{2 + 2} = 2, \quad a_{k+2} = \sqrt{2 + a_{k+1}} > \sqrt{2 + a_k} = a_{k+1}.$$

The claim is proved by induction. By Proposition 2.1.7, the sequence converges. Let L be the limit. Then by taking the limits on both sides of the equality $a_{n+1}^2 = 2 + a_n$ and applying the arithmetic rule, we get $L^2 = 2 + L$. Therefore $L = 2$ or -1 . Since $a_n > 0$, by the order rule, we must have $L \geq 0$. Therefore we conclude that $\lim_{n \rightarrow \infty} a_n = 2$. ■

Example 2.1.13 We will show that the limit

$$\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n$$

converges. We compare two consecutive terms by the binomial expansion,

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

A close examination shows that the sequence is increasing.

It remains to show that the sequence has upper bounded. By the expansion above, we have

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

By Proposition 2.1.7, the sequence converges. The limit is denoted by e

$$\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = e. \quad \blacksquare$$

Example 2.1.14 We give another argument that $\lim_{n \rightarrow \infty} a^n = 0$ for $0 < a < 1$.

Since $0 < a < 1$, the sequence $\{a^n\}$ is decreasing and satisfies $0 < a^n < 1$. Therefore the sequence converges to a limit L . On the other hand, the sequence $\{a^{n-1}\}$ is obtained by deleting the first term from $\{a^n\}$. Therefore the sequence $\{a^{n-1}\}$ should also converge to the same limit L . If $L \neq 0$, then we may apply the arithmetic rule to get

$$a = \lim_{n \rightarrow \infty} \frac{a^n}{a^{n-1}} = \frac{\lim_{n \rightarrow \infty} a^n}{\lim_{n \rightarrow \infty} a^{n-1}} = \frac{L}{L} = 1.$$

This contradicts to the assumption that $a < 1$. Therefore the limit L has to be 0. ■

Subsequences

By choosing infinitely many terms from a sequence $\{a_n\}$, we get a **subsequence**. The newly formed sequence usually is written as $\{a_{n_k}\}$, i.e.,

$$a_{n_1}, a_{n_2}, \dots, a_{n_k}, \dots$$

The indices satisfy

$$n_1 < n_2 < \dots < n_k < \dots,$$

and

$$n_k \geq k, \quad \text{for every } k.$$

For example, the following are subsequences of $\left\{a_n = \frac{1}{n}\right\}$, corresponding to the choices $n_k = 2k$, $n_k = 2k - 1$, $n_k = 2^k$, $n_k = k!$:

$$\begin{aligned} a_{2k} = \frac{1}{2k}: & \quad \frac{1}{2}, \frac{1}{4}, \frac{1}{6}, \dots, \frac{1}{2k}, \dots; \\ a_{2k-1} = \frac{1}{2k-1}: & \quad \frac{1}{1}, \frac{1}{3}, \frac{1}{5}, \dots, \frac{1}{2k-1}, \dots; \\ a_{2^k} = \frac{1}{2^k}: & \quad \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \dots, \frac{1}{2^k}, \dots; \\ a_{k!} = \frac{1}{k!}: & \quad \frac{1}{1}, \frac{1}{2}, \frac{1}{6}, \dots, \frac{1}{k!}, \dots \end{aligned}$$

Proposition 2.1.8 *If a sequence converges to L , then every subsequence converges to L .*

The property reflects the intuition that if something is close to L , then any part of it is also close to L .

Example 2.1.15 By $\lim_{n \rightarrow \infty} \frac{2n^2 + n}{n^2 - n + 1} = 2$, we also know

$$\lim_{n \rightarrow \infty} \frac{2(n^2 - 1)^2 + (n^2 - 1)}{(n^2 - 1)^2 - (n^2 - 1) + 1} = \lim_{n \rightarrow \infty} \frac{2n^4 - n^2 + 1}{n^4 - 3n^2 + 3} = 2,$$

and

$$\lim_{n \rightarrow \infty} \frac{2(n!)^2 + n!}{(n!)^2 - n! + 1} = 2. \quad \blacksquare$$

Example 2.1.16 Consider the sequence $\{a_n = (-1)^n\}$. The subsequence $\{a_{2k} = 1\}$ converges to 1, and the subsequence $\{a_{2k-1} = -1\}$ converges to -1 . Since the two subsequences have different limits, Proposition 2.1.8 tells us that the sequence $\{(-1)^n\}$ diverges. $\blacksquare$

The concept of subsequences allows us to state the partial converse to Proposition 2.1.5.

Theorem 2.1.9 (Bolzano-Weierstrass) theorem, Bolzano-Weierstrass Every bounded sequence has a convergent subsequence.

As a typical example, the sequence $\{(-1)^n\}$ is bounded. Although the whole sequence diverges, the sequence has two subsequences converging to 1 and -1 respectively.

Example 2.1.17 Let us list all the rational numbers in $(0, 1]$ as a sequence

$$a_1 = \frac{1}{1}, a_2 = \frac{1}{2}, a_3 = \frac{2}{2}, \frac{1}{3}, \frac{2}{3}, \frac{3}{3}, \dots, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n}{n}, \dots$$

The number $\pi^{-1} = 0.318309886184 \dots$ is the limit of the sequence

$$\frac{3}{10}, \frac{31}{100}, \frac{318}{1000}, \frac{3183}{10000}, \frac{31830}{100000}, \frac{318309}{1000000}, \dots,$$

which is clearly a subsequence of $\{a_n\}$. It is easy to see that any number in $[0, 1]$ is the limit of a convergent subsequence of $\{a_n\}$. $\blacksquare$

2.1.2 Rigorous Definition of Limit

The intuitive definition of limit is ambiguous because the meaning of the phrases “approaches L ” and “goes to infinity” is not precise. To arrive at a more precise definition, we study the statement $\lim_{n \rightarrow \infty} \frac{1}{n} = 0$ in more detail. When we say that $a_n = \frac{1}{n}$ approaches $L = 0$ as n goes to infinity, we mean an infinite collection of facts

$$\begin{aligned} n > 1 &\implies |a_n - L| < 1, \\ n > 10 &\implies |a_n - L| < 0.1, \\ n > 100 &\implies |a_n - L| < 0.01, \\ &\vdots \\ n > 1000000 &\implies |a_n - L| < 0.000001, \\ &\vdots \end{aligned}$$

In everyday life, we measure the largeness of quantity by comparing with some specific scale. For example, a city is considered as big if it has millions of people ($n > 1000000$), and the country is

considered as big if it has hundreds of millions of people ($n > 100000000$). On the other hand, a electric wire is considered thin if it is less than 1 millimeter in diameter ($d < 1$), and hair may be considered as thin if it is less than 0.05 millimeter in diameter ($d < 0.05$). Thus in ordinary language, we say when n is in the hundreds, then $|a_n - L|$ is in the hundredth, and when n is in the millions, then $|a_n - L|$ is in the millionth, etc. Of course for a different limit, the relation between the bigness of n and the smallness of $|a_n - L|$ may be different. For example, for $a_n = \frac{2^n}{n!}$ and $L = 0$, we have

$$\begin{aligned} n > 10 &\implies |a_n - L| < 0.0003, \\ n > 20 &\implies |a_n - L| < 0.00000000000005, \\ &\vdots \end{aligned}$$

But the key observation here is that a limit is an infinite collection of statements of the form “if n is larger than certain large number N , then $|a_n - L|$ is smaller than certain small number ϵ ”. In practice, we cannot verify all such statements one by one. Even if we have verified the truth of the first one million statements, there is no guarantee that the one million and the first statement is true. To mathematically establish the truth of all the statements, we have to establish the statements for all N and ϵ at once.

Which one statement do we need to establish? We note that $n > N$ is the cause of the effect $|a_n - L| < \epsilon$, and the precise relation between the cause and the effect varies from limit to limit. For some limit, n in thousands already guarantees that $|a_n - L| < 0.000001$. For some other limit, n has to be in the billions in order to make sure that $|a_n - L| < 0.000001$. But the key observation here is that no matter how small ϵ is, the “target” $|a_n - L| < \epsilon$ can always be guaranteed for sufficiently large n . So although $|a_n - L| < 0.000001$ may not be satisfied for n in the million range, it will probably be satisfied for n in the billions range (and if this fails again, perhaps in billions and billions range). In fact, for any given range $\epsilon > 0$, we can always find an integer N such that a_n is close to the limit L within the range ϵ whenever $n \geq N$, i.e.,

$$|a_n - L| < \epsilon, \quad \text{when } n \geq N.$$

In the case of $a_n = \frac{1}{n}$, we can take N to be any integer greater than $1/\epsilon$. One of such integer could be $[1/\epsilon] + 1$, where $[x]$ represents the integer part of the number x .¹

The concept of limits is characterized by the property that no matter how small (or close) the target is set, we can always achieve the target by going far enough in the sequence to achieve the target.

Definition 2.1.10 (Rigorous) A sequence $\{a_n\}$ converges to a finite number L if for any $\epsilon > 0$, there is an integer N , such that $n > N$ implies $|a_n - L| < \epsilon$.

Example 2.1.18 Consider the sequence $\left\{\frac{1}{n}\right\}$. For any $\epsilon > 0$, choose $N = \left[\frac{1}{\epsilon}\right] + 1$. When $n > N$, we then have

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

¹Any number x can be expressed as $x = [x] + (x)$, where $[x]$ is the integer part of x and (x) is the fractional part.

This proves $\lim_{n \rightarrow \infty} \frac{1}{n} = 0$ rigorously. ■

Example 2.1.19 Consider the sequence $\{a_n\}$, with $a_n = \frac{n^2 - 1}{n^2 + 1}$ for each n . For any $\epsilon > 0$, we can choose N to be any integer greater than $\sqrt{\frac{2}{\epsilon}}$. Whenever $n \geq N$,

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

Thus the sequence converges to 1.

How and why did we choose $\sqrt{\frac{2}{\epsilon}}$? The idea is that we need

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

This inequality is equivalent to the inequality

$$n > \sqrt{\frac{2}{\epsilon} - 1}.$$

Therefore, any integer greater than $\sqrt{\frac{2}{\epsilon} - 1}$ can be chosen to be N . Above, we chose $N \geq \sqrt{\frac{2}{\epsilon}}$. We can also choose $N \geq \sqrt{\frac{4}{\epsilon}}$.

However, the description above has an obvious flaw that requires to solve the inequality

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

Generally, this is not feasible, since some inequalities can be too difficult to solve. One way to overcome this is to “loose” the inequality until it can be solved easily. The key to do this is to keep track of every condition whenever you loose the inequality. In fact, we can solve the above inequality as follows. In the sequence of inequalities:

$$\left| \frac{n^2 - 1}{n^2 + 1} - 1 \right| = \frac{2}{n^2 + 1} \stackrel{\textcircled{1}}{<} \frac{2}{n} \stackrel{\textcircled{2}}{\leq} \frac{2}{N} \stackrel{\textcircled{3}}{\leq} \epsilon,$$

the inequality $\textcircled{1}$ holds for all natural integers n ; the inequality $\textcircled{2}$ holds under the condition $n \geq N$; the last inequality $\textcircled{3}$ holds under the condition that $N \geq \frac{2}{\epsilon}$. Thus, the whole sequential inequalities hold when

$$n \geq N, \quad N \text{ is an integer greater than } \frac{2}{\epsilon}.$$

Hence,

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

when

$$n \geq \left\lceil \frac{2}{\epsilon} \right\rceil + 1.$$

Therefore, by the definition, we have

$$\lim_{n \rightarrow \infty} \frac{n^2 - 1}{n^2 + 1} = 1.$$

The above “loose-and-track” approach is very useful in proving existence of limits. ■

Example 2.1.20 We demonstrate a similar argument once more by using the “loose-and-track” approach for the limit $\lim_{n \rightarrow \infty} (\sqrt{n^2 + 1} - n)$. In fact, for the sequence of inequalities,

$$\begin{aligned} \sqrt{n^2 + 1} - n &= (\sqrt{n^2 + 1} - n) \frac{(\sqrt{n^2 + 1} + n)}{(\sqrt{n^2 + 1} + n)} \\ &= \frac{1}{\sqrt{n^2 + 1} + n} \\ &\stackrel{\textcircled{1}}{<} \frac{1}{2n} \stackrel{\textcircled{2}}{\leq} \frac{1}{2N} \stackrel{\textcircled{3}}{\leq} \epsilon, \end{aligned}$$

the inequality $\textcircled{1}$ holds for all natural integers n ; the inequality $\textcircled{2}$ holds under the condition $n \geq N$; the last inequality $\textcircled{3}$ holds under the condition that $N \geq \frac{1}{2\epsilon}$. Thus, the whole sequential inequalities hold when

$$n \geq N, \quad N \text{ is an integer greater than } \frac{1}{2\epsilon}.$$

Hence,

$$|\sqrt{n^2 + 1} - n| < \epsilon$$

when

$$n \geq \left\lceil \frac{1}{2\epsilon} \right\rceil + 1.$$

Therefore, by the definition, we have

$$\lim_{n \rightarrow \infty} [\sqrt{n^2 + 1} - n] = 0. \quad \blacksquare$$

In the examples above, we saw that we often need to “loosen up” the difference between the sequence and the expected limit. Such estimation allows us to derive a simple choice of N that satisfies the definition. Of course the way of “loosen up” is not unique. For example, the estimation

$$\left| \frac{2n^2 + n}{n^2 - n + 1} - 2 \right| = \frac{|3n - 2|}{n^2 - n + 1} < \frac{3n}{n^2 - n + 1} < \frac{3n}{n^2 - \frac{n^2}{2}} = \frac{6}{n}$$

will lead to the choice $N = \frac{6}{\epsilon}$ and the estimation

$$|\sqrt{n+2} - \sqrt{n} - 0| = \frac{(n+2) - n}{\sqrt{n+2} + \sqrt{n}} < \frac{2}{\sqrt{n} + \sqrt{n}} = \frac{1}{\sqrt{n}}$$

will lead to the choice $N = \left\lceil \frac{1}{\epsilon^2} \right\rceil + 1$. So the choice of N is not unique. The only important thing here is that some N can be found to satisfy the requirement of the definition, and N depends only on ϵ (and especially independent of n).

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

Example 2.1.21 Assume $\lim_{n \rightarrow \infty} a_n = 1$. We prove that $\lim_{n \rightarrow \infty} \sqrt{a_n} = 1$.

The assumption tells us that for any $\epsilon > 0$, there is an integer N , such that $n > N$ implies $|a_n - 1| < \epsilon$. Then

$$n > N \implies |\sqrt{a_n} - 1| = \frac{|a_n - 1|}{\sqrt{a_n} + 1} \leq \frac{|a_n - 1|}{1} < \epsilon.$$

This completes the rigorous proof.

By the similar method, we can rigorously prove that $\lim_{n \rightarrow \infty} a_n = L > 0$ implies $\lim_{n \rightarrow \infty} \sqrt[k]{a_n} = \sqrt[k]{L}$ for any natural number k . In fact, we also have $\lim_{n \rightarrow \infty} a_n^p = L^p$ for any number p . But the proof would be more complicated. ■

Example 2.1.22 [Prove of $\lim_{n \rightarrow \infty} (a_n + b_n) = \lim_{n \rightarrow \infty} a_n + \lim_{n \rightarrow \infty} b_n$]

Assume $\lim_{n \rightarrow \infty} a_n = L$ and $\lim_{n \rightarrow \infty} b_n = K$. For any $\epsilon > 0$, we still have $\frac{\epsilon}{2} > 0$. Applying the definition of the limits of $\{a_n\}$ and $\{b_n\}$ to $\frac{\epsilon}{2}$, we find N_1 and N_2 , such that

$$n > N_1 \implies |a_n - L| < \frac{\epsilon}{2}, \quad n > N_2 \implies |b_n - K| < \frac{\epsilon}{2}.$$

Then we get

$$n > N = \max\{N_1, N_2\} \implies |(a_n + b_n) - (L + K)| \leq |a_n - L| + |b_n - K| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

This completes the proof that $\lim_{n \rightarrow \infty} (a_n + b_n) = L + K$.

If apply the definition to ϵ instead of $\frac{\epsilon}{2}$, then we will get $n > N$ implies $|(a_n + b_n) - (L + K)| < 2\epsilon$ in the end. It turns out that replacing $< \epsilon$ by $< 2\epsilon$ (or by $< \epsilon^2$, or by $< \sqrt{\epsilon}$, or by $\leq \epsilon$, etc) in (the final part of) the definition of limit gives a new definition that is equivalent to the old one. ■

Example 2.1.23 [Prove of Proposition 2.1.4] Assume $a_n \leq b_n \leq c_n$ and $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} c_n = L$. For any $\epsilon > 0$, there are integers N_1 and N_2 , such that

$$n > N_1 \implies |a_n - L| < \epsilon, \quad n > N_2 \implies |c_n - L| < \epsilon.$$

Then

$$n > N = \max\{N_1, N_2\} \implies L - \epsilon < a_n \leq b_n \leq c_n - L < L + \epsilon \implies |b_n - L| < \epsilon.$$

This proves that $\lim_{n \rightarrow \infty} b_n = L$. ■

Example 2.1.24 [Proof of Proposition 2.1.5] Assume $\lim_{n \rightarrow \infty} a_n = L$. By the definition of limit, for $\epsilon = 1$, there is an integer N , such that $n > N$ implies $|a_n - L| < 1$. By choosing larger N if necessary, we may further assume N to be a natural number. Then we have

$$|a_{N+1}| < |L| + 1, |a_{N+2}| < |L| + 1, |a_{N+3}| < |L| + 1, \dots$$

Take $B = \max\{|a_1|, \dots, |a_N|, |L| + 1\}$. Then we get $|a_n| \leq B$ for all n . Therefore the sequence is bounded by B .

Note that if in the definition of limit, N is not an integer (or natural number), then the new definition of limit is still equivalent to the old one. In the current example, it is more convenient to assume N to be a natural number here because we need to use N as the index of a term in a sequence. ■

Example 2.1.25 [Prove of $\lim_{n \rightarrow \infty} (a_n b_n) = (\lim_{n \rightarrow \infty} a_n) \cdot (\lim_{n \rightarrow \infty} b_n)$]

By Proposition 2.1.5 just proved, $\{a_n\}$ and $\{b_n\}$ are bounded. Assume $|a_n| < M_1$ and $|b_n| < M_2$. Assume

$$\lim_{n \rightarrow \infty} a_n = L_1, \quad \lim_{n \rightarrow \infty} b_n = L_2.$$

For any $\epsilon > 0$, there is an integer N_1 such that

$$n > N_1 \implies |a_n - L_1| < \frac{\epsilon}{2M_2},$$

and there is an integer N_2 such that

$$n \geq N_2 \implies |b_n - L_2| < \frac{\epsilon}{2|L_1| + 1}.$$

Choose $N = \max\{N_1, N_2\}$. Then, if $n \geq N$, we have

$$\begin{aligned} & |a_n b_n - L_1 L_2| \\ &= |a_n b_n - L_1 b_n + L_1 b_n - L_1 L_2| \\ &= |(a_n - L_1) b_n + L_1 (b_n - L_2)| \\ &\leq |b_n| \cdot |a_n - L_1| + |L_1| \cdot |b_n - L_2| \\ &< M_2 \frac{\epsilon}{2M_2} + |L_1| \frac{\epsilon}{2|L_1| + 1} < \epsilon. \end{aligned} \quad \blacksquare$$

Example 2.1.26 [Proof of Proposition 2.1.6] Assume $a_n \geq b_n$ and $\lim_{n \rightarrow \infty} a_n = L$ and $\lim_{n \rightarrow \infty} b_n = K$. For any $\epsilon > 0$, there are N_1 and N_2 , such that

$$n > N_1 \implies |a_n - L| < \epsilon, \quad n > N_2 \implies |b_n - K| < \epsilon.$$

Now pick a natural number n larger than both N_1 and N_2 . Then we have

$$L + \epsilon > a_n \geq b_n > K - \epsilon.$$

Therefore we proved that $L + \epsilon > K - \epsilon$ for any $\epsilon > 0$. It is easy to see that the property is the same as $L \geq K$.

Conversely, we assume $\lim_{n \rightarrow \infty} a_n = L$, $\lim_{n \rightarrow \infty} b_n = K$ and $L > K$. For $\epsilon = \frac{L - K}{2} > 0$, there are N_1 and N_2 , such that

$$n > N_1 \implies |a_n - L| < \epsilon, \quad n > N_2 \implies |b_n - K| < \epsilon.$$

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

$$a_n - b_n > (L - \epsilon) - (K + \epsilon) = L - K - 2\epsilon = 0.$$

Therefore $a_n > b_n$ as long as $n > N$. ■

Example 2.1.27 [Proof of Proposition 2.1.8]

By the definition, for any $\epsilon > 0$, there is an integer N so that $|a_n - L| < \epsilon$ whenever $n \geq N$. Thus, for any subsequence $\{a_{n_k}\}$, if $k \geq N$, since $n_k \geq k \geq N$, we have $|a_{n_k} - L| < \epsilon$. This, by the definition, implies that $\lim_{k \rightarrow \infty} a_{n_k} = L$. ■

Cauchy Sequence

The definition of limit makes explicit use of the limit value L . Therefore if we want to show the convergence of a sequence by the definition, we need to first find the limit L and then find suitable N for each ϵ . In many cases, however, it is very hard (or even impossible) to find the limit of a convergent sequence. For example, we know the world record for 100 meter dash must have a limit set by the human capability. However, nobody knows exactly what the human capability is.

Proposition 2.1.7 gives a special case that we know the convergence of a sequence without knowing the actual limit value. The following criterion provides a method for the general case.

Definition 2.1.11 A sequence $\{a_n\}$ is called a *Cauchy sequence* if for any $\epsilon > 0$, there is N , such that $m, n > N$ implies $|a_n - a_m| < \epsilon$.

Theorem 2.1.12 (Cauchy Criterion) A sequence converges if and only if it is a Cauchy sequence.

Proof. Assume $\{a_n\}$ converges to L . For any $\epsilon > 0$, there is N , such that $n \geq N$ implies $|a_n - L| < \frac{\epsilon}{2}$. Then

$$\begin{aligned} m, n > N &\implies |a_m - L| < \frac{\epsilon}{2}, \quad |a_n - L| < \frac{\epsilon}{2} \\ &\implies |a_n - a_m| = |(a_n - L) - (a_m - L)| \leq |a_n - L| + |a_m - L| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \end{aligned}$$

The converse is much more difficult. The convergence of a Cauchy sequence $\{a_n\}$ may be proved in the following steps.

1. The 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 .

Diligent readers may try to prove the first and the third statements. The Bolzano-Weierstrass Theorem (Proposition 2.1.9) is a very deep result that touches the essential difference between the real and rational numbers. ■

Example 2.1.28 Consider the sequence

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

If the sequence were convergent, the limit would be the sum of the infinite series

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

To show the convergence, we note that for $m > n$,

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

Thus for any ϵ , choose $N = \frac{1}{\epsilon}$. Then $m > n > N$ implies

$$|a_n - a_m| < \frac{1}{n} < \frac{1}{N} = \epsilon.$$

By the Cauchy criterion, the sequence $\{x_n\}$ converges.

The actual sum of the infinite series is $\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$. However, knowing the sum probably still gives you no clue how to prove the sequence converges. Moreover, it took mathematician great effort to prove that the sum of the series $\sum_{n=1}^{\infty} \frac{1}{n^3}$ is a transcendental number (not a root of any polynomial), and we do not even know whether $\sum_{n=1}^{\infty} \frac{1}{n^5}$ is transcendental or not. Yet we know that

$\sum_{n=1}^{\infty} \frac{1}{n^p}$ converges for any $p > 1$. ■

Example 2.1.29 We use the opposite of the Cauchy criterion to show that the sequence

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

diverges. Note that a sequence is not Cauchy if there is $\epsilon > 0$, such that for any N , we can find $m, n > N$ satisfying $|a_m - a_n| \geq \epsilon$.

For any n , we have

$$a_{2n} - a_n = \frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{2n} \geq \frac{1}{2n} + \frac{1}{2n} + \cdots + \frac{1}{2n} = \frac{1}{2}.$$

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

$$|a_m - a_n| = a_{2n} - a_n \geq \frac{1}{2} = \epsilon.$$

Therefore, the sequence fails the Cauchy criterion and diverges. ■

Exercises

2.1.1 Find limit of sequence:

1. $\sqrt{n}(\sqrt{n+1} - \sqrt{n});$

5. $\frac{\sin n!}{\sqrt{n}};$

2. $\frac{10^{10}n}{n^2 - 10};$

6. $1 + a + a^2 + \cdots + a^n, |a| < 1;$

3. $\frac{(2n-1)!}{(2n+1)!};$

7. $na^n, |a| < 1;$

4. $\frac{5^n - n \cdot 6^{n+1}}{3^{2n-1} - 2^{3n+1}};$

8. $\sqrt[n]{n}.$

2.1.2 Use the definition to prove the limit.

1. $\lim_{n \rightarrow \infty} \frac{1}{n^p} = 0, p > 0;$

4. $\lim_{n \rightarrow \infty} \frac{\pi^n}{n!} = 0;$

2. $\lim_{n \rightarrow \infty} \frac{\cos n}{n} = 0;$

5. $\lim_{n \rightarrow \infty} (\sqrt{n-1} - \sqrt{n+1}) = 0;$

3. $\lim_{n \rightarrow \infty} \frac{2n+3}{2n-\sqrt{n}-1} = 1;$

6. $\lim_{n \rightarrow \infty} (n^{\frac{1}{3}} - (n-1)^{\frac{1}{3}}) = 0;$

2.1.3 Determine convergence. Provide rigorous reason.

1. $\frac{(-1)^n n^2}{n^3 - 1}$;
2. $\frac{(-1)^n n^2}{n^2 - 1}$;
3. $\frac{(-1)^n n^2}{n - 1}$;
4. $\frac{\cos^2 n}{\sqrt{n}}$;
5. $\frac{2^n}{\sqrt{n!}}$;
6. $\frac{n!}{2^n}$;
7. $\frac{2n + \cos n^2}{n + (-1)^n \sqrt{n} + \sin n}$;
8. $\frac{1}{2}, \frac{2}{1}, \frac{2}{3}, \frac{3}{2}, \frac{3}{4}, \frac{4}{3}, \dots, \frac{n}{n+1}, \frac{n+1}{n}, \dots$;
9. $\frac{1}{2}, -\frac{2}{1}, \frac{2}{3}, -\frac{3}{2}, \frac{3}{4}, -\frac{4}{3}, \dots, \frac{n}{n+1}, -\frac{n+1}{n}, \dots$;
10. $1 + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \dots + \frac{1}{\sqrt{n}}$;
11. $1 - \frac{1}{2^3} + \frac{1}{3^3} - \dots + \frac{(-1)^{n+1}}{n^3}$;
12. $1 + \frac{1}{2!} + \frac{1}{3!} + \dots + \frac{1}{n!}$;
13. $a_n = 1 + \frac{2}{1!} + \frac{3}{2!} + \frac{4}{3!} + \dots + \frac{n}{(n-1)!}$;
14. $\frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2 \cdot 4 \cdot 6 \cdots (2n)}$;
15. $\frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{n!}$.

2.1.4 Suppose $\lim_{n \rightarrow \infty} a_n b_n = 0$. Can you conclude that either $\lim_{n \rightarrow \infty} a_n = 0$ or $\lim_{n \rightarrow \infty} b_n = 0$?

2.1.5 Prove that if $\lim_{n \rightarrow \infty} a_n = L$, then $\lim_{n \rightarrow \infty} |a_n| = |L|$.

2.1.6 Prove that $\lim_{n \rightarrow \infty} |a_n - L| = 0$ if and only if $\lim_{n \rightarrow \infty} a_n = L$.

2.1.7 Prove that if $\lim_{n \rightarrow \infty} a_n = L$, then $\lim_{n \rightarrow \infty} c a_n = cL$.

2.1.8 For $\alpha > 0$ and any $b > 0$, prove that the sequence defined by

$$a_0 = b, \quad a_{n+1} = \frac{1}{2} \left(a_n + \frac{\alpha}{a_n} \right)$$

satisfies $a_n \geq \sqrt{\alpha}$ for $n \geq 1$ and is decreasing. Then prove that the sequence converges to $\sqrt{\alpha}$. Finally discuss what happens when $b < 0$.

2.1.9 Define the sequence $\{x_n\}$ by $x_1 = \alpha$ and

$$x_{n+1} = \frac{x_n^2}{2} + \frac{1}{2}, \quad n = 1, 2, 3, \dots$$

Prove that $\lim_{n \rightarrow +\infty} x_n = 1$ for any real number α , with $0 \leq \alpha \leq 1$.

2.1.10 Let $a_1 > 1$ and $a_n = 2 - \frac{1}{a_{n-1}}$ for each positive integer $n > 1$. Show that the sequence $\{a_n\}$ is convergent. What is the limit of this sequence?

2.1.11 For $\alpha > 0$, consider the sequence

$$\sqrt{\alpha}, \sqrt{\alpha + \sqrt{\alpha}}, \sqrt{\alpha + \sqrt{\alpha + \sqrt{\alpha}}}, \sqrt{\alpha + \sqrt{\alpha + \sqrt{\alpha + \sqrt{\alpha}}}}, \dots$$

1. Assume the sequence converges, find the limit β .
2. Prove that β is the upper bound of the sequence.
3. Prove that the sequence is increasing.
4. Prove that the sequence indeed converges to β .

2.1.12 For any $a, b > 0$, define a sequence by

$$a_1 = a, \quad a_2 = b, \quad a_n = \frac{a_{n-1} + a_{n-2}}{2}.$$

Prove that the sequence converges.

2.1.13 The Fibonacci sequence

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

is defined by $a_0 = a_1 = 1$ and $a_{n+1} = a_n + a_{n-1}$. Consider the sequence $b_n = \frac{a_{n+1}}{a_n}$.

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

2.1.14 Prove that the sequences $\left\{\left(1 + \frac{1}{n}\right)^{n+1}\right\}$ and $\left\{\left(1 - \frac{1}{n}\right)^{-n}\right\}$ are decreasing, bounded, and converge to e .

2.1.15 Prove that for $n > k$, we have

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

The use Proposition 2.1.6 to show that

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

Finally, prove

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

2.1.16 If $\{a_n\}$ is Cauchy, can you conclude that $\{|a_n|\}$ is Cauchy? Why? What about the converse?

2.1.17 Suppose that the sequence $\{x_n\}$ satisfies

$$|x_{n+1} - x_n| \leq 1/n^2, \quad n = 1, 2, 3, \dots$$

Prove that $\{x_n\}$ is convergent.

2.1.18 If

$$a_n - a_{n-1} = \sin(\pi/n), \quad n = 1, 2, 3, \dots,$$

does the sequence $\{a_n\}$ converge? Explain.

2.1.19 Suppose a_n is a bounded sequence and $|q| < 1$. Prove that the sequence

$$x_n = a_0 + a_1q + a_2q^2 + \cdots + a_nq^n$$

is a Cauchy sequence and therefore converges.

2.2 LIMITS OF FUNCTIONS

Now we study limit of functions. We will see that it is close related to limit of sequences induced by functions.

2.2.1 Concept of Limit and Properties

Consider the functions $f(x) = x^2$, $g(x) = \frac{1}{x}$, and $h(x) = \sin \frac{1}{x}$ as in Fig. 2.1.

When x approaches 0, $f(x) = x^2$ also approaches 0, the absolute value of $g(x) = \frac{1}{x}$ gets larger and does not approach a fixed finite number. Moreover, the value of h swings, such as

$$h\left(\frac{1}{2n\pi}\right) = 0, \quad h\left(\frac{1}{(2n + \frac{1}{2})\pi}\right) = 1.$$

Therefore $h(x)$ does not approach any one specific finite number.

We saw similar behaviors for sequences when n approaches infinity. The functions $f(x)$, $g(x)$, $h(x)$ are comparable to $\left\{\frac{1}{n^2}\right\}$, $\{(-1)^n n\}$, $\{(-1)^n\}$. Only the first sequence converges. This leads to the similar definition of limit for functions.

Definition 2.2.1 (Non-rigorous) A function $f(x)$ converges to a finite number L at $x = a$ if $f(x)$ approaches L as x approaches a . We write

$$\lim_{x \rightarrow a} f(x) = L.$$

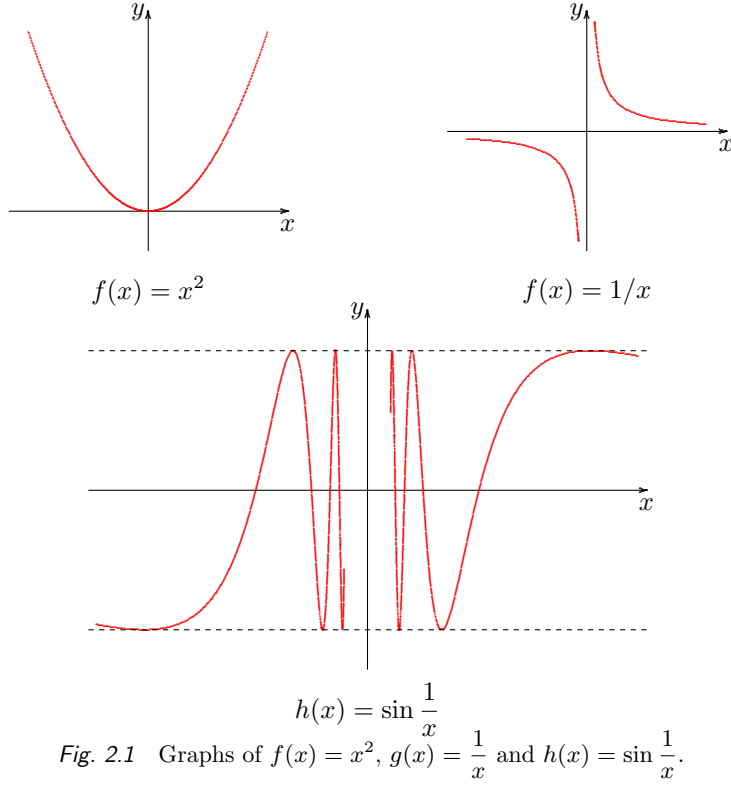


Fig. 2.1 Graphs of $f(x) = x^2$, $g(x) = \frac{1}{x}$ and $h(x) = \sin \frac{1}{x}$.

Thus the limit $\lim_{x \rightarrow 0} x^2 = 0$ converges, and the limits $\lim_{x \rightarrow 0} \frac{1}{x}$, $\lim_{x \rightarrow 0} \sin \frac{1}{x}$ diverge. The limit of function has properties similar to the limit of sequence.

Proposition 2.2.2 (Arithmetic Rule) Suppose $\lim_{x \rightarrow a} f(x) = L$ and $\lim_{x \rightarrow a} g(x) = K$. Then

$$\lim_{x \rightarrow a} (f(x) + g(x)) = \lim_{x \rightarrow a} f + \lim_{x \rightarrow a} g, \quad \lim_{x \rightarrow a} cf(x) = c \lim_{x \rightarrow a} f(x), \quad \lim_{x \rightarrow a} f(x)g(x) = \lim_{x \rightarrow a} f(x) \cdot \lim_{x \rightarrow a} g(x).$$

Moreover, if $\lim_{x \rightarrow a} g(x) \neq 0$, then

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \frac{\lim_{x \rightarrow a} f(x)}{\lim_{x \rightarrow a} g(x)}.$$

Proposition 2.2.3 If $\lim_{x \rightarrow a} f(x)$ converges, then $\lim_{x \rightarrow a} |f(x)|$ also converges and $\lim_{x \rightarrow a} |f(x)| = \left| \lim_{x \rightarrow a} f(x) \right|$. On the other hand, if $\lim_{x \rightarrow a} |f(x)| = 0$, then $\lim_{x \rightarrow a} f(x) = 0$.

Proposition 2.2.4 (Sandwich Rule) If $f(x) \leq g(x) \leq h(x)$ and $\lim_{x \rightarrow c} f(x) = \lim_{x \rightarrow c} h(x) = L$, then $\lim_{x \rightarrow c} g(x) = L$.

Proposition 2.2.5 (Order Rule) Suppose $\lim_{x \rightarrow a} f(x)$ and $\lim_{x \rightarrow a} g(x)$ converge.

1. If $f(x) \geq g(x)$, then $\lim_{x \rightarrow a} f(x) \geq \lim_{x \rightarrow a} g(x)$.
2. If $\lim_{x \rightarrow a} f(x) > \lim_{x \rightarrow a} g(x)$, then $f(x) > g(x)$ for x sufficiently close to a and not equal to a .

The convergence and the limit value of a sequence depends only on the terms with large index n . In other words, if finitely many terms in a sequence is dropped, added, or modified, then the new sequence converges if and only if the original sequence converges. Moreover, the limits of the two sequences are the same.

Similarly, the convergence and the limit value of a function $f(x)$ at a depends only on the value of $f(x)$ for x close to a and not equal to a . Moreover, the function does not even need to be defined at a in order for the limit to make sense. Therefore, if $g(x) = f(x)$ for x close but not equal to a , then $\lim_{x \rightarrow a} g(x)$ converges if and only if $\lim_{x \rightarrow a} f(x)$ converges. Moreover, the limits of the two functions are the same.

Example 2.2.1 We have $\lim_{x \rightarrow a} x = a$ and $\lim_{x \rightarrow a} c = c$. Then by the arithmetic rule, we have

$$\lim_{x \rightarrow a} x^2 = \lim_{x \rightarrow a} x \cdot \lim_{x \rightarrow a} x = a \cdot a = a^2, \quad \lim_{x \rightarrow a} cx^2 = c \lim_{x \rightarrow a} x^2 = ca^2.$$

By the similar idea, we get $\lim_{x \rightarrow a} cx^n = ca^n$. Then for any polynomial $p(x) = c_n x^n + c_{n-1} x^{n-1} + \cdots + c_1 x + c_0$, we get

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

Moreover, a rational function is the quotient of two polynomials $r(x) = \frac{p(x)}{q(x)}$ and is defined at a if $q(a) \neq 0$. Then further by the arithmetic rule, we have $\lim_{x \rightarrow a} r(x) = r(a)$ whenever r is defined at a . ■

Example 2.2.2 The function $\frac{x^3 - 1}{x - 1}$ is not defined at $x = 1$. Yet the function converges at 1

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

Example 2.2.3 By $\lim_{x \rightarrow 0} \sqrt[3]{x+1} = 1$ and the arithmetic rule, we get

$$\lim_{x \rightarrow 0} \frac{\sqrt[3]{x+1} - 1}{x} = \lim_{x \rightarrow 0} \frac{(x+1) - 1}{x((\sqrt[3]{x+1})^2 + \sqrt[3]{x+1} + 1)} = \frac{1}{(\lim_{x \rightarrow 0} \sqrt[3]{x+1})^2 + \lim_{x \rightarrow 0} \sqrt[3]{x+1} + 1} = \frac{1}{3}. \quad \blacksquare$$

Example 2.2.4 The limit $\lim_{x \rightarrow 0} x \sin \frac{1}{x}$ cannot be computed as follows

$$\lim_{x \rightarrow 0} x \sin \frac{1}{x} = \lim_{x \rightarrow 0} x \cdot \lim_{x \rightarrow 0} \sin \frac{1}{x},$$

because the second limit on the right diverges. However, if we use

$$-|x| \leq x \sin \frac{1}{x} \leq |x|,$$

and the sandwich rule, then we get $\lim_{x \rightarrow 0} x \sin \frac{1}{x} = 0$. ■

Another useful property of the limit of function is the following. The property can be compared with Proposition 2.1.8. The asterisk * indicates the statement is not quite true.

Proposition 2.2.6 (Composition Rule*) *If $\lim_{x \rightarrow a} f(x) = b$ and $\lim_{y \rightarrow b} g(y) = c$, then $\lim_{x \rightarrow a} g(f(x)) = c$.*

For example, the limit $\lim_{x \rightarrow 0} \sqrt[3]{x+1} = 1$ is really the composition of $\lim_{x \rightarrow 0} (x+1) = 1$ and $\lim_{y \rightarrow 1} \sqrt[3]{y} = 1$.

The composition rule can be understood as follows. Let $y = f(x)$ and $z = g(y)$. Then $z = g(f(x))$ is the composition. The assumptions $\lim_{x \rightarrow a} f(x) = b$ and $\lim_{y \rightarrow b} g(y) = c$ means that

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

When the two implications are combined, we get

$$x \rightarrow a \implies z \rightarrow c$$

which means $\lim_{x \rightarrow a} g(f(x)) = c$.

However, there is a slight error in the explanation above, what we really have are the following implications

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

To match the right side of the first implication to the left side of the second implication, we must either modify the first implication into

$$x \rightarrow a, \quad x \neq a \implies y \rightarrow b, \quad y \neq b,$$

which means that $\lim_{x \rightarrow a} f(x) = b$ and $f(x) \neq b$ for x close and not equal to a , or modify the second implication into

$$y \rightarrow b \implies z \rightarrow c,$$

which means that the limit $\lim_{y \rightarrow b} g(y) = c$ also includes the possibility of $y = b$. It is not hard to see that the second modification means $\lim_{y \rightarrow b} g(y) = g(b)$, or g is continuous at b .

Example 2.2.5 By $\lim_{x \rightarrow 0} x \sin \frac{1}{x} = 0$ and $\lim_{x \rightarrow 1} (x^2 - 1) = 0$, we get $\lim_{x \rightarrow 1} (x^2 - 1) \sin \frac{1}{x^2 - 1} = 0$. ■

Example 2.2.6 Consider $f(x) = x \sin \frac{1}{x}$ and $g(y) = \begin{cases} y, & \text{if } y \neq 0, \\ A, & \text{if } y = 0. \end{cases}$ We have

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

Although $\lim_{x \rightarrow 0} f(x) = 0$ and $\lim_{y \rightarrow 0} g(y) = 0$, the limit $\lim_{x \rightarrow 0} g(f(x))$ converges if and only if $A = 0$. Note that in case $A = 0$, we have $g(y) = y$ for all y and $\lim_{y \rightarrow 0} g(y) = g(0)$. ■

One Sided Limit

In the definition of the limit $\lim_{x \rightarrow a} f(x)$, x may approach a from either left or right side. Sometimes (for example, when the function is only defined on one side) we need to consider limit only on one side. This leads to one sided limits.

A function $f(x)$ converges to a finite limit L on the right side of $x = a$ if $f(x)$ approaches L as $x > a$ and approaches a . In this case, we say $f(x)$ has right limit at a and denote $\lim_{x \rightarrow a^+} f(x) = L$. The left limit at a is similarly defined.

Note that the the definition of the right limit only requires the function to be defined on the right and the left limit only requires to be defined on the left.

All the properties of the usual “two sided” limits applies to the ones sided limits. Intuitively, we should have the following relation between “two sided” and one sided limits.

Proposition 2.2.7 *The limit $\lim_{x \rightarrow a} f(x)$ converges to L if and only if both limits $\lim_{x \rightarrow a^+} f(x)$ and $\lim_{x \rightarrow a^-} f(x)$ converge to L .*

Example 2.2.7 The function $\sqrt{x}$ is only defined for $x \geq 0$. We have $\lim_{x \rightarrow 0^+} \sqrt{x} = 0$. More generally, we have $\lim_{x \rightarrow 0^+} x^p = 0$ for $p > 0$. ■

Example 2.2.8 The sign function

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

is shown in Fig. 2.2. We have $\lim_{x \rightarrow 0^+} f(x) = 1$ and $\lim_{x \rightarrow 0^-} f(x) = -1$. Since the two limits are not equal, the limit $\lim_{x \rightarrow 0} f(x)$ diverges. ■

Example 2.2.9 By applying the arithmetic rule to $\lim_{x \rightarrow 1} x = 1$, we know $\lim_{x \rightarrow 1} x^n = 1$ for any integer n . Now for any number p , we also wish to justify $\lim_{x \rightarrow 1} x^p = 1$.

We can find two integers m and n satisfying $m < p < n$. Then for $x > 1$, we have

$$x^m < x^p < x^n.$$

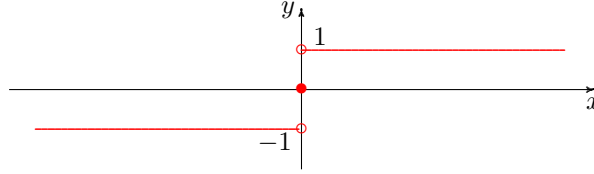


Fig. 2.2 A step function.

Since we already know $\lim_{x \rightarrow 1} x^m = \lim_{x \rightarrow 1} x^n = 1$, we get $\lim_{x \rightarrow 1^+} x^p = 1$ by the sandwich rule. Note that we only get the right limit because the inequality holds only on the right of 1.

For $0 < x < 1$, we have similar inequality $x^m > x^p > x^n$, and the sandwich rule again tells us $\lim_{x \rightarrow 1^-} x^p = 1$. Thus we have justified $\lim_{x \rightarrow 1} x^p = 1$.

In general, for any $a > 0$ and p , we have $\lim_{x \rightarrow a} \frac{x}{a} = 1$, so that $\lim_{x \rightarrow a} \left(\frac{x}{a}\right)^p = 1$ by what we just proved. Strictly speaking, we used the composition of $\lim_{x \rightarrow a} \frac{x}{a} = 1$ and $\lim_{y \rightarrow 1} y^p = 1$ here. Then

$$\lim_{x \rightarrow a} x^p = a^p \lim_{x \rightarrow a} \left(\frac{x}{a}\right)^p = a^p.$$

Finally by the composition rule, we know $\lim_{x \rightarrow a} f(x) = L > 0$ implies $\lim_{x \rightarrow a} f(x)^p = L^p$. ■

Example 2.2.10 The definition of trigonometric functions is illustrated in Fig. 2.3. It is clear that

$$0 < x < \frac{\pi}{2} \implies 0 < \sin x = \text{length of line } AB < \text{length of line } BC < \text{length of arc } BC = x.$$

By $\lim_{x \rightarrow 0^+} x = 0$ and the sandwich rule, we get $\lim_{x \rightarrow 0^+} \sin x = 0$. By $\sin(-x) = -\sin x$, we also know

$$-\frac{\pi}{2} < x < 0 \implies x < \sin x < 0.$$

Again the sandwich rule tells us $\lim_{x \rightarrow 0^-} \sin x = 0$. Therefore we conclude that

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

As for the limit of $\cos x$, for $0 < x < \pi$, we have

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

By the sandwich rule, we have $\lim_{x \rightarrow 0^+} \cos x = 1$. Similar argument can be made for $-\pi < x < 0$ to get the left limit, and we conclude

$$\lim_{x \rightarrow 0} \cos x = 1.$$

Finally, by the arithmetic rule, we have

$$\lim_{x \rightarrow 0} \tan x = \frac{\lim_{x \rightarrow 0} \sin x}{\lim_{x \rightarrow 0} \cos x} = 0. \quad \blacksquare$$

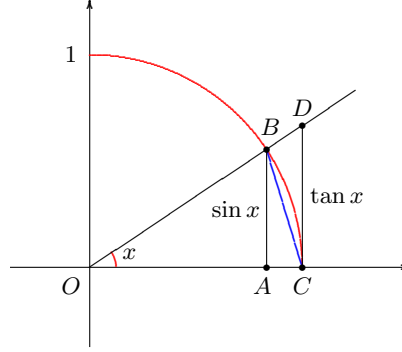


Fig. 2.3 Definition of $\sin x$ and $\tan x$ for $0 < x < \frac{\pi}{2}$.

Example 2.2.11 We study another important limit $\lim_{x \rightarrow 0} \frac{\sin x}{x}$ related to trigonometric functions. In Fig. 2.3, we have

$$\text{area of triangle } OBC < \text{area of fan } OBC < \text{area of triangle } ODC.$$

This means $\frac{\sin x}{2} < \frac{x}{2} < \frac{\tan x}{2}$. By $\tan x = \frac{\sin x}{\cos x}$, we get

$$0 < x < \frac{\pi}{2} \implies \cos x < \frac{\sin x}{x} < 1.$$

Since $\lim_{x \rightarrow 0} \cos x = 1$, by sandwich rule, we get $\lim_{x \rightarrow 0^+} \frac{\sin x}{x} = 1$.

For the left limit, we note that $x \rightarrow 0$ and $x < 0$ implies $-x \rightarrow 0$ and $-x > 0$. By the composition rule, we get

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

The limit can also be obtained by applying the sandwich rule to

$$-\frac{\pi}{2} < x < 0 \implies \cos x = \cos(-x) < \frac{\sin x}{x} = \frac{\sin(-x)}{(-x)} < 1.$$

In conclusion, we get

$$\lim_{x \rightarrow 0} \frac{\sin x}{x} = 1. \quad \blacksquare$$

Example 2.2.12 The function

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

has one sided limits

$$\lim_{x \rightarrow 0^-} f(x) = \lim_{x \rightarrow 0} \sin x = 0, \quad \lim_{x \rightarrow 0^+} f(x) = \lim_{x \rightarrow 0} x \sin \frac{1}{x} = 0.$$

Therefore $\lim_{x \rightarrow 0} f(x) = 0$. ■

Limit at Infinity

Another variation of the limit of function is the limit at infinity. A function $f(x)$ converges to a finite limit L at $+\infty$ if $f(x)$ approaches L as x is positive and gets larger and larger. In this case, we denote $\lim_{x \rightarrow +\infty} f(x) = L$.

The limit $\lim_{x \rightarrow -\infty} f(x) = L$ is similarly defined by considering large and negative x . And the limit $\lim_{x \rightarrow \infty} f(x) = L$ defined by considering (both positive and negative) x with large absolute value $|x|$.

For example, we have $\lim_{x \rightarrow \infty} \frac{1}{x} = 0$, $\lim_{x \rightarrow -\infty} 2^x = 1$, $\lim_{x \rightarrow +\infty} \frac{|x+4|}{x} = 1$, $\lim_{x \rightarrow -\infty} \frac{|x+4|}{x} = -1$. Moreover, the limits $\lim_{x \rightarrow \infty} x$ and $\lim_{x \rightarrow \infty} \sin x$ diverge.

The limits at infinity have property similar to the usual limit.

Example 2.2.13 By $\lim_{x \rightarrow \infty} c = c$, $\lim_{x \rightarrow \infty} \frac{1}{x} = 0$ and the arithmetic rule,

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

Example 2.2.14 For $x > 0$, we have

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

By the sandwich rule and $\lim_{x \rightarrow +\infty} \frac{2}{\sqrt{x}} = 0$, we get $\lim_{x \rightarrow +\infty} (\sqrt{x+2} - \sqrt{x}) = 0$. ■

Example 2.2.15 By $\lim_{x \rightarrow 0} \frac{\sin x}{x} = 1$, $\lim_{x \rightarrow \infty} \frac{1}{x} = 0$ and the composition rule, we have $\lim_{x \rightarrow \infty} x \sin \frac{1}{x} = 1$. ■

2.2.2 Rigorous Definition of Limit

The limit of sequence is the behavior as n approaches infinity, which is described by $n > N$. The limit of function is the behavior as x approaches a , which is described by $0 < |x - a| < \delta$ ($0 <$ means that $x \neq a$). The rigorous definition of the limit of function can be obtained from the rigorous definition of the limit of sequence by replacing $n > N$ with $0 < |x - a| < \delta$.

Definition 2.2.8 (Rigorous) A function $f(x)$ converges to a finite number L at $x = a$ if for any $\epsilon > 0$, there is $\delta > 0$, such that $0 < |x - a| < \delta$ implies $|f(x) - L| < \epsilon$.

If $0 < |x - a| < \delta$ is replaced by $0 < x - a < \delta$, then we have the definition of the right limit. If $0 < |x - a| < \delta$ is replaced by $-\delta < x - a < 0$, then we get the left limit.

The limit at infinity can be similarly defined.

Definition 2.2.9 (Rigorous) A function $f(x)$ converges to a finite number L at ∞ if for any $\epsilon > 0$, there is K , such that $|x| > K$ implies $|f(x) - L| < \epsilon$.

If $|x| > K$ is replaced by $x > K$, then we have the limit at $+\infty$. If $|x| > K$ is replaced by $x < K$ (K can be negative), then we get the limit at $-\infty$.

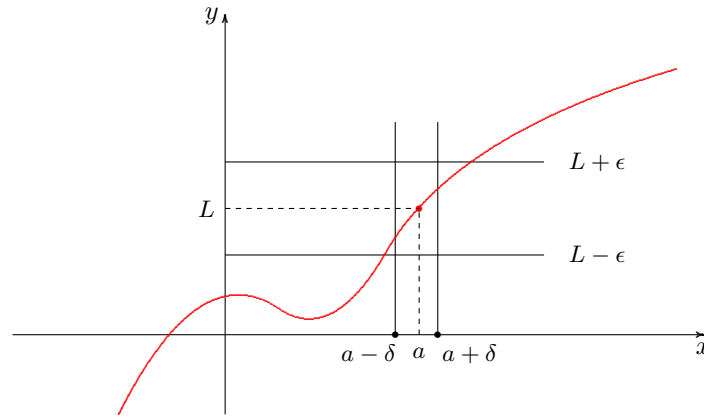


Fig. 2.4 Demonstration of the limit $\lim_{x \rightarrow a} f(x) = L$.

Example 2.2.16 We try to prove $\lim_{x \rightarrow 1} x^2 = 1$ rigorously.

Here is the analysis before we write down the proof. For any $\epsilon > 0$, we want $|x^2 - 1| < \epsilon$ when x is sufficiently close to 1. In other words, we need to find out how close x to 1 is so that $|x^2 - 1| < \epsilon$. We try to “loose” this inequality:

$$|x^2 - 1| = |x + 1| \cdot |x - 1| \leq (|x - 1| + 2)|x - 1|.$$

Thus, if $|x - 1| < \delta$, then

$$|x^2 - 1| < (\delta + 2)\delta.$$

Hence, if $0 < \delta < \sqrt{1 + \epsilon} - 1$, then $|x^2 - 1| < \epsilon$. Here $\sqrt{1 + \epsilon} - 1$ is a root of the equation $(\delta + 2)\delta = \epsilon$.

However, the above argument requires us to solve an algebraic equation. This is difficult in general. In fact, we can adapt the “loose-and-track” approach to bypass this difficulty, as follows. In the sequence of inequalities

$$|x^2 - 1| = |x + 1| \cdot |x - 1| \stackrel{\textcircled{1}}{\leq} (|x - 1| + 2) \cdot |x - 1| \stackrel{\textcircled{2}}{<} 3|x - 1| \stackrel{\textcircled{3}}{<} \epsilon,$$

the inequality ① holds for all real x ; the inequality ② holds under the condition $|x - 1| < 1$; the last inequality ③ holds under the condition that $|x - 1| < \epsilon/3$. Thus, the whole sequential inequalities hold when

$$|x - 1| < \min \{1, \epsilon/3\}.$$

In particular, they hold for

$$0 < |x - 1| < \min \{1, \epsilon/3\}.$$

Now we can write down the proof. For any $\epsilon > 0$, we choose $\delta = \min\{1, \epsilon/3\}$. It is obvious that $\delta > 0$. When $0 < |x - 1| < \delta$, we have $|x - 1| < 1$, and

$$|x^2 - 1| = |x + 1| \cdot |x - 1| \leq (|x - 1| + 2) \cdot |x - 1| < 3|x - 1| < \epsilon. \quad \blacksquare$$

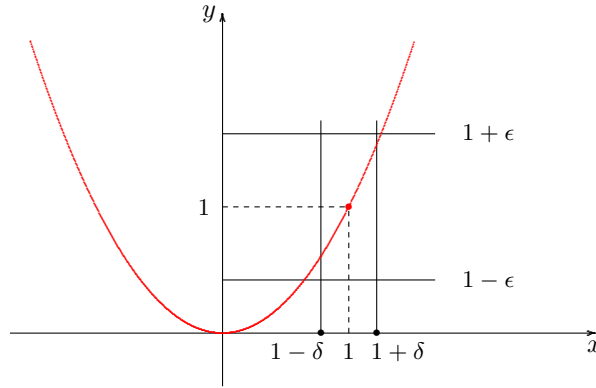


Fig. 2.5 $\lim_{x \rightarrow 1} x^2 = 1$.

Example 2.2.17 For $a > 0$, we try to rigorously prove $\lim_{x \rightarrow a} \sqrt{x} = \sqrt{a}$. Note that $0 < |x - a| < \delta$ implies that

$$|\sqrt{x} - \sqrt{a}| = \frac{|x - a|}{\sqrt{x} + \sqrt{a}} \leq \frac{|x - a|}{\sqrt{a}} < \frac{\delta}{\sqrt{a}}.$$

Therefore for any $\epsilon > 0$, we may choose $\delta = \sqrt{a}\epsilon > 0$. Then for $0 < |x - a| < \delta$, we have

$$|\sqrt{x} - \sqrt{a}| = \frac{|x - a|}{\sqrt{x} + \sqrt{a}} < \frac{\delta}{\sqrt{a}} = \epsilon. \quad \blacksquare$$

Example 2.2.18 To use the definition to prove that $\lim_{x \rightarrow +\infty} \frac{2x^2 + x}{x^2 - 2} = 2$, we may apply the “loose-and-track” approach. For the sequence of estimations, we have

$$\left| \frac{2x^2 + x}{x^2 - 2} - 2 \right| = \left| \frac{x + 4}{x^2 - 2} \right| \stackrel{\textcircled{1}}{=} \frac{|x| + 4}{x^2/2 + |x^2/2 - 2|} \stackrel{\textcircled{2}}{<} \frac{2x}{x^2/2} = \frac{4}{x} \stackrel{\textcircled{3}}{<} \epsilon,$$

where the equality ① holds for $x > 2$, the inequality ② holds under the condition $x > 4$; the last inequality ③ holds under the condition that $x > 4/\epsilon$. Thus, the whole sequential inequalities hold when

$$x > \max \{4, 4/\epsilon\}.$$

Now we can write down the rigorous proof of the limit $\lim_{x \rightarrow +\infty} \frac{2x^2 + x}{x^2 - 2} = 2$ as follows.

Given $\epsilon > 0$, take $K = \max \{4, 4/\epsilon\}$. When $x > K$, we have

$$\left| \frac{2x^2 + x}{x^2 - 2} - 2 \right| = \left| \frac{x + 4}{x^2 - 2} \right| = \frac{|x| + 4}{x^2/2 + |x^2/2 - 2|} < \frac{2x}{x^2/2} = \frac{4}{x} < \epsilon.$$

By the definition, we have

$$\lim_{x \rightarrow +\infty} \frac{2x^2 + x}{x^2 - 2} = 2. \quad \blacksquare$$

The example above is inspired by the limit of similar sequences. The subsequence examples are all inspired by the limit of similar sequences.

Example 2.2.19 For $a > 0$, we have $\lim_{n \rightarrow \infty} \sqrt[n]{a} = 1$. This suggests

$$\lim_{x \rightarrow \infty} a^{\frac{1}{x}} = 1,$$

or

$$\lim_{x \rightarrow 0} a^x = 1.$$

We prove the first statement by comparing x with natural numbers.

Let $a \geq 1$. For any $x > 1$, we have $n < x \leq n + 1$ for some (actually unique) natural number n . Then by $a \geq 1$, we get

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

We already know the left and right sides have limit 1 as $n \rightarrow \infty$. So the sandwich rule should imply that the limit of $a^{\frac{1}{x}}$ as $x \rightarrow +\infty$ should also be 1. However, we cannot quote the sandwich rule directly because the situation we have is a function sandwiched between two sequences. So we basically need to repeat the proof of the sandwich rule here.

Now we can present the formal proof. For any $\epsilon > 0$, by $\lim_{n \rightarrow \infty} a^{\frac{1}{n}} = 1$, there is N , such that

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

Then for $x > N + 1$, we find a natural number n satisfying $n < x \leq n + 1$ and have

$$\begin{aligned} x > N + 1 &\implies n \geq x - 1 > N \\ &\implies \left| a^{\frac{1}{n}} - 1 \right| < \epsilon, \quad \left| a^{\frac{1}{n+1}} - 1 \right| < \epsilon \\ &\implies 1 + \epsilon > a^{\frac{1}{n}} \geq a^{\frac{1}{x}} \geq a^{\frac{1}{n+1}} > 1 - \epsilon \\ &\implies \left| a^{\frac{1}{x}} - 1 \right| < \epsilon. \end{aligned}$$

This completes the proof that $\lim_{x \rightarrow +\infty} a^{\frac{1}{x}} = 1$ for $a \geq 1$. The case of the left limit and the case of $0 < a \leq 1$ can be converted to the proven case. ■

Example 2.2.20 For $|a| < 1$, the limit $\lim_{n \rightarrow \infty} a^n = 0$ suggests $\lim_{x \rightarrow +\infty} a^x = 0$. However, since x may not be integer, the number a in $\lim_{x \rightarrow +\infty} a^x = 0$ should be assumed to satisfy $0 \leq a < 1$.

Again a formal proof can be obtained by comparing the function and the sequence. For any $\epsilon > 0$, by $\lim_{n \rightarrow \infty} a^n = 0$, there is N , such that $n > N$ implies $a^n < \epsilon$. Then for $x > N + 1$, we find a natural number n satisfying $n < x \leq n + 1$ and have

$$x > N + 1 \implies n \geq x - 1 > N \implies a^x < a^n < \epsilon.$$

Similarly, we can prove $\lim_{x \rightarrow -\infty} a^x = 0$ when $a > 1$. ■

Example 2.2.21 The limit $\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = e$ suggests that

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

or

$$\lim_{x \rightarrow 0} (1 + x)^{\frac{1}{x}} = e.$$

We prove $\lim_{x \rightarrow +\infty} \left(1 + \frac{1}{x}\right)^x = e$ by comparing x with n . Note that if $n < x \leq n + 1$, then

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

Since

$$\begin{aligned} \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n+1}\right)^n &= \frac{\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n+1}\right)^{n+1}}{\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n+1}\right)} = e, \\ \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^{n+1} &= \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right) = e, \end{aligned}$$

a sandwich type argument can be made.

Here is the formal proof. For any $\epsilon > 0$, there is N , such that

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

Then for $x > N + 1$, we find a natural number n satisfying $n < x \leq n + 1$ and have

$$\begin{aligned}
 x > N + 1 &\implies n \geq x - 1 > N \\
 &\implies \left| \left(1 + \frac{1}{n+1}\right)^n - e \right| < \epsilon, \quad \left| \left(1 + \frac{1}{n}\right)^{n+1} - e \right| < \epsilon \\
 &\implies e - \epsilon < \left(1 + \frac{1}{n+1}\right)^n < \left(1 + \frac{1}{x}\right)^x < \left(1 + \frac{1}{n}\right)^{n+1} < e + \epsilon \\
 &\implies \left| \left(1 + \frac{1}{x}\right)^x - e \right| < \epsilon.
 \end{aligned}$$

The limit $\lim_{x \rightarrow -\infty} \left(1 + \frac{1}{x}\right)^x = e$ can be similarly proved by making use of $\lim_{n \rightarrow \infty} \left(1 - \frac{1}{n}\right)^{-n} = e$. ■

The rigorous definition also allows us to prove some properties of limits.

Example 2.2.22 [Proof of Proposition 2.2.7] Assume $\lim_{x \rightarrow a} f(x) = L$. 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

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

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

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

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

Therefore

$$0 < |x - a| < \min\{\delta^+, \delta^-\} \implies 0 < x - a < \delta^+ \text{ or } 0 > x - a > -\delta^- \implies |f(x) - L| < \epsilon.$$

This proves that $\lim_{x \rightarrow a} f(x) = L$. ■

Example 2.2.23 [Proof of Proposition 2.2.6] Assume $\lim_{x \rightarrow a} f(x) = b$ and $\lim_{y \rightarrow b} g(y) = g(b)$. We will prove $\lim_{x \rightarrow a} g(f(x)) = g(b)$.

By $\lim_{y \rightarrow b} g(y) = g(b)$, 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

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

By $\lim_{x \rightarrow a} f(x) = b$, for the $\mu > 0$ just found above, there is $\delta > 0$, such that

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

Combining this with the implication above (by taking $y = f(x)$), we get

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

This completes the proof that $\lim_{x \rightarrow a} g(f(x)) = g(b)$.

The proof of the composition rule under the other assumption is similar. ■

Relations with Limit of Sequence

We saw some limits of functions inspired by the similar limits of sequences. This suggests a relation between the two kinds of limits.

Proposition 2.2.10 *A function $f(x)$ converges to a finite limit L at $x = a$ if and only if for any sequence $\{x_n\}$ satisfying $x_n \neq a$ and $\lim_{n \rightarrow \infty} x_n = a$, we have $\lim_{n \rightarrow \infty} f(x_n) = L$.*

Example 2.2.24 By $\lim_{x \rightarrow \infty} \left(1 + \frac{1}{x}\right)^x = e$, we get $\lim_{n \rightarrow \infty} \left(1 + \frac{1}{\sqrt{n}}\right)^{\sqrt{n}} = e$. By $\lim_{x \rightarrow 0} \frac{\sin x}{x} = 1$, we get $\lim_{n \rightarrow \infty} n \sin \frac{1}{n} = 1$. By $\lim_{x \rightarrow 0} 2^x = 1$, we get $\lim_{n \rightarrow \infty} 2^{\frac{n}{n^2-1}} = 1$. ■

Example 2.2.25 The limit $\lim_{x \rightarrow 0} \cos \frac{1}{x}$ diverges because for the sequence $\left\{x_n = \frac{1}{n\pi}\right\}$ that converges to 0, the sequence $\{f(x_n) = (-1)^n\}$ diverges.

The limit $\lim_{x \rightarrow +\infty} x^{\cos x}$ diverges because the sequence $\{f(n\pi) = (n\pi)^{(-1)^n}\}$ diverges. ■

Example 2.2.26 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 can find a sequence of rational numbers $\{x_n\}$ and another sequence of irrational numbers $\{y_n\}$ such that both never equal to a and both converge to a . Then $\lim_{n \rightarrow \infty} f(x_n) = 1 \neq \lim_{n \rightarrow \infty} f(y_n) = 0$. This implies that $\lim_{x \rightarrow a} f(x)$ never converges. ■

2.2.3 Infinite Limit

The divergence of $\lim_{x \rightarrow 0} \frac{1}{x}$ and $\lim_{x \rightarrow 0} x \frac{1}{x}$ are of different nature. The first limit diverges because it gets larger and larger, indeed going to infinity. In contrast, the second limit is bounded and diverges

because it approaches to many finite targets, or the finite target is not unique. The behavior of the first type of divergence can be made more precise.

Definition 2.2.11 *A function diverges to infinity at $x = a$ if $f(x)$ approaches infinity as x approaches a . More precisely, for any B , there is $\delta > 0$, such that $0 < |x - a| < \delta$ implies $|f(x)| > B$. In this case, we denote*

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

If $|f(x)| > B$ is replaced by $f(x) > B$, we get the definition of $\lim_{x \rightarrow a} f(x) = +\infty$. If replaced by $f(x) < -B$, we get $\lim_{x \rightarrow a} f(x) = -\infty$. Moreover, the limits such as $\lim_{x \rightarrow -\infty} f(x) = +\infty$ can also be similarly defined.

The infinite limit has many properties similar to the usual finite limit. Here are some examples.

1. If $\lim_{x \rightarrow a} f(x) = +\infty$ and $\lim_{x \rightarrow a} g(x) = +\infty$, then $\lim_{x \rightarrow a} (f(x) + g(x)) = +\infty$. The property may be symbolically written as $(+\infty) + (+\infty) = +\infty$.
2. If $\lim_{x \rightarrow a} f(x) = +\infty$ and $\lim_{x \rightarrow a} g(x) < 0$, then $\lim_{x \rightarrow a} f(x)g(x) = -\infty$. This can be symbolically written as $(+\infty) \cdot (-L) = -\infty$ for $L > 0$.
3. If $\lim_{x \rightarrow a} f(x) = +\infty$ and $g(x) \geq f(x)$, then $\lim_{x \rightarrow a} g(x) = +\infty$. This is the sandwich rule.
4. If $\lim_{x \rightarrow a} f(x) = +\infty$ and $\lim_{y \rightarrow \infty} g(y) = L$, then $\lim_{x \rightarrow a} g(f(x)) = L$. This is the composition rule.

However, one needs to be careful in extending the usual properties, because equalities such as $\infty + \infty = \infty$ are not true.

Example 2.2.27 We have $\lim_{x \rightarrow +\infty} x = +\infty$. Then $\lim_{x \rightarrow +\infty} x^2 = \lim_{x \rightarrow +\infty} x \cdot \lim_{x \rightarrow +\infty} x = (+\infty) \cdot (+\infty) = +\infty$. In general, we have $\lim_{x \rightarrow +\infty} x^p = +\infty$ for $p > 0$. ■

Example 2.2.28 By Example 2.2.20, for $a > 1$, we have

$$\lim_{x \rightarrow +\infty} a^x = \frac{1}{\lim_{x \rightarrow +\infty} (a^{-1})^x} = \frac{1}{0^+} = +\infty,$$

Similarly, for $0 < a < 1$, we have $\lim_{x \rightarrow -\infty} a^x = +\infty$. ■

Example 2.2.29

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

Asymptotic Lines

Graphically, the limit $\lim_{x \rightarrow \infty} \frac{1}{x} = 0$ means that the function $\frac{1}{x}$ approaches the line $y = 0$ (the x -axis) when x goes to infinity. The x -axis is the horizontal asymptote of the function. Similarly, the limit

$\lim_{x \rightarrow 0} \frac{1}{x} = \infty$ means that the function $\frac{1}{x}$ approaches the line $x = 0$ (the y -axis) when x goes to 0. The y -axis is then the vertical asymptote of the function. In general, we can talk about any slant line being asymptote.

Definition 2.2.12 A line $y = ax + b$ is a **slant asymptote** of the graph $y = f(x)$ if one of the followings holds

$$\lim_{x \rightarrow +\infty} (f(x) - ax - b) = 0, \quad \lim_{x \rightarrow -\infty} (f(x) - ax - b) = 0.$$

When $a = 0$, $y = b$ is the **horizontal asymptote**. A line $x = a$ is a **vertical asymptote** of the graph $y = f(x)$ if one of the followings holds

$$\lim_{x \rightarrow a^+} f(x) = +\infty, \quad \lim_{x \rightarrow a^-} f(x) = +\infty, \quad \lim_{x \rightarrow a^+} f(x) = -\infty, \quad \lim_{x \rightarrow a^-} f(x) = -\infty.$$

The coefficients in a slant asymptote are calculated by (∞ is $+\infty$ or $-\infty$)

$$a = \lim_{x \rightarrow \infty} \frac{f(x)}{x}, \quad b = \lim_{x \rightarrow \infty} (f(x) - ax).$$

Example 2.2.30 Since $\lim_{x \rightarrow \frac{\pi}{2}^-} \tan x = +\infty$ and $\lim_{x \rightarrow -\frac{\pi}{2}^+} \tan x = -\infty$, the lines $x = \frac{\pi}{2}$ and $x = -\frac{\pi}{2}$ are vertical asymptotes of $\tan x$. ■

Example 2.2.31 The function $f(x) = \frac{x+1}{x-1}$ has limits $\lim_{x \rightarrow 1^+} f(x) = +\infty$, $\lim_{x \rightarrow 1^-} f(x) = -\infty$ and $\lim_{x \rightarrow \infty} f(x) = 1$. Therefore $x = 1$ is a vertical asymptote $x = 1$ and $y = 1$ is a horizontal asymptote. ■

Example 2.2.32 The function $f(x) = \frac{x^2+2}{x-1} = x+2 + \frac{2}{x-1}$ satisfies $\lim_{x \rightarrow \infty} \left(\frac{x^2+2}{x-1} - x - 2 \right) = 0$. Therefore the line $y = x + 2$ is an asymptote of $f(x)$. ■

Example 2.2.33 We may draw a rough graph of the function $f(x) = \frac{x(x-1)(x-2)}{x^3+1}$ by exhibiting x -intercepts and asymptotes. From the zeros of the numerator, we get the x -intercepts $x = 0, 1, 2$. By the limits

$$\lim_{x \rightarrow -1^-} f(x) = +\infty, \quad \lim_{x \rightarrow -1^+} f(x) = -\infty, \quad \lim_{x \rightarrow +\infty} f(x) = 1, \quad \lim_{x \rightarrow -\infty} f(x) = 1,$$

the graph has a vertical asymptote $x = -1$ and a horizontal asymptote $y = 1$. ■

Exercises

2.2.1 Find the limit:

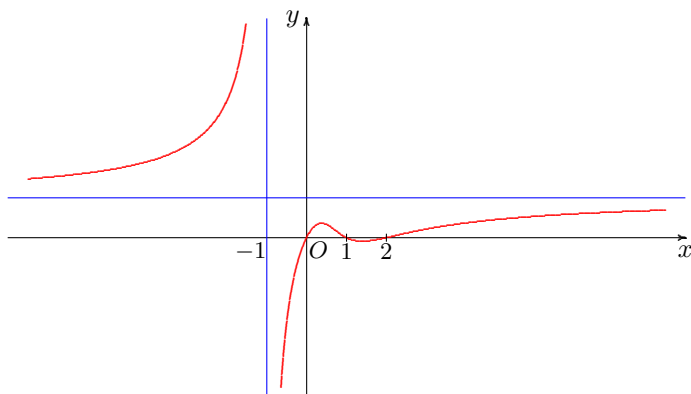


Fig. 2.6 Graph of $\frac{x(x-1)(x-2)}{x^3+1}$.

1. $\lim_{x \rightarrow 0} \frac{x^2 - 1}{x^2 + x - 2}$;
2. $\lim_{x \rightarrow 1} \frac{x^2 - 1}{x^2 + x - 2}$;
3. $\lim_{x \rightarrow +\infty} \frac{x^2 - 1}{x^2 + x - 2}$;
4. $\lim_{x \rightarrow a^+} \frac{\sqrt{x} - \sqrt{a} + \sqrt{x-a}}{\sqrt{x^2 - a^2}}$, where a is a positive real numbers;
5. $\lim_{x \rightarrow +\infty} \frac{(x+1)(x^2+1) \cdots (x^n+1)}{[(nx)^n + 1]^{\frac{n+1}{2}}}$;
6. $\lim_{x \rightarrow 1} \frac{x^n - 1}{x^m - 1}$;
7. $\lim_{x \rightarrow 1} \frac{x + x^2 + \cdots + x^n - n}{x - 1}$;
8. $\lim_{x \rightarrow +\infty} \left[\sqrt[3]{x^3 + x^2 + 1} - \sqrt[3]{x^3 - x^2 + 1} \right]$;
9. $\lim_{x \rightarrow +\infty} \left[\sqrt{(x+a)(x+b)} - x \right]$, where a and b are two real numbers;
10. $\lim_{x \rightarrow +\infty} \frac{\sin x}{x}$;
11. $\lim_{x \rightarrow +\infty} x^{1/2} \left(\sin \sqrt{x^2 + 1} - \sin x \right)$;
12. $\lim_{x \rightarrow 0^+} \frac{\tan(\sin x)}{\sin \sqrt{x}}$.

2.2.2 Use the definition to prove the following.

1. $\lim_{x \rightarrow 4} \sqrt{x} = 2$;
2. $\lim_{x \rightarrow 4} \frac{1}{\sqrt{x}} = \frac{1}{2}$;
3. $\lim_{x \rightarrow -\infty} \frac{x+1}{\sqrt{x^2+1}} = -1$;
4. $\lim_{x \rightarrow +\infty} [\sqrt{x+1} - \sqrt{x}] = 0$.

2.2.3 Let

$$R(x) = \frac{x^n + a_1x^{n-1} + \cdots + a_n}{x^m + b_1x^{m-1} + \cdots + b_m}.$$

Prove that

$$\lim_{x \rightarrow +\infty} R(x) = \begin{cases} +\infty, & \text{if } n > m, \\ 1, & \text{if } n = m, \\ 0, & \text{if } n < m. \end{cases}$$

2.2.4 Find two slant asymptotes of the function $f(x) = \sqrt{x^2 + x + 1}$.

2.2.5 For what values of a , b and c , does the function $y = \sqrt{ax^2 + bx + c}$ have an asymptote when x approaches $+\infty$? Please find the equation of the asymptote.

2.2.6 Determine all asymptotes of the function

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

2.2.7 Find the tangent line equation to the curve $y = x + \frac{2}{x}$, $x > 0$, that is perpendicular to the slant asymptote of the curve.

2.2.8 Suppose $f(x) \leq g(x)$ for all $x \in (a, b)$ except at a point $c \in (a, b)$. If both f and g have limits at $x = c$, then $\lim_{x \rightarrow c} f(x) \leq \lim_{x \rightarrow c} g(x)$.

2.2.9 Prove the Sandwich Theorem. (Proposition 2.2.4)

2.2.10 Prove that

$$\lim_{x \rightarrow +\infty} \frac{x + \sin x}{x + \cos x} = 1.$$

2.3 CONTINUOUS FUNCTIONS

Consider functions in Fig. 2.7. The functions f_1 and f_4 move continuously as x moves along the real line. The function f_2 is a linear function except not defined at $x = 1$, and thus is “broken” at $x = 1$. The function f_3 goes to infinity as x approaches 0, and skips $x = 0$ on the real line. The function f_5 jump from -1 to 1 at 0 .

The functions f_1 and f_4 are continuous in the sense that they do not break. The functions f_2 , f_3 and f_5 have discontinuity (i.e., breaks) at $x = 1$, $x = 0$ and $x = 0$ respectively. The reason for discontinuity varies. For f_2 , the limit $\lim_{x \rightarrow 1} f_2(x) = 1$ converges, but $f_2(1)$ is not defined. For f_3 , the limit $\lim_{x \rightarrow 0} f_3(x)$ diverges. In fact, both the left and right limits diverge. For f_5 , the limit $\lim_{x \rightarrow 0} f_5(x)$ diverges, although both the left and the right limit converge. For the continuous functions f_1 and f_4 , however, the limits $\lim_{x \rightarrow a} f_1(x)$ and $\lim_{x \rightarrow a} f_4(x) = f_4(a)$ converge and are equal to the value of the function at a .

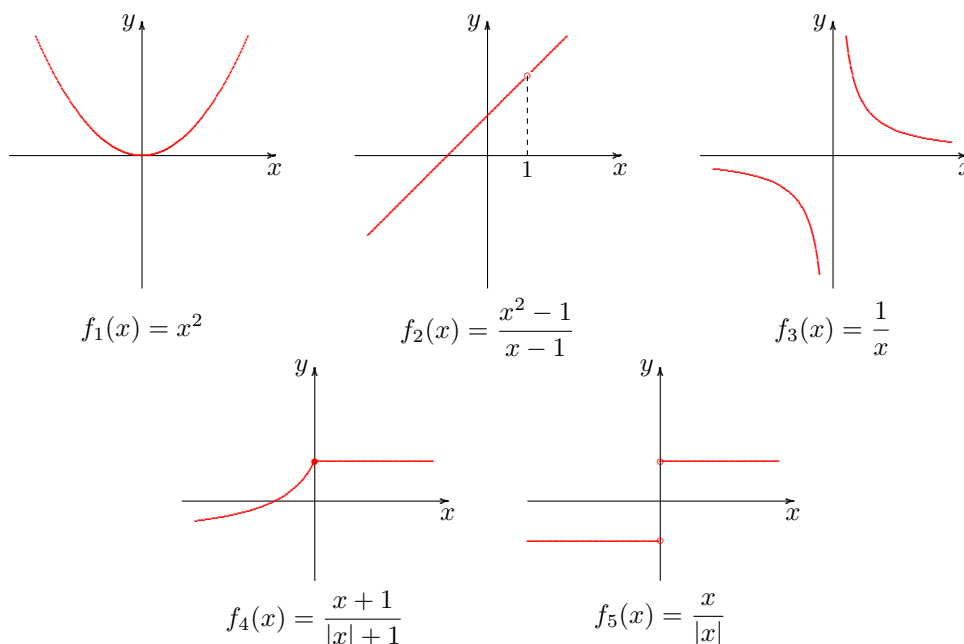


Fig. 2.7 Examples of functions.

Definition 2.3.1 A function f is **continuous** at a if $\lim_{x \rightarrow a} f(x) = f(a)$. In other words, for any $\epsilon > 0$, there is $\delta > 0$, such that $|x - a| < \delta$ implies $|f(x) - f(a)| < \epsilon$.

Implicit in the definition is that the function has to be defined at a , because the value of the function at a must be the value of the limit. Technically, the key difference from the definition of limit is that $0 < |x - a| < \delta$ is replaced by $|x - a| < \delta$. The difference is subtle but crucial.

If a function $f(x)$ is defined on $[a, b)$, we may also define the right continuity at a by requiring $\lim_{x \rightarrow a^+} f(x) = f(a)$. Similarly, a function $f(x)$ defined on $(b, a]$ is left continuous at a if $\lim_{x \rightarrow a^-} f(x) = f(a)$. A function is continuous at a point if and only if it is left and right continuous at the point.

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.

By Example 2.2.1, we know polynomials are continuous and rational functions are continuous wherever it is defined. By Examples 2.2.7 and 2.2.9, we know the power function x^p is continuous on $[0, \infty)$ for $p > 0$ and is continuous on $(0, \infty)$ for all p .

By properties of limit, we know the arithmetic operations of continuous functions are still continuous (wherever the new function is defined). Moreover, the compositions of continuous functions are continuous.

Example 2.3.1 In Example 2.2.10, we know the trigonometric functions $\sin x$, $\cos x$ and $\tan x$ are continuous at 0. What about the continuity at the other places?

For $|x - a| < \pi$, we have

$$|\sin x - \sin a| = 2 \left| \cos \frac{x+a}{2} \sin \frac{x-a}{2} \right| \leq 2 \left| \sin \frac{x-a}{2} \right| \leq |x-a|.$$

Based on this, it is easy to see that $\lim_{x \rightarrow a} \sin x = \sin a$. Therefore $\sin x$ is continuous everywhere. By the composition rule, $\cos x = \sin\left(\frac{\pi}{2} - x\right)$ is also continuous everywhere. Then by the arithmetic rule, $\tan x = \frac{\sin x}{\cos x}$ is continuous everywhere except at $\left(n + \frac{1}{2}\right)\pi$. ■

Example 2.3.2 By Example 2.2.19, the exponential function a^x is continuous at 0. What about the continuity at the other places?

By introducing $y = x - c$ (and essentially using the composition rule), we have

$$\lim_{x \rightarrow c} a^x = \lim_{y \rightarrow 0} a^{c+y} = \lim_{y \rightarrow 0} a^c a^y = a^c \lim_{y \rightarrow 0} a^y = a^c \cdot 1 = a^c.$$

Therefore the exponential function is continuous everywhere. ■

Example 2.3.3 The Dirichlet function in example 2.2.26 is not continuous everywhere, because the limit does not converge anywhere.

On the other hand, the function

$$xD(x) = \begin{cases} x, & \text{if } x \in \mathbb{Q}, \\ 0, & \text{if } x \notin \mathbb{Q} \end{cases}$$

also does not converge at any $a \neq 0$, by the same reason as the Dirichlet function. However, the function is continuous at 0:

$$\lim_{x \rightarrow 0, x \in \mathbb{Q}} xD(x) = \lim_{x \rightarrow 0, x \in \mathbb{Q}} x = 0 = 0D(0), \quad \lim_{x \rightarrow 0, x \notin \mathbb{Q}} xD(x) = \lim_{x \rightarrow 0, x \notin \mathbb{Q}} 0 = 0 = 0D(0). \quad \blacksquare$$

By Proposition 2.2.10, if $f(x)$ is continuous at a , then for any sequence $\{x_n\}$ satisfying $x_n \neq a$ and $\lim_{n \rightarrow \infty} x_n = a$, we have

$$\lim_{n \rightarrow \infty} f(x_n) = f(a) = f\left(\lim_{n \rightarrow \infty} x_n\right).$$

In fact, under the continuity assumption, the condition $x_n \neq a$ can be dropped. Therefore the continuity of a function means that the limit operation commutes with the evaluation of the function.

Similarly, Proposition 2.2.6 (composition rule) holds when the outside function is continuous. This means that if $f(x)$ is continuous at b and $\lim_{x \rightarrow a} g(x) = b$ (the notations f and g are exchanged from the proposition), then

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

This reinforces the interpretation that the continuity means that the limit and the evaluation can be exchanged.

Example 2.3.4 Since e^x and $\sin x$ are continuous, we have

$$\lim_{n \rightarrow \infty} e^{\frac{n}{n^2-1}} = e^{\lim_{n \rightarrow \infty} \frac{n}{n^2-1}} = e^0 = 1,$$

and

$$\lim_{x \rightarrow 0} \sin(\tan(2^x - 1)) = \sin\left(\lim_{x \rightarrow 0} \tan(2^x - 1)\right) = \sin\left(\tan \lim_{x \rightarrow 0} (2^x - 1)\right) = \sin(\tan 0) = 0. \quad \blacksquare$$

Intermediate Value Theorem

Suppose $f(x)$ is a continuous function on $[a, b]$. Then as x goes from a to b , the value $f(x)$ should “gradually” go from $f(a)$ to $f(b)$. In other words, there should not be any “jump” or “gap” in the values of the function. Consequently, any value between $f(a)$ and $f(b)$ should be reached.

Theorem 2.3.2 (Intermediate Value Theorem) *If a function $f(x)$ is continuous on $[a, b]$, the for any number γ between $f(a)$ and $f(b)$, there is $c \in [a, b]$, such that $f(c) = \gamma$.*

Example 2.3.5 The polynomial $p(x) = x^3 - 3x + 1$ satisfies $p(0) = 1$ and $p(1) = -1$. Since $p(x)$ is continuous on $[0, 1]$ and $1 > 0 > -1$, p will reach the intermediate value 0 somewhere on $[0, 1]$. In other words, the polynomial has at least one root on $(0, 1)$.

To get more precise location of the root, we may try to evaluate p at $0.1, 0.2, \dots, 0.9$ and find $p(0.3) = 0.727$ and $p(0.4) = -0.136$. Therefore we can further nail down the root to be between 0.3 and 0.4. ■

The following corollary is handy to determine existence of roots for continuous functions.

Corollary 2.3.3 *Assume f is a continuous function on $[a, b]$, with $f(a)$ and $f(b)$ having opposite signs. Then f has at least one root in (a, b) .*

Example 2.3.6 For any polynomial $p(x) = c_n x^n + c_{n-1} x^{n-1} + \dots + c_1 x + c_0$ with $c_n \neq 0$, we have

$$\lim_{n \rightarrow +\infty} p(x) = \lim_{n \rightarrow +\infty} x^n \left(c_n + \frac{c_{n-1}}{x} + \dots + \frac{c_1}{x^{n-1}} + \frac{c_0}{x^n} \right) = (+\infty) \cdot c_n = \begin{cases} +\infty, & \text{if } c_n > 0, \\ -\infty, & \text{if } c_n < 0. \end{cases}$$

Similarly, we have

$$\lim_{n \rightarrow -\infty} p(x) = \begin{cases} -\infty, & \text{if } c_n > 0, n \text{ odd;} \\ +\infty, & \text{if } c_n < 0, n \text{ even.} \end{cases}$$

In particular, if n is odd, then $\lim_{n \rightarrow +\infty} p(x)$ and $\lim_{n \rightarrow -\infty} p(x)$ are infinities of different sign. Therefore we can find large $R > 0$, such that $p(-R)$ and $p(R)$ have different signs. Then by the Intermediate Value Theorem, the polynomial must have a root on $(-R, R)$. This proves that any polynomial of odd degree must have at least one root.

Incidentally, for $p(x) = (x^3 - 3x + 1)$, we know $\lim_{n \rightarrow +\infty} p(x) = +\infty$ and $\lim_{n \rightarrow -\infty} p(x) = -\infty$. Combined with $p(0) > 0$ and $p(1) < 0$ and using the Intermediate Value Theorem, we conclude that $p(x)$ has one root < 0 , one root between 0 and 1, and one root > 1 . ■

Example 2.3.7 We know

$$\lim_{x \rightarrow \frac{\pi}{2}^-} \tan x = \frac{\lim_{x \rightarrow \frac{\pi}{2}^-} \sin x}{\lim_{x \rightarrow \frac{\pi}{2}^-} \cos x} = \frac{1}{0^+} = +\infty, \quad \lim_{x \rightarrow -\frac{\pi}{2}^+} \tan x = -\infty.$$

Therefore for any number γ , we can find $a > -\frac{\pi}{2}$ and very close to $-\frac{\pi}{2}$, such that $\tan a < \gamma$. We can also find $b < \frac{\pi}{2}$ and very close to $\frac{\pi}{2}$, such that $\tan b > \gamma$. Then $\tan x$ is continuous on $[a, b]$ and γ is a number between $\tan a$ and $\tan b$. By the Intermediate Value Theorem, we have $\gamma = \tan c$ for some $c \in [a, b]$. This shows that any number is the tangent of some angle between $-\frac{\pi}{2}$ and $\frac{\pi}{2}$. ■

Example 2.3.8 The function $f(x) = \begin{cases} x, & \text{if } -1 \leq x \leq 0, \\ x^2 + 1, & \text{if } 0 < x \leq 1 \end{cases}$ is shown in Fig. 2.8. It has values $f(-1) = -1$ and $f(1) = 2$. However, it does not take any number on $(0, 1]$ as value. The problem is that the function is not continuous at 0, where a jump in value misses the interval $(0, 1]$. Consequently, the conclusion of the Intermediate Value Theorem does not apply. ■

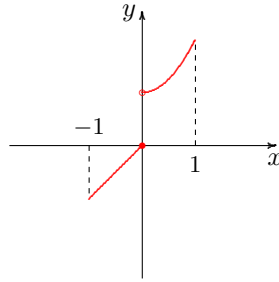


Fig. 2.8 A piecewise continuous function.

Inverse Function

The logarithmic function is the inverse of the exponential function. The trigonometric functions also have their inverses. The continuity of these functions are guaranteed by the following result.

Proposition 2.3.4 *If f is an increasing continuous function on an interval, then f is invertible and the inverse is increasing and continuous. The same holds for decreasing continuous functions.*

Proof. Assume f is increasing. It is known that f is one-to-one. Since $f(x)$ is continuous and increasing, we know that for any $x \in (a, b)$, $f(a) < f(x) < f(b)$. By the Intermediate Value Theorem, for any $y \in (f(a), f(b))$, there exists $x \in (a, b)$ such that $y = f(x)$. Thus f is onto from $[a, b]$ to $[f(a), f(b)]$. Hence the function $f : [a, b] \rightarrow [f(a), f(b)]$ is bijective, and therefore, the function f has the inverse function $g : [f(a), f(b)] \rightarrow [a, b]$.

To show that g is a continuous function, we need to show that for any $\epsilon > 0$, there exists $\delta > 0$ such that whenever $|y - y_0| < \delta$, we have $|g(y) - g(y_0)| < \epsilon$.

In fact, let $y_0 = f(x_0)$ and $y = f(x)$. Thus $g(y_0) = x_0$ and $g(y) = x$. Equivalently we need to find δ such that when $|f(x) - f(x_0)| < \delta$, we have $|x - x_0| < \epsilon$. Choose $\delta = \min\{f(x_0) - f(x_0 - \epsilon), f(x_0 + \epsilon) - f(x_0)\}$. When $|y - y_0| < \delta$, we have

$$f(x_0 - \epsilon) - f(x_0) \leq -\delta < y - y_0 < \delta \leq f(x_0 + \epsilon) - f(x_0),$$

which implies

$$f(x_0 - \epsilon) < f(x) < f(x_0 + \epsilon).$$

These give the inequalities

$$x_0 - \epsilon < x < x_0 + \epsilon,$$

since otherwise it will lead to a contradiction.

The proof on g being increasing leaves as an exercise. ■

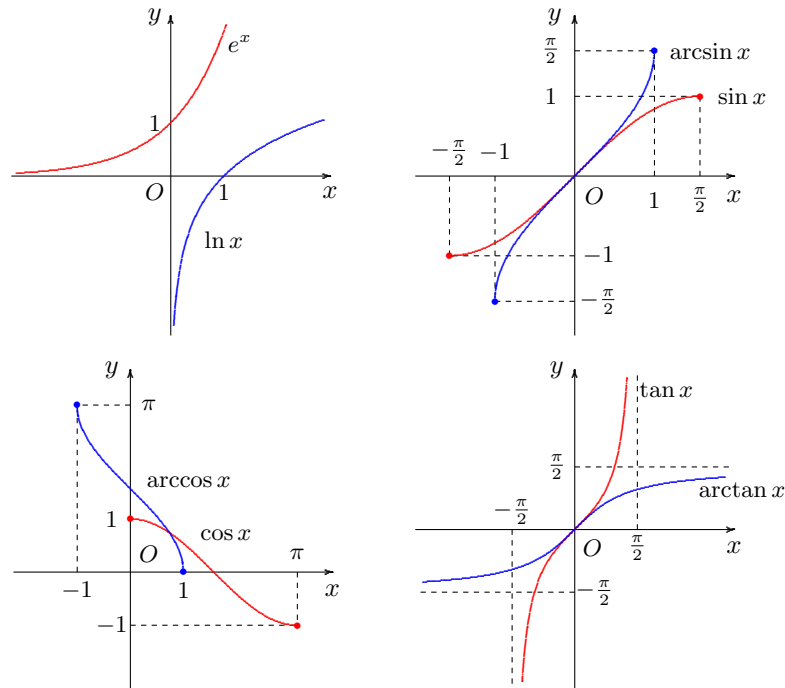


Fig. 2.9 The function e^x , $\sin x$, $\cos x$, $\tan x$ and their inverse functions $\ln x$, $\arcsin x$, $\arccos x$, $\arctan x$.

Example 2.3.9 For $a > 1$, the exponential function $a^x: \mathbb{R} \rightarrow (0, +\infty)$ is increasing and continuous. The inverse is the logarithmic function $\log_a x: (0, +\infty) \rightarrow \mathbb{R}$ is also continuous and increasing.

Moreover, by the monotonic property,

$$x > a^B \iff \log_a x > B.$$

This shows that $\lim_{x \rightarrow +\infty} \log_a x = +\infty$. Similar reason also shows $\lim_{x \rightarrow 0^+} \log_a x = -\infty$.

For $0 < a < 1$, the exponential a^x is decreasing and continuous. The corresponding logarithm is also decreasing and continuous.

For the special case $a = e = \lim_{x \rightarrow 0} (1+x)^{\frac{1}{x}}$, the logarithm is called the natural logarithm and denoted simply as $\ln x$. Since $\ln x$ is continuous, by the composition rule, we have

$$\lim_{x \rightarrow 0} \frac{\ln(x+1)}{x} = \lim_{x \rightarrow 0} \ln(1+x)^{\frac{1}{x}} = \ln \lim_{x \rightarrow 0} (1+x)^{\frac{1}{x}} = \ln e = 1.$$

Moreover, the function $x^x = e^{x \ln x}$ is continuous on $(0, \infty)$ by the following reason: Since x and $\ln x$ are continuous, $x \ln x$ is continuous. Since the exponential function e^y is also continuous, the composition $e^{x \ln x}$ is also continuous.

In general, if $f(x)$ and $g(x)$ are continuous, and $f(x) > 0$, then $f(x)^{g(x)}$ is continuous. ■

Example 2.3.10 The sine function

$$\sin x : \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \rightarrow [-1, 1]$$

is increasing and continuous. Its inverse function

$$\arcsin x : [-1, 1] \rightarrow \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$$

is also increasing and continuous.

The continuity of $\arcsin x$ at 0 tells us $\lim_{x \rightarrow 0} \arcsin x = \arcsin 0 = 0$. By the composition rule, we have

$$\lim_{x \rightarrow 0} \frac{\arcsin x}{x} = \lim_{y \rightarrow 0} \frac{y}{\sin y} = 1.$$

The cosine function $\cos x : [0, \pi] \rightarrow [-1, 1]$ is decreasing and continuous. Its inverse function $\arccos x : [-1, 1] \rightarrow [0, \pi]$ is also decreasing and continuous.

The tangent function $\tan x : \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \rightarrow (-\infty, +\infty)$ is increasing and continuous. Its inverse function $\arctan x : (-\infty, +\infty) \rightarrow \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ is also decreasing and continuous. Moreover, similar to the argument for the logarithm, we have

$$\lim_{x \rightarrow \frac{\pi}{2}^-} \arctan x = +\infty, \quad \lim_{x \rightarrow -\frac{\pi}{2}^+} \arctan x = -\infty. \quad \blacksquare$$

Exercises

2.3.1 Determine the region where the function is continuous:

1. $f(x) = x \sin \frac{1}{x};$

3. $f(x) = \begin{cases} e^{-1/x^2}, & \text{if } x \neq 0, \\ 0, & \text{if } x = 0; \end{cases}$

2. $f(x) = \ln \frac{x^2}{x^2 - 1};$

4. $f(x) = \lim_{n \rightarrow +\infty} \frac{1}{1 + x^n}.$

2.3.2 Use the definition to show that the function $f(x) = x^2$ is continuous at $x = 2$.

2.3.3 Use the definition to show that the function $f(x) = \sin \sqrt{x}$ is continuous at any $x_0 \geq 0$.

2.3.4 Please give two non-constant continuous functions f and g in $\mathbb{R}$ such that

$$\lim_{x \rightarrow 0} \frac{1 + f(0)g(x)}{1 + f(x)g(0)} \neq 1.$$

2.3.5 The following functions are not defined at $x = 0$. Please definite their values at $x = 0$ so that they are continuous.

1. $f(x) = \sin x \sin \frac{1}{x};$

2. $f(x) = \frac{\sqrt{x+1} - 1}{\sqrt[3]{x+1} - 1}.$

2.3.6 For what value of A , is the function

$$f(x) = \begin{cases} \frac{x^2 - 1}{x - 1}, & \text{if } x \neq 1, \\ A, & \text{if } x = 1 \end{cases}$$

continuous everywhere?

2.3.7 Is it possible to choose a value for A , such that the function

$$f(x) = \begin{cases} \frac{1}{(x-1)^2}, & \text{if } x \neq 1, \\ A, & \text{if } x = 1 \end{cases}$$

is continuous everywhere?

2.3.8 Assume that f and g are continuous. Then

$$\min\{f(x), g(x)\} \quad \text{and} \quad \max\{f(x), g(x)\}$$

are also continuous.

2.3.9 Show that the equation

$$x^3 - 3x + 1 = 0$$

has at least three real solutions.

2.3.10 Show that the function $f(x) = x^5 - 4x + 2$ has three real roots.

2.3.11 Let $f(x) : [0, 1] \rightarrow [0, 1]$ be a continuous function. Prove that there exists at least one $c \in [0, 1]$ such that $f(c) = c$.

2.3.12 Let f be an increasing function. Show that its inverse is also an increasing function.

2.3.13 Use the limit $\lim_{x \rightarrow 0} \frac{\sin x}{x} = 1$ to prove that $\sin x$ is continuous everywhere.

2.3.14 Show that for any $x_0 \in (-\infty, +\infty)$, the sequence $\{x_n\}$ generated by

$$x_n = \sqrt{\pi} + \frac{1}{2} \sin x_{n-1}, \quad n = 1, 2, \dots,$$

is a Cauchy sequence that converges to the unique positive solution of the equation

$$x = \sqrt{\pi} + \frac{1}{2} \sin x.$$

2.4 SUMMARY

Definitions

- A sequence $\{a_n\}$ is said to be **increasing** if $a_1 \leq a_2 \leq a_3 \leq \dots \leq a_n \leq a_{n+1} \leq \dots$; and **decreasing** if $a_1 \geq a_2 \geq a_3 \geq \dots \geq a_n \geq a_{n+1} \geq \dots$.
- A sequence $\{a_n\}$ is **convergent** to L if for any $\epsilon > 0$, there exists a positive integer N such that whenever $n \geq N$, we have $|a_n - L| < \epsilon$.
- The limit of the function f at $x = a$ equals L , denoted as $\lim_{x \rightarrow a} f(x) = L$, if for any $\epsilon > 0$, there exists $\delta > 0$ such that $|f(x) - L| < \epsilon$ whenever $0 < |x - a| < \delta$.
- We say $\lim_{x \rightarrow a} f(x) = +\infty$ if for any $K > 0$, there exists $\delta > 0$ such that

$$f(x) > K, \quad \text{whenever } 0 < |x - a| < \delta.$$

- $\lim_{x \rightarrow +\infty} f(x) = c$, where c is a finite number, if for any $\epsilon > 0$, there exists $K > 0$ such that

$$|f(x) - c| < \epsilon, \quad \text{whenever } x > K.$$

- A sequence $\{a_n\}$ is called a Cauchy sequence if for any $\epsilon > 0$, there exists a positive integer N such that

$$|a_n - a_m| < \epsilon, \quad \text{whenever } n \geq N, m \geq N.$$

- A line $y = L$ is called a **horizontal asymptote** of the graph $y = f(x)$ if at least one of the followings holds

$$\lim_{x \rightarrow +\infty} f(x) = L, \quad \lim_{x \rightarrow -\infty} f(x) = L.$$

Generally, a line $y = ax + b$ is called a **slant asymptote** of the graph $y = f(x)$ if at least one of the followings holds

$$\lim_{x \rightarrow +\infty} [f(x) - ax - b] = 0, \quad \lim_{x \rightarrow -\infty} [f(x) - ax - b] = 0,$$

where a and b are constants.

A line $x = c$ is called a **vertical asymptote** of the graph $y = f(x)$ if at least one of the followings holds

$$\begin{aligned} \lim_{x \rightarrow c} f(x) = +\infty, \quad \lim_{x \rightarrow c^-} f(x) = +\infty, \quad \lim_{x \rightarrow c^+} f(x) = +\infty, \\ \lim_{x \rightarrow c} f(x) = -\infty, \quad \lim_{x \rightarrow c^-} f(x) = -\infty, \quad \lim_{x \rightarrow c^+} f(x) = -\infty. \end{aligned}$$

- A function f defined on a domain D is **continuous** at $a \in D$ if $\lim_{x \in D, x \rightarrow a} f(x) = f(a)$.

Theorems

- **The Sandwich Theorem** Given two convergent sequences $\{a_n\}$ and $\{c_n\}$ with the same limit $\lim_{n \rightarrow \infty} a_n = L = \lim_{n \rightarrow \infty} c_n$. Let $\{b_n\}$ be a sequence. Suppose there exists a positive integer K such that for $n \geq K$, we have $a_n \leq b_n \leq c_n$. Then b_n converges to L .
- $|\sin x| \leq |x| \leq |\tan x|$ for $x \in (-\pi/2, \pi/2)$.
- Any bounded increasing or decreasing sequence is convergent.
- $\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = \lim_{x \rightarrow 0} (1 + x)^{\frac{1}{x}} = e$.
- $\lim_{x \rightarrow 0} \frac{\sin x}{x} = 1$.
- If $\{a_n\}$ converges to L , then every subsequence of $\{a_n\}$ converges to L .
- **Bolzano-Weierstrass Theorem** Every bounded sequence has a convergent subsequence.
- Any convergent sequence $\{a_n\}$ is bounded.
- **The Cauchy Criterion** A sequence $\{a_n\}$ is convergent if and only if it is a Cauchy sequence.
- **The Sandwich Theorem** Suppose $f(x) \leq g(x) \leq h(x)$ for all $x \in (a, b)$ except at a point $c \in (a, b)$. If $\lim_{x \rightarrow c} f(x)$ and $\lim_{x \rightarrow c} h(x)$ exist and are equal to L , then the limit of $g(x)$ at $x = c$ exists and equals L .
- Given a function $f(x)$ defined on $(a, c) \cup (c, b)$. The limit $\lim_{x \rightarrow c} f(x)$ exists if and only if $\lim_{x \rightarrow c^-} f(x)$ and $\lim_{x \rightarrow c^+} f(x)$ exist and are equal.
- The limit $\lim_{x \rightarrow c} f(x) = L$ iff for any sequence $\{x_n\}$ with $\lim_{n \rightarrow \infty} x_n = c$, $\lim_{n \rightarrow \infty} f(x_n) = L$. Here c can be taken to be ∞ .

- Given a function f defined over a domain D and a function g defined over a domain containing $f(D)$. Suppose f is continuous at $a \in D$ and g is continuous at $f(a)$, then the composite function $g \circ f$ is continuous at a .
- **The Intermediate Value Theorem** Given a continuous function $f(x)$ over $[a, b]$, for any number c between $f(a)$ and $f(b)$, there exists a number $\xi \in [a, b]$ such that $c = f(\xi)$.
- Assume f is a continuous function on $[a, b]$, with $f(a)$ and $f(b)$ having opposite signs. Then f has at least one root in (a, b) .
- If f is an increasing continuous function on an interval, then f is invertible and the inverse is increasing and continuous.

3

Differentiation

3.1 DERIVATIVE

The concept of differentiation is motivated by many real world problems. Let us look at two of them.

Instantaneous Speed in the Motion of Free Fall

If an object is initially located at the origin with zero velocity, then the displacement $s(t)$ in free fall is a function of time t

$$s(t) = \frac{1}{2}gt^2,$$

where g is the gravitational constant. The average velocity from 1 second to t seconds is

$$v_{[1,t]} = \frac{s(t) - s(1)}{t - 1} = \frac{\frac{1}{2}gt^2 - \frac{1}{2}g1^2}{t - 1} = \frac{1}{2}g(t + 1).$$

For instance, the average velocity from 1 second to 1.1 seconds is

$$v_{[1,1.1]} = \frac{s(1.1) - s(1)}{1.1 - 1} = 1.05g,$$

and from 1 second to 1.01 seconds is

$$v_{[1,1.01]} = \frac{s(1.01) - s(1)}{1.01 - 1} = 1.005g.$$

As t approaches 1, the average velocity $v_{[1,t]}$ should approach the instantaneous velocity at 1 second

$$v(1) = \lim_{t \rightarrow 1} v_{[1,t]} = g.$$

In general, the average free fall velocity from time t to t' is

$$v_{[t,t']} = \frac{s(t') - s(t)}{t' - t} = \frac{1}{2}g(t' + t).$$

As t' approaches t , we get the instantaneous velocity at time t

$$v(t) = \lim_{t' \rightarrow t} v_{[t,t']} = gt.$$

Tangent of Plane Curve

Consider a curve $y = f(x)$ on the plane. The tangent of the curve at $P = (a, f(a))$ is the limit of the straight line connecting P to a nearby point $Q = (x, f(x))$ as x approaches a . The slope of the straight line connecting P to Q is

$$\text{slope}[PQ] = \frac{f(x) - f(a)}{x - a}.$$

Therefore the slope of the tangent line is

$$\text{slope}(P) = \lim_{Q \rightarrow P} \text{slope}[PQ] = \lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a}.$$

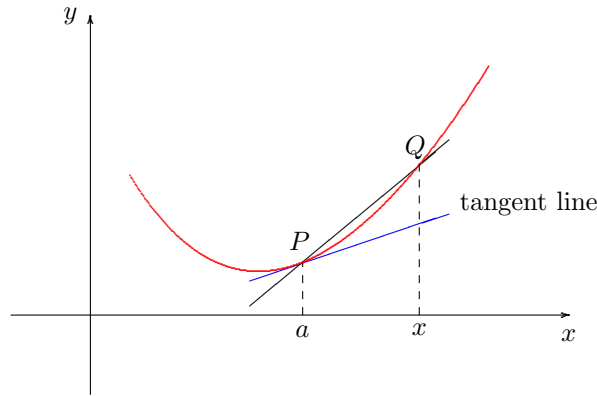


Fig. 3.1 Tangent line of plane curve

For example, the slope of the tangent line of the curve $y = x^2$ at $P = (2, 4)$ is

$$\lim_{x \rightarrow 2} \frac{x^2 - 2^2}{x - 2} = \lim_{x \rightarrow 2} (x + 2) = 4.$$

The equation of the tangent line is

$$y = 4 + 4(x - 2) = 4x - 8.$$

Definition of Derivative

Definition 3.1.1 Let $f(x)$ be a function defined near a . Its **derivative** at a is

$$f'(a) = \lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a} = \lim_{\Delta x \rightarrow 0} \frac{f(a + \Delta x) - f(a)}{\Delta x}.$$

In case the limit converges, we say the function $f(x)$ is **differentiable** at a . A function is differentiable on an open interval if it is differentiable at every point in the interval.

Note that the derivative is the limit $f'(a) = \lim_{x \rightarrow a} \frac{\Delta f}{\Delta x}$ of the quotient of the changes

$$\Delta f = f(x) - f(a) = f(a + \Delta x) - f(a), \quad \Delta x = x - a.$$

Therefore the derivative measures the instantaneous rate of the change.

If $f(x)$ is defined on $[a, b)$, then we may define the **right derivative** at a

$$f'_+(0) = \lim_{x \rightarrow a^+} \frac{f(x) - f(a)}{x - a} = \lim_{\Delta x \rightarrow 0^+} \frac{f(a + \Delta x) - f(a)}{\Delta x}.$$

If the function is defined on $(b, a]$, then we may similarly define the **left derivative** at a

$$f'_-(0) = \lim_{x \rightarrow a^-} \frac{f(x) - f(a)}{x - a} = \lim_{\Delta x \rightarrow 0^-} \frac{f(a + \Delta x) - f(a)}{\Delta x}.$$

A function is differentiable at a point if and only if it has equal left and right derivatives at the point.

Notice that $f'(0^+) = \lim_{x \rightarrow 0^+} f'(x)$. Hence, in general,

$$f'_+(0) \neq f'(0^+).$$

If $f(x)$ is differentiable at a , then

$$\lim_{x \rightarrow a} (f(x) - f(a)) = \lim_{x \rightarrow a} (x - a) \frac{f(x) - f(a)}{x - a} = 0 \cdot f'(a) = 0.$$

Therefore we conclude the following.

Theorem 3.1.2 Differentiability at a point implies continuity at the point.

The continuity does not imply differentiability. See Examples 3.1.6 and 3.1.7.

Sometimes, for the function $y = f(x)$, it is more convenient to use the Leibniz notation

$$\frac{df}{dx} \quad \text{or} \quad \frac{dy}{dx}$$

for the derivative $f'(x)$. The advantage of the notation is that the variable x is explicit, which is very useful when there are many variables around.

Example 3.1.1 The constant function $f(x) = c$ has trivial derivative

$$f'(a) = \lim_{x \rightarrow a} \frac{c - c}{x - a} = \lim_{x \rightarrow a} \frac{0}{x - a} = 0. \quad \blacksquare$$

Example 3.1.2 The cubic function $f(x) = x^3$ is differentiable, with derivative

$$f'(a) = \lim_{x \rightarrow a} \frac{x^3 - a^3}{x - a} = \lim_{x \rightarrow a} (x^2 + xa + a^2) = 3a^2.$$

Similar argument shows that for a natural number n , the power function x^n has derivative na^{n-1} at a . ■

Example 3.1.3 The reciprocal function $f(x) = \frac{1}{x}$ is differentiable at any $a \neq 0$, with derivative

$$f'(a) = \lim_{x \rightarrow a} \frac{\frac{1}{x} - \frac{1}{a}}{x - a} = \lim_{x \rightarrow a} \frac{\frac{a - x}{xa}}{x - a} = \lim_{x \rightarrow a} -\frac{1}{xa} = -\frac{1}{a^2}.$$

The derivative of the square root function $g(x) = \sqrt{x}$ at $a > 0$ is given by

$$g'(a) = \lim_{x \rightarrow a} \frac{\sqrt{x} - \sqrt{a}}{x - a} = \lim_{x \rightarrow a} \frac{1}{\sqrt{x} + \sqrt{a}} = \frac{1}{2\sqrt{a}}.$$

The two examples suggest that the formula $(x^n)' = nx^{n-1}$ should work for all the powers. ■

Example 3.1.4 The sine function $f(x) = \sin x$ is differentiable at 0, with derivative

$$(\sin x)'|_{x=0} = \lim_{\Delta x \rightarrow 0} \frac{\sin(0 + \Delta x) - \sin 0}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{\sin \Delta x}{\Delta x} = 1.$$

The cosine function is also differentiable at 0, with derivative

$$\begin{aligned} (\cos x)'|_{x=0} &= \lim_{\Delta x \rightarrow 0} \frac{\cos(0 + \Delta x) - \cos 0}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{\cos \Delta x - 1}{\Delta x} \\ &= -2 \lim_{\Delta x \rightarrow 0} \frac{\sin^2 \frac{\Delta x}{2}}{\Delta x} = -\frac{1}{2} \lim_{\Delta x \rightarrow 0} \Delta x \left(\frac{\sin \frac{\Delta x}{2}}{\frac{\Delta x}{2}} \right)^2 = 0. \end{aligned}$$

Example 3.1.5 The natural logarithmic function $f(x) = \ln x$ is differentiable at 1, with derivative

$$(\ln x)'|_{x=1} = \lim_{\Delta x \rightarrow 0} \frac{\ln(1 + \Delta x) - \ln 1}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{\ln(1 + \Delta x)}{\Delta x} = \ln \lim_{\Delta x \rightarrow 0} (1 + \Delta x)^{\frac{1}{\Delta x}} = \ln e = 1.$$

The exchange of the limit and the $\ln$ follows from the continuity of $\ln$. ■

Example 3.1.6 The absolute value function $f(x) = |x|$ is continuous everywhere. However, at 0, the function has the right derivative

$$f'_+(0) = \lim_{\Delta x \rightarrow 0^+} \frac{|0 + \Delta x| - |0|}{\Delta x} = \lim_{\Delta x \rightarrow 0^+} \frac{\Delta x}{\Delta x} = 1,$$

and the left derivative

$$f'_-(0) = \lim_{\Delta x \rightarrow 0^-} \frac{|0 + \Delta x| - |0|}{\Delta x} = \lim_{\Delta x \rightarrow 0^-} \frac{-\Delta x}{\Delta x} = -1.$$

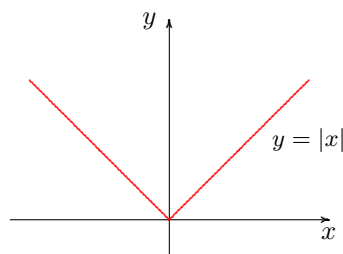


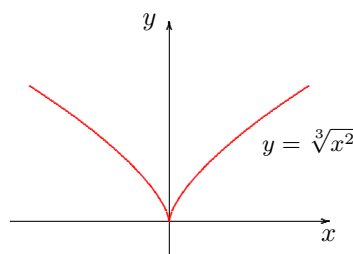
Fig. 3.2 Absolute value function

Since $f'_+(0) \neq f'_-(0)$, the function is not differentiable at 0. ■

Example 3.1.7 The function $f(x) = \sqrt[3]{x^2}$ is continuous everywhere. At 0, however, we have

$$f'_+(0) = \lim_{\Delta x \rightarrow 0^+} \frac{\sqrt[3]{(0 + \Delta x)^2} - \sqrt[3]{0^2}}{\Delta x} = \lim_{\Delta x \rightarrow 0^+} \frac{1}{\sqrt[3]{\Delta x}} = +\infty.$$

Similarly, we have $f'_-(0) = -\infty$. Either the left non-differentiability or the right non-differentiability can be the reason for the function not to be differentiable at 0. ■

Fig. 3.3 The function $\sqrt[3]{x^2}$

Exercises

3.1.1 After observing for t hours, the population of a bacterial colony is approximately

$$P(t) = P_0 + Rt + \frac{A}{2}t^2.$$

Find the rate at which the colony is growing after 5 hours.

3.1.2 Suppose a person's starting salary is \$100,000 per year and he gets a raise of \$8,000 each year.

1. Express the percentage rate of change of his salary as a function of time.
2. At what percentage rate will his salary be increasing after one year? After 5 years?
3. What will happen to the percentage rate of change of his salary in the long run?

3.1.3 A mass attached to a horizontal spring is at position $p(t) = A \sin \omega t$ at time t , where A is the amplitude of the oscillation and ω is a constant.

1. Find the velocity and acceleration.
2. Show that the acceleration is proportional to the displacement $p(t)$.
3. Show that the speed is at maximum when the acceleration is at minimum.

3.1.4 Find the place on the curve $y = x^2$ where the tangent line is parallel to the straight line $x + y = 1$.

3.1.5 Show that the area enclosed by the tangent line on the curve $xy = a^2$ and the coordinate axes is a constant.

3.1.6 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 .

3.1.7 Study the differentiability of the function $y = \sqrt{1 - \cos x}$ at $x = 0$.

3.1.8 Give an example to show that $f'_+(0) \neq f'(0^+)$.

3.1.9 Find the left and right derivatives at $x = 0$ and determine the differentiability

$$1. f(x) = \begin{cases} xe^{-\frac{1}{x}}, & \text{if } x > 0, \\ 0, & \text{if } x = 0; \end{cases} \qquad 2. f(x) = \begin{cases} x^2 \sin \frac{1}{x}, & \text{if } x \geq 0, \\ \cos x - x - 1, & \text{if } x < 0. \end{cases}$$

3.1.10 Suppose $g(x)$ is continuous and $g(a) \neq 0$. Find the left and the right derivatives of $f(x) = |x - a|g(x)$ and a and show that f is not differentiable at $x = a$.

3.1.11 Consider

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

Prove that

1. $f(x)$ is continuous at $x = 0$ if $p > 0$.
2. $f(x)$ is differentiable at $x = 0$ if $p > 1$.
3. $f'(x)$ is continuous at $x = 0$ if $p > 2$.

3.1.12 Let

$$f(x) = |\pi^2 - x^2| \cdot \sin^2 x.$$

Find $f'(\pi)$.

3.2 COMPUTATION OF DERIVATIVE

We already computed some basic examples of the derivative. More computations can be made by applying properties of derivative to the known examples.

Derivative of Arithmetic Combination

Theorem 3.2.1 Suppose functions f and g are differentiable. Then $f + g$, cf , fg and $\frac{f}{g}$ are differentiable, and

$$(f + g)' = f' + g', \quad (cf)' = cf', \quad (fg)' = f'g + fg', \quad \left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2}.$$

In Leibniz notation, the properties are

$$\frac{d(f + g)}{dx} = \frac{df}{dx} + \frac{dg}{dx}, \quad \frac{d(cf)}{dx} = c \frac{df}{dx}, \quad \frac{d(fg)}{dx} = \frac{df}{dx}g + f \frac{dg}{dx}, \quad \frac{d}{dx} \left(\frac{f}{g}\right) = \frac{1}{g^2} \left(\frac{df}{dx}g - f \frac{dg}{dx}\right).$$

The formula for the derivative of the product function is called the Leibniz rule.

Proof. Assume $f(x)$ and $g(x)$ are differentiable at a . Then the derivative of $f(x) + g(x)$ at a is

$$\lim_{x \rightarrow a} \frac{(f(x) + g(x)) - (f(a) + g(a))}{x - a} = \lim_{x \rightarrow a} \left(\frac{f(x) - f(a)}{x - a} + \frac{g(x) - g(a)}{x - a} \right) = f'(a) + g'(a).$$

The proof of $(cf)' = cf'$ is similar. The derivative of the product is

$$\begin{aligned} \lim_{x \rightarrow a} \frac{f(x)g(x) - f(a)g(a)}{x - a} &= \lim_{x \rightarrow a} \frac{(f(x)g(x) - f(a)g(x)) + (f(a)g(x) - f(a)g(a))}{x - a} \\ &= \lim_{x \rightarrow a} \left(\frac{f(x) - f(a)}{x - a} g(x) + f(a) \frac{g(x) - g(a)}{x - a} \right) \\ &= f'(a)g(a) + f(a)g'(a), \end{aligned}$$

where the differentiability of g at a implies $\lim_{x \rightarrow a} g(x) = g(a)$. The derivative of the quotient is

$$\begin{aligned} \lim_{x \rightarrow a} \frac{\frac{f(x)}{g(x)} - \frac{f(a)}{g(a)}}{x - a} &= \lim_{x \rightarrow a} \frac{f(x)g(a) - f(a)g(x)}{(x - a)g(x)g(a)} \\ &= \lim_{x \rightarrow a} \frac{(f(x)g(a) - f(a)g(a)) - (f(a)g(x) - f(a)g(a))}{(x - a)g(x)g(a)} \\ &= \lim_{x \rightarrow a} \frac{1}{g(x)g(a)} \left(\frac{f(x) - f(a)}{x - a} g(a) - f(a) \frac{g(x) - g(a)}{x - a} \right) \\ &= \frac{f'(a)g(a) - f(a)g'(a)}{g(a)^2}. \quad \blacksquare \end{aligned}$$

Example 3.2.1 We already know $(x^n)' = nx^{n-1}$ for natural numbers n . We also know the derivative of $x^0 = 1$ is $0 = 0x^{0-1}$. Then we have

$$\frac{dx^{-5}}{dx} = \frac{d}{dx} \left(\frac{1}{x^5} \right) = \frac{1' \cdot x^5 - 1 \cdot (x^5)'}{(x^5)^2} = \frac{0 \cdot x^5 - 1 \cdot 5x^4}{x^{10}} = -\frac{5}{x^6} = (-5)x^{(-5)-1}.$$

By the similar argument, we see that the formula $(x^n)' = nx^{n-1}$ holds for all integers n . ■

Example 3.2.2 All polynomials are differentiable

$$\begin{aligned}(c_n x^n + c_{n-1} x^{n-1} + \cdots + c_1 x + c_0)' &= c_n (x^n)' + c_{n-1} (x^{n-1})' + \cdots + c_1 (x)' + (c_0)' \\ &= nc_n x^{n-1} + (n-1)c_{n-1} x^{n-2} + \cdots + c_1.\end{aligned}$$

By the derivative of the quotient, any rational function is differentiable wherever it is defined. ■

Example 3.2.3 By the derivatives of the sine and cosine functions at 0, we can find the derivative of the tangent function at 0

$$(\tan x)'|_{x=0} = \left(\frac{\sin x}{\cos x} \right)' \bigg|_{x=0} = \frac{(\sin x)'|_{x=0} \cdot \cos 0 - \sin 0 \cdot (\cos x)'|_{x=0}}{\cos^2 0} = \frac{1 \cdot 1 - 0 \cdot 0}{1^2} = 1. \quad \blacksquare$$

Derivative of Composition

Theorem 3.2.2 (Chain Rule) Suppose $f(x)$ is differentiable at $x = a$ and $g(y)$ is differentiable at $y = f(a)$. Then the composition function $(g \circ f)(x) = g(f(x))$ is differentiable at a , with derivative

$$(g \circ f)'(a) = g'(f(a)) \cdot f'(a).$$

Given a composition $z = g(y) = g(f(x))$, a change

$$\Delta x = x - a$$

in the variable x causes a change

$$\Delta y = y - b = f(x) - f(a)$$

in the variable y , which further causes a change

$$\Delta z = z - c = g(y) - g(b) = g(f(x)) - g(f(a))$$

in the variable z . The corresponding rates of change are related by multiplication

$$\frac{\Delta z}{\Delta x} = \frac{\Delta z}{\Delta y} \cdot \frac{\Delta y}{\Delta x}.$$

As $\Delta x \rightarrow 0$, we also have $\Delta y \rightarrow 0$, and we get the relation for the instantaneous rate of change

$$\lim_{\Delta x \rightarrow 0} \frac{\Delta z}{\Delta x} = \lim_{\Delta y \rightarrow 0} \frac{\Delta z}{\Delta y} \cdot \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x}.$$

This is the formula in the chain rule.

In Leibniz notation, the chain rule is

$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}.$$

Example 3.2.4 The function $(x^2 - 1)^{10}$ is the composition of $z = y^{10}$ and $y = x^2 - 1$. Therefore

$$\frac{d(x^2 - 1)^{10}}{dx} = \frac{dy^{10}}{dy} \Big|_{y=x^2-1} \cdot \frac{d(x^2 - 1)}{dx} = 10(x^2 - 1)^9(2x) = 20x(x^2 - 1)^9. \quad \blacksquare$$

Example 3.2.5 The reciprocal $\frac{1}{f}$ of a function may be considered as the composition of $g(y) = \frac{1}{y}$ with the function $y = f(x)$. The chain rule tells us

$$\left(\frac{1}{f(x)}\right)' = \frac{d\left(\frac{1}{y}\right)}{dy} \Big|_{y=f(x)} \cdot \frac{dy}{dx} = \left(-\frac{1}{y^2}\right) \Big|_{y=f(x)} \cdot f'(x) = -\frac{f'(x)}{f^2(x)}. \quad \blacksquare$$

Example 3.2.6 We already know the derivatives of the trigonometric functions at 0. To find the derivative of $\sin x$ at a , we use

$$\sin(x + a) = \sin x \cos a + \cos x \sin a.$$

Taking the derivative of both sides at $x = 0$, and applying the chain rule to the function $\sin(x + a)$ (composition of $z = \sin y$ and $y = x + a$), we get

$$(\sin x)'|_{x=a} = (\sin x)'|_{x=a} \cdot (x + a)'|_{x=0} = (\sin x)'|_{x=0} \cos a + (\cos x)'|_{x=0} \sin a = \cos a.$$

In other words, $(\sin x)' = \cos x$.

Applying the chain rule to $\cos x = \sin\left(\frac{\pi}{2} - x\right)$, we get

$$(\cos x)'|_{x=a} = (\sin y)'|_{y=\frac{\pi}{2}-a} \cdot \left(\frac{\pi}{2} - x\right)' \Big|_{x=a} = \cos\left(\frac{\pi}{2} - a\right) \cdot (-1) = -\sin a.$$

Therefore $(\cos x)' = -\sin x$.

The derivatives of the sine and cosine functions give us the derivative of the tangent function

$$(\tan x)' = \left(\frac{\sin x}{\cos x}\right)' = \frac{(\sin x)' \cos x - \sin x (\cos x)'}{\cos^2 x} = \frac{\cos x \cos x - \sin x (-\sin x)}{\cos^2 x} = \frac{1}{\cos^2 x} = \sec^2 x,$$

and the derivative of the secant function

$$(\sec x)' = \left(\frac{1}{\cos x}\right)' = -\frac{(\cos x)'}{\cos^2 x} = \frac{\sin x}{\cos^2 x} = \sec x \tan x. \quad \blacksquare$$

Example 3.2.7 We already know the derivatives of the logarithmic function at 1. To find the derivative of $\ln x$ at $a > 0$, we take the derivative of both sides of $\ln ax = \ln a + \ln x$ at $x = 1$ and get

$$\frac{d \ln y}{dy} \Big|_{y=a} \cdot \frac{d(ax)}{dx} \Big|_{x=1} = \frac{d \ln x}{dx} \Big|_{x=1}.$$

In other words, we have $a(\ln x)'_{x=a} = 1$, or

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

Moreover, the derivative of $\ln |x|$ at $a < 0$ is the derivative of $\ln(-x)$ at a :

$$\left. \frac{d \ln(-x)}{dx} \right|_{x=a} = \left. \frac{d \ln y}{dy} \right|_{y=-a} \left. \frac{d(-x)}{dx} \right|_{x=a} = \frac{1}{-a}(-1) = \frac{1}{a}.$$

Combined with the derivative at positive numbers, we get

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

for all real numbers except 0. ■

Example 3.2.8 Let $u(x) > 0$ and $v(x)$ be differentiable. To find the derivative of $f(x) = u(x)^{v(x)}$, we take logarithm and get $\ln f = v \ln u$. Taking the derivative of both sides, we get

$$\frac{1}{f} f' = v' \ln u + v \frac{1}{u} u'.$$

Therefore

$$(u^v)' = u^v \left(v' \ln u + v \frac{1}{u} u' \right) = u^{v-1} (uv' \ln u + u'v).$$

This easily leads to the following derivatives:

$$\begin{aligned} (x^p)' &= px^{p-1}; & (a^x)' &= a^x \ln a; \\ (e^x)' &= e^x; & (x^x)' &= x^x (\ln x + 1). \end{aligned} \quad \blacksquare$$

Example 3.2.9 The unit circle $x^2 + y^2 = 1$ on the plane defines two functions $y = \sqrt{1-x^2}$ and $y = -\sqrt{1-x^2}$. We may certainly compute the derivative of each function by using the chain rule

$$\left(\sqrt{1-x^2} \right)' = \frac{1}{2} (1-x^2)^{-\frac{1}{2}} (-2x) = \frac{-x}{\sqrt{1-x^2}}.$$

On the other hand, we may use the fact that the two functions $y = y(x)$ satisfy the equation $x^2 + y(x)^2 = 1$. Taking the derivative of both sides of the equation with respect to x , we get $2x + 2yy' = 0$. Solving the equation, we get

$$y' = -\frac{x}{y}.$$

You may verify that the result is consistent with the direct computation for both functions. The method here is called implicit differentiation. ■

Example 3.2.10 Like the unit circle, the equation $2y - 2x^2 - \sin y + 1 = 0$ gives a curve on the plane, parts of the curve may define several functions $y = y(x)$. Unlike the circle, we cannot find an explicit formula for the functions. Yet we can still compute the derivatives of the functions.

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 $P = \left(\sqrt{\frac{\pi}{2}}, \frac{\pi}{2}\right)$ 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 $y = \sqrt{2\pi}x - \frac{\pi}{2}$. ■

Derivative of Inverse

Theorem 3.2.3 Suppose $f(x)$ is invertible near $x = a$. If f is differentiable at a and the inverse function f^{-1} is differentiable at $b = f(a)$, then

$$(f^{-1})'(b) = \frac{1}{f'(a)}.$$

By the definition of inverse function, we have $f^{-1}(f(x)) = x$. Taking derivative of both sides, we have

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

This proves the formula in the theorem.

In Leibniz notation, the formula is

$$\frac{dx}{dy} = \frac{1}{\frac{dy}{dx}}.$$

Example 3.2.11 The derivative of $\arcsin x$ at $x = a$ is

$$(\arcsin x)'|_{x=a} = \frac{1}{(\sin y)'|_{y=b}} = \frac{1}{\cos b},$$

where $b \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ satisfies $a = \sin b$. Therefore $\cos b \geq 0$ and $\cos b = \sqrt{1 - \sin^2 b} = \sqrt{1 - a^2}$. We conclude

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

The derivative for $\arccos x$ can be similarly obtained, or by using

$$\arccos x + \arcsin x = \frac{\pi}{2}.$$

The result is

$$(\arccos x)' = -\frac{1}{\sqrt{1-x^2}}. \quad \blacksquare$$

Example 3.2.12 To find the derivative of $\arctan x$, we let $y = \arctan x$, $x = \tan y$ and get

$$(\arctan x)' = \frac{dy}{dx} = \frac{1}{\frac{dx}{dy}} = \frac{1}{(\tan y)'} = \frac{1}{\sec^2 y} = \frac{1}{1 + \tan^2 y} = \frac{1}{1 + x^2}. \quad \blacksquare$$

High Order Derivative

Given a differentiable function $f(x)$ on an open interval (a, b) , the derivative $f'(x)$ is again 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 second order derivative function is yet again differentiable, then taking derivative one more time gives the third order derivative $f'''(x) = (f''(x))'$. The process can continue, and we have the concept of **n -th order derivative** $f^{(n)}(x)$.

The (first order) derivative $f'(x)$ measures the rate of change. Therefore the second order derivative measures the rate of the change of the rate of change. For example, the displacement of a free fall is given by $s(t) = \frac{1}{2}gt^2$. The derivative of the displacement

$$s'(t) = gt$$

is the instantaneous velocity. The second order derivative

$$s''(t) = g$$

is the acceleration of the instantaneous velocity. In general, Newton's second law says that the acceleration of the velocity of an object is proportional to the force exercised on the object.

The Leibniz notation of the high order derivative is

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

The rules for computing the derivatives can be extended to high order. For example, we have

$$(f + g)^{(n)} = f^{(n)} + g^{(n)}, \quad (cf)^{(n)} = cf^{(n)}.$$

Moreover, the Leibniz rule becomes

$$(f + g)'' = f'' + 2f'g' + g'', \quad (f + g)''' = f''' + 3f''g' + 3f'g'' + g''',$$

and the general formula

$$(fg)^{(n)} = \sum_{i=0}^n \frac{n!}{i!(n-i)!} f^{(i)} g^{(j)} = \sum_{i=0}^n \binom{n}{i} f^{(i)} g^{(j)}.$$

is similar to the expansion of $(x + y)^n$. The chain rule is more complicated

$$(g(f(x)))' = (g'(f(x)) \cdot f'(x))' = g''(f(x)) \cdot f'(x)^2 + g'(f(x)) \cdot f''(x).$$

Example 3.2.13 For the power function x^p , we have

$$\begin{aligned}(x^p)' &= px^{p-1}, \\(x^p)'' &= p(p-1)x^{p-2}, \\&\vdots \\(x^p)^{(n)} &= p(p-1)\cdots(p-n+1)x^{p-n}.\end{aligned}$$

In particular, if p is a natural number, then $(x^p)^{(n)} = 0$ for $p < n$. ■

Example 3.2.14 For the exponential function e^x , since

$$(e^x)' = e^x, \quad (e^x)'' = (e^x)' = e^x, \quad \dots, \quad (e^x)^{(n)} = e^x.$$

For the sine function, we have

$$(\sin x)' = \cos x, \quad (\sin x)'' = -\sin x, \quad (\sin x)''' = -\cos x, \quad (\sin x)^{(4)} = \sin x, \quad (\sin x)^{(5)} = \cos x, \quad \dots$$

and the pattern is periodic: $\sin^{(n+4)} x = \sin^{(n)} x$. The high order derivatives of the cosine function has the same periodic pattern

$$(\cos x)' = -\sin x, \quad (\cos x)'' = -\cos x, \quad (\cos x)''' = \sin x, \quad (\cos x)^{(4)} = \cos x, \quad (\cos x)^{(5)} = -\sin x, \quad \dots$$

Example 3.2.15 To compute the n -th order derivative of $x^2 e^{-x}$, we use the Leibniz rule and the fact that the derivatives of x^2 of order > 2 all vanish

$$(x^2 e^{-x})^{(n)} = x^2 (e^{-x})^{(n)} + n(x^2)'(e^{-x})^{(n-1)} + \frac{n(n-1)}{2}(x^2)''(e^{-x})^{(n-2)} = (-1)^n (x^2 - 2nx + n(n-1))e^{-x}.$$

Example 3.2.16 The high order derivatives of the function $\sin x^2$ may be computed by the chain rule

$$\begin{aligned}\frac{d(\sin x^2)}{dx} &= 2x \cos x^2, \\ \frac{d^2(\sin x^2)}{dx^2} &= 2 \cos x^2 + 2x(-\sin x^2)2x = 2 \cos x^2 - 4x^2 \sin x^2, \\ \frac{d^3(\sin x^2)}{dx^3} &= 2(-\sin x^2)2x - 8x \sin x^2 - 4x^2(\cos x^2)2x = -12x \sin x^2 - 8x^3 \cos x^2.\end{aligned}$$

Example 3.2.17 For the function $y = y(x)$ implicitly given by the unit circle $x^2 + y^2 = 1$, we have $y' = -\frac{x}{y}$. Therefore

$$y'' = -\frac{y - xy'}{y^2} = -\frac{y + x \cdot \frac{x}{y}}{y^2} = -\frac{x^2 + y^2}{y^3} = -\frac{1}{y^3}.$$

You may verify the result by directly computing the second order derivative of $y = \pm\sqrt{1-x^2}$. ■

Example 3.2.18 For the function $y = y(x)$ implicitly given by the equation $2y - 2x^2 - \sin y + 1 = 0$, we have $y' = \frac{4x}{2 - \cos y}$ from Example 3.2.10. Therefore

$$y'' = 4 \frac{(2 - \cos y) - xy' \sin y}{(2 - \cos y)^2}.$$

At the point $P = \left(\sqrt{\frac{\pi}{2}}, \frac{\pi}{2}\right)$, we know $y'|_P = \sqrt{2\pi}$. Therefore the second order derivative at the point is

$$y''|_P = 4 \frac{2 - \cos \frac{\pi}{2} - \sqrt{\frac{\pi}{2}} \sqrt{2\pi} \sin \frac{\pi}{2}}{\left(2 - \cos \frac{\pi}{2}\right)^2} = 2 - \pi. \quad \blacksquare$$

Exercises

3.2.1 Prove that if $f(x)$ is differentiable at a and $g(x)$ is not differentiable at a , then $f(x) + g(x)$ is not differentiable at a . What if both $f(x)$ and $g(x)$ are not differentiable at a ?

3.2.2 Compute the derivative:

- | | |
|---|--|
| 1. $\frac{1}{1+x^2};$ | 11. $\arctan \frac{1+x}{1-x};$ |
| 2. $x \ln x;$ | 12. $\ln(x + \sqrt{x^2 + a^2});$ |
| 3. $\sqrt[3]{\frac{1+x^3}{1-x^3}};$ | 13. $\sqrt{x+1} - \ln(1 + \sqrt{x+1});$ |
| 4. $\frac{x}{\sqrt{a^2 - x^2}};$ | 14. $x \ln(x + \sqrt{x^2 + 1}) - \sqrt{x^2 + 1};$ |
| 5. $\sqrt{x + \sqrt{x + \sqrt{x}}};$ | 15. $\ln \frac{1 - \sin x}{1 + \sin x};$ |
| 6. $\sqrt[3]{1 + \sqrt[3]{1 + \sqrt[3]{x}}};$ | 16. $x\sqrt{x^2 + a^2} + a^2 \ln(x + \sqrt{x^2 + a^2});$ |
| 7. $\frac{\sqrt{1+x} - \sqrt{1-x}}{\sqrt{1+x} + \sqrt{1-x}};$ | 17. $x\sqrt{a^2 - x^2} + a^2 \arcsin \frac{x}{a};$ |
| 8. $\arcsin \sqrt{\frac{1-x}{1+x}};$ | 18. $\frac{\arccos x}{x} + \frac{1}{2} \ln \frac{1 - \sqrt{1-x^2}}{1 + \sqrt{1-x^2}};$ |
| 9. $\arccos(\cos^2 x);$ | 19. $\frac{\arcsin x}{\sqrt{1-x^2}} + \frac{1}{2} \ln(1-x^2);$ |
| 10. $\arcsin \left(\frac{\sin a \sin x}{1 - \cos a \cos x} \right);$ | 20. $x^{a^a} + a^{x^a} + a^{a^x};$ |
| | 21. $x^{x^a} + x^{a^x} + a^{x^x};$ |

22. $(\sin x)^{\cos x} + (\cos x)^{\sin x};$

24. $\log_x e;$

25. $(\ln x)^x + x^{\ln x}.$

23. $\sqrt[x]{x};$

3.2.3 Use $\frac{d \ln f}{dx} = \frac{f'}{f}$ to compute the derivative of $(x - a_1)^{\alpha_1} (x - a_2)^{\alpha_2} \cdots (x - a_n)^{\alpha_n}.$

3.2.4 Use induction to derive the Leibniz formula for high order derivatives.

3.2.5 Compute the high order derivative:

1. $(\tan x)''';$

5. $(x^x)'';$

2. $(\sec x)''';$

6. $\frac{d^2}{dx^2} \left(1 + \frac{1}{x} \right)^x.$

3. $(\arcsin x)'';$

4. $(\arctan x)'';$

3.2.6 Compute the n -th order derivative:

1. $\log_a x;$

7. $\frac{1}{x^2 - 3x + 2};$

2. $a^x;$

3. $(ax + b)^p;$

8. $\frac{\ln x}{x};$

4. $(x^2 + 1) \sin x;$

5. $\frac{1-x}{1+x};$

9. $\frac{e^x}{x}.$

6. $\frac{1+x}{\sqrt{1-x}};$

3.2.7 Compute the second order derivative of the implicitly defined function:

1. $y^2 = 10x;$

4. $e^{x+y} - xy = 0;$

2. $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1;$

5. $x^2 + 2xy - y^2 = 2x.$

3. $\sqrt{x} + \sqrt{y} = \sqrt{a};$

3.2.8 A curve on the plane can also be describe by parametric equation $x = x(t)$, $y = y(t)$. Let $y = y(x)$ be the function given by the parametrized curve (i.e., the curve if the graph of the function). Show that the derivative of the function is given by

$$\frac{dy}{dx} = \frac{y'(t)}{x'(t)}.$$

The compute the second order derivative of the function $y = y(x)$ given by the following curves:

1. $x = \sin^2 t, y = \cos^2 t$;
2. $x = a(t - \sin t), y = a(1 - \cos t)$;
3. $x = \sin^2 t, y = \cos^2 t$;
4. $x = e^t \cos 2t, y = e^t \sin 2t$;
5. $x = (1 + \cos t) \cos t, y = (1 + \cos t) \sin t$.

3.3 DIFFERENTIAL

Linear Approximation

A linear function

$$L(x) = A + B(x - a)$$

is an approximation of $f(x)$ at a if the difference between $f(x)$ and L is significantly smaller than $x - a$: For any $\epsilon > 0$, there is $\delta > 0$, such that

$$|x - a| < \delta \implies |f(x) - L(x)| = |f(x) - A - B(x - a)| \leq \epsilon |x - a|.$$

Taking $x = a$, we get $A = f(a)$. Taking $x \neq a$, the definition becomes $\lim_{x \rightarrow a} \left| \frac{f(x) - f(a)}{x - a} - B \right| = 0$, which is the same as $B = f'(a)$. Therefore the linear approximation is exactly the tangent line

$$L(x) = f(a) + f'(a)(x - a)$$

of the function at a .

The observation on linear approximation can be rephrased as

$$f(x) = L(x) + o(x - a) = f(a) + f'(a)\Delta x + o(\Delta x),$$

or

$$\Delta f = f(x) - f(a) = f'(a)\Delta x + o(\Delta x),$$

where $o(D)$ denotes a function significantly smaller than D as x approaches a

$$\lim_{x \rightarrow a} \frac{o(D(x))}{D(x)} = 0.$$

Example 3.3.1 The function $f(x) = \sqrt{x}$ has derivative $f'(x) = \frac{1}{2\sqrt{x}}$. The linear approximation of the function at $x = 4$ is

$$L(x) = f(4) + f'(4)(x - 4) = 2 + \frac{1}{4}(x - 4).$$

Therefore we get the estimations of the value of the function near 4

$$\sqrt{3.96} \approx 2 + \frac{1}{4}(-0.04) = 1.99, \quad \sqrt{4.05} \approx 2 + \frac{1}{4}(0.05) = 2.0125.$$

The true values are $1.989974874213 \dots$ and $2.01246117975 \dots$, respectively. ■

Example 3.3.2 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 is caused by the error Δr of the radius by

$$\Delta W \approx \frac{dW}{dr} \Delta r = 4\rho\pi r^2 \Delta r.$$

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 $\left| \frac{\Delta r}{r} \right| \leq 1\%$, so that the relative error of the weight is $\left| \frac{\Delta W}{W} \right| \leq 3\%$. ■

Differential

For a differentiable function $f(x)$, the approximation

$$\Delta f \approx f'(a) \Delta x$$

indicates the almost proportional relation between the change of the function and the change of the variable. We denote such a relation by

$$df = f'(a) dx,$$

called the **differential** of the function f at a .

At the moment, the differential is only a notation obtained by formally changing Δ to d and $\approx$ to $=$. The precise definition has to wait until more advanced mathematics. The symbolic computation of the differential will become a very convenient tool, especially in the computation of integral and in the multi-variable calculus.

The rules for computing the derivatives may be rephrased in differential form

$$d(f + g) = df + dg, \quad d(cf) = c df, \quad d(fg) = gdf + f dg, \quad d\left(\frac{f}{g}\right) = \frac{g df - f dg}{g^2}.$$

The chain rule may be interpreted as

$$dz = A dy, \quad dy = B dx \implies dz = AB dx.$$

Basically, the differential forms can be manipulated like vectors in a vector space.

Newton's Method

Linear approximations can be used to find approximate solutions of equations. Suppose we try to find a solution a of the equation $f(x) = 0$ (i.e., a is a root of f). We start with a rough estimate x_0 of the root. Since $f(x)$ is very close to the linear approximation

$$L_0(x) = f(x_0) + f'(x_0)(x - x_0)$$

near x_0 , we expect the solution x_1 of $L_0(x) = 0$ to be very close to the solution a of $f(x) = 0$. The linear equation is very easy to solve, and we get

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

Although x_1 is not the actual solution, chances are it is an improvement of the initial estimate x_0 . In other words, we expect x_1 to be much closer to a than x_0 .

To get an approximate solution even better than x_1 , we repeat the process and get

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}.$$

The idea leads to a sequence inductively constructed by

$$x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})}.$$

We expect the sequence to rapidly converge to the real solution a .

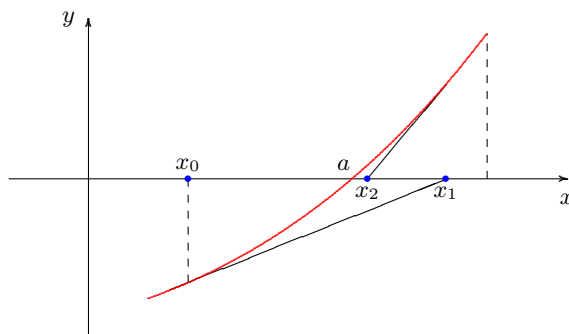


Fig. 3.4 Newton's method

The Newton's method will succeed only if the initial approximation x_0 is sufficiently close to α . Otherwise the sequence may diverge or converge to a root other than the one being sought. The following is one criterion that guarantees the convergence.

Theorem 3.3.1 Suppose f has continuous second order derivative on $[a, b]$. Assume the following conditions are satisfied.

1. $f(a)$ and $f(b)$ have different signs.
2. $f'(x) \neq 0$ on the interval.
3. $f''(x) \geq 0$ on the interval (or $f''(x) \leq 0$ on the interval).
4. $|f(a)| < |f'(a)|(b-a)$, $|f(b)| < |f'(b)|(b-a)$.

Then for any x_0 in $[a, b]$, Newton's method produces a sequence converging to the unique solution $f(x) = 0$ on $[a, b]$.

Example 3.3.3 The function $f(x) = x^3 + x + 2 \sin x - 1$ satisfies $f(0) = -1 < 0$, $f(1) = 1 + 2 \sin 1 > 0$, $f'(x) = 3x^2 + 1 + 2 \cos x > 0$ on $[0, 1]$, $f''(x) = 6x - 2 \sin x > 0$ on $[0, 1]$. Starting with $x_0 = 1$ and applying Newton's method, we get the following sequence.

n	x_n
0	1.0000000000000000
1	0.471924667505487
2	0.330968826345873
3	0.325645312076542
4	0.325639452734876
5	0.325639452727856
6	0.325639452727856

So the actual root should be $x = 0.325639452727856 \dots$. ■

Example 3.3.4 The function $f(x) = x^3 - 3x + 1$ should have a root between 0.3 and 0.4. Starting with $x_0 = 1$ and using Newton's method, we get

n	x_n
0	0.0000000000000000
1	0.3333333333333333
2	0.3472222222222222
3	0.347296353163868
4	0.347296355333861
5	0.347296355333861

So the actual root should be $x = 0.347296355333861 \dots$. ■

3.3.1 High Order Approximation

Linear approximation can be used to solve many problems. However, there are many other problems that require more refined approximations to solve. When approximation by linear functions is not enough, we may consider approximations by quadratic, cubic, or more generally, polynomial functions.

Linear approximation comes from the first order differentiability. The high order approximation comes from high order derivative.

Theorem 3.3.2 Suppose $f(x)$ has n -th order derivative at a . Then for the polynomial

$$T_n(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n,$$

we have

$$\lim_{x \rightarrow a} \frac{f(x) - T_n(x)}{(x-a)^n} = 0.$$

The function T_n is called the n -th order **Taylor expansion** of f , or simply **Taylor polynomial**. The approximation property can be rephrased as follows: For any $\epsilon > 0$, there is $\delta > 0$, such that

$$|x-a| < \delta \implies |f(x) - T_n(x)| \leq \epsilon |x-a|^n.$$

The choice of coefficients in $T_n(x)$ is very well motivated. The linear approximation $L(x)$ is determined by the properties $L(a) = f(a)$ and $L'(a) = f'(a)$. Similarly, the n -th order approximation should satisfy $T_n(a) = f(a)$, $T'_n(a) = f'(a)$, $T''_n(a) = f''(a)$, $\dots$, $T_n^{(n)}(a) = f^{(n)}(a)$. Since

$$T_n(a) = A_0 + A_1(x-a) + A_2(x-a)^2 + \cdots + A_n(x-a)^n \implies T_n^{(k)}(a) = k!A_k, \quad 0 \leq k \leq n,$$

we get $k!A_k = f^{(k)}(a)$.

Example 3.3.5 The function $f(x) = \sqrt{x}$ has quadratic approximation

$$Q(x) = f(4) + f'(4)(x-4) + \frac{f''(4)}{2}(x-4)^2 = 2 + \frac{1}{4}(x-4) - \frac{1}{64}(x-4)^2.$$

Therefore we get the estimations

$$\sqrt{3.96} \approx 2 + \frac{1}{4}(-0.04) - \frac{1}{64}(-0.04)^2 = 1.989975, \quad \sqrt{4.05} \approx 2 + \frac{1}{4}(0.05) - \frac{1}{64}(0.05)^2 = 2.0124609375.$$

The estimated values are much more precise than the ones obtained by the linear approximation. ■

Example 3.3.6 From Examples 3.2.10 and 3.2.18, the function $y = y(x)$ implicitly given by the equation $2y - 2x^2 - \sin y + 1 = 0$ has quadratic approximation

$$Q(x) = \frac{\pi}{2} + \sqrt{2\pi} \left(x - \sqrt{\frac{\pi}{2}} \right) + \frac{2-\pi}{2} \left(x - \sqrt{\frac{\pi}{2}} \right)^2$$

at $x = -\sqrt{\frac{\pi}{2}}$. ■

Example 3.3.7 From $\sin 0 = 0$, $(\sin x)'|_{x=0} = \cos 0 = 1$, $(\sin x)''|_{x=0} = -\sin 0 = 0$, $(\sin x)'''|_{x=0} = -\cos 0 = -1$, the cubic approximation of $\sin x$ at 0 is $T_3(x) = x - \frac{1}{6}x^3$. In other words, we have

$$\sin x = x - \frac{1}{6}x^3 + R(x), \quad \lim_{x \rightarrow 0} \frac{R(x)}{x^3} = 0.$$

This implies

$$\begin{aligned}\sin^2 x - \sin x^2 &= \left(x - \frac{1}{6}x^3 + R(x)\right)^2 - \left(x^2 - \frac{1}{6}x^6 + R(x^2)\right) \\ &= -\frac{1}{3}x^4 + 2xR(x) + \left(\frac{1}{36} - \frac{1}{6}\right)x^6 + \dots\end{aligned}$$

We have $\lim_{x \rightarrow 0} \frac{xR(x)}{x^4} = 0$, $\lim_{x \rightarrow 0} \frac{x^6}{x^4} = 0$, and the same happens to all the (finitely many) terms in "...". Therefore we conclude that

$$\lim_{x \rightarrow 0} \frac{\sin^2 x - \sin x^2}{x^4} = -\frac{1}{3}. \quad \blacksquare$$

Using the notation $o(D)$ for any function significantly smaller than D , we may write

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \dots + \frac{f^{(n)}(a)}{n!}(x-a)^n + o((x-a)^2).$$

In the example above, we used the properties such as

$$\lim_{x \rightarrow a} \frac{xR(D(x))}{xD(x)} = 0, \quad \lim_{x \rightarrow 0} \frac{R(D(x^2))}{D(x^2)} = 0.$$

Common sense can guide us carrying out computations with the symbol $o(D)$.

Example 3.3.8 The following are the high order approximations of some basic functions at 0

$$\begin{aligned}(1+x)^p &= 1 + px + \frac{p(p-1)}{2!}x^2 + \dots + \frac{p(p-1)\dots(p-n+1)}{n!}x^n + o(x^n), \\ e^x &= 1 + \frac{x}{1!} + \frac{x^2}{2!} + \dots + \frac{x^n}{n!} + o(x^n), \\ \ln(1+x) &= x - \frac{x^2}{2} + \frac{x^3}{3} - \dots + (-1)^{n+1} \frac{x^n}{n} + o(x^n), \\ \sin x &= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots + (-1)^{n+1} \frac{x^{2n-1}}{(2n-1)!} + o(x^{2n+1}), \\ \cos x &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \dots + (-1)^{n+1} \frac{x^{2n}}{(2n)!} + o(x^{2n+2}). \quad \blacksquare\end{aligned}$$

Example 3.3.9 We use the high order approximation and $o(D)$ notation to compute more limits. For example, by $(1+y)^{-1} = 1 - y + o(y)$, we get

$$\begin{aligned}\frac{1}{x} - \frac{1}{e^x - 1} &= \frac{1}{x} - \frac{1}{x + \frac{x^2}{2} + o(x^2)} = \frac{1}{x} \left(1 - \left(1 + \frac{x}{2} + o(x)\right)^{-1}\right) \\ &= \frac{1}{x} \left(1 - 1 + \left(\frac{x}{2} + o(x)\right) + o\left(\frac{x}{2} + o(x)\right)\right) = \frac{1}{2} + \frac{o(x)}{x}.\end{aligned}$$

This implies $\lim_{x \rightarrow 0} \left(\frac{1}{x} - \frac{1}{e^x - 1} \right) = \frac{1}{2}$. For another example, by

$$\sin x - \tan x = \tan x(\cos x - 1) = \frac{x + o(x)}{\cos x} \left(-\frac{x^2}{2} + o(x^2) \right) = -\frac{x^3 + o(x^3)}{2 \cos x},$$

we get $\lim_{x \rightarrow 0} \frac{\sin x - \tan x}{x^3} = -\frac{1}{2}$. ■

Exercises

3.3.1 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?

3.3.2 Explain the approximation for small x :

1. $(1+x)^4 \approx 1+4x$;

3. $\cos\left(x + \frac{\pi}{3}\right) \approx \frac{1}{2} + \frac{\sqrt{3}}{2}x$;

2. $\frac{1}{1+x} \approx 1-x$;

4. $\ln \frac{1+x}{1-x} \approx 2x$.

3.3.3 Derive

$$\sqrt[n]{a^n + x} \approx a + \frac{x}{na^{n-1}}, \quad (a > 0)$$

for small x . Use it to yield the approximations for the following values:

1. $\sqrt[4]{15}$;

2. $\sqrt{46}$;

3. $\sqrt[5]{39}$;

4. $\sqrt[7]{127}$.

3.3.4 Find approximate value of $\tan 4^\circ$.

3.3.5 Using linearization to derive

$$\sin 46^\circ \approx \frac{1}{\sqrt{2}} \left(1 + \frac{\pi}{180} \right).$$

3.3.6 Use Newton's method to find the unique positive root of $f(x) = e^x - x - 2$.

3.3.7 Use Newton's method to find all the roots of $f(x) = x^2 - \cos x$.

3.3.8 Let

$$f(x) = \begin{cases} e^{-\frac{1}{x^2}}, & \text{if } x \neq 0, \\ 0, & \text{if } x = 0. \end{cases}$$

Use $\lim_{x \rightarrow +\infty} x^p e^{-x} = 0$ to prove that all the high order derivatives of the function vanish at 0.

3.4 SUMMARY

Definitions

- The **derivative** of $f(x)$ at a is

$$f'(a) = \lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a} = \lim_{\Delta x \rightarrow 0} \frac{f(a + \Delta x) - f(a)}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{\Delta f}{\Delta x}.$$

$f(x)$ is **differentiable** at a if the limit converges.

- If $f(x)$ is differentiable at a , then

$$L(x) = f(a) + f'(a)(x - a)$$

is the **linear approximation** of $f(x)$ near a .

- The high order derivatives are obtained by repeatedly taking derivatives. If $f(x)$ has n -th order derivative at a , then

$$T_n(x) = f(a) + f'(a)(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x - a)^n$$

is the **n -th order approximation** of $f(x)$ near a .

- **Newton's method** starts from some x_0 and produces a sequence by the recursive relation

$$x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})}.$$

The sequence often converges to a solution of the equation $f(x) = 0$.

Theorems

- Differentiability implies continuity.

- Arithmetic Rule: $(f + g)' = f' + g'$, $(cf)' = cf'$, $(fg)' = f'g + fg'$, $\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2}$.

- Arithmetic Rule for High Order Derivative: $(f + g)^{(n)} = f^{(n)} + g^{(n)}$, $(cf)^{(n)} = cf^{(n)}$, $(fg)^{(n)} = \sum_{i+j=n} \frac{n!}{i!j!} f^{(i)} g^{(j)}$.

- Chain Rule: $(g(f(x)))' = g'(f(x))f'(x)$. If we denote the composition by $y = f(x)$ and $z = g(x)$, then the chain rule can also be written as $\frac{dz}{dx} = \frac{dz}{dy} \frac{dy}{dx}$.

- Inverse Rule: If f is invertible at a and $f'(a) \neq 0$, then f^{-1} is also differentiable at $f(a)$, and $(f^{-1})'(f(a)) = \frac{1}{f'(a)}$. If we denote $y = f(x)$ and $x = f^{-1}(y)$, then the inverse rule can also be written as $\frac{dx}{dy} = \frac{1}{\frac{dy}{dx}}$.

Applications of Differentiation

4.1 MAXIMUM AND MINIMUM

For a function $f(x)$ defined on D , we say f has (global) **maximum** at a if $f(x) \leq f(a)$ for any $x \in D$. We say f has (global) **minimum** at a if $f(x) \geq f(a)$ for any $x \in D$.

We say f has a **local maximum** at a if $f(a)$ is the largest value near a . In other words, there is $\delta > 0$, such that

$$x \in D, |x - a| < \delta \implies f(x) \leq f(a).$$

We say f has a **local minimum** at c if there is $\delta > 0$, such that

$$x \in D, |x - a| < \delta \implies f(x) \geq f(a).$$

Global extremes are also local extremes. Maximum and minimum are called **extremes**. Local maximum and minimum are **local extremes**.

Example 4.1.1 The continuous function x has global and local extremes on $[0, 1]$. However, the same function has neither global extreme nor local extreme on $(0, 1)$.

On the whole real line, the function $\sin x$ has (infinitely many) global maxima at $\left(2n + \frac{1}{2}\right)\pi$ and global minima at $\left(2n - \frac{1}{2}\right)\pi$. These are also all the local extremes.

On the interval $[-1, 2]$, the function x^2 has a global minimum at 0, a global maximum at 2, and also a local maximum at -1 . On the other hand, the same function on the interval $(-1, 2)$ has a global minimum at 0 and has no global or local maximum.

On the interval $[-1, 2]$, the function $f(x) = \begin{cases} x^2, & \text{if } x \neq 0, \\ 1, & \text{if } x = 0 \end{cases}$ has a global maximum at 2, two local maxima at -1 and 0, and has no global or local minimum.

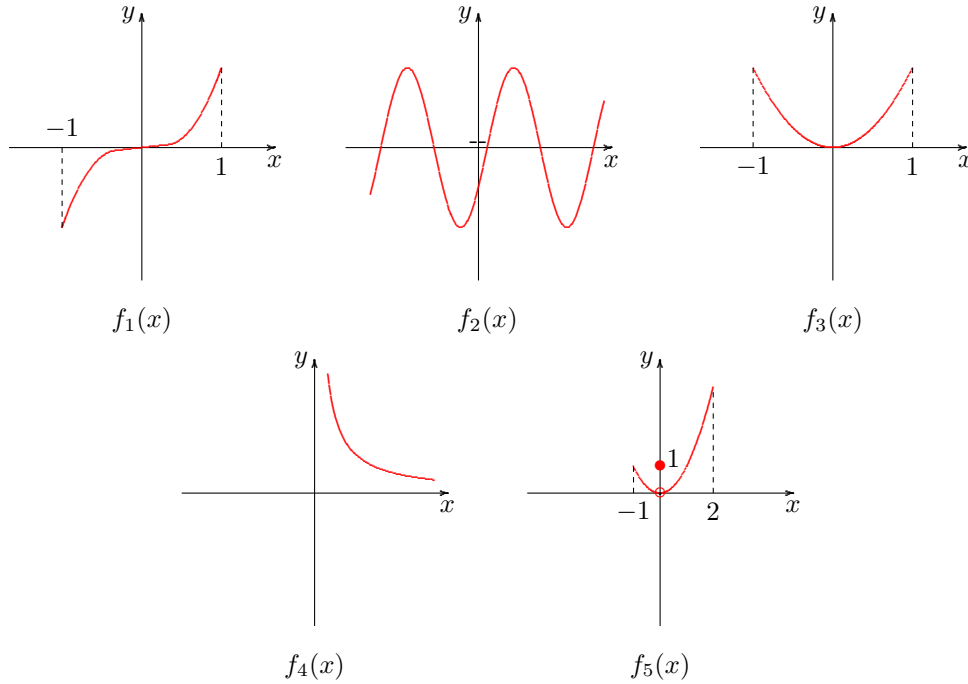


Fig. 4.1 Global and local extremes

Theorem 4.1.1 Any continuous function on a bounded closed interval has global maximum and global minimum.

Suppose $f(x)$ is differentiable at a . Then $f(x)$ is approximated by $L(x) = f(a) + f'(a)(x - a)$. If $f'(a) \neq 0$, then the linear function $L(x)$ does not have local extreme at a . Since L approximates f near a , it is also unlikely that f has local extreme at a . This motivates the following criterion for finding local extremes.

Theorem 4.1.2 If $f(x)$ has a local extreme at $x = a$ where f is differentiable, then $f'(a) = 0$.

Proof. We prove the theorem by a contradiction. If $f'(a) \neq 0$, then, without loss of generality, we may assume that $f'(a) > 0$. By $\lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a} = f'(a) > 0$, we know that $\frac{f(x) - f(a)}{x - a} > 0$ for x sufficiently close to a . In other words, there is $\delta > 0$, such that

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

This leads to

$$a - \delta < x < a \implies f(x) < f(a),$$

and

$$a + \delta > x > a \implies f(x) > f(a).$$

This shows that a is not a local extreme of $f(x)$. ■

Since the proof makes explicit use of the left and the right side of a , the criterion does not work for left and right derivatives. In particular, it can only be applied to interior points of intervals.

For a function $f(x)$ defined on an interval, the following are the possible candidates for the local extremes.

- Ends points of the interval.
- Points inside the interval where $f(x)$ is not differentiable.
- Points inside the interval where $f(x)$ is differentiable and has derivative 0.

Note that the three cases are only the possibilities candidates for the local extremes. Whether or not these candidates are actually local extrema need to be studied case by case.

Example 4.1.2 The function x is always differentiable, but the derivative never vanishes. Therefore on an interval, the local extremes must be the end points. Indeed, on $[0, 1]$, the function has local extrema at the end points 0 and 1 of the closed interval. On the other hand, the function has no local extrema on $(0, 1)$ because the interval does not include the ends.

The function $\sin x$ is always differentiable, and $(\sin x)' = \cos x = 0$ exactly when $x = (n + \frac{1}{2})\pi$. We have $\sin x = \pm 1$ at these points. Since $-1 \leq \sin x \leq 1$ for all x , these are all the extremes of the sine function.

The function x^2 is always differentiable, and the derivative vanishes only at 0. Since $x^2 \geq 0^2$ for all x , 0 is a minimum of the function. If we consider the function on the closed intervals $[-1, 2]$, then the end points -1 and 2 are the only other possible local extrema. Since

$$-1 \leq x \leq 0 \implies (-1)^2 \geq x^2, \quad -1 \leq x \leq 2 \implies 2^2 \geq x^2,$$

-1 is a local maximum and 2 is a global maximum. If we consider the function on the open interval $(-1, 2)$, however, there is no more candidates for local extrema, so that 0 is the only local extreme of the function on $(-1, 2)$.

The function $f(x) = \begin{cases} x^2, & \text{if } x \neq 0, \\ 1, & \text{if } x = 0 \end{cases}$ is differentiable everywhere except 0, and the derivative never vanishes. Therefore on $[-1, 2]$, the only candidates for the local extrema are 0 and the end points -1 and 2 . Since

$$-1 < x < 1 \implies f(x) < 1 = f(0),$$

we see that 0 is a local maximum. The end points are also local maxima. The function has no local minimum. ■

Example 4.1.3 The cubic function x^3 is differentiable everywhere and the derivative vanishes only at 0. Moreover, we have $x^3 < 0^3$ for $x < 0$ and $x^3 > 0^3$ for $x > 0$. Therefore 0 is not a local extreme, despite it satisfies the criterion. ■

Example 4.1.4 The function

$$f(x) = \begin{cases} \frac{3}{2} + \left(x - \frac{1}{2}\right) \sin(1 - 2x) + \sin\left(x - \frac{1}{2}\right), & \text{if } -1 \leq x \leq 0, \\ -\frac{1}{2} - x \sin 2x + \sin x, & \text{if } 0 < x \leq 1 \end{cases}$$

is shown in Fig. 4.2. Six points $-1, x_1, 0, x_2, x_3$ and 1 , ordered from the left to the right, are candidates for local extrema. The function is not differentiable at 0 , and 0 is a global maximum. The function also has global minimum at x_3 . ■

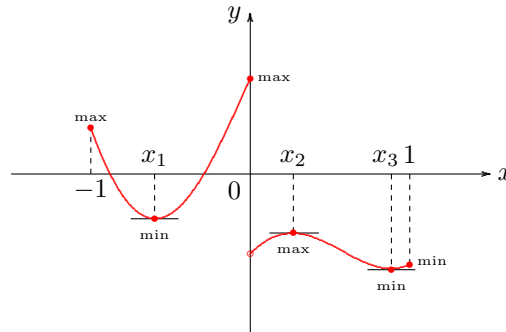


Fig. 4.2 The function has a maximum at a non-differentiable point.

Optimization Problem

Optimization problem typically try to maximize or minimize some quantity. Such problems can often be interpreted as finding the global extrema of a function on some interval. Since global extrema are also local extrema, we may often first find candidates for local extrema. Then we may often select the largest or the smallest values at such local extrema. In case the intervals are open, the limit of the function at the end points may also need to be considered.

Example 4.1.5 We try to enclose a rectangular region by a fence of total length 400 meters. What is the maximal region we can enclose? The problem is illustrated on the left of Fig. 4.3.

Assume that one side of the rectangular region is x . Then the area of the enclosed rectangle is

$$f(x) = x(200 - x).$$

From practical consideration, we try to find the global minimum of $f(x)$ on the interval $[0, 200]$. By $f'(x) = 200 - 2x$, the candidates for the local extrema are $0, 100$ and 200 . Since global extrema are also local extrema, one of the three must be the global extreme. Moreover, the global maximum exists because we have a continuous function on a closed bounded interval. By

$$f(0) = 0, \quad f(100) = 10,000, \quad f(200) = 0,$$

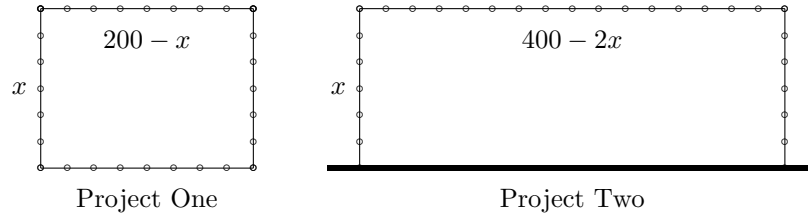


Fig. 4.3 The maximization projects

We see that the maximal area 10,000 is reached when $x = 100$. The maximal rectangle is actually a square.

If we already have a wall on one side and only need to build fences on three sides, then the enclosed area is

$$g(x) = x(400 - 2x).$$

where x is the side of the fence perpendicular to the wall. See the right of Fig. 4.3. Since $g(x) = 2f(x)$, the maximal area 20,000 is again reached when $x = 100$. ■

Example 4.1.6 We try to make a box with square bottom and top, such that the volume of the box is V . We wish to use the least amount of material to make such a box.

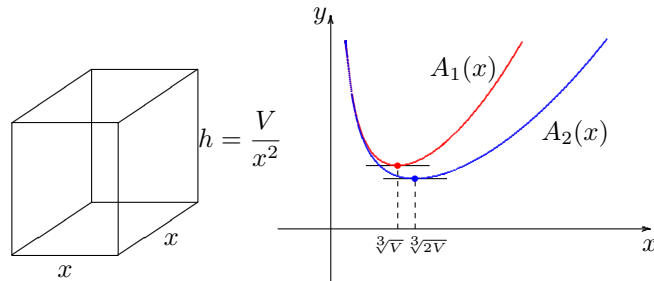


Fig. 4.4 Two designs of metal boxes.

The amount of material is measured by the surface area

$$f(x) = 2x^2 + 4xh$$

of the box, where x is the side length of the box and h is the height. Since the volume $V = x^2h$ is fixed, the problem becomes the global minimum of

$$f(x) = 2x^2 + 4\frac{V}{x}, \quad 0 < x < +\infty.$$

The function has derivative

$$f'(x) = 4x - 4\frac{V}{x^2}$$

on the interval, so that the only candidate for the local extreme is $a = \sqrt[3]{V}$.

Is this candidate a global minimum? By

$$\lim_{x \rightarrow 0^+} f(x) = \lim_{x \rightarrow +\infty} f(x) = +\infty,$$

we can find $\delta, N > 0$, such that

$$x \leq \delta \text{ or } x \geq N \implies f(x) > f(a).$$

Now on the closed and bounded interval $[\delta, N]$, the only candidates for local extrema are a and the ends points δ, N , where the values $f(\delta) > f(a)$ and $f(N) > f(a)$. Therefore a is a global minimum of the function on $[\delta, N]$. Combined with $f(x) > f(a)$ on $(0, +\infty) - [\delta, N]$, we conclude that $f(x) > f(a)$ for any $0 < x < +\infty$. In other words, $a = \sqrt[3]{V}$ is indeed a global minimum. Note that since $h = \frac{V}{a} = \sqrt[3]{V}$ in this case, we see that the most economical way of building the box is to keep the ratio of the sides as

$$\text{length} : \text{width} : \text{height} = 1 : 1 : 1.$$

Now we modify the problem a little bit by building a box with bottom but without the top. Then the problem becomes minimizing

$$g(x) = x^2 + 4\frac{V}{x}, \quad 0 < x < +\infty.$$

By

$$g'(x) = 2x - 4\frac{V}{x^2}, \quad \lim_{x \rightarrow 0^+} g(x) = \lim_{x \rightarrow +\infty} g(x) = +\infty,$$

we find the global minimum is reached at $a = \sqrt[3]{2V}$. In this case, the ratio of the sides is

$$\text{length} : \text{width} : \text{height} = 2 : 2 : 1. \quad \blacksquare$$

Exercises

4.1.1 A rectangle is inscribed in an isosceles triangle in Fig. 4.5. Show that the biggest area possible is half of the area of the triangle.

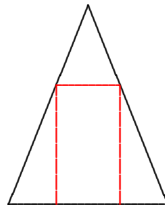


Fig. 4.5 A rectangle is inscribed in an isosceles triangle.

4.1.2 Among all the rectangles with area A , which one has the smallest perimeter?

4.1.3 Among all the rectangles with perimeter L , which one has the biggest area?

4.1.4 A rectangle is inscribed in a circle of radius R . When does the rectangle have the biggest area?

4.1.5 Determine the dimensions of the biggest rectangle that is inscribed in the ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$.

4.1.6 A right circular cone is inscribed in a sphere of radius R in Fig. 4.6. What is the biggest volume?

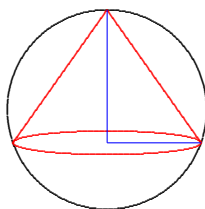


Fig. 4.6 A right circular cone inscribed in a sphere.

4.1.7 Find the volume of the biggest right circular cone with a given slant height l .

4.1.8 What is the shortest distance from the point $(2, 1)$ to the parabola $y = 2x^2$?

4.1.9 Find the shortest distance from the point (x_0, y_0) to the plane $Ax + Bx + C = 0$.

4.2 MEAN VALUE THEOREM

If the average speed of a train between two cities is 100km per hour, we expect that the speed reaches exactly 100km per hour at some time during the trip. Let $s(t)$ be the distance traveled by the time t . Then the average speed from the time $t = a$ to $t = b$ is

$$v_{[a,b]} = \frac{s(b) - s(a)}{b - a}.$$

Our expectation can be interpreted as

$$\frac{s(b) - s(a)}{b - a} = s'(c)$$

for some $c \in (a, b)$.

Theorem 4.2.1 (Mean Value Theorem) If $f(x)$ is continuous on a closed and bounded interval $[a, b]$ and is differentiable on the open interval (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

$$f(a + h) - f(a) = f'(a + \theta h)h, \text{ for some } 0 < \theta < 1.$$

Geometrically, the theorem means that the tangent of the function is parallel to the straight line connecting the two ends $(a, f(a))$ and $(b, f(b))$ somewhere along the interval. We also note that there is no need to insist $a < b$ in the conclusion.

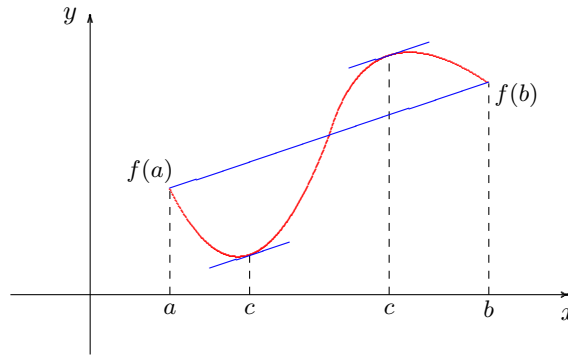


Fig. 4.7 Mean Value Theorem

Proof. From the picture, it appears that the parallel tangent is reached when the distance between the function and the straight line connecting the two ends becomes extremal.

The straight line connecting the two ends is

$$y = f(a) + \frac{f(b) - f(a)}{b - a}(x - a).$$

The distance between the function and the straight line is

$$d(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b - a}(x - a).$$

By the assumption on $f(x)$, the distance is continuous on $[a, b]$ and differentiable on (a, b) .

By Theorem 4.1.1, $d(x)$ must have global maximal and global minimum. If the a global extreme is reached at $c \in (a, b)$, then by Theorem 4.1.2, we get

$$d'(c) = f'(c) - \frac{f(b) - f(a)}{b - a} = 0.$$

If there is no global extreme in the interior of $[a, b]$, then the only global extrema are at the ends. Since $d(a) = d(b) = 0$, we see that 0 is the maximal as well as the minimal value of $d(x)$ on $[a, b]$. This implies that $d(x) = 0$ throughout $[a, b]$. In this case, we have $f'(x) = \frac{f(b) - f(a)}{b - a}$ throughout the interval. ■

The Mean Value Theorem has the following important consequences.

Corollary 4.2.2 *If $f'(x) = 0$ for all $x \in (a, b)$, then $f(x)$ is a constant on (a, b) .*

Proof. For $x_1, x_2 \in (a, b)$, by applying the Mean Value Theorem to the function on the interval $[x_1, x_2]$ (and using differentiability implies continuity), there is c between x_1 and x_2 , such that

$$f(x_2) - f(x_1) = f'(c)(x_2 - x_1) = 0. \quad \blacksquare$$

Corollary 4.2.3 *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) .*

The corollary may be proved by applying Corollary 4.2.2 to $f(x) - g(x)$.

Example 4.2.1 Applying the Mean Value Theorem to $\ln x$, we get

$$\ln(1+x) - \ln 1 = \frac{1}{1+\theta x}x, \quad \text{for some } 0 < \theta < 1.$$

Since

$$\frac{x}{1+x} \leq \frac{1}{1+\theta x}x \leq x$$

for $x > -1$, we conclude that

$$\frac{x}{1+x} \leq \ln(1+x) \leq x. \quad \blacksquare$$

Example 4.2.2 For the function $|x|$ on $[-1, 1]$, there is no $c \in (-1, 1)$ satisfying

$$f(1) - f(-1) = 0 = f'(c)(1 - (-1)).$$

We do not have the conclusion of the Mean Value Theorem because $|x|$ is not differentiable at 0. ■

Example 4.2.3 We know the function $f(x) = e^x$ satisfies $f' = f$. Is there any other function also satisfying $f' = f$?

Suppose $f' = f$. Then

$$(e^{-x}f(x))' = (e^{-x})'f(x) + e^{-x}(f(x))' = -e^{-x}f(x) + e^{-x}f'(x) = e^{-x}(-f(x) + f'(x)) = 0.$$

Therefore $e^{-x}f(x) = C$ is a constant, and $f(x) = Ce^{-x}$. ■

Cauchy's Mean Value Theorem

The Mean Value Theorem may be extended as follows.

Theorem 4.2.4 (Cauchy's Mean Value Theorem) *If $f(x)$ and $g(x)$ are continuous on a closed and bounded interval $[a, b]$ and are differentiable on the open interval (a, b) , such that $g'(x) \neq 0$ on (a, b) , then there is $c \in (a, b)$, such that*

$$\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(c)}{g'(c)}.$$

The theorem can be proved similar to the Mean Value Theorem, simply by changing $b - a$ to $g(b) - g(a)$ and $x - a$ to $g(x) - g(a)$. Geometrically, Cauchy's Mean Value Theorem means that for a differentiable parametrized curve

$$x = f(t), \quad y = g(t), \quad a \leq t \leq b$$

on the plane, some tangent vector $(f'(c), g'(c))$ along the curve is parallel to the straight line connecting the end points $(f(a), g(a))$ and $(f(b), g(b))$ of the curve.

Example 4.2.4 Cauchy's Mean Value Theorem may be used to prove Theorem 3.3.2 on the high order approximation. Here we present the argument in case $n = 2$.

Suppose $f(x)$ has first order derivative on $(a - \delta, a + \delta)$ and has second order derivative at a . The difference (called the **remainder**)

$$R(x) = f(x) - f(a) - f'(a)(x - a) - \frac{f''(a)}{2}(x - a)^2$$

between $f(x)$ and its quadratic approximation satisfies

$$R(a) = R'(a) = R''(a) = 0.$$

The approximation theorem basically says $\lim_{x \rightarrow a} \frac{R(x)}{(x - a)^2} = 0$.

By taking $f(x)$ to be $R(x)$ and $g(x) = (x - a)^2$ in Cauchy's Mean Value Theorem, we get

$$\frac{R(x)}{(x - a)^2} = \frac{R(x) - R(a)}{g(x) - g(a)} = \frac{R'(c)}{g'(c)}$$

for some $c \in (a, x)$. Since $R'(a) = 0$, we further have

$$\lim_{x \rightarrow a} \frac{R'(c)}{g'(c)} = \lim_{c \rightarrow a} \frac{R'(c) - R'(a)}{2(c - a)} = \frac{1}{2}R''(a) = 0,$$

where the first equality is due to $x \rightarrow a$ implying $c \rightarrow a$, and the second equality is the definition of the second order derivative. ■

Example 4.2.5 We can use Cauchy's Mean Value Theorem to prove the following inequality:

$$\ln(1 + x) < \frac{x}{\sqrt{1 + x}}, \quad x > 0.$$

In fact, for any $x > 0$, there is a number $\xi \in (0, x)$ such that

$$\frac{\ln(1+x)}{\frac{x}{\sqrt{1+x}}} = \frac{\frac{1}{1+\xi}}{(1+\xi)^{-1/2} + (-\frac{1}{2})\xi(1+\xi)^{-3/2}} = \frac{(1+\xi)^{1/2}}{1+\xi/2} = \left(\frac{1+\xi}{1+\xi+\xi^2/4}\right)^{1/2} < 1.$$

This yields the inequality. ■

Exercises

4.2.1 Prove that if the derivative of a function is constant on an interval, then the function is linear on the interval.

4.2.2 Solve the differential equation on an interval:

1. $f'(x) = \lambda f(x)$;

3. $f'(x) = [f(x)]^2$;

2. $f'(x) = xf(x)$;

4. $f'(x) = \frac{1}{f(x)}$.

4.2.3 Prove that $3 \arccos x - \arccos(3x - 4x^3) = \pi$ on $\left[-\frac{1}{2}, \frac{1}{2}\right]$.

4.2.4 Show that for $x \neq 0$,

$$\arctan x + \arctan x^{-1} = \frac{\pi}{2}.$$

4.2.5 Show that

$$\begin{aligned} \arctan \frac{x+a}{1-ax} - \arctan x &= \arctan a, & \text{if } ax < 1; \\ \arctan \frac{x+a}{1-ax} - \arctan x &= \arctan a - \pi, & \text{if } ax > 1. \end{aligned}$$

4.2.6 Use the Mean Value Theorem to prove the inequality:

1. $|\sin a - \sin b| \leq |a - b|$;

2. $\frac{a-b}{a} < \ln \frac{a}{b} < \frac{a-b}{b}$ for $a > b > 0$;

3. $\arctan b - \arctan a \leq 2 \arctan \frac{b-a}{2}$ for $b > a > 0$.

4.2.7 Use Cauchy's Mean Value Theorem to prove the inequality: for $x > 0$

$$\frac{x^2}{2(1+x)} < x - \ln(1+x) < \frac{x^2}{2}.$$

What about for $-1 < x < 0$?

4.3 L'HOSPITAL'S RULE

L'Hospital's rule is a powerful method for finding limits of **indeterminate forms** such as

$$\lim_{x \rightarrow 0} \frac{\ln(1+x)}{x}, \quad \lim_{x \rightarrow +\infty} \frac{x^2}{e^x}, \quad \lim_{x \rightarrow 0} \frac{\sin x - \tan x}{x^3}.$$

These are the limits of quotients in which both the numerator and the denominator tend to 0 or ∞ , and usual arithmetic rule cannot be applied.

The indeterminate forms of quotient type are either $\frac{0}{0}$ type or $\frac{\infty}{\infty}$ type. The following are the other types of indeterminate forms

$$\begin{array}{ccccc} \lim_{x \rightarrow 0} x \ln x, & \lim_{x \rightarrow 0} \left(\frac{1}{x} - \frac{1}{e^x - 1} \right), & \lim_{x \rightarrow 0^+} x^x, & \lim_{x \rightarrow 0} (1 + \sin x)^x, & \lim_{x \rightarrow +\infty} x^{\frac{1}{x}}. \\ \text{(type } 0 \cdot \infty) & \text{(type } \infty - \infty) & \text{(type } 0^0) & \text{(type } 1^\infty) & \text{(type } \infty^0) \end{array}$$

These types can often be converted to either $\frac{0}{0}$ type or $\frac{\infty}{\infty}$ type.

Theorem 4.3.1 (L'Hospital's Rule) Suppose $f(x)$ and $g(x)$ are differentiable functions on (a, b) , with $g'(x) \neq 0$. Suppose

1. Either $\lim_{x \rightarrow a^+} f(x) = \lim_{x \rightarrow a^+} g(x) = 0$ or $\lim_{x \rightarrow a^+} f(x) = \lim_{x \rightarrow a^+} g(x) = \infty$.
2. $\lim_{x \rightarrow a^+} \frac{f'(x)}{g'(x)}$ converges.

Then $\lim_{x \rightarrow a^+} \frac{f(x)}{g(x)}$ also converges, and

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

Of course we also have the left limit version of the theorem, and the two sides can be combined to give the two side limit version. Moreover, we allow a or $\lim_{x \rightarrow a^+} \frac{f'(x)}{g'(x)}$ to be any kind of infinity.

Proof. We only prove the simple case that $\lim_{x \rightarrow a^+} f(x) = \lim_{x \rightarrow a^+} g(x) = 0$ and a is a finite number.

Given the assumptions, the functions

$$\tilde{f}(x) = \begin{cases} f(x), & \text{if } a < x < b, \\ 0, & \text{if } x = a, \end{cases} \quad \tilde{g}(x) = \begin{cases} g(x), & \text{if } a < x < b, \\ 0, & \text{if } x = a \end{cases}$$

are continuous on $[a, y]$ and differentiable on (a, y) for any $a < y < b$. Applying Cauchy's mean value theorem to $\tilde{f}(x)$ and $\tilde{g}(x)$ on $[a, y]$, we get

$$\frac{f(y)}{g(y)} = \frac{\tilde{f}(y) - \tilde{f}(a)}{\tilde{g}(y) - \tilde{g}(a)} = \frac{\tilde{f}'(c)}{\tilde{g}'(c)} = \frac{f'(c)}{g'(c)},$$

for some $a < c < y$. Since $y \rightarrow a^+$ implies $c \rightarrow a^+$, the convergence of $\lim_{c \rightarrow a^+} \frac{f'(c)}{g'(c)}$ implies that $\lim_{y \rightarrow a^+} \frac{f(y)}{g(y)}$ also converges to the same limit. ■

Example 4.3.1 By applying L'Hospital's rule repeatedly, for $a > 1$ we get

$$\lim_{x \rightarrow +\infty} \frac{x^2}{a^x} = \lim_{x \rightarrow +\infty} \frac{2x}{a^x \ln a} = \lim_{x \rightarrow +\infty} \frac{2}{a^x (\ln a)^2} = 0.$$

Here is the precise logic behind the computation. First because $\lim_{x \rightarrow +\infty} 2x = \lim_{x \rightarrow +\infty} a^x \ln a = \infty$ and the last limit converges, we know the second limit converges. Then because $\lim_{x \rightarrow +\infty} x^2 = \lim_{x \rightarrow +\infty} a^x = \infty$ and the second limit converges, we know the first limit also converges. Moreover, L'Hospital's rule tells us that all the limits are the same. ■

Example 4.3.2 By $\lim_{x \rightarrow 0} \frac{\sin x}{x} = 1$ and the L'Hospital's Rule, we have

$$\lim_{x \rightarrow 0} \frac{1 - \cos x}{x^2} = \lim_{x \rightarrow 0} \frac{(1 - \cos x)'}{(x^2)'} = \lim_{x \rightarrow 0} \frac{\sin x}{2x} = \frac{1}{2}.$$

It is tempting to further use L'Hospital's Rule to get

$$\lim_{x \rightarrow 0} \frac{\sin x}{x} = \lim_{x \rightarrow 0} \frac{(\sin x)'}{(x)'} = \lim_{x \rightarrow 0} \cos x = 1.$$

However, such argument is logically circular because we have already used the limit $\lim_{x \rightarrow 0} \frac{\sin x}{x}$ to derive the derivative of $\sin x$, and yet such derivative is used back in the computation of the limit. Such logical problem also occurs if you try to use L'Hospital's Rule to compute $\lim_{x \rightarrow 0} \frac{\ln(1+x)}{x}$, which is actually the definition of $(\ln x)'|_{x=1}$. ■

Example 4.3.3 The following limit was computed in Example 3.3.9.

$$\lim_{x \rightarrow 0} \frac{\sin x - \tan x}{x^3} = \lim_{x \rightarrow 0} \frac{\cos x - \sec^2 x}{2x^2} = \lim_{x \rightarrow 0} \frac{\cos^3 x - 1}{2x^2} = \lim_{x \rightarrow 0} \frac{-3 \cos^2 x \sin x}{6x} = -\frac{1}{2}.$$

In the second equality, we simplified the limit so that the derivative of the numerator is easier to compute. L'Hospital's Rule is used in the first and the third equalities. ■

Example 4.3.4 The limit $\lim_{x \rightarrow \infty} \frac{x + \sin x}{x}$ is of $\frac{\infty}{\infty}$ type. However, the L'Hospital's Rule

$$\lim_{x \rightarrow \infty} \frac{x + \sin x}{x} = \lim_{x \rightarrow \infty} (1 + \cos x)$$

does not appear to work because the left converges to 1 and the right diverges. The reason is because the second condition in Theorem 4.3.1 is not satisfied.

The lesson is that if the first condition is satisfied, then we can use the convergence of $\lim_{x \rightarrow a^+} \frac{f'}{g'}$ to conclude the convergence of $\lim_{x \rightarrow a^+} \frac{f}{g}$ but not vice versa. ■

Example 4.3.5 Many indeterminate forms can be converted to the $\frac{0}{0}$ or $\frac{\infty}{\infty}$ type and then computed by L'Hospital's rule. For example, the limit $\lim_{x \rightarrow 0^+} x \ln x$ is of $0 \cdot \infty$ type. We have

$$\lim_{x \rightarrow 0^+} x \ln x = \lim_{x \rightarrow 0^+} \frac{\ln x}{\frac{1}{x}} = \lim_{x \rightarrow 0^+} \frac{(\ln x)'}{\left(\frac{1}{x}\right)'} = \lim_{x \rightarrow 0^+} \frac{\frac{1}{x}}{-\frac{1}{x^2}} = \lim_{x \rightarrow 0^+} -x = 0.$$

By $x^x = e^{x \ln x}$, we further get

$$\lim_{x \rightarrow 0^+} x^x = 1,$$

and

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

By similar argument, we have

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

Taking a positive power of the limit, we further get

$$\lim_{x \rightarrow 0^+} x^p (\ln x)^q = 0, \text{ for any } p, q > 0.$$

By converting x to $\frac{1}{x}$, we also have

$$\lim_{x \rightarrow +\infty} \frac{(\ln x)^q}{x^p} = 0, \text{ for any } p, q > 0. \quad \blacksquare$$

Example 4.3.6 Here is the computation of a $\infty - \infty$ type limit

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

The limit was already computed in Example 3.3.9. Can you see that the two computations are essentially the same? ■

Exercises

4.3.1 Compute limit. You may combine l'Hospital's rule, high order approximation, and known limits:

1. $\lim_{x \rightarrow 0} x^p e^{-x^q};$
2. $\lim_{x \rightarrow +\infty} x^p e^{-x^q};$
3. $\lim_{x \rightarrow +\infty} x^p \ln x;$
4. $\lim_{x \rightarrow 1} \frac{x^x - x}{\ln x - x + 1};$
5. $\lim_{x \rightarrow 0+} \frac{a^x - a^{\sin x}}{x^3}, \text{ for } a > 0;$
6. $\lim_{x \rightarrow 0+} \frac{\ln(\sin ax)}{\ln(\sin bx)}, \text{ for } a, b > 0;$
7. $\lim_{x \rightarrow 0+} \frac{\ln(\cos ax)}{\ln(\cos bx)};$
8. $\lim_{x \rightarrow 0} \frac{\cos(\sin x) - \cos x}{x^4};$
9. $\lim_{x \rightarrow 0} \frac{\arcsin 2x - 2 \arcsin x}{x^3};$
10. $\lim_{x \rightarrow 0} \frac{1 - \cos x^2}{x^2 \sin x^2};$
11. $\lim_{x \rightarrow \frac{\pi}{6}} \frac{1 - 2 \sin x}{\cos 3x};$
12. $\lim_{x \rightarrow 0+} (x^{x^x} - 1);$
13. $\lim_{x \rightarrow 0+} x^{x^{x-1}};$
14. $\lim_{x \rightarrow 0} x^{\sin x};$
15. $\lim_{x \rightarrow 0+} (-\ln x)^x;$
16. $\lim_{x \rightarrow 1} \left(\frac{1}{\ln x} - \frac{1}{x-1} \right);$
17. $\lim_{x \rightarrow 0} \left(\frac{1}{\ln(x + \sqrt{1+x^2})} - \frac{1}{\ln(1+x)} \right);$
18. $\lim_{x \rightarrow \infty} x^3 \left(\sin \frac{1}{x} - \frac{1}{2} \sin \frac{2}{x} \right);$
19. $\lim_{x \rightarrow 1+} \ln x \tan \frac{\pi x}{2};$
20. $\lim_{x \rightarrow a} \frac{a^x - x^a}{x - a}, a > 0;$
21. $\lim_{x \rightarrow 0} \frac{(a+x)^x - a^x}{x^2};$
22. $\lim_{x \rightarrow 0} \left(e^{-1} (1+x)^{\frac{1}{x}} \right)^{\frac{1}{x}};$
23. $\lim_{x \rightarrow 0} (x^{-1} \arcsin x)^{\frac{1}{x^2}};$
24. $\lim_{x \rightarrow 0} \left(\frac{2}{\pi} \arccos x \right)^{\frac{1}{x}};$
25. $\lim_{x \rightarrow 0} \left(\frac{\cos x}{1 + \sin x} \right)^{\frac{1}{x}}.$

4.3.2 Suppose $f(x)$ has second order derivative near a . Prove that

$$f''(a) = \lim_{h \rightarrow 0} \frac{f(a+h) + f(a-h) - 2f(a)}{h^2}.$$

4.3.3 Show that the limit converges but one cannot be computed by L'Hospital's Rule:

1. $\lim_{x \rightarrow 0} \frac{x^2 \sin \frac{1}{x}}{\sin x};$
2. $\lim_{x \rightarrow \infty} \frac{x - \sin x}{x + \sin x}.$

4.3.4 For

$$f(x) = \begin{cases} e^{-1/x^2}, & \text{if } x \neq 0, \\ 0, & \text{if } x = 0, \end{cases}$$

show that $f^{(n)}(0) = 0$, $n = 1, 2, \dots$

4.4 INCREASING AND DECREASING

A function f is increasing if $x_1 > x_2$ implies $f(x_1) > f(x_2)$. If $x_1 > x_2$ implies $f(x_1) \geq f(x_2)$, the function f is nondecreasing. The function is decreasing if $x_1 > x_2$ implies $f(x_1) < f(x_2)$. Similarly, if $x_1 > x_2$ implies $f(x_1) \leq f(x_2)$, the function f is nonincreasing.

If two functions are close to each other, then it is very likely that one function is increasing implies that the other function is also increasing.

If $f(x)$ is differentiable, then it is approximated by the linear function $L(x) = f(a) + f'(a)(x - a)$ near a . Since $L(x)$ is increasing if and only if the slope $f'(a) > 0$, it is very likely that $f'(a) > 0$ implies that $f(x)$ is increasing near a . This is the intuition behind the following criterion for determining the monotonic property of functions.

Theorem 4.4.1 Suppose $f(x)$ is continuous on an interval and differentiable in the interior of the interval. If $f'(x) \geq 0$ on the interval, then the function is nondecreasing. If $f'(x) > 0$ on the interval, then the function is increasing.

Similar statements hold for decreasing functions.

Proof. Suppose $x_1 > x_2$. Then by applying the mean value theorem to the function on the interval $[x_2, x_1]$, we get

$$f(x_1) - f(x_2) = f'(c)(x_2 - x_1), x_1 > c > x_2.$$

If $f' \geq 0$ everywhere, then $f(x_1) \geq f(x_2)$. If $f' > 0$ everywhere, then $f(x_1) > f(x_2)$. ■

Example 4.4.1 The function $f(x) = x^3 - 3x + 1$ has derivative $f'(x) = 3x^2 - 3 = 3(x + 1)(x - 1)$. The sign of the derivative and the implication on the monotonic property is illustrated below.

interval	sign of f'	monotonicity
$(-\infty, -1]$	+	↗
$[-1, 1]$	−	↘
$[1, +\infty)$	+	↗

Moreover, we know $f'(-1) = f'(1) = 0$, which makes ± 1 the candidates for local extremes. From the table above, we see that

$$x < -1 \implies f(x) < f(-1), \quad -1 < x < 1 \implies f(-1) < f(x).$$

Therefore -1 is indeed a local maximum. Similarly, 1 is a local minimum.

For a more sophisticated example, consider the function $g(x) = 12x^5 - 45x^4 + 40x^3 + 5$, which has derivative

$$g'(x) = 60x^4 - 180x^3 + 120x^2 = 60x^2(x - 1)(x - 2).$$

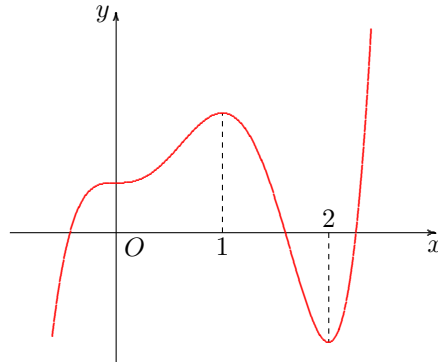


Fig. 4.8 The graph of $12x^5 - 45x^4 + 40x^3 + 5$

Then we can make the following table.

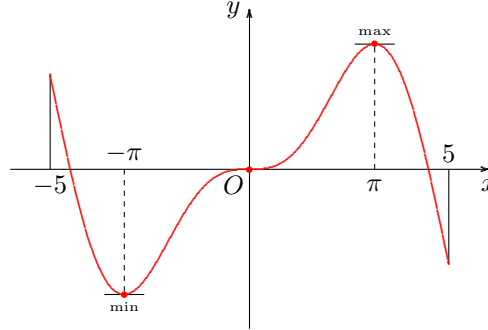
interval	sign of g'	monotonicity
$(-\infty, 0]$	+	$\nearrow$
0	0	not local extreme
$[0, 1]$	+	$\nearrow$
1	0	local max
$[1, 2]$	-	$\searrow$
2	0	local min
$[2, +\infty)$	+	$\nearrow$

The candidate 0 for the local extreme is actually not a local extreme because $g(0)$ is larger than the left side and smaller than the right side. ■

Example 4.4.2 Consider the function $f(x) = \sin x - x \cos x$ on $[-5, 5]$. The behavior of its derivative $f'(x) = x \sin x$ is illustrated below.

interval	sign of f'	monotonicity
-5	end	local max
$[-5, -\pi]$	-	$\searrow$
$-\pi$	0	local min
$[-\pi, 0]$	+	$\nearrow$
0	0	not local extreme
$[0, \pi]$	+	$\nearrow$
π	0	local max
$[\pi, 5]$	-	$\searrow$
5	end	local min

Example 4.4.3 Let us compare the trigonometric functions and their high order approximations at 0 in Example 3.3.8. We already know the comparison $\sin x < x$ for $0 < x < \frac{\pi}{2}$. In fact, since

Fig. 4.9 The graph of $\sin x - x \cos x$

$(x - \sin x)' = 1 - \cos x > 0$ everywhere except $x = n\pi$, we see that $x - \sin x$ is increasing. Therefore $x > 0$ implies $x - \sin x > 0 - \sin 0 = 0$, and we get

$$\sin x < x \quad \text{for } x > 0.$$

To compare $\cos x$ with the quadratic approximation $1 - \frac{x^2}{2!}$, we note that for $x > 0$,

$$\left(1 - \frac{x^2}{2!} - \cos x\right)' = -x + \sin x < 0.$$

Therefore $1 - \frac{x^2}{2!} - \cos x$ is decreasing for $x \geq 0$, so that $1 - \frac{x^2}{2!} - \cos x < 1 - \frac{0^2}{2!} - \cos 0 = 0$ for $x > 0$. Thus,

$$\cos x > 1 - \frac{x^2}{2!} \quad \text{for } x > 0.$$

Now we compare $\sin x$ with its cubic approximation $x - \frac{x^3}{3!}$. By

$$\left(x - \frac{x^3}{3!} - \sin x\right)' = 1 - \frac{x^2}{2!} - \cos x < 0 \quad \text{for } x > 0$$

and $0 - \frac{0^3}{3!} - \sin 0 = 0$, we get

$$\sin x > x - \frac{x^3}{3!} \quad \text{for } x > 0.$$

The comparison can continue with all the high order approximations of $\sin x$ and $\cos x$. ■

Example 4.4.4 We prove the inequality

$$(a^p + b^p)^{\frac{1}{p}} < (a^q + b^q)^{\frac{1}{q}}$$

for $a, b > 0$ and $p > q > 0$.

By taking $c = \frac{b}{a}$, the claim is the same as the function $f(x) = (1 + c^x)^{\frac{1}{x}}$ being decreasing on $(0, +\infty)$. This is further equivalent to $\ln f(x) = \frac{\ln(1 + c^x)}{x}$ being decreasing.

We have

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

This implies that $\ln f$ is indeed decreasing. ■

Determination of Local Extrema

In Section 4.1, we found the candidates of local extrema. In Examples 4.4.1 and 4.4.2, we further use the criterion for the monotonicity (Theorem 4.4.1) to determine whether the candidates are indeed local extrema. The method can be summarized as the **First Derivative Test**: Suppose that $f(x)$ is continuous at a and differentiable near a (but not necessarily differentiable at a).

1. If $f' \geq 0$ on the left of a and $f' \leq 0$ on the right of a , then a is a local maximum.
2. If $f' \leq 0$ on the left of a and $f' \geq 0$ on the right of a , then a is a local minimum.
3. If f' does not change sign at a (i.e., $f' > 0$ on both sides of a , or $f' < 0$ on both sides of a), then a is not a local extreme.

Suppose $f(x)$ has second order derivative at a candidate a for the local extreme. Then $f(x)$ should have first order derivative near a , and $f'(a) = 0$. If $f''(a) > 0$, then by $f'(a) = 0$, we get $f'(x) < 0$ on the left of a and $f'(x) > 0$ on the right of a (see proof of Theorem 4.1.2, taking f' here as the function f in the proof). Therefore by the criterion above, a is a local minimum. This proves the following test for local extrema.

Theorem 4.4.2 (Second Derivative Test) Suppose $f(x)$ has second order derivative at a .

1. If $f'(a) = 0$ and $f''(a) < 0$, then a is a local maximum.
2. If $f'(a) = 0$ and $f''(a) > 0$, then a is a local minimum.

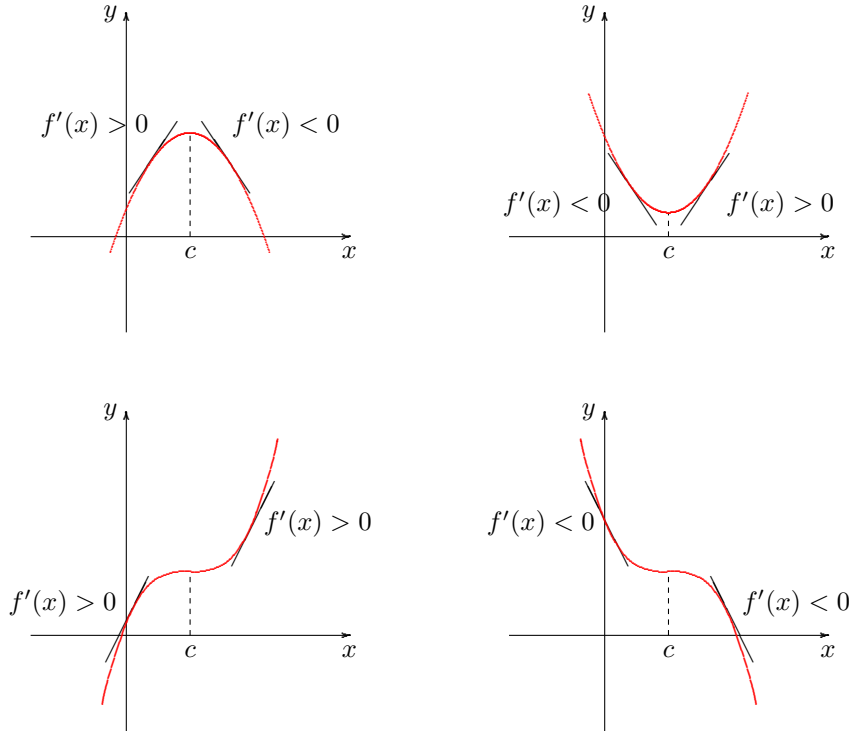
Example 4.4.5 In Example 4.4.1, we found the possible local extrema ± 1 for $f(x) = x^3 - 3x + 1$ and 0, 1, 2 for $g(x) = 12x^5 - 45x^4 + 40x^3 + 5$. By

$$f'' = 6x, \quad f''(-1) < 0, \quad f''(1) > 0,$$

we see that -1 is local maximum and 1 is local minimum for f . By

$$g'' = 60x(4x^2 - 9x + 4), \quad g''(0) = 0, \quad g''(1) < 0, \quad g''(2) > 0,$$

we see that 1 is local maximum and 2 is local minimum for g . That fact $g''(0) = 0$ does not tell us whether 0 is a local extrema. Further study is needed in order to determine the nature of 0 . ■

Fig. 4.10 The relation between extremum and sign change of $f'(x)$

Example 4.4.6 In Example 4.4.2. We found the possible local extrema $\pm 5, \pm\pi, 0$ for the function $f(x) = \sin x - x \cos x$ on $[-5, 5]$. We have

$$f''(x) = \sin x + x \cos x, \quad f''(-\pi) = \pi, \quad f''(0) = 0, \quad f''(\pi) = -\pi.$$

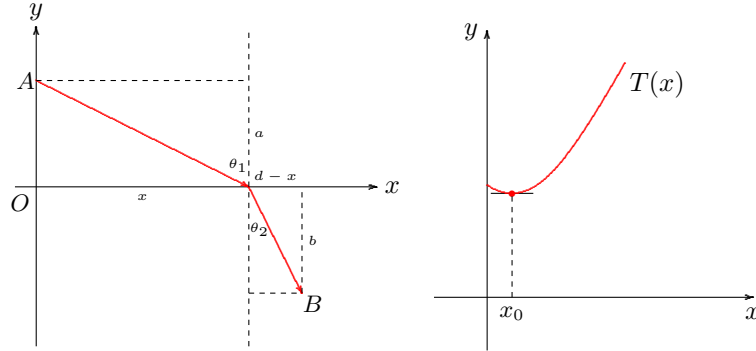
This tells us that $-\pi$ is a local minimum and π is a local maximum. However, we cannot draw any conclusion for $x = 0$ and the end points ± 5 from the second order derivative at these points. ■

Example 4.4.7 Fermat's principle says that light travels along the path of shortest traveling time. In Fig. 4.11, we consider a light ray traveling from point A located in one medium to point B in another medium. The media are separated by the x -axis. Assume that the speed of light is respectively c_1 and c_2 in the media. Denote by θ_1 the angle of incidence and θ_2 the angle of refraction.

The total traveling time from A to B is

$$T(x) = \frac{\sqrt{a^2 + x^2}}{c_1} + \frac{\sqrt{b^2 + (d-x)^2}}{c_2},$$

The problem is to minimize $T(x)$ on $[0, d]$.

Fig. 4.11 Fraction of light from A to B

We have

$$T'(x) = \frac{x}{c_1 \sqrt{a^2 + x^2}} - \frac{d-x}{c_2 \sqrt{b^2 + (d-x)^2}}.$$

By

$$T'(0) = -\frac{d}{c_2 \sqrt{b^2 + d^2}} < 0, \quad T'(d) = \frac{d}{c_1 \sqrt{a^2 + d^2}} > 0,$$

and the intermediate value theorem, the derivative must vanish somewhere on $[0, d]$. Moreover, by

$$T''(x) = \frac{a^2}{c_1 (a^2 + x^2)^{\frac{3}{2}}} + \frac{b^2}{c_2 (b^2 + (d-x)^2)^{\frac{3}{2}}} > 0,$$

the derivative is increasing, so that the place where the derivative vanishes is unique.

Let $x_0 \in (0, d)$ be the unique place where $T(x_0) = 0$. Since $T'(x_0) = 0$ and T' is increasing, we have $T' < 0$ on $(0, x_0)$ and $T' > 0$ on (x_0, d) . Therefore T is decreasing on $[0, x_0]$ and increasing on $[x_0, d]$. This implies that x_0 is the global minimum on $[0, d]$.

The equality $T'(x_0) = 0$ means

$$\frac{\sin \theta_1}{c_1} = \frac{x_0}{c_1 \sqrt{a^2 + x_0^2}} = \frac{d-x_0}{c_2 \sqrt{b^2 + (d-x_0)^2}} = \frac{\sin \theta_2}{c_2}.$$

This is Snell's law of refraction. ■

Example 4.4.8 The profit of a company makes from producing and selling x amount of a product is

$$p(x) = r(x) - c(x),$$

where $r(x)$ is the revenue from selling the product and $c(x)$ is the cost of producing the product. If $p(x) > 0$, the company makes money. Otherwise it loses money.

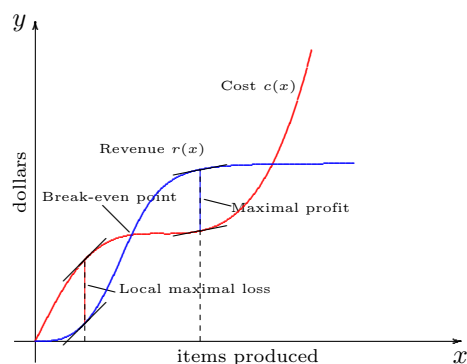


Fig. 4.12 A typical financial situation of a company

Typically, the company initially operates at a loss due to set-up cost of the production line and the advertisement. Later on, the company operates at a profit, illustrated by the cost function curve surpassing the revenue curve at the break even point. The maximal profit occurs at the point where

$$c'(x) = r'(x).$$

Eventually, the company may lose money due to a combination of rising labor and material cost and market saturation. Fig. 4.12 gives a schematic demonstration of the changes of the revenue and the cost of a typical company.

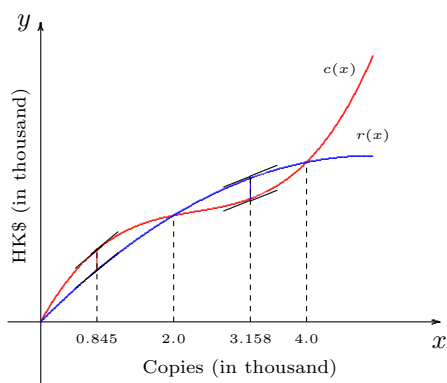


Fig. 4.13 Financial analysis of a publisher on one magazine.

Consider the specific financial situation of a magazine publisher. Suppose that before reaching 5,000 copies, the cost and revenue for printing and selling x thousand copies are given by (in the unit of 1,000 hong Kong dollars, for example)

$$c(x) = x^3 - 7x^2 + 18x, \quad r(x) = -x^2 + 10x.$$

The equation $r(x) = c(x)$ has two solutions 2 and 4 on $(0, 5]$, which indicates that the publisher begins to make money after selling 2,000 copies but then lose money after reaching 4,000 copies. The equation $r'(x) = c'(x)$ has two solutions $\frac{2}{3}(3 - \sqrt{3}) \approx 0.845$ and $\frac{2}{3}(3 + \sqrt{3}) \approx 3.158$. At these two solutions, we have $p''\left(\frac{2}{3}(3 - \sqrt{3})\right) = 4\sqrt{3} > 0$ and $p''\left(\frac{2}{3}(3 + \sqrt{3})\right) = -4\sqrt{3} < 0$. Therefore the company suffers the biggest loss when it sells 845 copies and earns the biggest profit when it sells 3,158 copies. Fig. 4.13 shows the financial analysis of the company. ■

Exercises

4.4.1 Determine the monotonicity property and find local and global extrema:

1. $x^3 - 3x + 2$;
2. $-x^4 + 2x^2 - 1$;
3. $\frac{x}{x^2 + 1}$;
4. $x + \frac{1}{x}$, $x \neq 0$;
5. $\frac{x + a}{x^2 + 1}$;
6. $\sqrt{3 + 2x - x^2}$, $-1 \leq x \leq 3$;
7. $x^2 e^x$;
8. $x^p a^x$, $a > 0$;
9. $x \ln x$, $x > 0$;
10. $x^3 + 3 \ln x$, $x > 0$;
11. $x - \ln(1 + x)$, $x > -1$;
12. $e^{-x} \sin x$;
13. $x - \sin x$, $0 \leq x \leq 2\pi$;
14. $2 \sin x + \sin 2x$, $0 \leq x \leq 2\pi$;
15. $2x - 4 \sin x + \sin 2x$, $0 \leq x \leq 2\pi$;
16. $|x^2 - 3x + 2|$, $0 \leq x \leq 4$;
17. $|x - \sin x|$, $-\frac{\pi}{2} \leq x \leq \frac{\pi}{2}$;
18. $\frac{1}{1 + \sin^2 x}$, $0 \leq x \leq 2\pi$;
19. $|x|e^{-|x-1|}$, $-2 \leq x \leq 2$.

4.4.2 Prove inequality:

1. $e^x > 1 + x$ for $x \neq 0$;
2. $e^x > 1 + x + \frac{x^2}{2}$ for $x > 0$;
3. $x - \frac{x^2}{2} < \ln(1 + x) < x$ for $x > 0$;
4. $\sin x > \frac{2}{\pi} x$ for $0 < x < \frac{\pi}{2}$;
5. $\frac{1}{2^{p-1}} \leq x^p + (1 - x)^p \leq 1$ for $0 \leq x \leq 1$, $p > 1$;
6. $\frac{\sqrt{3}}{6 + 2\sqrt{3}} \leq \frac{1 + x}{2 + x^2} \leq \frac{\sqrt{3}}{6 - 2\sqrt{3}}$;
7. $\left(1 + \frac{1}{x}\right)^x < e < \left(1 + \frac{1}{x}\right)^{x+1}$ for $x > 0$.

4.4.3 Prove that if $f(x)$ and $g(x)$ are continuous for $x \geq 0$, $f'(x) > g'(x)$ for $x > 0$ and $f(0) = g(0)$, then $f(x) > g(x)$ for $x > 0$.

4.4.4 Prove that the equation

$$1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} = ae^x$$

has only one positive solution, where $n \geq 1$ is an integer and $0 < a < 1$.

4.4.5 Find the largest value in the sequence $\{\sqrt[n]{n}\}$.

4.4.6 Find the smallest positive value A and the largest negative value B such that

$$Bx^{-1/2} \leq \ln x \leq A\sqrt{x}, \quad x > 0.$$

4.4.7 A quantity x was measured n times, yielding the measurements $x_1, x_2, \dots, x_n$. Find the estimate value $\hat{x}$ of x such that it minimizes the squared error:

$$(\hat{x} - x_1)^2 + (\hat{x} - x_2)^2 + \cdots + (\hat{x} - x_n)^2.$$

4.4.8 As in Fig. 4.14, assume that a L -shaped corridor is a and b in width and c in height (to ceiling). What are the largest possible dimensions of a rectangular whiteboard that can be moved through the corridor? Justify your argument.

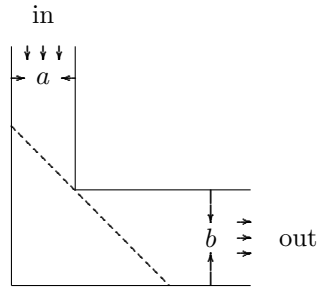


Fig. 4.14 A whiteboard is moved through a corridor.

4.5 CONVEXITY

A differentiable function is **concave upward** or **convex** if the graph of the function always lies above its tangent lines. In other words, we have

$$f(x) \geq f(a) + f'(a)(x - a)$$

for any x and a . The function is **concave downward** or simply **concave** if the graph lies below the tangent lines.

Theorem 4.5.1 *A differentiable function is convex if and only if its derivative is nondecreasing. It is concave if and only if its derivative is nonincreasing.*

By Theorem 4.4.1, we further conclude that if $f(x)$ has second order derivative, then f is convex if and only if $f'' \geq 0$ and f is concave if and only if $f'' \leq 0$.

Proof. Assume $f(x)$ is convex. Then for any $a < b$, we have

$$f(b) \geq f(a) + f'(a)(b - a), \quad f(a) \geq f(b) + f'(b)(a - b).$$

Therefore

$$f'(a) \leq \frac{f(b) - f(a)}{b - a} \leq f'(b).$$

This proves that the derivative is nondecreasing.

Conversely, assume the derivative is nondecreasing. Then for $x > a$, we have

$$f(x) - f(a) = f'(c)(x - a) \geq f'(a)(x - a),$$

where the equality is by the mean value theorem, with $x > c > a$, and the inequality is due to $f'(c) \geq f'(a)$ and $x - a > 0$. Similarly, for $x < a$, we have

$$f(x) - f(a) = f'(c)(x - a) \geq f'(a)(x - a), \quad x < c < a. \quad \blacksquare$$

When the behavior of a function f changes from increasing to decreasing at a , we get a local maximum. If f is differentiable near a , this means f' changes sign from positive to negative at a . Therefore, local extrema can be tested by the change of sign of the first order derivative.

Similarly, a function may be convex on one interval and becomes concave on another interval. The places where the function changes from convex to concave (or vice versa) is called an **inflection point**. If the function has second order derivative, then the inflection points are the places where the second order derivative changes sign.

Example 4.5.1 The function x^p has second order derivative $(x^p)' = p(p-1)x^{p-2}$ for $x > 0$, which has the same sign as $p(p-1)$. Therefore on $(0, +\infty)$, x^p is convex if $p \geq 1$ or $p < 0$, and is concave if $0 < p \leq 1$. ■

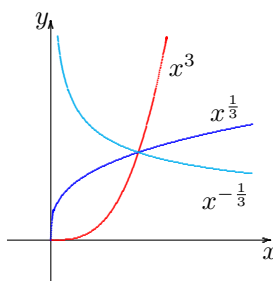


Fig. 4.15 Convexity of x^p .

Example 4.5.2 The function $f(x) = \frac{x-1}{x^2+1}$ has the derivatives

$$f' = \frac{-x^2 + 2x + 1}{(x^2 + 1)^2}, \quad f'' = \frac{2(x+1)(x^2 - 4x + 1)}{(x^2 + 1)^3}.$$

The sign of the second order derivative and the implication on the convexity is illustrated below. ■

interval	sign of f''	convexity
$(-\infty, -1]$	−	⌒
−1	0	inflection
$[-1, 2 - \sqrt{3}]$	+	⌒
$2 - \sqrt{3}$	0	inflection
$[2 - \sqrt{3}, 2 + \sqrt{3}]$	−	⌒
$2 + \sqrt{3}$	0	inflection
$[2 + \sqrt{3}, +\infty)$	+	⌒

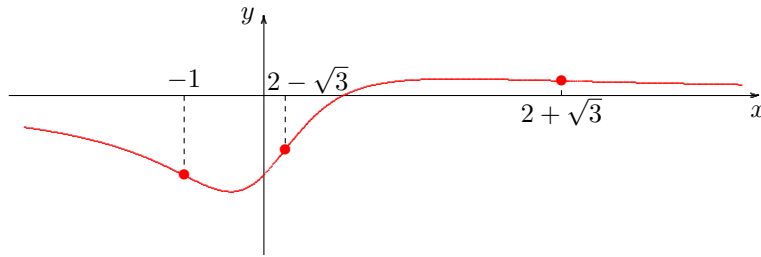


Fig. 4.16 The graph of $\frac{x-1}{x^2+1}$

Convexity can also be characterized by the following property: The straight line segment connecting any two points on the graph lies above the graph. Note that a point between x and y can be expressed as $z = \lambda x + (1 - \lambda)y$ for some $0 \leq \lambda \leq 1$. Moreover, if $L(t) = at + b$ is the straight line connecting $(x, f(x))$ and $(y, f(y))$, then

$$\begin{aligned} L(z) &= a(\lambda x + (1 - \lambda)y) + b = \lambda(ax + b) + (1 - \lambda)(ay + b) \\ &= \lambda L(x) + (1 - \lambda)L(y) = \lambda f(x) + (1 - \lambda)f(y). \end{aligned}$$

Therefore the characterization can be summarized as follows.

Theorem 4.5.2 A differentiable function is convex if and only if

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y), \quad 0 \leq \lambda \leq 1.$$

A differentiable function is concave if and only if

$$f(\lambda x + (1 - \lambda)y) \geq \lambda f(x) + (1 - \lambda)f(y), \quad 0 \leq \lambda \leq 1.$$

Proof. Suppose $f(x)$ is a differentiable convex function. Then for $z = \lambda x + (1 - \lambda)y$ with $0 \leq \lambda \leq 1$, we have

$$f(x) \geq f(z) + f'(z)(x - z), \quad f(y) \geq f(z) + f'(z)(y - z).$$

Therefore

$$\lambda f(x) + (1 - \lambda)f(y) \geq f(z) + f'(z)(\lambda x + (1 - \lambda)y - z) = f(z).$$

Conversely, suppose $f(x)$ is a differentiable and satisfies $f(z) \leq \lambda f(x) + (1 - \lambda)f(y)$ for $z = \lambda x + (1 - \lambda)y$ with $0 \leq \lambda \leq 1$. Then

$$f(y) - f(z) \geq \lambda(f(y) - f(x)), \quad f(x) - f(z) \geq (1 - \lambda)(f(y) - f(x)).$$

If $x < y$ and $0 < \lambda < 1$, then substituting $\lambda = \frac{y - z}{y - x}$ gives us

$$\frac{f(y) - f(z)}{y - z} \geq \frac{f(y) - f(x)}{y - x} \geq \frac{f(x) - f(z)}{x - z}.$$

Let $z \rightarrow y^-$ in the first inequality and $z \rightarrow x^+$ in the second inequality, we get

$$f'(y) = f'_-(y) \geq \frac{f(y) - f(x)}{y - x} \geq f'_+(x) = f'(x).$$

This proves that the derivative is nondecreasing. ■

If we repeatedly make use of the characterization in Theorem 4.5.2, for a convex function f and $0 \leq \lambda, \mu \leq 1$, we have

$$\begin{aligned} & f(\lambda x + (1 - \lambda)\mu y + (1 - \lambda)(1 - \mu)z) = f(\lambda x + (1 - \lambda)(\mu y + (1 - \mu)z)) \\ & \leq \lambda f(x) + (1 - \lambda)(\mu f(y) + (1 - \mu)f(z)) = \lambda f(x) + (1 - \lambda)\mu f(y) + (1 - \lambda)(1 - \mu)f(z). \end{aligned}$$

The inequality may be rephrased as

$$f(\lambda x + \mu y + \nu z) \leq \lambda f(x) + \mu f(y) + \nu f(z), \quad \lambda + \mu + \nu = 1, \quad 0 \leq \lambda, \mu, \nu \leq 1.$$

In general, we have Jensen's inequality for convex functions

$$f(\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_n x_n) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2) + \cdots + \lambda_n f(x_n),$$

where

$$\lambda_1 + \lambda_2 + \cdots + \lambda_n = 1, \quad 0 \leq \lambda_1, \lambda_2, \dots, \lambda_n \leq 1.$$

Jensen's inequality for concave functions reverses the direction of the inequality.

Example 4.5.3 The logarithmic function is concave because $(\ln x)'' = -\frac{1}{x^2} < 0$. Therefore we get

$$\ln(\lambda x + (1 - \lambda)y) \geq \lambda \ln x + (1 - \lambda) \ln y.$$

Taking exponential, we get Young's inequality

$$\lambda x + (1 - \lambda)y \geq x^\lambda y^{1-\lambda}.$$

In the special case $\lambda = \frac{1}{2}$, we have $\frac{x+y}{2} \geq \sqrt{xy}$. ■

Example 4.5.4 For $p > 1$ and $x > 0$, the function x^p is convex because $(x^p)'' = p(p-1)x^{p-2} > 0$. By taking $\lambda_i = \frac{1}{n}$ in Jensen's inequality, we have

$$\left(\frac{x_1 + x_2 + \cdots + x_n}{n} \right)^p \leq \frac{x_1^p + x_2^p + \cdots + x_n^p}{n}.$$

For $p > q > 0$, replace p by $\frac{p}{q}$ and replace x_i by x_i^q . Then we get

$$\left(\frac{x_1^p + x_2^p + \cdots + x_n^p}{n} \right)^{\frac{1}{p}} \leq \left(\frac{x_1^q + x_2^q + \cdots + x_n^q}{n} \right)^{\frac{1}{q}}.$$

In case $n = 2$, we get

$$(x^p + y^p)^{\frac{1}{p}} \leq 2^{\frac{1}{p} - \frac{1}{q}} (x^q + y^q)^{\frac{1}{q}}.$$

This implies the inequality in Example 4.4.4. ■

Curve Sketching

One of useful applications of differential calculus is curve sketching. For a given function f , elementary mathematics learnt in high schools can help us to determine (a) the domain of f ; (b) whether f is an even, or odd, or periodic function; (c) the locations of intercepts in x and y axis¹. Together with the knowledge of (d) the asymptotes; (e) the extrema; (f) the monotonicity; (g) the convexity and the inflection points enables us to draw fairly accurate graphs of functions.

Example 4.5.5 The function $f(x) = \frac{x}{x^2 + 1}$ is defined on the whole real line. Its intercepts on both axes are 0. The function is odd: $f(-x) = -f(x)$, which means that the graph is symmetric with respect to the origin. Moreover, the limit

$$\lim_{x \rightarrow \infty} \frac{x}{x^2 + 1} = 0$$

tells us that the x -axis is the horizontal asymptote in both infinity directions.

From the derivative

$$f'(x) = -\frac{(x+1)(x-1)}{(x^2+1)^2},$$

we find the monotonicity and the extrema.

interval	sign of f'	monotonicity
$(-\infty, -1]$	—	$\searrow$
-1	0	local min
$[-1, 1]$	+	$\nearrow$
1	0	local max
$[1, +\infty)$	—	$\searrow$

¹Newton's Method can be also used to find the intercepts approxiamtely.

The extreme values are $f(-1) = -\frac{1}{2}$ and $f(1) = \frac{1}{2}$. From the second derivative

$$f''(x) = \frac{2x(x^2 - 3)}{(x^2 + 1)^3},$$

we find the convexity and the inflection.

interval	sign of f''	convexity
$(-\infty, -\sqrt{3}]$	—	$\frown$
$-\sqrt{3}$	0	inflection
$[-\sqrt{3}, 0]$	+	$\smile$
0	0	inflection
$[0, \sqrt{3}]$	—	$\frown$
$\sqrt{3}$	0	inflection
$[\sqrt{3}, +\infty)$	+	$\smile$

The values at the inflections are $f(-\sqrt{3}) = -\frac{\sqrt{3}}{4}$ and $f(\sqrt{3}) = \frac{\sqrt{3}}{4}$.

The function $y = f(x)$ is sketched as in Fig. 4.17. ■

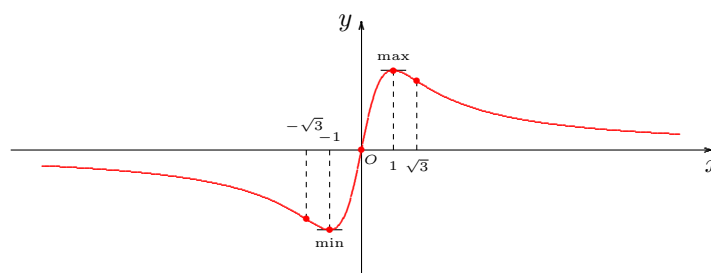


Fig. 4.17 The graph of $\frac{x}{x^2 + 1}$

Example 4.5.6 The function $f(x) = \frac{x^3}{x^2 - 1}$ is defined everywhere except ± 1 . Its intercepts on both axis are 0. The function is odd. The limits

$$\lim_{x \rightarrow -1^-} \frac{x^3}{x^2 - 1} = \lim_{x \rightarrow 1^-} \frac{x^3}{x^2 - 1} = -\infty, \quad \lim_{x \rightarrow -1^+} \frac{x^3}{x^2 - 1} = \lim_{x \rightarrow 1^+} \frac{x^3}{x^2 - 1} = +\infty$$

show that $x = \pm 1$ are vertical asymptotes. Moreover, the limit

$$\lim_{x \rightarrow \infty} \left(\frac{x^3}{x^2 - 1} - x \right) = \lim_{x \rightarrow \infty} \frac{x}{x^2 - 1} = 0$$

shows that $y = x$ is a slant asymptote along both infinity directions. We also see that the graph of f is above the asymptote for $x > 0$ and below the asymptote for $x < 0$. The left graph in Fig. 4.18 shows an initial sketch based on the study of the asymptotes.

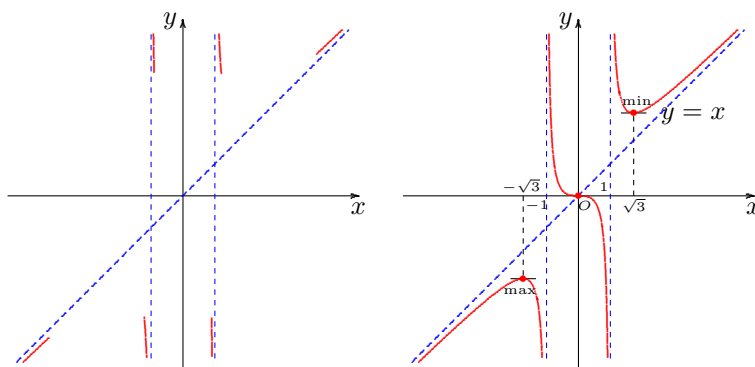


Fig. 4.18 The graph of $\frac{x^3}{x^2 - 1}$.

The derivatives

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

tell us the monotonicity

interval	sign of f'	monotonicity
$(-\infty, -\sqrt{3}]$	+	$\nearrow$
$[\sqrt{3}, -1)$	-	$\searrow$
$(-1, 0]$	-	$\searrow$
$[0, 1)$	-	$\searrow$
$(1, \sqrt{3}]$	-	$\searrow$
$[\sqrt{3}, +\infty)$	+	$\nearrow$

and convexity.

interval	sign of f''	convexity
$(-\infty, -1)$	-	$\cap$
$(-1, 0]$	+	$\cup$
$[0, 1)$	-	$\cap$
$(1, +\infty)$	+	$\cup$

The function has one local maximum $f(-\sqrt{3}) = -\frac{3\sqrt{3}}{2}$, one local minimum $f(\sqrt{3}) = \frac{3\sqrt{3}}{2}$, and one inflection $f(0) = 0$.

With the all the above information, we sketch the function $y = f(x)$ as the right graph in Fig. 4.18. ■

Exercises

4.5.1 Study the convexity and sketch the graph of the functions in Exercise 4.4.1.

4.5.2 Study the convexity:

- | | |
|---------------------|---------------------|
| 1. $x^2 + ax + b$; | 3. $(x^2 + 1)e^x$; |
| 2. $x^3 + ax + b$; | 4. $\ln(x^2 + 1)$. |

4.5.3 The generalizes of the inequality in Example 4.4.4.

1. Show that $\ln x$ is concave.
2. For positive $x_1, x_2, \dots, x_n$, prove that

$$\frac{x_1 \ln x_1 + x_2 \ln x_2 + \dots + x_n \ln x_n}{x_1 + x_2 + \dots + x_n} \leq \ln \frac{x_1^2 + x_2^2 + \dots + x_n^2}{x_1 + x_2 + \dots + x_n} \leq \ln(x_1 + x_2 + \dots + x_n).$$

3. For positive $x_1, x_2, \dots, x_n$, and $p > q > 0$, prove that

$$(x_1^p + x_2^p + \dots + x_n^p)^{\frac{1}{p}} \leq (x_1^q + x_2^q + \dots + x_n^q)^{\frac{1}{q}}.$$

4.5.4 Sketch the following functions:

- | | |
|--|--|
| 1. $y = x^3 - 3x + 2$; | 9. $y = x - \ln(1 + x), \quad x > -1$; |
| 2. $y = -x^4 + 2x^2 - 1$; | 10. $y = \frac{\sin x}{e^x}, \quad x > 0$; |
| 3. $y = \frac{x}{x^2 + 1}$; | 11. $y = x - \sin x, \quad 0 \leq x \leq 2\pi$; |
| 4. $y = x + \frac{1}{x}, \quad x \neq 0$; | (k) $y = x^2 - 3x + 2 , \quad 0 \leq x \leq 4$; |
| 5. $y = \frac{x^3}{(x-1)^2}, \quad x \neq 0$; | 12. $y = x - \sin x , \quad -\pi/2 \leq x \leq \pi/2$; |
| 6. $y = \sqrt{3 + 2x - x^2}, \quad -1 \leq x \leq 3$; | 13. $y = \frac{1}{1 + \sin^2 x}, \quad 0 \leq x \leq 2\pi$; |
| 7. $y = x^2 e^{-x}$; | 14. $y = x e^{- x-1 }, \quad -2 \leq x \leq 2$. |
| 8. $y = x \ln x, \quad x > 0$; | |

4.6 SUMMARY

Definitions

- A **maximum** of a function f is a point a such that $f(x) \leq f(a)$ for all x in the domain of f . The point a is a **local maximum** if $f(x) \leq f(a)$ for all x near a in the domain of f . The

concepts of **minimum** and **local minimum** are defined similarly. Maxima and minima are also called **extrema**.

- A differentiable function is **convex** if its graph is always above its tangent lines. In other words, we have $f(x) \geq f(a) + f'(a)(x - a)$ for any x and a . The concept of **concave** is defined similarly.
- An **inflection point** is the place where a function changes from convex to concave or vice versa.
- In sketching the curve of a function, it is often helpful to study the following aspects: Domain, intercepts, symmetry (parity, periodicity, etc.), asymptote, monotonicity, local extrema (and the value at the local extrema), convexity, inflection points (and the value at the inflection), etc.

Theorems

- Any continuous function on a bounded closed interval has global maximum and global minimum.
- **Mean Value Theorem** If f is continuous on $[a, b]$ and differentiable on (a, b) , then there is $c \in (a, b)$, such that $\frac{f(b) - f(a)}{b - a} = f'(c)$.
 - If $f'(x) = 0$ for all $x \in (a, b)$, then $f(x)$ is a constant on (a, b) .
 - 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) .
- **Cauchy's Mean Value Theorem** If f and g are continuous on $[a, b]$ and differentiable on (a, b) , such that $g' \neq 0$ on (a, b) , then there is $c \in (a, b)$, such that $\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(c)}{g'(c)}$.
- **L'Hospital's Rule** Suppose $\lim_{x \rightarrow a} f(x) = \lim_{x \rightarrow a} g(x) = 0$ or $\lim_{x \rightarrow a} f(x) = \lim_{x \rightarrow a} g(x) = \infty$. Then $\lim_{x \rightarrow a} \frac{f'(x)}{g'(x)} = L$ implies $\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = L$. In L'Hospital's rule, a can be a^+ , a^- or any kind of ∞ , and L can be finite or infinite.
- **Candidates for Local Extrema:** If f is differentiable at a local extreme a , then $f'(a) = 0$. Therefore the local extrema a of a function on an interval are of three types: (1) a is an end point; (2) a is an interior point, f is not differentiable at a ; (3) a is an interior point, $f'(a) = 0$.
- **Criterion for Monotonicity:** If $f' \geq 0$ on an interval, then f is nondecreasing in the interval. If $f' > 0$ on an interval, then f is increasing on the interval. Similar criterion works for the decreasing functions.
- **First Derivative Test** for Local Extrema: Suppose that $f(x)$ is continuous at a and differentiable near a (but not necessarily differentiable at a).

- If $f' \geq 0$ on the left of a and $f' \leq 0$ on the right of a , then a is a local maximum.
- If $f' \leq 0$ on the left of a and $f' \geq 0$ on the right of a , then a is a local minimum.
- If f' does not change sign at a (i.e., $f' > 0$ on both sides of a , or $f' < 0$ on both sides of a), then a is not a local extreme.

• **Second Derivative Test** for Local Extrema:

- If $f'(a) = 0$ and $f''(a) < 0$, then a is a local maximum.
- If $f'(a) = 0$ and $f''(a) > 0$, then a is a local minimum.
- **Criterion for Convexity:** A differentiable function f is convex on an interval if and only if f' is nondecreasing, or $f'' \geq 0$ in case f has second order derivative. Similar criterion applies to concave functions.
- **Jensen's Inequality:** A function is convex if and only if $f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$ for $0 \leq \lambda \leq 1$.

5

Integration

5.1 RIEMANN INTEGRATION

Let $f(x)$ be a function on a closed bounded interval $[a, b]$. If $f(x)$ is non-negative, then the graph of $f(x)$, the x -axis and the two vertical lines $x = a$ and $x = b$ encloses a region on the plane. See Fig. 5.1.

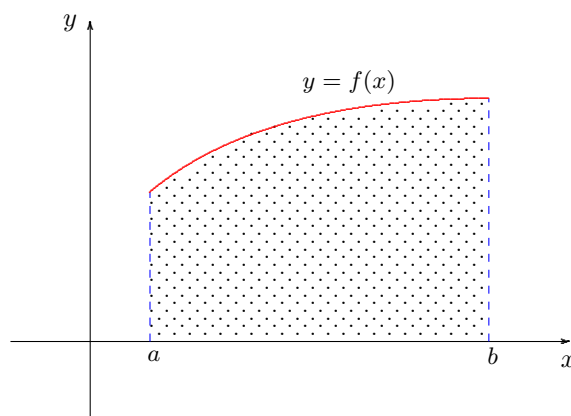


Fig. 5.1 The area under the function $f(x)$

To find the area of the region, we approximate the area by a series of rectangles. First we divide the interval into smaller intervals, called a **partition**

$$P: a = x_0 < x_1 < x_2 < \cdots < x_n = b.$$

Next we choose one sample point x_i^* in each interval $[x_{i-1}, x_i]$. Then the rectangles $[x_{i-1}, x_i] \times [0, f(x_i^*)]$ together, indicated by the shaded region in Fig. 5.2, form a good approximation of the region below the graph of f . The total area of the rectangles is

$$S(P, f) = \sum_{i=1}^n f(x_i^*)(x_i - x_{i-1}) = \sum_{i=1}^n f(x_i^*) \Delta x_i, \quad \Delta x_i = x_i - x_{i-1},$$

called the **Riemann sum** of the function $f(x)$ with respect to the partition P and sample points x_i^* . Note that the sample points do not explicitly appear in the notation $S(P, f)$ for the Riemann sum.

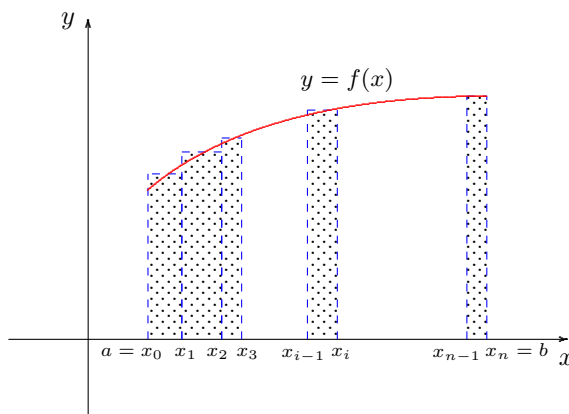


Fig. 5.2 The area under $f(x)$ is approximated by a sum of rectangular areas

The size of the partition P is measured by

$$\|P\| = \max\{\Delta x_1, \Delta x_2, \dots, \Delta x_n\}.$$

We expect that as the size of P gets smaller and smaller, the rectangles to get closer and closer to the region below f , and the Riemann sum to get closer and closer to the area of the region below f . The intuition can be illustrated by the equality

$$\lim_{\|P\| \rightarrow 0} S(P, f) = \text{area under the graph of } f.$$

Example 5.1.1 Consider a constant function $f(x) = h$. For any partition P and any choice of the sample points x_i^* , we always have

$$S(P, f) = \sum h \Delta x_i = h \sum \Delta x_i = h(b - a).$$

Therefore as the size of the partition approaches 0, the limit of the Riemann sum is $h(b - a)$. Indeed, the region below the graph of the constant function is a rectangle of base $b - a$ and height h , and has $h(b - a)$. ■

Example 5.1.2 The function

$$\delta_c(x) = \begin{cases} 0, & \text{if } x \neq c, \\ 1, & \text{if } x = c \end{cases}$$

is 0 everywhere except at c . The Riemann sum ($c = x_j$ in the third case)

$$S(P, \delta_c) = \begin{cases} 0, & \text{if all } x_i^* \neq c, \\ \Delta x_j, & \text{if one } x_j^* = c, \\ \Delta x_j + \Delta x_{j+1}, & \text{if two } x_j^* = x_{j+1}^* = c. \end{cases}$$

Therefore $|S(P, \delta_c)| \leq 2\|P\|$, and the limit of the Riemann sum is 0. Indeed the region below the graph is one vertical line, which has area 0. ■

Example 5.1.3 Consider the identity function $f(x) = x$ on $[0, 1]$. For any partition P and the choice of the middle sample points $x_i^* = \frac{x_i + x_{i-1}}{2}$, we have

$$S(P, f) = \sum x_i^* \Delta x_i = \frac{1}{2} \sum (x_i + x_{i-1})(x_i - x_{i-1}) = \frac{1}{2} \sum (x_i^2 - x_{i-1}^2) = \frac{1}{2} (x_n^2 - x_0^2) = \frac{1}{2}.$$

The limit as $\|P\| \rightarrow 0$ is $\frac{1}{2}$, which is indeed the area of the region below f (the rectangle connecting three points $(0, 0)$, $(1, 0)$, $(1, 1)$ on the plane). ■

Example 5.1.4 To compute the area below the parabola $f(x) = x^2$ over the interval $[0, 1]$, we divide the interval evenly into n parts. In other words, we have $x_i = \frac{i}{n}$. Moreover, we take the sample points $x_i^* = \frac{i}{n}$ to be the right ends of the intervals in the partition. The corresponding Riemann sum is

$$\sum_{i=1}^n \left(\frac{i}{n}\right)^2 \frac{1}{n} = \frac{1}{n^3} (1^2 + 2^2 + \cdots + n^2) = \frac{1}{6n^3} n(n+1)(2n+1) = \frac{(n+1)(2n+1)}{6n^2}.$$

As $n \rightarrow \infty$, the size of the partition approaches 0, and the Riemann sum approaches $\frac{1}{3}$. Thus the area below the parabola should be $\frac{1}{3}$. ■

Rigorous Definition

The examples suggests the following general definition.

Definition 5.1.1 A function $f(x)$ has **Riemann integral** I on a bounded and closed interval $[a, b]$, if for any $\epsilon > 0$, there is $\delta > 0$, such that

$$\|P\| < \delta \implies |S(P, f) - I| < \epsilon.$$

The Riemann integral is denoted

$$I = \int_a^b f(x) \, dx.$$

Examples 5.1.1 and 5.1.2 tell us

$$\int_a^b h \, dx = h(b-a), \quad \int_a^b \delta_c(x) \, dx = 0$$

in the rigorous way. Examples 5.1.3 and 5.1.4 suggest that

$$\int_0^1 x \, dx = \frac{1}{2}, \quad \int_0^1 x^2 \, dx = \frac{1}{3}.$$

However, the argument was not rigorous because not all possible partitions and sample points have been considered. The following is a rigorous argument for the integral of $f(x) = x$ on $[0, 1]$.

Example 5.1.5 In Example 5.1.3, we already know the Riemann sum $S_m(P, f) = \frac{1}{2}$ in case the sample points are the middle points $x_{m,i} = \frac{x_{i-1} + x_i}{2}$ (the subscript m means “middle”). Now for arbitrary choice of the sample points x_i^* , we have $|x_i^* - x_{m,i}| \leq \frac{1}{2}\Delta x_i$, so that

$$\begin{aligned} \left| S(P, f) - \frac{1}{2} \right| &= |S(P, f) - S_m(P, f)| = \left| \sum x_i^* \Delta x_i - \sum x_{m,i} \Delta x_i \right| \\ &\leq \sum |x_i^* - x_{m,i}| \Delta x_i \leq \frac{1}{2} \sum \Delta x_i^2 \leq \frac{1}{2} \|P\| \sum \Delta x_i = \frac{1}{2} \|P\|. \end{aligned}$$

This implies that as $\|P\|$ approaches 0, the general Riemann sum $S(P, f)$ approaches $\frac{1}{2}$. ■

Integrability

Examples 5.1.3 and 5.1.5 illustrate the rigorous argument for Riemann integral is by no means trivial. After all, the Riemann integral is defined as a limit, and we know that many limits diverge.

If there is a number I that fits into the definition of the Riemann integral $\int_a^b f(x) \, dx$, then we say $f(x)$ is **Riemann integrable** on $[a, b]$. Otherwise, we say $f(x)$ is not integrable. We already see that the constant function $f(x) = h$ and the identity function $f(x) = x$ are integrable.

Example 5.1.6 Consider the function

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

For any partition P of $[0, 1]$, if $x_1^* \neq 0$, then the Riemann sum

$$S(P, f) = \sum_{i=1}^n \frac{1}{x_i^*} \Delta x_i = \frac{1}{x_1^*} \Delta x_1 + \sum_{i=2}^n \frac{1}{x_i^*} \Delta x_i.$$

Since we allow x_1^* to be any point in $(x_0, x_1] = (0, x_1]$, the first term can get arbitrarily large. In particular, the Riemann sum will not converge as $\|P\| \rightarrow 0$. Therefore the function is not integrable on $[0, 1]$. ■

The example can be easily generalized.

Proposition 5.1.2 *Riemann integrable functions are bounded.*

However, the following example shows that the bounded property does not guarantee integrability.

Example 5.1.7 The Dirichlet function in Example 2.2.26 is bounded. For any partition P of $[a, b]$, if we choose the sample points x_i^* to be rational, then we get $S(P, D) = \sum 1 \cdot \Delta x_i = b - a$. If we choose the sample points x_i^* to be irrational, then we get $S(P, D) = \sum 0 \cdot \Delta x_i = 0$. Therefore the function is not integrable. ■

The Dirichlet function is not integrable because it is an extremely bad function. Reasonably good functions are expected to be integrable.

Theorem 5.1.3 *Continuous functions are integrable.*

For example, the functions x and x^2 are integrable on $[0, 1]$. The computation of their integrals on $[0, 1]$ in Examples 5.1.3 and 5.1.4 are justified.

The most general criterion for integrability can be obtained by considering the Cauchy criterion for the convergence of the Riemann sum: For any $\epsilon > 0$, there is $\delta > 0$, such that

$$\|P\| < \delta, \|Q\| < \delta \implies |S(P, f) - S(Q, f)| < \epsilon.$$

At least it is easy to see that the criterion is necessary, by a proof similar to the proof of Cauchy criterion for sequence limit (Theorem 2.1.12). Specializing the Cauchy criterion above to the case $P = Q$ and making the selection of sample points (which may be different for P and Q) explicit, we get the necessary condition: For any $\epsilon > 0$, there is $\delta > 0$, such that

$$\|P\| < \delta \implies \left| \sum (f(x_i^*) - f(x_i^{**})) \Delta x_i \right| < \epsilon.$$

By suitably choosing the sample points, the difference $f(x_i^*) - f(x_i^{**})$ can be as large as the **oscillation**

$$\omega_{[x_{i-1}, x_i]}(f) = \sup_{[x_{i-1}, x_i]}(f) - \inf_{[x_{i-1}, x_i]}(f) = \sup\{|f(x) - f(x')| : x, x' \in [x_{i-1}, x_i]\}$$

of the function on the interval $[x_{i-1}, x_i]$. Therefore the condition becomes the following: For any $\epsilon > 0$, there is $\delta > 0$, such that

$$\|P\| < \delta \implies \sum \omega_{[x_{i-1}, x_i]}(f) \Delta x_i < \epsilon.$$

This proves the necessary part of the following criterion.

Theorem 5.1.4 *The following are equivalent for a bounded function $f(x)$ on a bounded closed interval $[a, b]$.*

1. $f(x)$ is Riemann integrable on $[a, b]$.
2. For any $\epsilon > 0$, there is $\delta > 0$, such that $\|P\| < \delta$ implies $\sum \omega_{[x_{i-1}, x_i]}(f) \Delta x_i < \epsilon$.
3. For any $\epsilon > 0$, there is a partition P satisfying $\sum \omega_{[x_{i-1}, x_i]}(f) \Delta x_i < \epsilon$.

The full proof of the theorem is rather complicated and will not be given. Theorem 5.1.3 follows from this theorem and the fact that any continuous function on bounded and closed interval must be *uniformly* continuous.

Example 5.1.8 We show that the function

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

is integrable on $[0, 1]$.

For any $\epsilon > 0$, $f(x)$ is continuous on $[\epsilon, 1]$. By Theorem 5.1.3, it is integrable on $[\epsilon, 1]$. Therefore there is a partition P' of $[\epsilon, 1]$, such that $\sum_{P'} \omega_{[x_{i-1}, x_i]}(f) \Delta x_i < \epsilon$. Now we add 0 to P' to form a partition P of $[0, 1]$. By $|f(x)| \leq 1$, on the first interval $[0, \epsilon]$ of the partition, we have

$$\omega_{[0, \epsilon]}(f) \leq 2.$$

Therefore

$$\sum_P \omega_{[x_{i-1}, x_i]}(f) \Delta x_i = \omega_{[0, \epsilon]}(f) \epsilon + \sum_{P'} \omega_{[x_{i-1}, x_i]}(f) \Delta x_i < 2\epsilon + \epsilon = 3\epsilon.$$

This verifies the last criterion in Theorem 5.1.4.

In general, if $f(x)$ is bounded on $[a, b]$ and is integrable on $[a + \epsilon, b]$ for any $\epsilon > 0$, then $f(x)$ is integrable on $[a, b]$. ■

Example 5.1.9 By

$$||f(x)| - |f(x')|| \leq |f(x) - f(x')|,$$

we have $\omega_{[x_{i-1}, x_i]}(|f|) \leq \omega_{[x_{i-1}, x_i]}(f)$. Therefore

$$\sum \omega_{[x_{i-1}, x_i]}(|f|) \Delta x_i \leq \sum \omega_{[x_{i-1}, x_i]}(f) \Delta x_i.$$

Then by Theorem 5.1.4, we conclude that the integrability of $f(x)$ implies the integrability of $|f(x)|$. ■

Example 5.1.10 We show that if $f(x)$ is integrable, then $[f(x)]^2$ is also integrable.

By Theorem 5.1.2, we have $|f(x)| < B$ for a constant B . Then by

$$|[f(x)]^2 - [f(x')]^2| = |f(x) + f(x')| \cdot |f(x) - f(x')| \leq 2B \cdot |f(x) - f(x')|,$$

we have

$$\sum \omega_{[x_{i-1}, x_i]}(f^2) \Delta x_i \leq \sum 2B \omega_{[x_{i-1}, x_i]}(f) \Delta x_i.$$

Then by Theorem 5.1.4, we conclude that the integrability of $f(x)$ implies the integrability of $[f(x)]^2$. ■

The example above can be extended to the integrability of the functions such as $[f(x)]^3$ or $\sin f(x)$. In general, the argument applies to the composition $g(f(x))$, with $g(y)$ satisfying the **Lipschitz condition**

$$|g(y) - g(y')| \leq C|y - y'|,$$

where C is a constant independent of y and y' . We note that by the Mean Value Theorem, a differentiable function with bounded derivative satisfies the Lipschitz condition. In particular, $\sin x$, $\cos x$, $\frac{1}{1+x^2}$ are Lipschitz on the whole real line, and e^x , x^2 are Lipschitz on any bounded interval.

Exercises

5.1.1 Use the definition to compute the integral.

1. $\int_0^a (2x - 3) \, dx;$
2. $\int_{-1}^2 x^2 \, dx;$
3. $\int_0^1 a^x \, dx;$
4. $\int_a^b \frac{dx}{x^2}$ (Hint: take $x_i^* = \sqrt{x_{i-1}x_i}$).

5.1.2 Explain the identity.

1. $\lim_{n \rightarrow \infty} \left(\frac{1}{n^2} + \frac{2}{n^2} + \cdots + \frac{n-1}{n^2} \right) = \int_0^1 x \, dx$
2. $\lim_{n \rightarrow \infty} \left(\frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{n+n} \right) = \int_0^1 \frac{dx}{1+x}.$
3. $\lim_{n \rightarrow \infty} \left(\frac{n}{n^2+1^2} + \frac{n}{n^2+2^2} + \cdots + \frac{n}{n^2+n^2} \right) = \int_0^1 \frac{dx}{1+x^2}.$
4. $\lim_{n \rightarrow \infty} \frac{1}{n} \left(\sin \frac{\pi}{n} + \sin \frac{2\pi}{n} + \cdots + \sin \frac{(n-1)\pi}{n} \right) = \int_0^1 \sin x \, dx.$

5.1.3 Express the limit as integration.

1. $\lim_{n \rightarrow \infty} \frac{1^p + 2^p + \cdots + n^p}{n^{p+1}}$
2. $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n f \left(a + k \frac{b-a}{n} \right)$

5.1.4 Show that the integrability of $|f(x)|$ does not imply the integrability of $f(x)$.

5.2 PROPERTIES OF INTEGRATION

For the same partition P and the same choice of the sample points x_i^* , the Riemann sum satisfies

$$S(P, f + g) = \sum (f(x_i^*) + g(x_i^*)) \Delta x_i = \sum f(x_i^*) \Delta x_i + \sum g(x_i^*) \Delta x_i = S(P, f) + S(P, g).$$

We also have $S(P, cf) = cS(P, f)$. Similar to the arithmetic rule for the limit of sequences and functions, we then get the arithmetic rule for the integration.

Proposition 5.2.1 *If $f(x)$ and $g(x)$ are integrable on $[a, b]$, then $f(x) + g(x)$ is integrable on $[a, b]$, and*

$$\int_a^b (f(x) + g(x)) \, dx = \int_a^b f(x) \, dx + \int_a^b g(x) \, dx, \quad \int_a^b cf(x) \, dx = c \int_a^b f(x) \, dx.$$

For $f \geq g$, the Riemann sum also satisfies

$$S(P, f) = \sum f(x_i^*) \Delta x_i \geq \sum g(x_i^*) \Delta x_i = S(P, g).$$

Similar to the limit of sequences and functions, we get the order rule for the integration.

Proposition 5.2.2 *If f and g are integrable on $[a, b]$ and $f(x) \geq g(x)$, then*

$$\int_a^b f(x) \, dx \geq \int_a^b g(x) \, dx.$$

Applying the order rule to $-|f(x)| \leq f(x) \leq |f(x)|$ (see Example 5.1.9 for the integrability of $|f(x)|$), we get

$$\left| \int_a^b f(x) \, dx \right| \leq \int_a^b |f(x)| \, dx.$$

If $m \leq f(x) \leq M$, then by Example 5.1.1 and the order rule, we get

$$m(b - a) \leq \int_a^b f(x) \, dx \leq M(b - a).$$

Suppose an interval $[a, b]$ is divided into two intervals $[a, c]$ and $[c, b]$ by $a < c < b$. Then a partition P' of $[a, c]$ and a partition P'' on $[c, b]$ can be combined to form a partition P of $[a, b]$. Moreover, for any function f on $[a, b]$, we have

$$S(P, f) = S(P', f) + S(P'', f).$$

In other words, the Riemann sum of the function on $[a, b]$ is the sum of the Riemann sums of the function on $[a, c]$ and on $[c, b]$. This leads to the following result. The full proof contains some subtleties and will not be given here.

Proposition 5.2.3 *If $a < c < b$, then a function f is integrable on $[a, b]$ if and only if it is integrable on $[a, c]$ and $[c, b]$. Moreover,*

$$\int_a^b f(x) \, dx = \int_a^c f(x) \, dx + \int_c^b f(x) \, dx.$$

Intuitively, the proposition means that for a non-negative function, the region below the graph on $[a, b]$ is the union of the region below the graph on $[a, c]$ and the region below the graph on $[c, b]$. Correspondingly, the area of the region over $[a, b]$ is the sum of the area over $[a, c]$ and the area over $[c, b]$.

By combining Propositions 5.2.1 and 5.2.3, we see that in general, the integral $\int_a^b f(x) dx$ is the area between the non-negative part of f and the x -axis subtracting the area between the negative part of f and the x -axis. For example, in the situation depicted in Fig. 5.3, we have

$$\int_a^b f(x) dx = A_1 - A_2 + A_3.$$

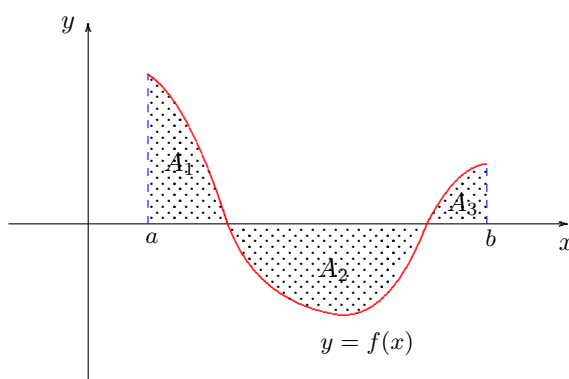


Fig. 5.3 Integration is the difference of areas

Proposition 5.2.3 suggests us to define

$$\int_a^a f(x) dx = 0, \quad \int_a^b f(x) dx = - \int_b^a f(x) dx$$

in case $a < b$ does not hold. Then Propositions 5.2.1 and 5.2.3 still hold without specifying the order among the limits of integration.

Example 5.2.1 Suppose f and g are integrable. By Proposition 5.2.1, the sum $f + g$ is also integrable. By Example 5.1.10, the function $(f + g)^2$ is again integrable. Then by Proposition 5.2.1, since

$$f(x) \cdot g(x) = \frac{1}{2} \{ [f(x) + g(x)]^2 - [f(x)]^2 - [g(x)]^2 \},$$

the product $f \cdot g$ is integrable. ■

Example 5.2.2 Suppose f is bounded and is continuous everywhere on $[a, b]$ except at c . By the argument in Example 5.1.8, f is integrable on $[a, c]$ and $[c, b]$. Then Proposition 5.2.3 further implies that f is integrable on $[a, b]$.

In general, any bounded function that is continuous everywhere except at finitely many places is integrable. Such functions are depicted in Fig. 5.4. ■

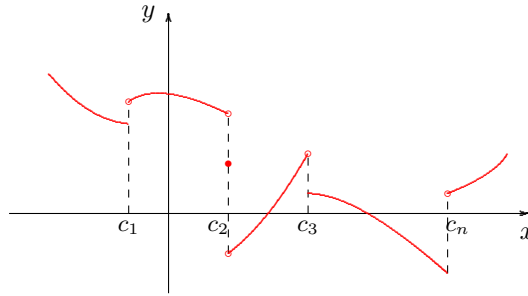


Fig. 5.4 Continuous everywhere except at finitely many places

Example 5.2.3 If two functions $f(x)$ and $g(x)$ are equal everywhere except at c , then

$$f(x) = g(x) + a\delta_c(x), \quad a = f(c) - g(c),$$

where the function $\delta_c(x)$ is defined in Example 5.1.2. Since $\delta_c(x)$ is integrable, by Theorem 5.2.1, we know that $f(x)$ is integrable if and only if $g(x)$ is integrable. Moreover, since $\int_a^b \delta_c(x) dx = 0$, we also have

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

We just proved that the integrability and the value of the integral are not changed if a function is modified at finitely many places. ■

Example 5.2.4 (Integral Mean Value Theorem) Suppose f is continuous on $[a, b]$. Then it is integrable (by Theorem 5.1.3) and reaches its maximum M and minimum m (by Theorem 4.1.1)

$$f(c_1) = m, \quad f(c_2) = M.$$

By $m \leq f(x) \leq M$, Example 5.1.1 and Proposition 5.2.2, we get

$$m(b-a) \leq \int_a^b f(x) dx \leq M(b-a).$$

This is the same as

$$f(c_1) = m \leq \frac{1}{b-a} \int_a^b f(x) dx \leq M = f(c_2).$$

By applying the Intermediate Value Theorem to the continuous function f on $[c_1, c_2]$, we have

$$\frac{1}{b-a} \int_a^b f(x) dx = f(c)$$

for some $c \in [c_1, c_2] \subset [a, b]$. Thus we conclude

$$\int_a^b f(x) \, dx = f(c)(b - a). \quad \blacksquare$$

Exercises

5.2.1 Find the relation between the Riemann sums of $f(x)$ on $[a, b]$ and $f(-x)$ on $[-b, -a]$. Then prove that $\int_{-b}^{-a} f(-x) \, dx = \int_a^b f(x) \, dx$.

5.2.2 Prove that if $f(x)$ is an even continuous function $f(-x) = f(x)$, then $\int_{-a}^a f(x) \, dx = 2 \int_0^a f(x) \, dx$. Prove that if $f(x)$ is an odd continuous function $f(-x) = -f(x)$, then $\int_{-a}^a f(x) \, dx = 0$. In fact, by using Exercise 5.2.1, $f(x)$ only needs to be integrable instead of continuous.

5.2.3 Find the relation between the Riemann sums of $f(x+T)$ on $[a, b]$ and $f(x)$ on $[a+T, b+T]$. Then prove that $\int_a^b f(x+T) \, dx = \int_{a+T}^{b+T} f(x) \, dx$.

5.2.4 Prove that if $f(x)$ is a continuous periodic function $f(x+T) = f(x)$, then

$$\lim_{b \rightarrow \infty} \frac{1}{b} \int_0^b f(x) \, dx = \frac{1}{T} \int_0^T f(x) \, dx.$$

In fact, by using Exercise 5.2.3, $f(x)$ only needs to be integrable instead of continuous.

5.2.5 Suppose $f(x)$ is continuous and non-negative on $[a, b]$. Prove that if $f(x)$ is not identically zero, then $\int_a^b f(x) \, dx > 0$.

5.2.6

5.3 FUNDAMENTAL THEOREM OF CALCULUS

Theorem 5.3.1 (Fundamental Theorem of Calculus) *If f is integrable, then*

$$F(x) = \int_a^x f(t) \, dt$$

is continuous. Moreover, if f is continuous at c , then

$$F'(c) = f(c).$$

Proof. If $f(x)$ is integrable, then by Theorem 5.1.2, we have $|f(x)| < B$ for some constant B . By Propositions 5.2.2 and 5.2.3, this implies

$$|F(x) - F(c)| = \left| \int_a^x f(t) dt - \int_a^c f(t) dt \right| = \left| \int_c^x f(t) dt \right| \leq \left| \int_c^x B dt \right| = B|x - c|.$$

This further implies $\lim_{x \rightarrow c} F(x) = F(c)$.

Now we further assume $f(x)$ is continuous at c . For any $\epsilon > 0$, there is $\delta > 0$, such that $|f(x) - f(c)| < \epsilon$ for $|x - c| < \delta$. By Propositions 5.2.1, 5.2.2 and 5.2.3, for $|x - c| < \delta$,

$$\begin{aligned} |F(x) - F(c) - f(c)(x - c)| &= \left| \int_a^x f(t) dt - \int_a^c f(t) dt - f(c)(x - c) \right| \\ &= \left| \int_c^x f(t) dt - \int_c^x f(c) dt \right| = \left| \int_c^x (f(t) - f(c)) dt \right| \\ &\leq \left| \int_c^x \epsilon dt \right| = \epsilon|x - c|. \end{aligned}$$

This means

$$\lim_{x \rightarrow c} \frac{F(x) - F(c)}{x - c} = f(c). \quad \blacksquare$$

A function $F(x)$ is an **antiderivative** of a (continuous) function $f(x)$ if

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

Theorem 5.3.1 shows that

$$G(x) = \int_a^x f(t) dt$$

is one such antiderivative. Any two anti-derivatives $F(x)$ and $G(x)$ satisfy

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

By Corollary 4.2.2, we find $F(x) - G(x)$ is a constant. Therefore any two antiderivatives differ by a constant. In particular, any antiderivative of a continuous function $f(x)$ is of the form

$$F(x) = \int_a^x f(t) dt + C$$

for some constant $C = F(a)$. By substituting $x = a$ and $x = b$ into the last formula, we obtain the so-called **Newton-Leibnitz formula**:

$$\int_a^b f(t) dt = F(b) - F(a).$$

In other words, we can use anti-derivative to evaluate integration.

Example 5.3.1 By

$$(hx)' = h, \quad \left(\frac{x^2}{2}\right)' = x, \quad \left(\frac{x^3}{3}\right)' = x^2,$$

we get antiderivatives hx , $\frac{x^2}{2}$ and $\frac{x^3}{3}$ for h , x and x^2 . Then by Newton-Leibnitz formula, we get the integrals in Examples 5.1.1, 5.1.3 and 5.1.4

$$\int_a^b h \, dx = hb - ha = h(b-a), \quad \int_0^1 x \, dx = \frac{1^2}{2} - \frac{0^2}{2} = \frac{1}{2}, \quad \int_0^1 x^2 \, dx = \frac{1^3}{3} - \frac{0^3}{3} = \frac{1}{3}.$$

In general, $\frac{x^{p+1}}{p+1}$ is an antiderivative of x^p in case $p \neq -1$, and we get

$$\int_a^b x^p \, dx = \frac{b^{p+1} - a^{p+1}}{p+1}.$$

As for the case $p = -1$, the antiderivative of $\frac{1}{x}$ is $\ln|x|$, and in case a and b have the same sign,

$$\int_a^b \frac{dx}{x} = \ln|b| - \ln|a| = \ln \frac{b}{a}. \quad \blacksquare$$

Example 5.3.2 The antiderivatives of $\sin x$, $\cos x$, $\tan x$, $\frac{1}{1+x^2}$ and e^x are $-\cos x$, $\sin x$, $-\ln|\cos x|$, $\arctan x$ and e^x . Therefore

$$\begin{aligned} \int_0^\pi \sin x \, dx &= -\cos \pi + \cos 0 = 2; \\ \int_0^\pi \cos x \, dx &= \sin \pi - \sin 0 = 0; \\ \int_0^{\frac{\pi}{4}} \tan x \, dx &= -\ln \cos \frac{\pi}{4} + \ln \cos 0 = \frac{\ln 2}{2}; \\ \int_0^1 \frac{dx}{1+x^2} &= \arctan 1 - \arctan 0 = \frac{\pi}{4}; \\ \int_0^a e^x \, dx &= e^a - e^0 = e^a - 1. \end{aligned} \quad \blacksquare$$

Example 5.3.3 The graphs of the functions $\sin x$ and $\cos x$ enclose many regions. One such region is over the interval from $\frac{\pi}{4}$ to $\frac{3\pi}{4}$. Since $\sin x \geq \cos x$ on the interval, the area of the region is

$$\int_{\frac{\pi}{4}}^{\frac{3\pi}{4}} (\sin x - \cos x) \, dx = \left(-\cos \frac{3\pi}{4} - \sin \frac{3\pi}{4}\right) - \left(-\cos \frac{\pi}{4} - \sin \frac{\pi}{4}\right) = 2\sqrt{2}.$$

The computation makes use of the antiderivative $-\cos x - \sin x$ of the function $\sin x - \cos x$. $\blacksquare$

In the example above, we used the fact that if $f(x) \geq g(x)$ on $[a, b]$, then the area of the region between f and g over $[a, b]$ is $\int_a^b (f(x) - g(x)) \, dx$. In case both $f(x)$ and $g(x)$ are non-negative, it is clear that the area between the two functions is the area $\int_a^b f(x) \, dx$ of the region between f and the x -axis subtracting the area $\int_a^b g(x) \, dx$ of the region between g and the x -axis. Therefore

$$\int_a^b f(x) \, dx - \int_a^b g(x) \, dx = \int_a^b (f(x) - g(x)) \, dx$$

is the area between f and g . In general, the area between f and g is the same as the area between $f + A$ and $g + A$. For sufficiently large A , both $f + A$ and $g + A$ are non-negative, and the area between them is

$$\int_a^b ((f(x) + A) - (g(x) + A)) \, dx = \int_a^b (f(x) - g(x)) \, dx.$$

Integration by Parts

The Riemann integral can be computed by the antiderivative, which is the reverse of the process of taking the derivative. Consequently, the properties of the derivative should translate into properties of the Riemann integral. For example, if $F(x)$ and $G(x)$ are the antiderivatives of $f(x)$ and $g(x)$, then by

$$(F(x) + G(x))' = F'(x) + G'(x) = f(x) + g(x)$$

and the Newton-Leibniz formula, we get

$$\begin{aligned} \int_a^b (f(x) + g(x)) \, dx &= (F(b) + G(b)) - (F(a) + G(a)) \\ &= (F(b) - F(a)) + (G(b) - G(a)) = \int_a^b f(x) \, dx + \int_a^b g(x) \, dx. \end{aligned}$$

We already know this property directly from the definition of Riemann integral.

The Leibniz rule

$$(f(x)g(x))' = f'(x)g(x) + f(x)g'(x)$$

means that $f(x)g(x)$ is the antiderivative of $f'(x)g(x) + f(x)g'(x)$. By the Newton-Leibniz formula, we conclude that

$$f(b)g(b) - f(a)g(a) = \int_a^b (f'(x)g(x) + f(x)g'(x)) \, dx.$$

This is the formula for the integration by parts.

In applying the integration by parts (and the change of variable) in concrete computations, it is often convenient to use the differential notation

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

For example,

$$x^n dx = \frac{1}{n+1} d(x^{n+1}), \quad \frac{dx}{x} = d \ln x, \quad \sin x dx = -d \cos x, \quad \cos x dx = d \sin x, \quad e^x dx = de^x.$$

In other words, in the differential form $g(x)dx$, the function $g(x)$ can be “moved” into the “differential” d to become $dG(x)$, with $G(x)$ being any antiderivative of $g(x)$.

Using the differential notation, the integration by parts becomes the following.

Theorem 5.3.2 (Integration by Parts) Suppose $f(x)$ and $g(x)$ are continuously differentiable (i.e., the derivatives $f'(x)$ and $g'(x)$ are continuous). Then

$$\int_a^b f(x) dg(x) = f(b)g(b) - f(a)g(a) - \int_a^b g(x) df(x).$$

Example 5.3.4 By taking $f(x) = x$ and $g(x) = \sin x$ in the integration by parts, we get

$$\begin{aligned} \int_0^\pi x \cos x dx &= \int_0^\pi x d \sin x = \pi \sin \pi - 0 \sin 0 - \int_0^\pi \sin x dx \\ &= - \int_0^\pi \sin x dx = \cos \pi - \cos 0 = -2. \end{aligned}$$

A key step is to move $\cos x$ inside the differential to become $\cos x dx = d \sin x$, by $(\sin x)' = \cos x$.

By applying integration by parts twice, we get

$$\begin{aligned} \int_0^\pi x^2 \cos x dx &= \int_0^\pi x^2 d \sin x = \pi^2 \sin \pi - 0^2 \sin 0 - \int_0^\pi \sin x dx^2 = - \int_0^\pi 2x \sin x dx \\ &= \int_0^\pi 2x d \cos x = 2(\pi \cos \pi - 0 \cos 0) - 2 \int_0^\pi \cos x dx \\ &= -2\pi - 2(\sin \pi - \sin 0) = -2\pi. \end{aligned}$$

Example 5.3.5 The integration of $\arctan x$ can also be computed by using integration by parts

$$\begin{aligned} \int_0^1 \arctan x dx &= 1 \arctan 1 - 0 \arctan 0 - \int_0^1 x d \arctan x = \frac{\pi}{4} - \int_0^1 \frac{x dx}{1+x^2} \\ &= \frac{\pi}{4} - \frac{1}{2} \ln(1+x^2) \Big|_{x=0}^{x=1} = \frac{\pi}{4} - \frac{1}{2} \ln 2. \end{aligned}$$

Here we used the convenient notation

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

Example 5.3.6 The function $x(1-x)^n$ can be integrated by expanding $(1-x)^n$ as long as n is a small natural number. But the expansion becomes complicated when n becomes large, and the method does not work if n is not an integer. Instead, we may use integration by parts. For $n \geq 0$,

$$\begin{aligned} \int_0^1 x(1-x)^n dx &= - \int_0^1 x d \left(\frac{(1-x)^{n+1}}{n+1} \right) = \frac{(1-x)^{n+1}}{n+1} \Big|_{x=0}^{x=1} + \int_0^1 \frac{(1-x)^{n+1}}{n+1} dx \\ &= \frac{(1-x)^{n+2}}{(n+1)(n+2)} \Big|_{x=0}^{x=1} = \frac{1}{(n+1)(n+2)}. \end{aligned}$$

Example 5.3.7 To integrate $e^x \sin x$, we use $e^x dx = de^x$ to get

$$\int_0^{\frac{\pi}{2}} e^x \sin x dx = \int_0^{\frac{\pi}{2}} \sin x de^x = e^{\frac{\pi}{2}} \sin \frac{\pi}{2} - e^0 \sin 0 - \int_0^{\frac{\pi}{2}} e^x \cos x dx = e^{\frac{\pi}{2}} - \int_0^{\frac{\pi}{2}} e^x \cos x dx.$$

Let

$$I = \int_0^{\frac{\pi}{2}} e^x \sin x dx, \quad J = \int_0^{\frac{\pi}{2}} e^x \cos x dx.$$

Then we get $I = e^{\frac{\pi}{2}} - J$. By similarly applying the integration by parts to $\int_0^{\frac{\pi}{2}} e^x \cos x dx$, we also get $J = -1 + I$. Solving the system

$$I + J = e^{\frac{\pi}{2}}, \quad I - J = 1,$$

we get

$$\int_0^{\frac{\pi}{2}} e^x \sin x dx = \frac{e^{\frac{\pi}{2}} + 1}{2}, \quad \int_0^{\frac{\pi}{2}} e^x \cos x dx = \frac{e^{\frac{\pi}{2}} - 1}{2}.$$

The interested reader may get the same relation between I and J by using $\sin x dx = -d \cos x$ and $\cos x dx = d \sin x$ in the integration by parts. ■

Example 5.3.8 Let

$$I_p = \int_0^{\frac{\pi}{2}} \cos^p x dx.$$

Moreover, for $p > 1$, we have

$$\begin{aligned} I_p &= \int_0^{\frac{\pi}{2}} \cos^{p-1} x d \sin x = \cos^{p-1} \frac{\pi}{2} \sin \frac{\pi}{2} - \cos^{p-1} 0 \sin 0 - (p-1) \int_0^{\frac{\pi}{2}} \sin^2 x \cos^{p-2} x dx \\ &= -(p-1) \int_0^{\frac{\pi}{2}} (1 - \cos^2 x) \cos^{p-2} x dx = -(p-1)(I_p - I_{p-2}). \end{aligned}$$

This implies that

$$I_p = \frac{p-1}{p} I_{p-2}.$$

By $I_0 = \frac{\pi}{2}$ and $I_1 = 1$, for natural number n , we get

$$\begin{aligned} I_{2n} &= \frac{(2n-1)(2n-3) \cdots 1}{2n(2n-2) \cdots 2} \frac{\pi}{2} = \frac{(2n)!}{2^{2n+1}(n!)^2} \pi, \\ I_{2n+1} &= \frac{2n(2n-2) \cdots 2}{(2n+1)(2n-1) \cdots 3} = \frac{2^{2n}(n!)^2}{(2n+1)!}. \end{aligned}$$

■

Change of Variable

It remains to translate the chain rule. Suppose $F(x)$ is an antiderivative of $f(x)$ and $\phi(t)$ is differentiable. Then

$$(F(\phi(t)))' = F'(\phi(t))\phi'(t) = f(\phi(t))\phi'(t)$$

tells us that $F(\phi(t))$ is the antiderivative of $f(\phi(t))\phi'(t)$. By the Newton-Leibniz formula,

$$\int_a^b f(\phi(t))\phi'(t) dt = F(\phi(b)) - F(\phi(a)) = \int_{\phi(a)}^{\phi(b)} f(x) dx.$$

Strictly speaking, the argument requires all the functions in the integral to be continuous. So we have proved the following result.

Theorem 5.3.3 (Change of Variable) *Suppose ϕ is continuously differentiable on $[a, b]$ (i.e., the derivative ϕ' is continuous). Suppose $f(x)$ is continuous on $\phi([a, b])$. Then*

$$\int_{\phi(a)}^{\phi(b)} f(x) dx = \int_a^b f(\phi(t))\phi'(t) dt.$$

By using the differential notation, the change of variable formula can also be written as

$$\int_{\phi(a)}^{\phi(b)} f(x) dx = \int_a^b f(\phi(t)) d\phi(t).$$

Example 5.3.9 To compute $\int_0^3 \frac{x dx}{1 + \sqrt{1+x}}$, we introduce $t = \sqrt{1+x}$, or $x = \phi(t) = t^2 - 1$. Then

$$\int_{a^2-1}^{b^2-1} \frac{x dx}{1 + \sqrt{1+x}} = \int_a^b \frac{t^2-1}{1+t} (t^2-1)' dt = \int_a^b (2t^2 - 2t) dt.$$

Taking $a = 1$ and $b = 2$, we get

$$\int_0^3 \frac{x dx}{1 + \sqrt{1+x}} = \int_1^2 (2t^2 - 2t) dt = \left(\frac{2t^3}{3} - t^2 \right) \Big|_{t=1}^{t=2} = \frac{5}{3}. \quad \blacksquare$$

Example 5.3.10 To compute $\int_0^1 \frac{x dx}{1+x^2}$, we notice that $x dx = \frac{1}{2} dx^2$ and the integral becomes the form $\int_0^1 f(x^2) dx^2$, and we may think about x^2 as the new variable. Therefore

$$\int_0^1 \frac{x dx}{1+x^2} = \frac{1}{2} \int_0^1 \frac{dx^2}{1+x^2} = \frac{1}{2} \int_{0^2}^{1^2} \frac{dt}{1+t} = \frac{1}{2} \int_0^1 \frac{d(1+t)}{1+t} = \frac{1}{2} \ln(1+t) \Big|_{t=0}^{t=1} = \frac{1}{2} \ln 2. \quad \blacksquare$$

At the end, we further think of $1+t$ as a new variable.

Example 5.3.11 The function $x(1-x)^n$ can also be integrated by introducing

$$t = 1 - x, \quad x = 1 - t, \quad dx = -dt.$$

Since x going from 0 to 1 corresponds to t going from 1 to 0, we get

$$\int_0^1 x(1-x)^n dx = - \int_1^0 (1-t)t^n(-dt) = \int_1^0 (t^n - t^{n+1})dt = \left(\frac{t^{n+1}}{n+1} - \frac{t^{n+2}}{n+2} \right) \Big|_{t=1}^{t=0} = \frac{1}{(n+1)(n+2)}.$$

An alternative change of variable is

$$t = (1-x)^n, \quad x = 1 - t^{\frac{1}{n}}, \quad dx = -\frac{1}{n}t^{\frac{1}{n}-1} dt.$$

The detailed computation is left to the reader. ■

Example 5.3.12 The disk of radius R is the region between functions $\sqrt{R^2 - x^2}$ and $-\sqrt{R^2 - x^2}$ over the interval $[-R, R]$. The area of the disk is

$$\int_{-R}^R \left(\sqrt{R^2 - x^2} - (-\sqrt{R^2 - x^2}) \right) dx = 2 \int_{-R}^R \sqrt{R^2 - x^2} dx.$$

To compute the integral, we let $x = R \sin t$ or $t = \arcsin \frac{x}{R}$, with $x \in [-R, R]$ corresponding to $t \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. Then

$$\begin{aligned} 2 \int_{-R}^R \sqrt{R^2 - x^2} dx &= 2 \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sqrt{R^2 - R^2 \sin^2 t} (R \sin t)' dt = 2 \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} R^2 \cos^2 t dt \\ &= 2R^2 \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{1 + \cos 2t}{2} dt = 2R^2 \left(1 + \frac{\sin 2t}{2} \right) \Big|_{t=-\frac{\pi}{2}}^{t=\frac{\pi}{2}} = \pi R^2. \end{aligned} \quad \blacksquare$$

Example 5.3.13 Let $x = \sin t$. Then we have

$$\int_0^1 (1-x^2)^p dx = \int_0^{\frac{\pi}{2}} (1-\sin^2 t)^p \cos t dt = \int_0^{\frac{\pi}{2}} \cos^{2p+1} t dt.$$

By Example 5.3.8, for natural number n , we get

$$\begin{aligned} \int_0^1 (1-x^2)^{n-\frac{1}{2}} dx &= \frac{(2n-1)(2n-3)\cdots 1}{2n(2n-2)\cdots 2} \frac{\pi}{2} = \frac{(2n)!}{2^{2n+1}(n!)^2} \pi, \\ \int_0^1 (1-x^2)^n dx &= \frac{2n(2n-2)\cdots 2}{(2n+1)(2n-1)\cdots 3} = \frac{2^{2n}(n!)^2}{(2n+1)!}. \end{aligned} \quad \blacksquare$$

Example 5.3.14 To compute the integral $I = \int_0^\pi \frac{x \sin x}{1 + \cos^2 x} dx$, we introduce $t = \pi - x$. Then

$$\begin{aligned} I &= \int_\pi^0 \frac{(\pi-t) \sin(\pi-t)}{1 + \cos^2(\pi-t)} d(\pi-t) = - \int_\pi^0 \frac{(\pi-t) \sin t}{1 + \cos^2 t} dt \\ &= \pi \int_0^\pi \frac{\sin t}{1 + \cos^2 t} dt - \int_0^\pi \frac{t \sin t}{1 + \cos^2 t} dt = \pi \int_0^\pi \frac{\sin t}{1 + \cos^2 t} dt - I. \end{aligned}$$

Therefore

$$I = \frac{\pi}{2} \int_0^\pi \frac{\sin t}{1 + \cos^2 t} dt = -\frac{\pi}{2} \int_0^\pi \frac{d \cos t}{1 + \cos^2 t} = -\frac{\pi}{2} \int_1^{-1} \frac{du}{1 + u^2} = -\frac{\pi}{2} \arctan u \Big|_{u=1}^{u=-1} = \frac{\pi^2}{4}. \quad \blacksquare$$

Exercises

5.3.1 Compute integrals:

1. $\int_0^1 \sqrt[4]{1-x} dx;$

5. $\int_0^\pi \sin\left(\frac{\pi}{2} - x\right) dx;$

2. $\int_0^1 x \sqrt[n]{1-x^2} dx;$

6. $\int_0^{\frac{\pi}{2}} \sin^3 x \cos^2 x dx;$

3. $\int_0^1 \left(\frac{x-1}{x+1}\right)^4 dx;$

7. $\int_a^{2a} \frac{\sqrt{x^2 - a^2}}{x^4} dx.$

4. $\int_0^2 (e^x - e^{-x})^2 dx;$

5.3.2 Prove that the antiderivative of an odd function is even. What about the antiderivative of an even function?

5.3.3 Suppose f is an integrable periodic function with period T . Show that

$$\int_a^{a+T} f(x) dx = \int_0^T f(x) dx$$

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

5.3.4 Explain why the Newton-Leibnitz formula cannot be used to evaluate $\int_{-1}^1 \frac{dx}{x^2}$.

5.3.5 Suppose $f(x)$ is continuous. Compute the derivatives:

1. $\frac{d}{dx} \int_a^x f(t^2) dt;$

3. $\frac{d}{dx} \int_0^{x^3} f(t) dt;$

2. $\frac{d}{dx} \int_x^b f(t) dt;$

4. $\frac{d}{dx} \int_x^{x^3} f(t) dt.$

5.3.6 Suppose that $f(x)$ is continuous on $[0, 1]$. Prove the following equalities:

1. $\int_0^{\frac{\pi}{2}} f(\sin x) dx = \int_0^{\frac{\pi}{2}} f(\cos x) dx;$

$$2. \int_0^\pi x f(\sin x) \, dx = \frac{\pi}{2} \int_0^\pi f(\sin x) \, dx.$$

$$5.3.7 \text{ Prove } \lim_{\epsilon \rightarrow 0} \int_0^{2\pi} \frac{dx}{1 + \epsilon \cos x} = 2\pi.$$

5.4 ANTIDERIVATIVE

The computation of the Riemann integral in terms of the antiderivative suggests us to use the notation $\int f(x) \, dx$ to denote all the antiderivatives of $f(x)$. Therefore

$$\begin{aligned} \int x^p \, dx &= \frac{x^{p+1}}{p+1} + C, & p \neq -1; & \int \frac{dx}{x} &= \ln |x| + C; \\ \int \sin x \, dx &= -\cos x + C; & \int \cos x \, dx &= \sin x + C; \\ \int \tan x \, dx &= -\ln |\cos x| + C; & \int \sec^2 x \, dx &= \tan x + C; \\ \int e^x \, dx &= e^x + C; & \int \ln x \, dx &= x \ln x - x + C, \end{aligned}$$

where C denotes an arbitrary constant. Since the “integral” $\int f(x) \, dx$ does not have a range $[a, b]$ and has no definite numerical value, we call $\int f(x) \, dx$ the **indefinite integral**. Correspondingly, the usual Riemann integral $\int_a^b f(x) \, dx$ is the **definite integral**.

The fundamental theorem of calculus tells us that if $f(x)$ is continuous, then

$$\int f(x) \, dx = \int_a^x f(t) \, dt + C.$$

Moreover, the derivative equalities translate into the properties of the indefinite integral

$$\begin{aligned} \int (f(x) + g(x)) \, dx &= \int f(x) \, dx + \int g(x) \, dx, \\ \int cf(x) \, dx &= c \int f(x) \, dx, \\ \int f(x)g'(x) \, dx &= f(x)g(x) - \int g(x)f'(x) \, dx, \\ \int f(\phi(t))\phi'(t) \, dt &= \left(\int f(x) \, dx \right) \Big|_{x=\phi(t)}. \end{aligned}$$

The third equality is the integration by parts and can also be written as

$$\int f \, dg = fg - \int g \, df$$

by using the differential notation. The fourth equality is the change of variable, which can also be written as

$$\int f(\phi(t)) \, d\phi(t) = \int f(x) \, dx$$

by using the differential notation. Note that after computing $\int f(x) \, dx$ on the right side, x should be substituted by $\phi(t)$. Equivalently, if ϕ is invertible, then we have

$$\int f(x) \, dx = \left(\int f(\phi(t)) \phi'(t) \, dt \right) \Big|_{t=\phi^{-1}(x)},$$

where after computing the integral on the right, the variable t should be substituted by $\phi^{-1}(x)$.

Example 5.4.1 For $p \neq -1$, the function $x^p \ln x$ can be integrated by using integration by parts

$$\begin{aligned} \int x^p \ln x \, dx &= \frac{1}{p+1} \int \ln x \, dx^{p+1} = \frac{1}{p+1} x^{p+1} \ln x - \frac{1}{p+1} \int x^{p+1} \, d \ln x \\ &= \frac{1}{p+1} x^{p+1} \ln x - \frac{1}{p+1} \int x^{p+1} \frac{dx}{x} = \frac{1}{p+1} x^{p+1} \ln x - \frac{1}{(p+1)^2} x^{p+1} + C. \end{aligned}$$

For $p = -1$, we have

$$\int \frac{\ln x}{x} \, dx = \int \ln x \, d \ln x = \frac{\ln^2 x}{2} + C. \quad \blacksquare$$

Example 5.4.2 To integrate a function of $(a + bx)^p$, we may introduce $t = a + bx$ and get

$$\int f((a + bx)^p) \, dx = \frac{1}{b} \int f(t^p) \, dt.$$

For example, by $t = 2x - 1$, we have

$$\begin{aligned} \int x \sqrt{2x-1} \, dx &= \int \frac{t+1}{2} \sqrt{t} \frac{1}{2} \, dt = \frac{1}{4} \int (t^{\frac{3}{2}} + t^{\frac{1}{2}}) \, dt = \frac{1}{4} \left(\frac{2}{5} t^{\frac{5}{2}} + \frac{2}{3} t^{\frac{3}{2}} \right) + C \\ &= \frac{1}{30} (3(2x-1)^{\frac{5}{2}} + 5(2x-1)^{\frac{3}{2}}) + C = \frac{1}{15} (2x-1)^{\frac{3}{2}} (3x+1) + C. \end{aligned}$$

For another example,

$$\begin{aligned} \int \frac{x^5 \, dx}{\sqrt{1+x^2}} &= \frac{1}{2} \int \frac{x^4 \, dx^2}{\sqrt{1+x^2}} = \frac{1}{2} \int \frac{t^2 \, dt}{\sqrt{1+t}} = \frac{1}{2} \int \frac{(u-1)^2 \, du}{\sqrt{u}} = \frac{1}{2} \int (u^{\frac{5}{2}} - 2u^{\frac{3}{2}} + u^{-\frac{1}{2}}) \, du \\ &= \frac{1}{2} \left(\frac{3}{8} u^{\frac{8}{2}} - 2 \frac{3}{5} u^{\frac{5}{2}} + \frac{3}{2} u^{\frac{2}{2}} \right) + C = \frac{1}{80} (1+x^2)^{\frac{2}{3}} (5x^4 + 2x^2 + 17) + C. \quad \blacksquare \end{aligned}$$

Example 5.4.3 The function xe^{x^2} may be integrated by using change of variable

$$\int xe^{x^2} \, dx = \frac{1}{2} \int e^{x^2} \, dx^2 = \frac{1}{2} e^{x^2} + C.$$

Here we computed $\int \phi'(x)f(\phi(x)) \, dx = \int f(\phi(x)) \, d\phi(x)$ for $f(y) = e^y$ and $\phi(x) = x^2$.

Similarly, we have

$$\int x^3 e^{x^2} \, dx = \frac{1}{2} \int x^2 e^{x^2} \, dx^2 = \frac{1}{2} \int t e^t \, dt, \quad t = x^2.$$

Then $t e^t$ may be integrated by using integration by parts

$$\int t e^t \, dt = \int t \, d e^t = t e^t - \int e^t \, dt = t e^t - e^t + C.$$

Therefore

$$\int x^3 e^{x^2} \, dx = \frac{1}{2}(t e^t - e^t) + C = \frac{1}{2}(x^2 - 1)e^{x^2} + C.$$

In general, we may integrate $x^n e^{x^2}$ for odd natural numbers n . However, the integrals $\int e^{x^2} \, dx$ and $\int x^2 e^{x^2} \, dx$ do not have elementary expressions. ■

Example 5.4.4 We compute the integral of some trigonometric functions.

$$\begin{aligned} \int \cos 2x \sin 3x \, dx &= \int \frac{1}{2}(\sin 5x + \sin x) \, dx = \frac{1}{2} \left(-\frac{1}{5} \cos 5x - \cos x \right) + C \\ &= -\frac{1}{10} \cos 5x - \frac{1}{2} \cos x + C, \\ \int \cos^2 x \sin 3x \, dx &= \int \frac{1}{2}(1 + \cos 2x) \sin 3x \, dx = \int \frac{1}{2} \left(\sin 3x + \frac{1}{2}(\sin 5x + \sin x) \right) \, dx \\ &= -\frac{1}{20} \cos 5x - \frac{1}{6} \cos 3x - \frac{1}{4} \cos x + C. \end{aligned}$$

In general, the antiderivative of the product of several $\sin ax$ and $\cos bx$ can be computed by using the trigonometric equalities to rewrite the product as a linear combination of several $\sin Ax$ and $\cos Bx$. ■

Example 5.4.5 The antiderivative of $\sin^m x \cos^n x$ can be easily computed if m, n are integers and one of them is odd.

$$\begin{aligned} \int \sin^4 x \cos^7 x \, dx &= \int \sin^4 x (1 - \sin^2 x)^3 \, d \sin x = \int t^4 (1 - t^2)^3 \, dt \\ &= \frac{1}{11} t^{11} - \frac{1}{3} t^9 + \frac{3}{7} t^7 - \frac{1}{5} t^5 + C \\ &= \frac{1}{11} \sin^{11} x - \frac{1}{3} \sin^9 x + \frac{3}{7} \sin^7 x - \frac{1}{5} \sin^5 x + C, \\ \int \sec x \, dx &= \int \frac{dx}{\cos x} = \int \frac{d \sin x}{1 - \sin^2 x} = \int \frac{dt}{1 - t^2} = \frac{1}{2} \int \left(\frac{1}{t+1} - \frac{1}{t-1} \right) \, dt \\ &= \frac{1}{2} \ln \left| \frac{t+1}{t-1} \right| + C = \frac{1}{2} \ln \frac{1 + \sin x}{1 - \sin x} + C \\ &= \frac{1}{2} \ln \frac{(1 + \sin x)^2}{1 - \sin^2 x} + C = \ln \frac{1 + \sin x}{|\cos x|} + C = \ln |\sec x + \tan x| + C. \end{aligned}$$

Note that the second integral is $\int \sin^0 x \cos^{-1} x \, dx$. In general, the antiderivative of $\tan^m x \sec^n x$ may be computed if one of m and $m+n$ is odd. ■

Example 5.4.6 The antiderivative of $\tan^m x \sec^n x$ can be computed also by using

$$\sec^2 x = 1 + \tan^2 x, \quad d \tan x = \sec^2 x \, dx, \quad d \sec x = \tan x \sec x \, dx.$$

For examples,

$$\begin{aligned} \int \tan^3 x \sec x \, dx &= \int (\sec^2 x - 1) d \sec x = \frac{1}{3} \sec^3 x - \sec x + C, \\ \int \tan^3 x \sec^2 x \, dx &= \int \tan^3 x \, d \tan x = \frac{1}{4} \tan^4 x + C, \\ \int \sec^4 x \, dx &= \int \sec^2 x \, d \tan x = \int (\tan^2 x + 1) d \tan x = \frac{1}{3} \tan^3 x + \tan x + C. \end{aligned}$$

In general, we may easily compute the antiderivative of $\tan^m x \sec^n x$ when m is odd or n is even. ■

Example 5.4.7 The antiderivative of $\sec^3 x$ can be computed as $\int \sin^0 x \cos^{-3} x \, dx$, by Example 5.4.4. Here is a more clever (but less systematic) way

$$\begin{aligned} \int \sec^3 x \, dx &= \int \sec x \, d \tan x = \sec x \tan x - \int \tan x \, d \sec x = \sec x \tan x - \int \tan^2 x \sec x \, dx \\ &= \sec x \tan x - \int (\sec^2 x - 1) \sec x \, dx = \sec x \tan x - \int \sec^3 x \, dx + \int \sec x \, dx. \end{aligned}$$

By the second computation in Example 5.4.5, we get

$$\int \sec^3 x \, dx = \frac{1}{2} \sec x \tan x + \frac{1}{2} \int \sec x \, dx = \frac{1}{2} \sec x \tan x + \frac{1}{2} \ln |\sec x + \tan x| + C.$$

Similar idea can be used to reduce the computation of $\int \tan^3 x \, dx$ to $\int \tan x \, dx$. ■

Example 5.4.8 Let

$$I = \int e^x \sin x \, dx, \quad J = \int e^x \cos x \, dx.$$

Using integration by parts, we have

$$\begin{aligned} I &= \int \sin x \, de^x = e^x \sin x - \int e^x \cos x \, dx = e^x \sin x - J, \\ J &= \int \cos x \, de^x = e^x \cos x + \int e^x \sin x \, dx = e^x \cos x + I. \end{aligned}$$

Solving the system, we get

$$\int e^x \sin x \, dx = \frac{1}{2} e^x (\sin x - \cos x) + C, \quad \int e^x \cos x \, dx = \frac{1}{2} e^x (\sin x + \cos x) + C.$$

The antiderivative of $e^x \sin x$ and $e^x \cos x$ helps us to find the antiderivatives of $xe^x \sin x$ and $xe^x \cos x$.

$$\begin{aligned}\int xe^x \sin x \, dx &= \int x \, d\left(\frac{1}{2}e^x(\sin x - \cos x)\right) = \frac{1}{2}xe^x(\sin x - \cos x) - \frac{1}{2} \int e^x(\sin x - \cos x) \, dx \\ &= \frac{1}{2}xe^x(\sin x - \cos x) - \frac{1}{2}(I - J) = \frac{1}{2}e^x(x \sin x - x \cos x + \cos x) + C.\end{aligned}$$

In general, we may find the antiderivative of $x^n e^{ax} \sin bx$ and $x^n e^{ax} \cos bx$ by the similar idea. ■

Example 5.4.9 To integrate functions of $\sqrt{a^2 - x^2}$, $a > 0$, we may introduce $x = a \sin t$, $-\frac{\pi}{2} \leq t \leq \frac{\pi}{2}$. Then $\cos t \geq 0$, and we have

$$\sqrt{a^2 - x^2} = \sqrt{a^2(1 - \sin^2 t)} = a \cos t, \quad dx = a \cos t \, dt.$$

Therefore

$$\int f(\sqrt{a^2 - x^2}) \, dx = \int f(a \cos t) a \cos t \, dt.$$

For example,

$$\int \sqrt{a^2 - x^2} \, dx = a^2 \int \cos^2 t \, dt = a^2 \int \frac{1}{2}(1 + \cos 2t) \, dt = \frac{a^2}{2}t + \frac{a^2}{4} \sin 2t + C.$$

By $\sin 2t = 2 \sin t \cos t = \frac{2}{a^2}x\sqrt{a^2 - x^2}$, we conclude that

$$\int \sqrt{a^2 - x^2} \, dx = \frac{a^2}{2} \arcsin \frac{x}{a} + \frac{1}{2}x\sqrt{a^2 - x^2} + C. \quad \blacksquare$$

Example 5.4.10 For $\sqrt{x^2 + a^2}$, we may introduce $x = a \tan t$ and get

$$\int f(\sqrt{x^2 + a^2}) \, dx = \int f(a \sec t) a \sec^2 t \, dt.$$

For $\sqrt{x^2 - a^2}$, we may introduce $x = a \sec t$ and get

$$\int f(\sqrt{x^2 - a^2}) \, dx = \int f(a \tan t) a \sec t \tan t \, dt.$$

For example,

$$\int \frac{dx}{\sqrt{x^2 + a^2}} = \int \frac{a \sec^2 t \, dt}{\sqrt{a^2 \tan^2 t + a^2}} = \int \frac{a \sec^2 t \, dt}{a \sec t} = \int \sec t \, dt.$$

By the second integral in Example 5.4.5, we get

$$\int \frac{dx}{\sqrt{x^2 + a^2}} = \ln |\sec t + \tan t| + C = \ln(\sqrt{x^2 + a^2} + x) + C.$$

Similarly, we have

$$\begin{aligned}\int \frac{dx}{\sqrt{x^2 - a^2}} &= \int \frac{a \sec t \tan t \, dt}{\sqrt{a^2 \sec^2 t - a^2}} = \int \frac{a \sec t \tan t \, dt}{a \tan t} = \int \sec t \, dt \\ &= \ln |\sec t + \tan t| + C = \ln |x + \sqrt{x^2 - a^2}| + C.\end{aligned}$$

Example 5.4.11 Let

$$I_p = \int (x^2 + a^2)^p \, dx.$$

Using integration by parts, we have

$$\begin{aligned}I_p &= x(x^2 + a^2)^p - \int x \, d(x^2 + a^2)^p = x(x^2 + a^2)^p - 2p \int x^2(x^2 + a^2)^{p-1} \, dx \\ &= x(x^2 + a^2)^p - 2p \int ((x^2 + a^2)^p - a^2(x^2 + a^2)^{p-1}) \, dx = x(x^2 + a^2)^p - 2p(I_p - a^2 I_{p-1}).\end{aligned}$$

This gives the recursive formula

$$(2p + 1)I_p = 2pa^2 I_{p-1} + x(x^2 + a^2)^p.$$

For $p = \frac{1}{2}$, by Example 5.4.10, we have

$$\int \sqrt{x^2 + a^2} \, dx = \frac{1}{2 \cdot \frac{1}{2} + 1} \left(2 \cdot \frac{1}{2} a^2 \int \frac{dx}{\sqrt{x^2 + a^2}} + x \sqrt{x^2 + a^2} \right) = \frac{1}{2} a^2 \ln(x + \sqrt{x^2 + a^2}) + \frac{1}{2} x \sqrt{x^2 + a^2} + C.$$

Note that

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

For $p = -1$, we get

$$\int \frac{dx}{(x^2 + 1)^2} = I_{-2} = \frac{1}{2} \left(I_{-1} + \frac{x}{x^2 + 1} \right) = \frac{1}{2} \arctan x + \frac{x}{2(x^2 + 1)} + C.$$

We also have

$$\int \frac{dx}{(x^2 + 1)^3} = I_{-3} = \frac{1}{4} \left(3I_{-2} + \frac{x}{(x^2 + 1)^2} \right) = \frac{3}{8} \arctan x + \frac{x}{4(x^2 + 1)^2} + \frac{3x}{8(x^2 + 1)} + C.$$

Finally, we remark that similar recursive relations can be found for $\int (x^2 - a^2)^p \, dx$ and $\int (a^2 - x^2)^p \, dx$, and the integrals may be computed by reducing the the special cases of p . ■

Example 5.4.12 Suppose we want to find the antiderivative of a function $f(\sqrt{ax^2 + bx + c})$ of the square root of a quadratic function, we may use completing the square to get rid of the first order term and reduce to the antiderivative of a function of $\sqrt{a^2 - x^2}$, $\sqrt{x^2 + a^2}$ or $\sqrt{x^2 - a^2}$.

For examples,

$$\begin{aligned}
 \int \frac{dx}{\sqrt{x(a-x)}} &= \int \frac{dx}{\sqrt{\frac{a^2}{4} - \left(x - \frac{a}{2}\right)^2}} = \int \frac{d\left(\frac{a}{2} \sin t\right)}{\sqrt{\frac{a^2}{4} - \left(\frac{a}{2} \sin t\right)^2}} = \int \frac{\frac{a}{2} \cos t \, dt}{\frac{a}{2} \cos t} \\
 &= t + C = \arcsin\left(\frac{2}{a}x - 1\right) + C, \\
 \int \frac{x \, dx}{\sqrt{5-2x+x^2}} &= \int \frac{x \, dx}{\sqrt{4+(x-1)^2}} = \int \frac{(2 \tan t + 1) \, d(2 \tan t + 1)}{\sqrt{4+4 \tan^2 t}} = \int (2 \tan t + 1) \sec t \, dt \\
 &= 2 \sec t + \ln |\sec t + \tan t| + C \\
 &= 2 \sqrt{\left(\frac{x-1}{2}\right)^2 + 1} + \ln \left| \sqrt{\left(\frac{x-1}{2}\right)^2 + 1} + \frac{x-1}{2} \right| + C \\
 &= \sqrt{5-2x+x^2} + \ln(\sqrt{5-2x+x^2} + x-1) + C. \quad \blacksquare
 \end{aligned}$$

Example 5.4.13 To integrate $\frac{1}{\sqrt{1+e^x}}$, we introduce $t = \sqrt{1+e^x}$. Then $t^2 = 1+e^x$, $2t \, dt = e^x \, dx$, and

$$\begin{aligned}
 \int \frac{dx}{\sqrt{1+e^x}} &= \int \frac{\frac{2t}{t^2-1} \, dt}{t} = \int \frac{2}{t^2-1} \, dt = \int \left(\frac{1}{t-1} - \frac{1}{t+1} \right) \, dt \\
 &= \ln \left| \frac{t-1}{t+1} \right| + C = \ln \frac{\sqrt{1+e^x}-1}{\sqrt{1+e^x}+1} + C.
 \end{aligned}$$

The integral can also be computed by introducing $t = e^{-\frac{x}{2}}$

$$\begin{aligned}
 \int \frac{dx}{\sqrt{1+e^x}} &= \int \frac{e^{-\frac{x}{2}} \, dx}{\sqrt{e^{-x}+1}} = -2 \int \frac{de^{-\frac{x}{2}}}{\sqrt{e^{-x}+1}} = -2 \int \frac{dt}{\sqrt{t^2+1}} \\
 &= -2 \ln(t + \sqrt{t^2+1}) + C = -2 \ln(e^{-\frac{x}{2}} + \sqrt{e^{-x}+1}) + C
 \end{aligned}$$

The reader may check that the two answers are the same. ■

Exercises

5.4.1 Compute the following antiderivatives:

- | | |
|---|------------------------------------|
| 1. $\int \frac{x-1}{\sqrt{x}} \, dx;$ | 3. $\int \frac{x^2 \, dx}{1+x^2};$ |
| 2. $\int \left(\frac{1-x}{x} \right)^2 \, dx;$ | 4. $\int \frac{x^2 \, dx}{1-x^2};$ |

$$5. \int (2^x + 3^x)^2 dx;$$

$$6. \int \frac{2^{x+1} - 3^{x-1}}{6^x} dx;$$

$$7. \int \frac{x dx}{\sqrt{1-x^2}};$$

$$8. \int \frac{x^2 dx}{\sqrt{1-x^2}};$$

$$9. \int \frac{dx}{(x^2 + a^2)^{\frac{3}{2}}};$$

$$10. \int \frac{\ln x dx}{x\sqrt{1+\ln x}};$$

$$11. \int \sqrt{(x-a)(b-x)} dx;$$

$$12. \int \sqrt{(x-a)(x-b)} dx;$$

$$13. \int \sqrt{\frac{x-a}{b-x}} dx;$$

$$14. \int \sqrt{\frac{x-a}{x-b}} dx;$$

$$15. \int e^{\sqrt{x}} dx;$$

$$16. \int (\ln x)^2 dx;$$

$$17. \int (x^2 - 1)e^{-x} dx;$$

$$18. \int \frac{xe^x dx}{(x+1)^2};$$

$$19. \int x \ln \frac{1+x}{1-x} dx;$$

$$20. \int \arccos x dx;$$

$$21. \int x^2 \arccos x dx;$$

$$22. \int \frac{\arcsin x}{x^2} dx;$$

$$23. \int \arctan x dx;$$

$$24. \int \frac{(\arctan x)^2}{1+x^2} dx;$$

$$25. \int x^2 e^{-x} \sin 3x dx;$$

$$26. \int \sin x \sin 2x \sin 3x dx;$$

$$27. \int x \sin^2 x \cos 2x dx;$$

$$28. \int \sin^2 x \cos 2x dx;$$

$$29. \int \sin^2 x \cos^3 x dx;$$

$$30. \int \frac{1-2\sin x}{\cos^2 x} dx;$$

$$31. \int \cos^3 x \sqrt[3]{\sin x} dx;$$

$$32. \int \frac{a \cos x + b \sin x}{\sin 2x} dx;$$

$$33. \int \frac{dx}{a \sin x + b \cos x};$$

$$34. \int \frac{dx}{a^2 \sin^2 x + b^2 \cos^2 x};$$

$$35. \int \frac{dx}{\sin^2 x \cos^2 x};$$

$$36. \int \cot x dx;$$

$$37. \int \csc x dx;$$

$$38. \int \tan^3 x dx ;$$

$$39. \int \tan^5 x \sec^7 x \, dx;$$

$$40. \int \tan^6 x \sec^4 x \, dx;$$

$$41. \int \cot^6 x \csc^4 x \, dx;$$

$$42. \int \frac{dx}{1 + \tan x};$$

$$43. \int \sin x \ln \tan x \, dx;$$

$$44. \int \frac{x \, dx}{\cos^2 x};$$

$$45. \int \frac{x \ln(x + \sqrt{1 + x^2})}{\sqrt{1 + x^2}} \, dx;$$

$$46. \int \left(\ln(x + \sqrt{1 + x^2}) \right)^2 \, dx;$$

$$47. \int e^{\sin x} \left(\cos^2 x + \frac{1}{\cos^2 x} \right) \, dx;$$

$$48. \int \frac{x^3 \arccos x}{\sqrt{1 - x^2}} \, dx.$$

5.4.2 Derive the recursive formula for each of the following problems:

$$1. \int (\ln x)^n \, dx = x(\ln x)^n - n \int (\ln x)^{n-1} \, dx;$$

$$2. \int x^n e^x \, dx = x^n e^x - n \int x^{n-1} e^x \, dx;$$

$$3. \int \sin^n x \, dx = -\frac{1}{n} \cos x \sin^{n-1} x + \frac{n-1}{n} \int \sin^{n-2} x \, dx, \, n \geq 2.$$

5.4.3 Derive the formula

$$\int \frac{dx}{\sqrt{ax^2 + bx + c}} = \begin{cases} \frac{1}{\sqrt{a}} \ln \left(\frac{2ax + b}{\sqrt{a}} + \sqrt{ax^2 + bx + c} \right) + C, & \text{if } a > 0, \\ \frac{1}{\sqrt{-a}} \arcsin \frac{-2ax - b}{\sqrt{b^2 - 4ac}} + C, & \text{if } a < 0. \end{cases}$$

5.4.4 Find the antiderivative of

$$f(x) = \begin{cases} x^2, & \text{if } x \leq 0, \\ \sin x, & \text{if } x > 0. \end{cases}$$

5.4.5 Find the antiderivative of

$$f(x) = \begin{cases} 1 - x^2, & \text{if } |x| \leq 1, \\ \sin(1 - |x|), & \text{if } |x| > 1. \end{cases}$$

5.5 SUMMARY

Definitions

- Given a **partition** $P: a = x_0 < x_1 < \cdots < x_n = b$ of the interval $[a, b]$ and a choice of **sample points** $x_i^* \in [x_{i-1}, x_i]$, the **Riemann sum** of a function $f(x)$ is $S(P, f) = \sum_{i=1}^n f(x_i^*) \Delta x_i$, where $\Delta x_i = x_i - x_{i-1}$.
- The function $f(x)$ has **Riemann integral** $I = \int_a^b f(x) dx$ if for any $\epsilon > 0$, there is $\delta > 0$, such that $\|P\| = \max\{\Delta x_i\} < \delta$ implies $|S(P, f) - I| < \epsilon$. In other words, the Riemann integral is the limit of the Riemann sum as the size of partition goes to 0: $\int_a^b f(x) dx = \lim_{\|P\| \rightarrow 0} S(P, f) = \lim_{\max\{\Delta x_i\} \rightarrow 0} \sum_{i=1}^n f(x_i^*) \Delta x_i$.
- For $a < b$, we also defined $\int_b^a f(x) dx = - \int_a^b f(x) dx$ and $\int_a^a f(x) dx = 0$.
- A function $F(x)$ is an **antiderivative** of a function $f(x)$ if $F'(x) = f(x)$.
- All the antiderivatives of $f(x)$ is denoted $\int f(x) dx$ and called **indefinite integral**. The Riemann integral $\int_a^b f(x) dx$ is called **definite integral**.

Theorems

- Integrable functions are bounded.
- Continuous functions are integrable. In fact, a function that is continuous everywhere except at finitely many points is also integrable.
- Arithmetic Rule: If $f(x)$ and $g(x)$ are integrable, then $f(x) + g(x)$, $cf(x)$, $f(x)g(x)$ are integrable. Moreover, we have $\int_a^b (f(x) + g(x)) dx = \int_a^b f(x) dx + \int_a^b g(x) dx$ and $\int_a^b cf(x) dx = c \int_a^b f(x) dx$.
- Order Rule: If $a < b$ and $f(x) \geq g(x)$, then $\int_a^b f(x) dx \geq \int_a^b g(x) dx$.
- Addition of Intervals: If $a < b < c$, then $f(x)$ is integrable if and only if $f(x)$ is integrable on $[a, b]$ and $[b, c]$. Moreover, we have $\int_a^c f(x) dx = \int_a^b f(x) dx + \int_b^c f(x) dx$.

- **Fundamental Theorem of Calculus** If $f(x)$ is integrable and is continuous at c , then $F(x) = \int_a^c f(t) \, dt$ is differentiable at c , with $F'(c) = f(c)$.
- **Newton-Leibniz formula** If $F(x)$ is an antiderivative of a continuous function $f(x)$, then $\int_a^b f(x) \, dx = F(b) - F(a)$.
- **Integration by Parts:** If $f(x)$ and $g(x)$ are continuously differentiable, then $\int_a^b f(x)g'(x) \, dx = f(b)g(b) - f(a)g(a) - \int_a^b f'(x)g(x) \, dx$.
- **Change of Variable:** If $\phi(t)$ is continuously differentiable on $[a, b]$ and $f(x)$ is continuous on $\phi([a, b])$, then $\int_{\phi(a)}^{\phi(b)} f(x) \, dx = \int_a^b f(\phi(t))\phi'(t) \, dt$.

6

Topics of Integration

6.1 INTEGRATION OF RATIONAL FUNCTIONS

There is a general process for integrating rational functions. We illustrate the process by the following examples.

Example 6.1.1 To compute $\int \frac{dx}{(x+1)(x+2)(x+3)}$, we express the integrand as a sum of simpler quotients

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

Then

$$\int \frac{dx}{(x+1)(x+2)(x+3)} = \frac{1}{2} \ln|x+1| - \ln|x+2| + \frac{1}{2} \ln|x+3| + C = \frac{1}{2} \ln \left| \frac{(x+1)(x+3)}{(x+2)^2} \right| + C. \blacksquare$$

Example 6.1.2 To integrate $\frac{x^3-1}{x^3+1}$, we first divide the numerator by the denominator

$$\frac{x^3-1}{x^3+1} = 1 - \frac{2}{x^3+1}.$$

Then by $x^3+1 = (x+1)(x^2-x+1)$, we expect

$$\frac{2}{x^3+1} = \frac{A}{x+1} + \frac{Bx+C}{x^2-x+1}.$$

To find the coefficients A, B, C , we note that the above is the same as $2 = A(x^2-x+1) + (Bx+C)(x+1)$. By comparing the coefficients of powers of x , we get

$$A+B=0, \quad -A+B+C=0, \quad A+C=2.$$

The solution of the system is $A = \frac{2}{3}$, $B = -\frac{2}{3}$, $C = \frac{4}{3}$. Therefore

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

The last integral may be computed by “completing the square”

$$\begin{aligned} \int \frac{(x-2) dx}{x^2-x+1} &= \int \frac{(x-2) dx}{\left(x-\frac{1}{2}\right)^2 + \frac{3}{4}} = \int \frac{\frac{1}{2}(\sqrt{3}t-3)\frac{\sqrt{3}}{2} dt}{\frac{3}{4}t^2 + \frac{3}{4}} = \int \frac{(t-\sqrt{3}) dt}{t^2+1} \\ &= \int \frac{\frac{1}{2}d(t^2)}{t^2+1} + \sqrt{3} \int \frac{dt}{t^2+1} = \frac{1}{2} \ln(t^2+1) + \sqrt{3} \arctan t + C \\ &= \frac{1}{2} \ln(x^2-x+1) - \sqrt{3} \arctan \frac{2x-1}{\sqrt{3}} + C, \end{aligned}$$

where we used the substitution

$$t = \frac{2}{\sqrt{3}} \left(x - \frac{1}{2}\right), \quad x = \frac{1}{2}(\sqrt{3}t + 1)$$

in the computation. We conclude that

$$\int \frac{x^3 - 1}{x^3 + 1} dx = x - \frac{2}{3} \ln|x+1| + \frac{1}{3} \ln(x^2-x+1) - \frac{2}{\sqrt{3}} \arctan \frac{2x-1}{\sqrt{3}} + C. \quad \blacksquare$$

In general, consider a rational function $f(x) = \frac{p(x)}{q(x)}$. Like the division of integers, we may divide $p(x)$ by $q(x)$ to get

$$p(x) = h(x)q(x) + r(x),$$

where $h(x)$ is the quotient polynomial and $r(x)$ is the remainder polynomial. The important thing here is that the degree $r(x)$ is strictly smaller than the degree of $q(x)$. Then

$$\int f(x) dx = \int h(x) dx + \int \frac{r(x)}{q(x)} dx.$$

The integration $\int h(x) dx$ of a polynomial is easy to compute, so that the problem becomes the integration of the quotient $\frac{r(x)}{q(x)}$.

The Fundamental Theorem of Algebra tells us that any real polynomial $q(x)$ is a product of real linear and quadratic polynomials

$$q(x) = A(x+a_1)^{m_1} \cdots (x+a_k)^{m_k} (x^2+b_1x+c_1)^{n_1} \cdots (x^2+b_lx+c_l)^{n_l},$$

such that the quadratic factors do not have real roots

$$4c_j > b_j^2.$$

Then the division theory of polynomials tells us that, because the degree of $r(x)$ is strictly smaller than the degree of $q(x)$, the quotient $\frac{r(x)}{q(x)}$ can be written as a summation of quotients with $(x + a_i)^m$ or $(x^2 + b_jx + c_j)^n$ as denominators

$$\frac{r(x)}{q(x)} = \sum_{i=1}^k \left(\frac{A_{i1}}{x + a_i} + \cdots + \frac{A_{im_i}}{(x + a_i)^{m_i}} \right) + \sum_{j=1}^l \left(\frac{B_{j1}x + C_{j1}}{x^2 + b_jx + c_j} + \cdots + \frac{B_{jn_j}x + C_{jn_j}}{(x^2 + b_jx + c_j)^{n_j}} \right).$$

Therefore the computation of the integral is reduced to the following types

$$\int \frac{dx}{(x + a)^m}, \quad \int \frac{(Bx + C) dx}{(x^2 + bx + c)^n}.$$

The first type is easy to compute

$$\int \frac{dx}{(x + a)^m} = \begin{cases} \ln|x + a| + C, & \text{if } m = 1, \\ -\frac{1}{(m-1)(x + a)^{m-1}} + C, & \text{if } m > 1. \end{cases}$$

The second type is

$$\int \frac{(Bx + C) dx}{(x^2 + bx + c)^n} = \frac{B}{2} \int \frac{d(x^2 + bx + c)}{(x^2 + bx + c)^n} + \left(C - \frac{B}{2} \right) \int \frac{dx}{(x^2 + bx + c)^n}.$$

We have

$$\int \frac{d(x^2 + bx + c)}{(x^2 + bx + c)^n} = \begin{cases} \ln(x^2 + bx + c) + C, & \text{if } n = 1, \\ -\frac{1}{(n-1)(x^2 + bx + c)^{n-1}} + C, & \text{if } n > 1. \end{cases}$$

The second part may be computed by completing the square

$$x^2 + bx + c = \left(x + \frac{b}{2} \right)^2 + \frac{4c - b^2}{4} = D^2(t^2 + 1), \quad t = \frac{1}{D} \left(x + \frac{b}{2} \right), \quad D = \frac{\sqrt{4c - b^2}}{2}.$$

Then

$$\int \frac{dx}{(x^2 + bx + c)^n} = \frac{1}{D^{n-1}} \int \frac{dt}{(t^2 + 1)^n},$$

which may be computed from the recursive relation in Example 5.4.11.

Example 6.1.3 To integrate the rational function

$$f(x) = \frac{x^4 + x^3 + x^2 - 2x + 1}{x^5 + x^4 - 2x^3 - 2x^2 + x + 1},$$

we find the factorization $x^5 + x^4 - 2x^3 - 2x^2 + x + 1 = (x-1)^2(x+1)^3$ of the denominator and write

$$f(x) = \frac{A_1}{x-1} + \frac{A_2}{(x-1)^2} + \frac{B_1}{x+1} + \frac{B_2}{(x+1)^2} + \frac{B_3}{(x+1)^3}.$$

This is the same as

$$\begin{aligned} x^4 + x^3 + x^2 - 2x + 1 &= A_1(x-1)(x+1)^3 + A_2(x+1)^3 \\ &\quad + B_1(x-1)^2(x+1)^2 + B_2(x-1)^2(x+1) + B_3(x-1)^2. \end{aligned}$$

Then we get the following equations

$$\begin{aligned} 2 &= 8A_2, && \text{by taking } x = 1; \\ 4 &= 4B_3, && \text{by taking } x = -1; \\ 7 &= 8A_1 + 12A_2, && \text{by taking } \frac{d}{dx} \text{ at } x = 1; \\ -5 &= 4B_2 - 4B_3, && \text{by taking } \frac{d}{dx} \text{ at } x = -1; \\ 1 &= A_1 + A_2 + B_1 + B_2 + B_3, && \text{by computing the coefficient of } x^4. \end{aligned}$$

It is easy to solve the equations and get

$$\begin{aligned} \int f(x) \, dx &= \int \left(\frac{1}{2(x-1)} + \frac{1}{4(x-1)^2} + \frac{1}{2(x+1)} - \frac{1}{4(x+1)^2} + \frac{1}{(x+1)^3} \right) dx \\ &= \frac{1}{2} \ln|x-1| - \frac{1}{4(x-1)} + \frac{1}{2} \ln|x+1| + \frac{1}{4(x+1)} - \frac{1}{2(x+1)^2} + C \\ &= \frac{1}{2} \ln|x^2-1| - \frac{x}{(x-1)(x+1)^2} + C. \end{aligned} \quad \blacksquare$$

Example 6.1.4 To integrate the rational function $\frac{(x^2+1)^4}{(x^3-1)^2}$, we first divide $(x^2+1)^4$ by $(x^3-1)^2$ to get

$$\frac{(x^2+1)^4}{(x^3-1)^2} = x^2 + 4 + \frac{2x^5 + 6x^4 + 8x^3 + 3x^2 - 3}{(x^3-1)^2}.$$

By the factorization $x^3-1 = (x-1)(x^2+x+1)$, we write

$$\frac{2x^5 + 6x^4 + 8x^3 + 3x^2 - 3}{(x^3-1)^2} = \frac{A_1}{x-1} + \frac{A_2}{(x-1)^2} + \frac{B_1x + C_1}{x^2+x+1} + \frac{B_2x + C_2}{(x^2+x+1)^2}.$$

This is the same as

$$\begin{aligned} 2x^5 + 6x^4 + 8x^3 + 3x^2 - 3 &= A_1(x-1)(x^2+x+1)^2 + A_2(x^2+x+1)^2 \\ &\quad + (B_1x + C_1)(x-1)^2(x^2+x+1) + (B_2x + C_2)(x-1)^2. \end{aligned}$$

Then we get the following equations

$$\begin{array}{ll}
 -3 &= -A_1 + A_2 + C_1 + C_2, & \text{by taking } x = 0; \\
 16 &= 9A_2, & \text{by taking } x = 1; \\
 -4 &= -2A_1 + A_2 + 4(-B_1 + C_1) + 4(-B_2 + C_2), & \text{by taking } x = -1; \\
 2 &= A_1 + B_1, & \text{by computing the coefficient of } x^5; \\
 6 &= A_1 + A_2 - B_1 + C_1, & \text{by computing the coefficient of } x^4; \\
 32 &= 9A_1 + 18A_2, & \text{by taking } \frac{d}{dx} \text{ at } x = 1.
 \end{array}$$

The solution of the system is

$$A_2 = \frac{32}{9}, \quad A_1 = \frac{16}{9}, \quad B_1 = -\frac{14}{9}, \quad C_1 = -\frac{8}{9}, \quad B_2 = 0, \quad C_2 = -\frac{1}{3}.$$

Therefore

$$\begin{aligned}
 \int \frac{(x^2 + 1)^4}{(x^3 - 1)^2} dx &= \frac{1}{3}x^3 + 4x + \int \frac{32}{9(x-1)} dx + \int \frac{16}{9(x-1)^2} dx \\
 &\quad - \int \frac{14x + 8}{9(x^2 + x + 1)} dx - \int \frac{1}{3(x^2 + x + 1)^2} dx \\
 &= \frac{1}{3}x^3 + 4x + \frac{32}{9} \ln|x-1| - \frac{16}{9(x-1)} \\
 &\quad - \int \frac{7d(x^2 + x + 1) + dx}{9(x^2 + x + 1)} - \int \frac{dx}{3(x^2 + x + 1)^2} \\
 &= \frac{1}{3}x^3 + 4x + \frac{32}{9} \ln|x-1| - \frac{16}{9(x-1)} - \frac{7}{9} \ln(x^2 + x + 1) \\
 &\quad - \int \frac{dx}{9(x^2 + x + 1)} - \int \frac{dx}{3(x^2 + x + 1)^2}.
 \end{aligned}$$

By $x^2 + x + 1 = \left(x + \frac{1}{2}\right)^2 + \frac{3}{4}$, we introduce

$$t = \frac{2}{\sqrt{3}} \left(x + \frac{1}{2}\right) = \frac{2x+1}{\sqrt{3}}, \quad x = \frac{1}{2}(\sqrt{3}t - 1).$$

Then by the computation of $\int \frac{dt}{(t^2 + 1)^2}$ in Example 5.4.11,

$$\int \frac{dx}{x^2 + x + 1} = \int \frac{\frac{\sqrt{3}}{2} dt}{\frac{3}{4}(t^2 + 1)} = \frac{2}{\sqrt{3}} \arctan t + C = \frac{2}{\sqrt{3}} \arctan \frac{2x+1}{\sqrt{3}} + C,$$

$$\begin{aligned}
\int \frac{dx}{(x^2 + x + 1)^2} &= \int \frac{\frac{\sqrt{3}}{2} dt}{\frac{3^2}{4^2}(t^2 + 1)^2} = \frac{8}{3\sqrt{3}} \left(\frac{1}{2} \arctan t + \frac{t}{2(t^2 + 1)} \right) + C \\
&= \frac{4}{3\sqrt{3}} \arctan \frac{2x+1}{\sqrt{3}} + \frac{2x+1}{3(x^2 + x + 1)} + C.
\end{aligned}$$

Combining everything and substituting x back, we get

$$\begin{aligned}
\int \frac{(x^2 + 1)^4}{(x^3 - 1)^2} dx &= \frac{1}{3}x^3 + 4x + \frac{32}{9} \ln|x-1| - \frac{16}{9(x-1)} - \frac{7}{9} \ln(x^2 + x + 1) \\
&\quad - \frac{2}{3\sqrt{3}} \arctan \frac{2x+1}{\sqrt{3}} - \frac{2x+1}{9(x^2 + x + 1)} + C. \\
&= \frac{1}{3}x^3 + 4x - \frac{6x^2 + 5x + 5}{3(x^3 - 1)} + \frac{13}{3} \ln|x-1| - \frac{7}{9} \ln|x^3 - 1| \\
&\quad - \frac{2}{3\sqrt{3}} \arctan \frac{2x+1}{\sqrt{3}} + C. \quad \blacksquare
\end{aligned}$$

Integrating $\int R\left(x, \sqrt[n]{\frac{ax+b}{cx+d}}\right) dx$

If $R(x, y)$ is a rational function in x, y , then $\int R\left(x, \sqrt[n]{\frac{ax+b}{cx+d}}\right) dx$ can be changed to an integral of rational functions.

Introduce $t = \sqrt[n]{\frac{ax+b}{cx+d}}$. Then

$$x = \frac{dt^n - b}{-ct^n + a}, \quad \frac{dx}{dt} = \frac{n(ad - bc)t^{n-1}}{(ct^n - a)^2},$$

and

$$\int R\left(x, \sqrt[n]{\frac{ax+b}{cx+d}}\right) dx = n(ad - bc) \int R\left(\frac{dt^n - b}{-ct^n + a}, t\right) \frac{t^{n-1}}{(ct^n - a)^2} dt.$$

The right side is the integration of a rational function.

Example 6.1.5 By $t = \sqrt[3]{x+1}$, we have

$$\begin{aligned}
\int \frac{(x-1) dx}{x \sqrt[3]{x+1}} &= \int \frac{(t^3 - 1) - 1}{(t^3 - 1)t} 3t^2 dt = \int \frac{3t(t^3 - 2)}{t^3 - 1} dt \\
&= \int \left(3t - \frac{1}{t-1} + \frac{t-1}{t^2 + t + 1} \right) dt \\
&= \frac{3}{2}t^2 - \ln|t-1| + \frac{1}{2} \int \frac{d(t^2 + t + 1)}{t^2 + t + 1} - \frac{3}{2} \int \frac{dt}{t^2 + t + 1}
\end{aligned}$$

$$\begin{aligned}
&= \frac{3}{2}t^2 - \ln|t-1| + \frac{1}{2}\ln|t^2+t+1| - \frac{3}{2}\int \frac{d\left(t-\frac{1}{2}\right)}{\left(t-\frac{1}{2}\right)^2 + \frac{3}{4}} \\
&= \frac{3}{2}t^2 - \ln|t-1| + \frac{1}{2}\ln\left|\frac{t^3-1}{t-1}\right| - \frac{3}{2}\frac{2}{\sqrt{3}}\arctan\frac{2}{\sqrt{3}}\left(t-\frac{1}{2}\right) + C \\
&= \frac{3}{2}\sqrt[3]{(x+1)^2} + \frac{1}{2}\ln|x| - \frac{3}{2}\ln|\sqrt[3]{x+1}-1| - \sqrt{3}\arctan\frac{2\sqrt[3]{x+1}-1}{\sqrt{3}} + C. \quad \blacksquare
\end{aligned}$$

Example 6.1.6 To compute

$$\int \frac{2x+3}{\sqrt{x^2+x}} dx = \int \frac{2x+3}{|x|} \sqrt{\frac{x}{x+1}} dx,$$

we introduce

$$t = \sqrt{\frac{x}{x+1}}, \quad x = \frac{t^2}{1-t^2}, \quad \frac{dx}{dt} = \frac{2t}{(1-t^2)^2}$$

For $x > 0$, we get

$$\begin{aligned}
\int \frac{2x+3}{\sqrt{x^2+x}} dx &= \int \left(2 + 3\frac{1-t^2}{t^2}\right) t \frac{2t}{(1-t^2)^2} dt = \int \frac{2(3-t^2)}{(1-t^2)^2} dt \\
&= \int \left(\frac{2}{t+1} - \frac{2}{t-1} + \frac{1}{(t+1)^2} + \frac{1}{(t-1)^2}\right) dt \\
&= 2\ln|t+1| - 2\ln|t-1| - \frac{1}{t+1} - \frac{1}{t-1} + C \\
&= 2\ln\left|\frac{t+1}{t-1}\right| - \frac{2t}{t^2-1} + C \\
&= 2\sqrt{x^2+x} + 2\ln|2x+1+2\sqrt{x^2+x}| + C.
\end{aligned}$$

For $x < -1$, we have the same computation for the integration in terms of t , except adding a negative sign. After substituting x back and taking into account of the sign, we still get

$$\int \frac{2x+3}{\sqrt{x^2+x}} dx = 2\sqrt{x^2+x} + 2\ln|2x+1+2\sqrt{x^2+x}| + C. \quad \blacksquare$$

Integrating $\int R(\sin x, \cos x) dx$

For a rational function $R(x, y)$, the integral $\int R(\sin x, \cos x) dx$ can also be reduced to an integral of rational functions.

Introduce $t = \tan \frac{x}{2}$. Then

$$\sin x = \frac{2t}{1+t^2}, \quad \cos x = \frac{1-t^2}{1+t^2}, \quad \frac{dx}{dt} = \frac{2}{1+t^2}.$$

Therefore

$$\int R(\sin x, \cos x) dx = \int R\left(\frac{2t}{1+t^2}, \frac{1-t^2}{1+t^2}\right) \frac{2}{1+t^2} dt.$$

Example 6.1.7 By $t = \tan \frac{x}{2}$, we have

$$\int \frac{dx}{1 + \sin x + \cos x} = \int \frac{\frac{2}{1+t^2} dt}{1 + \frac{2t}{1+t^2} + \frac{1-t^2}{1+t^2}} = \int \frac{dt}{1+t} = \ln|1+t| + C = \ln\left|1 + \frac{x}{2}\right| + C.$$

We also have

$$\begin{aligned} & \int \frac{\sin x \cos x dx}{\sin x + \cos x} \\ &= \int \frac{dx}{\frac{1}{\sin x} + \frac{1}{\cos x}} = \int \frac{\frac{2}{1+t^2} dt}{\frac{1}{\frac{2t}{1+t^2}} + \frac{1}{\frac{1-t^2}{1+t^2}}} = \int \frac{2t(t^2-1) dt}{(t^2+1)(t^2-2t-1)} \\ &= \int \left(\frac{t+1}{t^2+1} + \frac{\sqrt{2}+1}{2} \frac{1}{t-\sqrt{2}-1} - \frac{\sqrt{2}-1}{2} \frac{1}{t+\sqrt{2}-1} \right) dt \\ &= \frac{1}{2} \ln(t^2+1) + \arctan t + \frac{\sqrt{2}+1}{2} \ln|t-\sqrt{2}-1| - \frac{\sqrt{2}-1}{2} \ln|t+\sqrt{2}-1| + C \\ &= \frac{1}{2} \ln\left(\tan^2 \frac{x}{2} + 1\right) + 2x + \frac{\sqrt{2}+1}{2} \ln\left|\tan \frac{x}{2} - \sqrt{2} - 1\right| - \frac{\sqrt{2}-1}{2} \ln\left|\tan \frac{x}{2} + \sqrt{2} - 1\right| + C. \quad \blacksquare \end{aligned}$$

If $R(x, -y) = -R(x, y)$, then the substitution $t = \sin x$ can be used to change $\int R(\sin x, \cos x) dx$ to the integral of a rational function. If $R(-x, y) = -R(x, y)$, then the substitution $t = \cos x$ can be used to change $\int R(\sin x, \cos x) dx$ to the integral of a rational function. If $R(-x, -y) = R(x, y)$, then the substitution $t = \tan x$ can be used to change $\int R(\sin x, \cos x) dx$ to the integral of a rational function.

Example 6.1.8 One may try to use $t = \tan \frac{x}{2}$ to compute the integral $\int \frac{4 \sin x + 3 \cos x}{\sin x + \cos x} dx$. Noticing the integrand satisfies $R(-x, -y) = R(x, y)$, we may introduce $t = \tan x$ to get

$$\int \frac{4 \sin x + 3 \cos x}{\sin x + \cos x} dx = \int \frac{4 \tan x + 3}{\tan x + 1} dx = \int \frac{(4t+3) dt}{(t+1)(t^2+1)}$$

$$\begin{aligned}
&= \int \left(-\frac{1}{2(t+1)} + \frac{t+7}{2(t^2+1)} \right) dt \\
&= -\frac{1}{2} \ln|t+1| + \frac{1}{4} \ln(t^2+1) + \frac{7}{2} \arctan t + C \\
&= -\frac{1}{2} \ln|\tan x + 1| + \frac{1}{4} \ln(\tan^2 x + 1) + \frac{7}{2}x + C \\
&= \frac{7}{2}x + \frac{1}{2} \ln \left| \frac{\sec x}{\tan x + 1} \right| + C = \frac{7}{2}x - \frac{1}{2} \ln|\sin x + \cos x| + C. \quad \blacksquare
\end{aligned}$$

Exercises

6.1.1 Compute the following integrals:

- | | |
|---|--|
| 1. $\int \frac{x^2 dx}{1+x};$ | 13. $\int \frac{dx}{1+\sqrt{x}};$ |
| 2. $\int \frac{(1+x)^2 dx}{1+x^2};$ | 14. $\int \frac{dx}{\sqrt{x}(1+x)};$ |
| 3. $\int \frac{(2-x)^2 dx}{2-x^2};$ | 15. $\int \frac{x\sqrt[3]{2+x}}{x+\sqrt[3]{2+x}} dx;$ |
| 4. $\int \frac{dx}{x^2-x-2};$ | 16. $\int \frac{x^3}{\sqrt[3]{x^2+1}} dx;$ |
| 5. $\int \frac{dx}{(x+a)^2(x+b)^2};$ | 17. $\int \frac{dx}{\sqrt[3]{x} + \sqrt[4]{x}} dx;$ |
| 6. $\int \frac{dx}{(x+1)(x^2+1)};$ | 18. $\int \frac{dx}{\sqrt[4]{x^3(a-x)}};$ |
| 7. $\int \frac{dx}{x^3+1};$ | 19. $\int \frac{\sqrt{x+1} + \sqrt{x-1}}{\sqrt{x+1} - \sqrt{x-1}} dx;$ |
| 8. $\int \frac{x^2 dx}{1-x^4};$ | 20. $\int \frac{1-r^2}{1-2r \cos x + r^2} dx, \text{ for } 0 < r < 1;$ |
| 9. $\int \frac{dx}{x^4+x^2+1};$ | 21. $\int \frac{dx}{a+\sin x};$ |
| 10. $\int \frac{dx}{(x^2+a^2)(x^2+b^2)};$ | 22. $\int \frac{dx}{\cos x + \tan x};$ |
| 11. $\int \frac{x^3 dx}{(x^2+1)^2};$ | 23. $\int \frac{dx}{\sin x + \tan x};$ |
| 12. $\int \frac{dx}{(x^2+4x+6)^2};$ | |

$$24. \int \frac{dx}{2 \sin x + \sin 2x};$$

$$25. \int \frac{\sin^2 x}{1 + \sin^2 x} dx;$$

$$26. \int \frac{dx}{\sin(x+a) \sin(x+b)};$$

$$27. \int \frac{dx}{(1 + \cos^2 x)(2 + \sin^2 x)};$$

$$28. \int \frac{1 - \tan x}{1 + \tan x} dx;$$

$$29. \int \frac{dx}{\sin^2 x \cos^2 x}.$$

6.2 NUMERICAL INTEGRATION

Many integrals cannot be expressed as combinations of elementary functions. Some examples are

$$\int \sqrt{x^3 + 1} dx, \quad \int e^{x^2} dx, \quad \int \sin x^2 dx, \quad \int \sqrt{1 - k^2 \sin^2 x} dx \quad (0 < k < 1).$$

Even when the integrals can be expressed in terms of elementary functions, the formula may be so complicated that evaluating the formula becomes very difficult. For many practical purposes, an approximate value of the Riemann integral is enough, and numerical computations is often much more efficient than the computation of the antiderivatives.

Rectangular Rule

The Riemann integral is the limit of the Riemann sum as the size of the partition approaches 0. In particular, by evenly dividing the interval $[a, b]$ into n parts, we get a partition P_n with partition points $x_i = a + ih$, where $h = \frac{b-a}{n}$ is the **step size**. Denote

$$y_i = f(x_i) = f(a + ih).$$

Then the Riemann sum with the right end sample points $x_i^* = x_i$ is

$$R_n = h(y_1 + y_2 + \cdots + y_n),$$

and the Riemann sum with the left end sample points $x_i^* = x_{i-1}$ is

$$L_n = h(y_0 + y_1 + \cdots + y_{n-1}).$$

These **rectangular rules** give approximations of the Riemann integral $\int_a^b f(x) dx$.

Trapezoidal and Midpoint Rules

A more efficient scheme for approximating the Riemann integral is to take the average of the two rectangular rules

$$T_n = \frac{R_n + L_n}{2} = \frac{h}{2}(y_0 + 2y_1 + 2y_2 + \cdots + 2y_{n-1} + y_n).$$

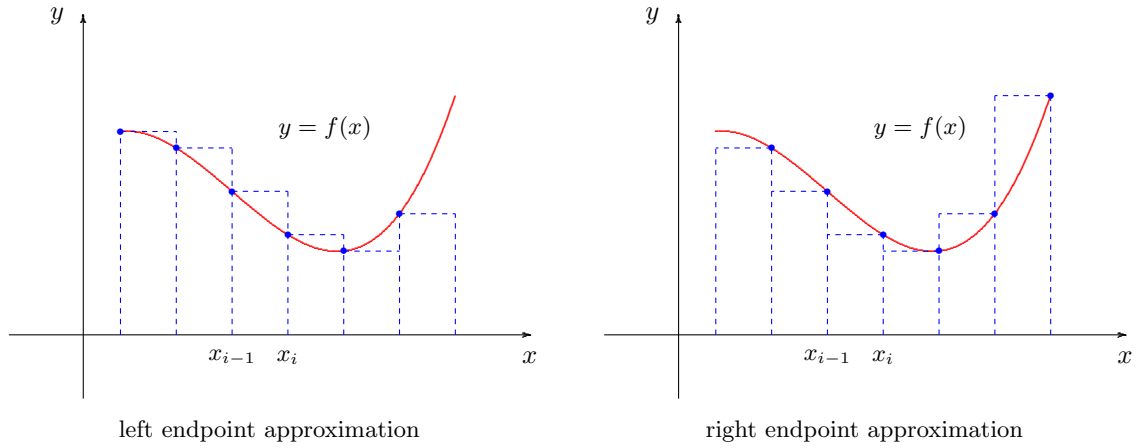


Fig. 6.1 Rectangular rules

This is called the **trapezoidal rule** because it is the sum of the trapezoids over the partition intervals.

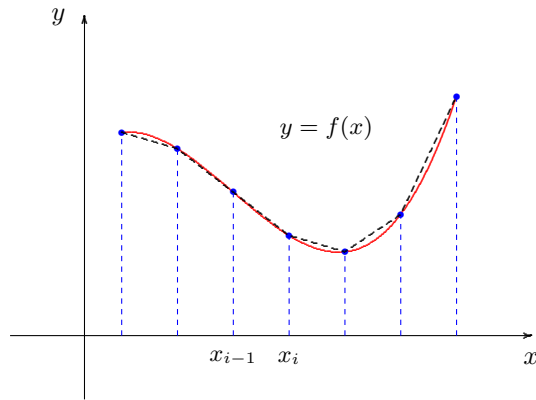


Fig. 6.2 Trapezoidal rule

Another way of improving the left and right end sample points is to take the middle points as the sample points. This gives the **midpoint rule**

$$M_n = h(\bar{y}_1 + \bar{y}_2 + \cdots + \bar{y}_n), \quad \bar{y}_i = f(\bar{x}_i), \quad \bar{x}_i = \frac{x_{i-1} + x_i}{2} = a + \frac{2i-1}{2}h.$$

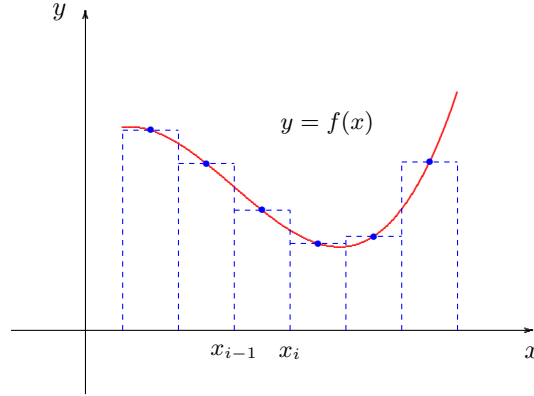


Fig. 6.3 Midpoint rule

Example 6.2.1 We use various rules to estimate the integral $\int_0^1 \frac{dx}{1+x^2}$. Take $n = 4$, $h_4 = 0.25$. The partition is

$$P_4: \quad 0 < 0.25 < 0.5 < 0.75 < 1.$$

The values at the partition points are

$$y_0 = 1.000000, \quad y_1 = 0.941176, \quad y_2 = 0.800000, \quad y_3 = 0.640000, \quad y_4 = 0.500000.$$

The middle points are

$$0.125 < 0.375 < 0.625 < 0.875,$$

and the values at the middle points are

$$\bar{y}_1 = 0.984615, \quad \bar{y}_2 = 0.876712, \quad \bar{y}_3 = 0.719101, \quad \bar{y}_4 = 0.566372.$$

Then we get the approximate values of $\int_0^1 \frac{dx}{1+x^2}$ according to various rules

$$R_4 = 0.25 \times (0.941176 + 0.800000 + 0.640000 + 0.500000) \approx 0.720294,$$

$$L_4 = 0.25 \times (1.000000 + 0.941176 + 0.800000 + 0.640000) \approx 0.845294,$$

$$\begin{aligned} T_4 &= \frac{0.25}{2} \times (1.000000 + 2 \times 0.941176 + 2 \times 0.800000 + 2 \times 0.640000 + 0.500000) \\ &\approx 0.782794, \end{aligned}$$

$$M_4 = 0.25 \times (0.984615 + 0.876712 + 0.719101 + 0.566372) \approx 0.786700.$$

If we double the number of partition points by taking $n = 8$, then $h_8 = 0.125$, and we have

i	x_i	y_i	$\bar{x}_i$	$\bar{y}_i$
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

$$R_8 \approx 0.753497, \quad L_8 \approx 0.815997, \quad T_8 \approx 0.784747, \quad M_8 \approx 0.785721. \quad \blacksquare$$

We actually know the exact value of the integral in the example

$$I = \int_0^1 \frac{dx}{1+x^2} = \frac{\pi}{4} \approx 0.785398.$$

We may compare the approximate values with the actual value.

error	$n = 4$	$n = 8$
$I - R_n$	0.065104	0.031901
$I - L_n$	-0.059896	-0.030599
$I - T_n$	0.002604	0.000651
$I - M_n$	-0.001302	-0.000323

We can make the following observations from the table.

1. All rules produce better approximation as n increases.
2. The errors in the rectangular rules decrease by a factor of 2 when n is doubled.
3. The errors in the trapezoidal and midpoint rules decrease by a factor of 4 when n is doubled.
4. The error in the midpoint rule is about half of the error in the trapezoidal rule.

The following gives the estimation of the error in the trapezoidal and midpoint rules.

Theorem 6.2.1 Suppose $f''(x)$ is continuous and bounded by K_2 on $[a, b]$, then

$$\left| \int_a^b f(x) dx - T_n \right| \leq \frac{K_2(b-a)^3}{12n^2}, \quad \left| \int_a^b f(x) dx - M_n \right| \leq \frac{K_2(b-a)^3}{24n^2}.$$

Proof. We prove the estimation for the error of the midpoint rule. First of all, consider the Taylor expansion of the function $f(x)$ at the middle point $c = \frac{a+b}{2}$ of $[a, b]$

$$f(x) = f(c) + f'(c)(x-c) + R(x).$$

Since $f(x)$ has continuous second order derivative, by a later result, the remainder $R(x) = \frac{f''(y)}{2}(x-c)^2$ for some y between x and c . In particular,

$$|f(x) - f(c) - f'(c)(x-c)| = \frac{|f''(y)|}{2}(x-c)^2 \leq \frac{K}{2}(x-c)^2.$$

This implies

$$\left| \int_a^b (f(x) - f(c) - f'(c)(x-c)) \, dx \right| \leq \int_a^b |f(x) - f(c) - f'(c)(x-c)| \, dx \leq \frac{K}{2} \int_a^b (x-c)^2 \, dx.$$

Since

$$\begin{aligned} \int_a^b (f(x) - f(c) - f'(c)(x-c)) \, dx &= \int_a^b f(x) \, dx - f(c)(b-a) - f'(c) \frac{(b-c)^2 - (a-c)^2}{2} \\ &= \int_a^b f(x) \, dx - f\left(\frac{a+b}{2}\right)(b-a), \end{aligned}$$

and

$$\frac{K}{2} \int_a^b (x-c)^2 \, dx = \frac{K}{2} \frac{(b-c)^3 - (a-c)^3}{3} = \frac{K}{24}(b-a)^3,$$

we conclude that

$$\left| \int_a^b f(x) \, dx - f\left(\frac{a+b}{2}\right)(b-a) \right| \leq \frac{K}{24}(b-a)^3.$$

Now in the midpoint rule, we apply the inequality to each interval $[x_{i-1}, x_i]$ to get

$$\left| \int_{x_{i-1}}^{x_i} f(x) \, dx - f(\bar{x}_i)h \right| \leq \frac{K}{24}h^3 = \frac{K(b-a)^3}{24n^3}.$$

Adding all the inequalities together, we get

$$\left| \int_a^b f(x) \, dx - M_n \right| \leq \sum_{i=1}^n \left| \int_{x_{i-1}}^{x_i} f(x) \, dx - f(\bar{x}_i)h \right| \leq n \frac{K(b-a)^3}{24n^3} = \frac{K(b-a)^3}{24n^2}. \quad \blacksquare$$

Simpson's Rule

The trapezoidal rule approximates the function on the interval $[x_{i-1}, x_i]$ by connecting the two ends of the interval by a straight line. Therefore the rule is obtained by linear approximation. We expect quadratic approximation to give even better approximation to the integral.

A quadratic curve is determined by three points. Therefore the “quadratic rule” will approximate $f(x)$ on the interval $[x_{i-1}, x_{i+1}]$ by a function $Q(x) = A(x-x_i)^2 + B(x-x_i) + C$ satisfying

$$\begin{aligned} y_{i-1} &= f(x_{i-1}) = Q(x_{i-1}) = Ah^2 - Bh + C, \\ y_i &= f(x_i) = Q(x_i) = C, \\ y_{i+1} &= f(x_{i+1}) = Q(x_{i+1}) = Ah^2 + Bh + C. \end{aligned}$$

The integral of $f(x)$ on $[x_{i-1}, x_{i+1}]$ is then approximated by

$$\int_{x_{i-1}}^{x_{i+1}} Q(x) \, dx = \int_{-h}^h (At^2 + Bt + C) dt = \frac{2}{3}Ah^3 + 2Ch = \frac{h}{3}(y_{i-1} + 4y_i + y_{i+1}).$$

Now suppose n is even, then we may apply quadratic approximations to $[x_0, x_2]$, $[x_2, x_4]$, $\dots$, $[x_{n-2}, x_n]$. Adding such approximations together, we get an approximation of $\int_a^b f(x) \, dx$

$$S_n = \frac{h}{3}(y_0 + 4y_1 + 2y_2 + 4y_3 + 2y_4 + \dots + 2y_{n-2} + 4y_{n-1} + y_n).$$

This is **Simpson's rule**.

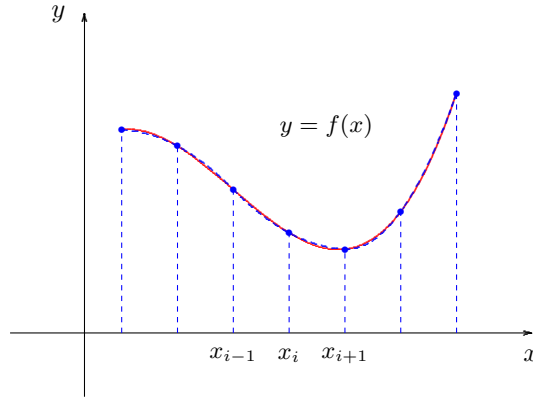


Fig. 6.4 Simpson's rule

Example 6.2.2 Applying Simpson's rule to $\int_0^1 \frac{dx}{1+x^2}$ for $n = 4$, we get

$$S_4 = \frac{0.25}{3} \times (1.000000 + 4 \times 0.941176 + 2 \times 0.800000 + 4 \times 0.640000 + 0.500000) \approx 0.785392.$$

The error $I - S_n = 0.000540$ is comparable to the trapezoidal and midpoint rule at $n = 8$ and is much better than all the rules at $n = 4$. ■

The errors in Simpson's rule is estimated as follows.

Theorem 6.2.2 Suppose $f^{(4)}(x)$ is continuous and bounded by K_4 on $[a, b]$, then

$$\left| \int_a^b f(x) \, dx - S_n \right| \leq \frac{K_4(b-a)^5}{180n^4}.$$

Example 6.2.3 We try to determine the number of partition points needed in order for the Simpson's rule to produce an approximate value of $\int_0^1 \frac{dx}{1+x^2}$ accurate up to the 6-th digit.

We have

$$f^{(4)}(x) = \frac{24(5x^4 - 10x^2 + 1)}{(1+x^2)^5}, \quad f^{(5)}(x) = -\frac{240x(x^2-3)(3x^2-1)}{(1+x^2)^6}.$$

From $f^{(5)}(x)$, we see that the maximum of $|f^{(4)}(x)|$ on $[0, 1]$ can only be reached at 0, $\frac{1}{\sqrt{3}}$ 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

$$\frac{24}{180n^4} \leq 10^{-6}.$$

Therefore $n \geq 19.1$. Considering n should be an even integer, we need $n \geq 20$.

Similar estimation can be made for the trapezoidal and midpoint rules. We can find $K_2 = |f''(0)| = 2$. Then the question respectively becomes

$$\frac{2}{12n^2} \leq 10^{-6}, \quad \frac{2}{24n^2} \leq 10^{-6}.$$

The answer are respectively $n \geq 409$ and $n \geq 289$. ■

Exercises

6.2.1 Use the trapezoidal, midpoint and Simpson's rules to evaluate the following integrals:

- | | |
|--|---|
| 1. $\int_1^2 \frac{dx}{x}$, with $n = 4$; | 4. $\int_0^\pi \cos x^2 dx$, with $n = 8$; |
| 2. $\int_0^\pi x \sin x dx$, with $n = 5$; | 5. $\int_0^{\pi/2} \sqrt{1 - \frac{1}{4} \sin^2 x} dx$, with $n = 6$; |
| 3. $\int_0^\pi \sin x dx$, with $n = 6$; | 6. $\int_0^{\pi/2} \frac{\sin x}{x} dx$, with $n = 8$. |

6.2.2 Determine the number of partition points needed to get an approximate value of the following integrals accurate up to 10^{-4} , by using trapezoidal, midpoint and Simpson's rules:

$$1. \int_1^2 \frac{dx}{x};$$

$$4. \int_0^\pi \cos x^2 dx;$$

$$2. \int_0^\pi x \sin x dx;$$

$$5. \int_0^{\pi/2} \sqrt{1 - \frac{1}{4} \sin^2 x} dx;$$

$$3. \int_0^\pi \sin x dx;$$

$$6. \int_0^{\pi/2} \frac{\sin x}{x} dx.$$

6.3 IMPROPER INTEGRAL

The definition of Riemann integral requires that the domain of integration is a bounded interval and the integrand function is a bounded function. However, in many applications, it is necessary to consider the integral over unbounded intervals or the integral of unbounded functions. In such cases, we get **improper integrals**.

6.3.1 Integral on Unbounded Interval

The integral on unbounded interval can be computed by using the bounded intervals to approximate the unbounded interval.

Example 6.3.1 To integrate e^{-x} on the unbounded interval $[0, +\infty)$, 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

$$\lim_{b \rightarrow +\infty} \int_0^b e^{-x} dx = \lim_{b \rightarrow +\infty} (1 - e^{-b}) = 1.$$

Therefore the improper integral $\int_0^{+\infty} e^{-x} dx$ 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 6.3.2 To compute the improper integral $\int_1^{+\infty} \frac{dx}{x}$, we first compute the integral on a bounded interval

$$\int_1^b \frac{dx}{x} = \ln b.$$

Since $\lim_{b \rightarrow +\infty} \ln b = +\infty$, we may write $\int_1^{+\infty} \frac{dx}{x} = +\infty$, and the improper integral diverges. ■

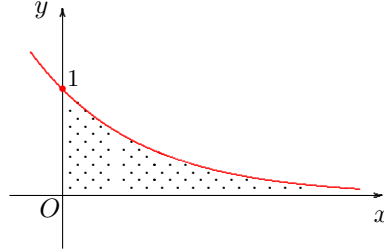


Fig. 6.5 The unbounded region has area 1.

Example 6.3.3 To compute the improper integral $\int_{-\infty}^{+\infty} \frac{dx}{x^2 + 1}$, we consider the integral on a bounded interval

$$\int_a^b \frac{dx}{x^2 + 1} = \arctan b - \arctan a.$$

Since $\lim_{a \rightarrow -\infty} \arctan a = -\frac{\pi}{2}$ and $\lim_{b \rightarrow +\infty} \arctan b = \frac{\pi}{2}$, the improper integral converges to the value

$$\int_{-\infty}^{+\infty} \frac{dx}{x^2 + 1} = \lim_{a \rightarrow -\infty, b \rightarrow +\infty} \int_a^b \frac{dx}{x^2 + 1} = \pi. \quad \blacksquare$$

In general, assume $f(x)$ is integrable on $[a, b]$ for any large b . Then we may consider the **improper integral** $\int_a^{+\infty} f(x) dx$. If the limit

$$\lim_{b \rightarrow +\infty} \int_a^b f(x) dx$$

converges, then we say the improper integral **converges** and has value

$$\int_a^{+\infty} f(x) dx = \lim_{b \rightarrow +\infty} \int_a^b f(x) dx.$$

Otherwise we say the improper integral **diverges**.

Similar discussion applies to the improper integral $\int_{-\infty}^a f(x) dx$. Moreover, if $f(x)$ is Riemann integrable on any bounded interval, then the improper integral $\int_{-\infty}^{+\infty} f(x) dx$ converges if and only if both $\int_{-\infty}^a f(x) dx$ and $\int_a^{+\infty} f(x) dx$ converge, and the value of the integral is

$$\int_{-\infty}^{+\infty} f(x) dx = \int_{-\infty}^a f(x) dx + \int_a^{+\infty} f(x) dx.$$

Example 6.3.4 For $a > 0$, consider the improper integral $\int_a^{-\infty} x^p dx$. If $p \neq -1$, then we have

$$\int_a^b x^p dx = \frac{b^{p+1} - a^{p+1}}{p+1}.$$

As $b \rightarrow +\infty$, the above converges if and only if $p+1 < 0$, and the limit is $-\frac{a^{p+1}}{p+1}$. Combined with the case $p = -1$ studied in Example 6.3.2, we see that

$$\int_a^{-\infty} x^p dx = -\frac{a^{p+1}}{p+1} \text{ if } p < -1,$$

and the improper integral diverges when $p \geq -1$. ■

Example 6.3.5 Since

$$\lim_{b \rightarrow +\infty} \int_0^b \cos x dx = \lim_{b \rightarrow +\infty} \sin b$$

diverges, the improper integral $\int_0^{+\infty} \cos x dx$ diverges. ■

Example 6.3.6 To compute the improper integral $\int_{-\infty}^0 xe^x dx$, we start with integration by parts on a bounded interval

$$\int_b^0 xe^x dx = \int_b^0 x de^x = -be^b - \int_b^0 e^x dx = -be^b - 1 + e^b.$$

Taking $b \rightarrow -\infty$ on both sides, we get

$$\int_{-\infty}^0 xe^x dx = -1. \quad \blacksquare$$

The example shows that the integration by parts can be extended to improper integrals

$$\int_a^{+\infty} f(x)g'(x) dx = f(+\infty)g(+\infty) - f(a)g(a) - \int_a^{+\infty} f'(x)g(x) dx.$$

The equality means that if the improper integral $\int_a^{+\infty} f'(x)g(x) dx$ converges, and

$$f(+\infty)g(+\infty) = \lim_{x \rightarrow +\infty} f(x)g(x)$$

converge, then the improper integral $\int_a^{+\infty} f(x)g'(x) dx$ also converges and can be computed by the integration by parts formula.

Example 6.3.7 For $a > 1$, consider the improper integral $\int_a^{+\infty} \frac{dx}{x(\ln x)^p}$. We have

$$\int_a^b \frac{dx}{x(\ln x)^p} = \int_a^b \frac{d(\ln x)}{(\ln x)^p} = \int_{\ln a}^{\ln b} \frac{dx}{x^p}.$$

Taking $b \rightarrow +\infty$ on both sides, we get

$$\int_a^{+\infty} \frac{dx}{x(\ln x)^p} = \int_{\ln a}^{+\infty} \frac{dx}{x^p}.$$

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 6.3.4, we see that the improper integral $\int_a^{+\infty} \frac{dx}{x(\ln x)^p}$ converges if and only if $p < -1$, and

$$\int_a^{+\infty} \frac{dx}{x(\ln x)^p} = -\frac{(\ln a)^{p+1}}{p+1} \text{ if } p < -1. \quad \blacksquare$$

The example shows that the change of variable can also be extended to improper integrals

$$\int_{\phi(a)}^{+\infty} f(\phi(x))\phi'(x) dx = \int_a^{+\infty} f(x) dx$$

as long as $\lim_{x \rightarrow +\infty} \phi(x) = +\infty$.

Of course the arithmetic property can also be extended to improper integrals

$$\int_a^{+\infty} (f(x) + g(x)) dx = \int_a^{+\infty} f(x) dx + \int_a^{+\infty} g(x) dx, \quad \int_a^{+\infty} cf(x) dx = c \int_a^{+\infty} f(x) dx.$$

The improper integral is defined by taking the limits. Therefore there is always the problem of convergence, even for many improper integrals that we cannot evaluate.

The convergence of the improper integral $\int_a^{+\infty} f(x) dx$ means the convergence of

$$I(b) = \int_a^b f(x) dx$$

as $b \rightarrow +\infty$. The Cauchy criterion for the convergence of $\lim_{b \rightarrow +\infty} I(b)$ is the following: for any $\epsilon > 0$, there is N , such that

$$b, c > N \implies |I(c) - I(b)| < \epsilon.$$

Since

$$I(c) - I(b) = \int_a^c f(x) dx - \int_a^b f(x) dx = \int_b^c f(x) dx,$$

we conclude the following criterion for the convergence of improper integral.

Proposition 6.3.1 *The improper integral $\int_a^{+\infty} f(x) dx$ converges if and only if for any $\epsilon > 0$, there is N , such that*

$$b, c > N \implies \left| \int_b^c f(x) dx \right| < \epsilon.$$

In the special case that $f(x) \geq 0$ (or $f(x) \geq 0$ for sufficiently large x), the integral $I(b) = \int_a^b f(x) dx$ is a nondecreasing function of b . Therefore the convergence of $\int_a^{+\infty} f(x) dx$ is the same as the boundedness of $I(b)$, and the divergence is the same as $\lim_{b \rightarrow +\infty} I(b) = +\infty$. Thus the convergence of the improper integral can also be expressed by the inequality

$$\int_a^{+\infty} f(x) dx < +\infty.$$

Theorem 6.3.2 (Comparison Test) *If $|f(x)| \leq g(x)$ for sufficiently large x and $\int_a^{+\infty} g(x) dx$ converges, then $\int_a^{+\infty} f(x) dx$ also converges.*

We say $\int_a^{+\infty} f(x) dx$ **absolutely converges** if $\int_a^{+\infty} |f(x)| dx$ converges. Under the assumption for the comparison test, we actually also know that $f(x)$ is absolutely integrable. Moreover, by taking $g(x) = |f(x)|$ in the comparison test, we know that absolute convergence implies the convergence.

Proof. The convergence of $\int_a^{+\infty} g(x) dx$ implies that $\int_a^{+\infty} g(x) dx$ satisfies the Cauchy criterion in Proposition 6.3.1. Therefore for any $\epsilon > 0$, we have N , such that

$$c > b > N \implies \int_b^c g(x) dx < \epsilon.$$

The assumption $|f(x)| \leq g(x)$ implies that for $c > b$,

$$\left| \int_b^c f(x) dx \right| \leq \int_b^c |f(x)| dx \leq \int_b^c g(x) dx.$$

Combined with the implication above, we get

$$c > b > N \implies \left| \int_b^c f(x) dx \right| \leq \int_b^c g(x) dx < \epsilon.$$

This verifies the Cauchy criterion for the convergence of $\int_a^{+\infty} f(x) dx$. ■

Example 6.3.8 The improper integrals $\int_0^{+\infty} \frac{x+1}{x^3-2x+3} dx$, $\int_1^{+\infty} \frac{\ln x}{x^2} dx$, $\int_{-\infty}^0 \frac{x \arctan x}{\sqrt{x^6+x^2+1}} dx$ converge by (for sufficiently large x)

$$\left| \frac{x+1}{x^3-2x+3} \right| \leq \frac{1}{2x^2}, \quad \left| \frac{\ln x}{x^2} \right| \leq \frac{1}{x^{\frac{3}{2}}}, \quad \left| \frac{x \arctan x}{\sqrt{x^6+x^2+1}} \right| \leq \frac{\pi}{2x^2},$$

the comparison test, and Example 6.3.4. Moreover, by

$$\frac{\ln x}{\sqrt{x^2+1}} \geq \frac{1}{\sqrt{x^2+1}} \geq \frac{1}{2x},$$

and the divergence of $\int_1^{+\infty} \frac{dx}{x}$, the comparison test tells us that $\int_1^{+\infty} \frac{\ln x}{\sqrt{x^2+1}} dx$ diverges. ■

Example 6.3.9 We have $e^{-x^2} \leq e^{-x}$ for $x \geq 1$. Since $\int_0^{+\infty} e^{-x} dx$ converges, by the comparison test, the improper integral $\int_0^{+\infty} e^{-x^2} dx$ also converges. It is known (by using integration of two variable function, for example) that

$$\int_0^{+\infty} e^{-x^2} dx = \frac{\sqrt{\pi}}{2}.$$

By changing x to $-x$, we also find $\int_{-\infty}^0 e^{-x^2} dx = \frac{\sqrt{\pi}}{2}$, so that

$$\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}.$$

More generally, for any natural number n , we have

$$x^n e^{-x^2} = x^n e^{-\frac{x^2}{2}} \cdot e^{-\frac{x^2}{2}} \leq e^{-\frac{x^2}{2}} \leq e^{-x}$$

for sufficiently large x . By the comparison test, the improper integral

$$I_n = \int_0^{+\infty} x^n e^{-x^2} dx$$

converges. Then we may apply the integration by parts to get

$$I_n = -\frac{1}{2} \int_0^{+\infty} x^{n-1} d e^{-x^2} = -\frac{1}{2} x^{n-1} e^{-x^2} \Big|_{x=0}^{x=+\infty} + \frac{n-1}{2} \int_0^{+\infty} x^{n-2} e^{-x^2} dx = \frac{n-1}{2} I_{n-2}$$

for $n \geq 2$. The recursive relation and

$$I_0 = \frac{\sqrt{\pi}}{2}, \quad I_1 = \int_0^{+\infty} x e^{-x^2} dx = \frac{1}{2} \int_0^{+\infty} e^{-x} dx = \frac{1}{2}$$

enable us to compute I_n for all natural numbers n . ■

Example 6.3.10 By $\left| \frac{\sin x}{x^p} \right| \leq \frac{1}{x^p}$, the comparison test, and Example 6.3.4, we know $\int_1^{+\infty} \frac{\sin x}{x^p} dx$ (absolutely) converges for $p > 1$. By the similar reason, $\int_1^{+\infty} \frac{\cos x}{x^p} dx$ also converges for $p > 1$.

The argument fails for the case $p = 1$. In fact, the improper integral $\int_1^{+\infty} \left| \frac{\sin x}{x} \right| dx$ diverges. The reason is that $|\sin x| \geq \frac{1}{2}$ on the disjoint intervals $[a_n, b_n] = \left[n\pi + \frac{\pi}{6}, n\pi + \frac{5\pi}{6} \right]$, so that

$$\int_{\pi}^{(k+1)\pi} \left| \frac{\sin x}{x} \right| dx \geq \sum_{n=1}^k \int_{a_n}^{b_n} \left| \frac{\sin x}{x} \right| dx \geq \sum_{n=1}^k \int_{a_n}^{b_n} \frac{dx}{2b_n} \geq \sum_{n=1}^k \frac{b_n - a_n}{2b_n} \geq \frac{1}{3} \sum_{n=1}^k \frac{1}{n+1}.$$

By Example 2.1.29, the right side diverges to $+\infty$. By the remark after Proposition 6.3.1, we conclude that $\int_1^{+\infty} \left| \frac{\sin x}{x} \right| dx$ diverges.

Although $\int_1^{+\infty} \frac{\sin x}{x} dx$ does not converge absolutely (which means the comparison test cannot be used), it does not immediately follow that the integral itself diverges. In fact, we will show that, without taking the absolute value, the improper integral $\int_1^{+\infty} \frac{\sin x}{x} dx$ converges.

Using integrating by parts, we have

$$\int_1^b \frac{\sin x}{x} dx = - \int_1^b \frac{1}{x} d \cos x = - \frac{\cos b}{b} + \cos 1 - \int_1^b \frac{\cos x}{x^2} dx.$$

Since the improper integral $\int_1^{+\infty} \frac{\cos x}{x^2} dx$ converges, the right side converges as $b \rightarrow +\infty$. This proves that $\int_1^{+\infty} \frac{\sin x}{x} dx$ converges. ■

The idea used in deriving the convergence of improper integrals in the Example above can be elaborated to the following useful tests.

Theorem 6.3.3 (Dirichlet Test) Suppose $\int_a^b f(x) dx$ is bounded for $b \in [a, +\infty)$. Suppose $g(x)$ is monotonic and $\lim_{x \rightarrow +\infty} g(x) = 0$. Then $\int_a^{+\infty} f(x)g(x) dx$ converges.

Theorem 6.3.4 (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.

Proof. The proof is basically replacing $\sin x$ and $\frac{1}{x}$ in the last part of Example 6.3.10 by $f(x)$ and $g(x)$. Since we will use integration by parts, the following argument technically requires that $f(x)$ is continuous and $g(x)$ is continuously differentiable. The technical requirement can be removed by using more general version of the integration by parts.

Let $F(x) = \int_a^x f(t)dt$. Then $F(a) = 0$, and by integration by parts,

$$\int_a^b f(x)g(x) dx = \int_a^b g(x) dF(x) = g(b)F(b) - \int_a^b F(x)g'(x) dx.$$

Under the assumption of the Dirichlet test, we have $\lim_{b \rightarrow +\infty} g(b)F(b) = 0$, and $|F(x)| < B$ for some constant B and all $x \geq a$. Assume the monotonic function $g(x)$ is nondecreasing. Then $g'(x) \geq 0$, and

$$|F(x)g'(x)| \leq Bg'(x).$$

Since

$$\int_a^{+\infty} g'(x) dx = \lim_{b \rightarrow +\infty} \int_a^b g'(x) dx = \lim_{b \rightarrow +\infty} (g(b) - g(a)) = -g(a)$$

converges, by the comparison test, the improper integral

$$\int_a^{+\infty} F(x)g'(x) dx = \lim_{b \rightarrow +\infty} \int_a^b F(x)g'(x) dx$$

converges. Therefore $\lim_{b \rightarrow +\infty} \int_a^b f(x)g(x) dx$ converges. The proof in the case $g(x)$ is decreasing is similar.

Under the assumption of the Abel test, we know both $\lim_{b \rightarrow +\infty} F(b)$ and $\lim_{b \rightarrow +\infty} g(b)$ converge. Therefore $F(x)$ is bounded, and we may apply the comparison test as before. Moreover, the convergence of $\lim_{b \rightarrow +\infty} g(b)$ implies the convergence of $\int_a^{+\infty} g'(x) dx$. We conclude again that $\lim_{b \rightarrow +\infty} \int_a^b f(x)g(x) dx$ converges. ■

6.3.2 Integral of Unbounded Function

An integral can also become improper when the integrand is unbounded. Such improper integrals may be computed by first consider intervals on which the function is bounded and then taking the limit.

Example 6.3.11 The integral $\int_0^1 \ln x dx$ is improper at (the right side of) 0. The integrand $\ln x$ is bounded on $[\epsilon, 1]$ for any $\epsilon > 0$, and

$$\int_{\epsilon}^1 \ln x dx = (x \ln x - x) \Big|_{x=\epsilon}^{x=1} = -1 - \epsilon \ln \epsilon + \epsilon.$$

Since the right side converges to -1 as $\epsilon \rightarrow 0^+$, the improper integral converges and has value

$$\int_0^1 \ln x dx = \lim_{\epsilon \rightarrow 0^+} \int_{\epsilon}^1 \ln x dx = -1.$$

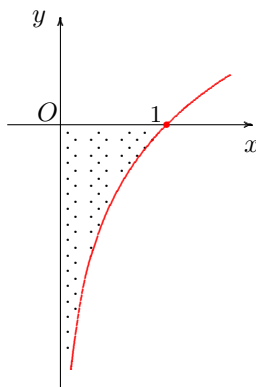


Fig. 6.6 The unbounded region has area 1.

Geometrically, this means that the area of the unbounded region between the graph of $\ln x$ and the x -axis over the interval $[0, 1]$ is 1. ■

Example 6.3.12 The integral $\int_0^1 x^p dx$ is improper at 0 when $p < 0$. If $p \neq -1$, for any $\epsilon > 0$, we have

$$\int_{\epsilon}^1 x^p dx = \frac{1 - \epsilon^{p+1}}{p+1}.$$

The right side converges as $\epsilon \rightarrow 0^+$ if and only if $p > -1$, and we have

$$\int_0^1 x^p dx = \lim_{\epsilon \rightarrow 0^+} \int_{\epsilon}^1 x^p dx = \frac{1}{p+1} \text{ if } p > -1.$$

For $p = -1$, the improper integral

$$\int_0^1 \frac{dx}{x} = \lim_{\epsilon \rightarrow 0^+} \int_{\epsilon}^1 \frac{dx}{x} = \lim_{\epsilon \rightarrow 0^+} (-\ln \epsilon) = +\infty$$

still diverges.

By the same argument, for $a < b$, the improper integrals $\int_a^b (x-a)^p dx$ and $\int_a^b (b-x)^p dx$ converges if and only if $p > -1$. ■

Example 6.3.13 A naive application of the fundamental theorem of calculus would tell us

$$\int_{-1}^2 \frac{dx}{x} = (\ln |x|) \Big|_{x=-1}^{x=2} = \ln 2 - \ln 1 = \ln 2.$$

However, the computation is wrong since the integrand $\frac{1}{x}$ is not continuous on $[-1, 2]$, which is a condition in the fundamental theorem. In fact, the integral $\int_{-1}^2 \frac{dx}{x}$ is improper on both sides of 0, and we need both improper integrals $\int_{-1}^0 \frac{dx}{x}$ and $\int_0^2 \frac{dx}{x}$ to converge and then get

$$\int_{-1}^2 \frac{dx}{x} = \int_{-1}^0 \frac{dx}{x} + \int_0^2 \frac{dx}{x}.$$

Since we know from Example 6.3.12 that both improper integrals on the right diverge, the improper integral $\int_{-1}^2 \frac{dx}{x}$ diverges. ■

The properties of the improper integrals over unbounded intervals also hold for improper integrals of unbounded functions.

Example 6.3.14 The integral $\int_0^1 \frac{dx}{\sqrt{x(1-x)}}$ is improper at 0 and 1. Near 0, we have

$$\lim_{x \rightarrow 0^+} \frac{\sqrt{x}}{\sqrt{x(1-x)}} = 1.$$

This implies that $\frac{1}{\sqrt{x(1-x)}} \leq \frac{2}{\sqrt{x}}$ for sufficiently small and positive x . Then by the convergence of $\int_0^1 \frac{dx}{\sqrt{x}}$ (see Example 6.3.12) and the comparison test, we see that $\int_0^{\frac{1}{2}} \frac{dx}{\sqrt{x(1-x)}}$ converges (the upper limit $\frac{1}{2}$ is chosen to avoid the other improper place). Similarly, the limit

$$\lim_{x \rightarrow 1^-} \frac{\sqrt{1-x}}{\sqrt{x(1-x)}} = 1$$

and the convergence of $\int_0^1 \frac{dx}{\sqrt{1-x}}$ implies the convergence of $\int_{\frac{1}{2}}^1 \frac{dx}{\sqrt{x(1-x)}}$. We conclude that the integral $\int_0^1 \frac{dx}{\sqrt{x(1-x)}}$ converges at all its improper places and is therefore convergent. ■

Example 6.3.15 The integral $\int_0^{\frac{\pi}{2}} \ln \sin x \, dx$ is improper at 0. We have

$$\lim_{x \rightarrow 0^+} \frac{\ln \sin x}{\ln x} = \lim_{x \rightarrow 0^+} \frac{\frac{\cos x}{\sin x}}{\frac{1}{x}} = \lim_{x \rightarrow 0^+} \frac{x \cos x}{\sin x} = 1.$$

This implies that $|\ln \sin x| \leq 2|\ln x|$ for x close to 0^+ . By the convergence of $\int_0^1 |\ln x| dx = -\int_0^1 \ln x dx$ in Example 6.3.11 and the comparison test, we see that $\int_0^{\frac{\pi}{2}} \ln \sin x dx$ converges.

The value of the improper integral can be computed as follows

$$\begin{aligned} \int_0^{\frac{\pi}{2}} \ln \sin x dx &= \int_0^{\frac{\pi}{4}} \ln \sin x dx + \int_{\frac{\pi}{4}}^{\frac{\pi}{2}} \ln \sin x dx = \int_0^{\frac{\pi}{4}} \ln \sin x dx - \int_{\frac{\pi}{4}}^0 \ln \cos x dx \\ &= \int_0^{\frac{\pi}{4}} (\ln \sin x + \ln \cos x) dx = \int_0^{\frac{\pi}{4}} \ln \left(\frac{1}{2} \sin 2x \right) dx \\ &= \int_0^{\frac{\pi}{4}} \ln \sin 2x dx - \frac{\pi}{4} \ln 2 = \frac{1}{2} \int_0^{\frac{\pi}{2}} \ln \sin x dx - \frac{\pi}{4} \ln 2. \end{aligned}$$

Note that all the deductions are legitimate because all the improper integrals involved converge. Thus we conclude that

$$\int_0^{\frac{\pi}{2}} \ln \sin x dx = -\frac{\pi}{2} \ln 2. \quad \blacksquare$$

Example 6.3.16 The integral $\int_0^1 \frac{1}{x} \sin \frac{1}{x} dx$ is improper at 0. The change of variable from x to $\frac{1}{x}$ gives us

$$\int_0^1 \frac{1}{x} \sin \frac{1}{x} dx = - \int_{+\infty}^1 (x \sin x) \frac{1}{x^2} dx = \int_1^{+\infty} \frac{\sin x}{x} dx.$$

By Example 6.3.10, we know the improper integral on the right converges. Therefore $\int_0^1 \frac{1}{x} \sin \frac{1}{x} dx$ also converges.

The change of variable is justified by

$$\int_{\epsilon}^1 \frac{1}{x} \sin \frac{1}{x} dx = \int_1^b \frac{\sin x}{x} dx, \quad b = \frac{1}{\epsilon},$$

and the limit of the left as $\epsilon \rightarrow 0^+$ converges if and only if the limit of the right as $b \rightarrow +\infty$ converges. $\blacksquare$

Exercises

6.3.1 Determine the convergence and compute the value of convergent improper integrals:

1. $\int_1^{+\infty} \frac{dx}{x+1};$

3. $\int_0^{+\infty} \frac{x^2 dx}{x^3+1};$

2. $\int_1^{+\infty} \frac{dx}{\sqrt[3]{x+1}};$

4. $\int_0^{+\infty} \frac{x^2 dx}{(x^3+1)^2};$

5. $\int_0^1 x \ln x \, dx;$
6. $\int_0^1 (\ln x)^n \, dx;$
7. $\int_0^{+\infty} e^{-ax} \cos bx \, dx, a > 0;$
8. $\int_0^{+\infty} e^{-ax} \sin bx \, dx, a > 0;$
9. $\int_0^{+\infty} \frac{dx}{x(x+1) \cdots (x+n)};$
10. $\int_1^9 \frac{dx}{\sqrt[3]{x-9}};$
11. $\int_0^\pi \sec x \, dx;$
12. $\int_0^{+\infty} \frac{x \arctan x}{(1+x^2)^2} \, dx;$
13. $\int_1^{+\infty} \frac{\ln x}{x^2} \, dx;$
14. $\int_0^1 \frac{\ln x}{\sqrt{x}} \, dx;$
15. $\int_0^1 \frac{x^n \, dx}{\sqrt{1-x^2}}.$

6.3.2 Determine the convergence of improper integrals:

1. $\int_{-\infty}^{+\infty} \frac{dx}{1+x^2};$
2. $\int_0^{+\infty} \frac{dx}{x^3+1};$
3. $\int_3^{+\infty} \frac{dx}{x^2-3x+2};$
4. $\int_0^{+\infty} \frac{x^2+1}{x^3+1} \, dx;$
5. $\int_0^{+\infty} \frac{x^2+1}{x^4+1} \, dx;$
6. $\int_0^1 \frac{x^p \, dx}{\sqrt{1-x^q}}, p, q > 0;$
7. $\int_{-\infty}^{+\infty} \frac{dx}{x^2-x+1};$
8. $\int_0^{+\infty} \frac{\ln(1+x)}{x^n} \, dx;$
9. $\int_0^{+\infty} e^{-ax} \cos bx^2 \, dx, a > 0;$
10. $\int_0^{+\infty} \frac{\cos ax}{1+x^p} \, dx, p > 0;$
11. $\int_0^{+\infty} \frac{\sin^2 x}{x} \, dx;$
12. $\int_0^\pi \frac{\ln \sin x}{\sqrt{x}} \, dx;$
13. $\int_0^{10} \frac{dx}{\sqrt{|x^2-4x+3|}};$
14. $\int_2^{+\infty} \frac{dx}{x^p(\ln x)^q};$
15. $\int_0^1 \frac{dx}{x^p |\ln x|^q};$
16. $\int_0^{\frac{\pi}{2}} \frac{dx}{x^p \sin x};$
17. $\int_0^{\frac{\pi}{2}} \frac{dx}{\sqrt{\cos x}};$
18. $\int_0^{+\infty} \frac{\sin x}{x^p} \, dx;$

$$19. \int_0^{+\infty} \sin x^2 \, dx;$$

$$22. \int_{-\infty}^{+\infty} \frac{dx}{|x-1|^p |x-2|^q};$$

$$20. \int_1^{+\infty} \frac{\sin x \arctan x}{x^p} \, dx;$$

$$23. \int_{-\infty}^{+\infty} \frac{dx}{|x-a_1|^{p_1} |x-a_2|^{p_2} \cdots |x-a_n|^{p_n}}.$$

$$21. \int_0^1 \frac{e^{-2x}}{\sqrt{x}} \, dx;$$

6.3.3 Prove that if $\int_a^{+\infty} f^2(x) dx$ and $\int_a^{+\infty} g^2(x) dx$ converge, then $\int_a^{+\infty} f(x)g(x) dx$ and $\int_a^{+\infty} (f(x)+g(x))^2 dx$ converge.

6.3.4 Construct a function $f(x)$ such that $|f(x)| = 1$ and $\int_0^{+\infty} f(x) \, dx$ converges.

6.4 SUMMARY

Numerical Integration

A **numerical scheme** for computing an integration on $[a, b]$ starts with an even partition $P: x_i = a + ih$, where $h = \frac{b-a}{n}$ is the **step size**. Let $y_i = f(x_i)$.

- The **rectangular rules** take either the left or right sample points in the Riemann sum

$$\begin{aligned} R_n &= h(y_1 + y_2 + \cdots + y_n), \\ L_n &= h(y_0 + y_1 + \cdots + y_{n-1}). \end{aligned}$$

- The **trapezoidal rule** takes the average of the two rectangular rules

$$T_n = \frac{R_n + L_n}{2} = \frac{h}{2}(y_1 + 2y_2 + 2y_3 + \cdots + 2y_{n-1} + y_n).$$

- The **midpoint rule** takes the the middle sample points in the Riemann sum

$$M_n = h(\bar{y}_1 + \bar{y}_2 + \cdots + \bar{y}_n), \quad \bar{y}_i = f(\bar{x}_i), \quad \bar{x}_i = \frac{x_{i-1} + x_i}{2}.$$

- The **Simpson's rule** approximates the function by quadratic functions on any two adjacent intervals (this requires even number of partition intervals)

$$S_n = \frac{h}{3}(y_0 + 4y_1 + 2y_2 + 4y_3 + 2y_4 + \cdots + 2y_{n-2} + 4y_{n-1} + y_n).$$

The more sophisticated rules has more accurate computation.

- If $f''(x)$ is continuous and bounded by K_2 , then

$$\left| \int_a^b f(x) \, dx - T_n \right| \leq \frac{K_2(b-a)^3}{12n^2}, \quad \left| \int_a^b f(x) \, dx - M_n \right| \leq \frac{K_2(b-a)^3}{24n^2}.$$

- If $f^{(4)}(x)$ is continuous and bounded by K_4 , then

$$\left| \int_a^b f(x) \, dx - S_n \right| \leq \frac{K_4(b-a)^5}{180n^4}$$

Improper Integral

An integral become **improper** either when the interval becomes unbounded or the integrand function becomes unbounded.

- If $f(x)$ is integrable on $[a, b]$ for any $b \in [a, +\infty)$, then

$$\int_a^{+\infty} f(x) \, dx = \lim_{b \rightarrow +\infty} \int_a^b f(x) \, dx.$$

The improper integral $\int_a^{+\infty} f(x) \, dx$ **converges** if the limit converges.

- If $f(x)$ is integrable on $[c, b]$ for any $c \in (a, b]$ and $f(x)$ is not bounded when $x \rightarrow a^+$, then

$$\int_a^b f(x) \, dx = \lim_{c \rightarrow a^+} \int_c^b f(x) \, dx.$$

The improper integral $\int_a^b f(x) \, dx$ **converges** if the limit converges.

- An integral may become improper at several places ($+\infty$, $-\infty$, left of a , right of a , for possibly several a 's). An improper integral converges if it converges in all its improper places.
- An improper integral $\int_a^b f(x) \, dx$ **absolutely converges** if $\int_a^b |f(x)| \, dx$ converges.

Like the usual integral, the improper integral has the properties such as arithmetic property, integration by parts and change of variable, as long as all the integrals involved converges. For an integral $\int_a^{+\infty} f(x) \, dx$ that is improper only at $+\infty$, the convergence can be determined as follows.

- Cauchy Criterion: $\int_a^{+\infty} f(x) \, dx$ converges if and only if for any $\epsilon > 0$, there is N , such that $b, c > N$ implies $\left| \int_b^c f(x) \, dx \right| < \epsilon$.

- Absolute Convergence: If $\int_a^{+\infty} |f(x)| \, dx$ converges, then $\int_a^{+\infty} f(x) \, dx$ converges. Moreover, the convergence of $\int_a^{+\infty} |f(x)| \, dx$ means that $\int_a^{+\infty} |f(x)| \, dx < +\infty$. In other words, $\int_a^b |f(x)| \, dx$ is bounded for $b \in [a, +\infty)$.
- Comparison Test: If $|f(x)| \leq g(x)$ and $\int_a^{+\infty} g(x) \, dx$ converges, then $\int_a^{+\infty} f(x) \, dx$ converges.
- Dirichlet Test: If $\int_a^b f(x) \, dx$ is bounded for $b \in [a, +\infty)$, and $g(x)$ is monotonic and $\lim_{x \rightarrow +\infty} g(x) = 0$, then $\int_a^{+\infty} f(x)g(x) \, dx$ converges.
- Abel Test: If $\int_a^{+\infty} f(x) \, dx$ converges, and $g(x)$ is monotonic and bounded on $[a, +\infty)$, then $\int_a^{+\infty} f(x)g(x) \, dx$ converges.

Similar results hold for other types of improper integrals.

Applications of Integration

7.1 APPLICATIONS IN GEOMETRY

7.1.1 Arc Length

A **parametrized curve** in the Euclidean plane is given by a pair of functions

$$x = x(t), \quad y = y(t), \quad t \in [a, b].$$

The variable t is the parameter. The point $\phi(a)$ is the beginning of the curve and the $\phi(b)$ is the end of the curve.

A parametrized curve in the Euclidean space is given by a triple of functions

$$x = x(t), \quad y = y(t), \quad z = z(t), \quad t \in [a, b].$$

More generally, we have parametrized curves in high dimensional Euclidean spaces.

Example 7.1.1 The graph of a function $f(x)$ on $[a, b]$ can be considered as a parametrized curve

$$x = t, \quad y = f(t), \quad t \in [a, b]. \quad \blacksquare$$

Example 7.1.2 The equation $x^2 + y^2 = r^2$ defined a circle of radius r . The circle is a not yet parametrized curve, and can be parametrized as

$$x = r \cos \theta, \quad y = r \sin \theta, \quad \theta \in [0, 2\pi].$$

It can also be parametrized as

$$x = r \cos 2t, \quad y = r \sin 2t, \quad t \in [0, \pi],$$

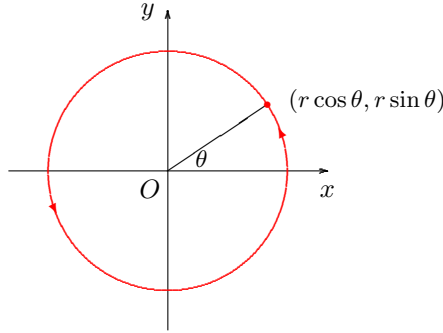


Fig. 7.1 The circle $x^2 + y^2 = r^2$ parametrized

or even parametrized as

$$x = r \cos 2u, \quad y = r \sin 2u, \quad u \in [0, 2\pi].$$

Note that the θ -parametrization and t -parametrization are the same in the sense that both “wraps around” the circle once. They are different from the u -parametrization, which wraps around the circle twice. ■

Example 7.1.3 The helix is a curve in Euclidean space that can be parametrized as

$$x = r \cos \theta, \quad y = r \sin \theta, \quad z = h\theta.$$

From the viewpoint of the (x, y) -plane, the curve is still a circle. From the viewpoint of the z -axis, the curve moves upward at the speed of h . ■

To find the length of a continuously differentiable parametrized curve $(x(t), y(t))$, we let $s(t)$ be the length of the curve from a to t . When the parameter t is changed by Δt , the corresponding change of length

$$\Delta s = s(t + \Delta t) - s(t)$$

is the length of the segment from t to $t + \Delta$. The segment is approximated by the straight line segment connecting $(x(t), y(t))$ to $(x(t + \Delta t), y(t + \Delta t))$. Therefore $|\Delta s|$ is approximated by the length of the straight line segment

$$|\Delta s| \approx \sqrt{\Delta x^2 + \Delta y^2}, \quad \Delta x = x(t + \Delta t) - x(t), \quad \Delta y = y(t + \Delta t) - y(t).$$

By the linear approximations $\Delta x \approx x'(t)\Delta t$ and $\Delta y \approx y'(t)\Delta t$, we further get

$$|\Delta s| \approx \sqrt{x'(t)^2 + y'(t)^2} |\Delta t|.$$

Considering the signs of Δs and Δt , we get

$$\frac{\Delta s}{\Delta t} \approx \sqrt{x'(t)^2 + y'(t)^2}.$$

Taking the limit, we have

$$\frac{ds}{dt} = \sqrt{x'(t)^2 + y'(t)^2}.$$

In other words, the length function $s(t)$ is the antiderivative of the function $\sqrt{x'(t)^2 + y'(t)^2}$, which is simply the length of the tangent vector $(x'(t), y'(t))$ of the curve. In the differential form, the formula can also be expressed as

$$ds^2 = dx^2 + dy^2,$$

which may be considered as the infinitesimal form of the relation $\Delta s^2 \approx \Delta x^2 + \Delta y^2$.

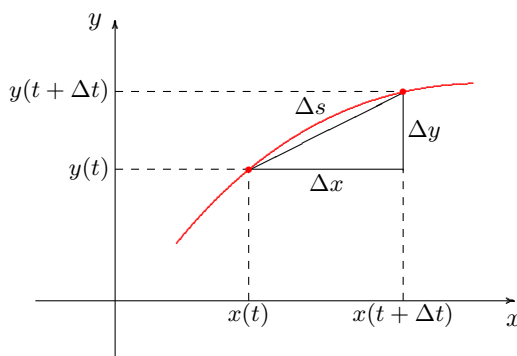


Fig. 7.2 Change of arc length

By the fundamental theorem of calculus, we get

$$s(t) = \int_a^t \sqrt{x'(t)^2 + y'(t)^2} dt.$$

by taking $t = b$, we get the length of the parametrized curve.

The length of a parametrized curves in the Euclidean space is given by

$$s(t) = \int_a^t \sqrt{x'(t)^2 + y'(t)^2 + z'(t)^2} dt.$$

Similar formula holds for curves in high dimensional Euclidean spaces.

Example 7.1.4 The graph of a function $f(x)$ on $[a, b]$ has tangent vector $(t', f'(t)) = (1, f'(t))$. Therefore the length of the graph is

$$\int_a^b \sqrt{1 + f'(x)^2} dx. \quad \blacksquare$$

Example 7.1.5 The parametrized circle

$$(x, y) = (r \cos \theta, r \sin \theta), \quad \theta \in [0, 2\pi]$$

of radius r has tangent vector $(x', y') = (-r \sin \theta, r \cos \theta)$ and length

$$\int_0^{2\pi} \sqrt{(-r \sin \theta)^2 + (r \cos \theta)^2} d\theta = \int_0^{2\pi} r \sqrt{\sin^2 \theta + \cos^2 \theta} d\theta = \int_0^{2\pi} r d\theta = 2\pi r.$$

The other parametrizations

$$\begin{aligned}(x, y) &= (r \cos 2t, r \sin 2t), & t \in [0, \pi]; \\(x, y) &= (r \cos 2u, r \sin 2u), & u \in [0, 2\pi].\end{aligned}$$

give the respective lengths

$$\begin{aligned}\int_0^\pi \sqrt{(-2r \sin 2t)^2 + (2r \cos 2t)^2} dt &= \int_0^\pi 2r dt = 2\pi r, \\ \int_0^{2\pi} \sqrt{(-2r \sin u)^2 + (2r \cos u)^2} du &= \int_0^{2\pi} 2r dt = 4\pi r. \quad \blacksquare\end{aligned}$$

Note that in the computation of the arc length of the circle, the t -parametrization gives the same length as the θ -parametrization, because both parametrizations represent wrapping around the circle once. However, the u -parametrization gives twice the length as the t - and θ -parametrizations because it represents wrapping around the circle twice.

The t - and θ -parametrizations are related by **reparametrization** in the sense that $\theta = 2t$ changes the parametrization

$$(x, y) = (r \cos \theta, r \sin \theta), \quad \theta \in [0, 2\pi]$$

to the parametrization

$$(x, y) = (r \cos 2t, r \sin 2t), \quad t \in [0, \pi]$$

including the range for the parameters. In general, a reparametrization of

$$(x, y) = (x(t), y(t)), \quad t \in [a, b]$$

is an invertible function $t = \phi(u)$ that changes the curve to

$$(x, y) = (x(\phi(u)), y(\phi(u))), \quad u \in [\phi^{-1}(a), \phi^{-1}(b)].$$

To keep the continuous differentiability, we also assume that both $\phi(u)$ and $\phi^{-1}(t)$ are continuously differentiable.

Therefore are two types of reparametrizations: ϕ is increasing, or ϕ is decreasing (the range of u should be $[\phi^{-1}(a), \phi^{-1}(b)]$ in this case). Geometrically, the first type preserves the “direction” of the curve, and the second type reverses the direction.

As we have seen in the example above, a reparametrization should not change the length. The following is the proof in case ϕ is increasing

$$\begin{aligned}\int_{\phi^{-1}(a)}^{\phi^{-1}(b)} \sqrt{x(\phi(u))'^2 + y(\phi(u))'^2} du &= \int_{\phi^{-1}(a)}^{\phi^{-1}(b)} \sqrt{x'(\phi(u))^2 \phi'(u)^2 + y'(\phi(u))^2 \phi'(u)^2} du \\ &= \int_{\phi^{-1}(a)}^{\phi^{-1}(b)} \sqrt{x'(\phi(u))^2 + y'(\phi(u))^2} \phi'(u) du \\ &= \int_a^b \sqrt{x'(t)^2 + y'(t)^2} dt.\end{aligned}$$

The last equality makes use of the change of variable formula for the integration. The proof for the case ϕ is decreasing is similar.

Example 7.1.6 To find the perimeter of the astroid

$$x^{\frac{2}{3}} + y^{\frac{2}{3}} = a^{\frac{2}{3}},$$

we use the parameterization

$$x = a \cos^3 t, \quad y = a \sin^3 t, \quad t \in [0, 2\pi].$$

The range $[0, 2\pi]$ for the parameter is chosen so that the curve moves along the astroid exactly once. By

$$x' = -3a \cos^2 t \sin t, \quad y' = 3a \sin^2 t \cos t,$$

The perimeter is

$$L = \int_0^{2\pi} 3a |\sin t \cos t| dt = \frac{3a}{2} \int_0^{2\pi} |\sin 2t| dt = 6a.$$

Although it is possible to parametrize the astroid in many different ways, they all give the same length as long as the parametrization wraps around exactly once. ■

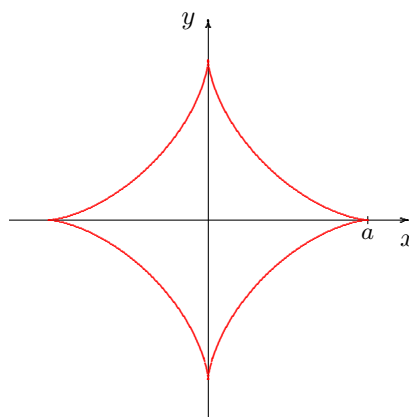


Fig. 7.3 Astroid $x^{\frac{2}{3}} + y^{\frac{2}{3}} = a^{\frac{2}{3}}$.

Example 7.1.7 To find the perimeter of the ellipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \quad b \geq a > 0,$$

we use the parametrization

$$x = a \cos t, \quad y = b \sin t, \quad t \in [0, 2\pi]$$

and get the perimeter

$$\int_0^{2\pi} \sqrt{a^2 \sin^2 t + b^2 \cos^2 t} dt = b \int_0^{2\pi} \sqrt{1 - k^2 \sin^2 t} dt, \quad k = \sqrt{1 - \frac{a^2}{b^2}}.$$

The integral is an **elliptic integral of the second kind**, and can not be evaluated by elementary means except the trivial case $k = 0$ (when the ellipse is a circle).

The perimeter of the ellipse is the same as the length of the graph of one period of the rescaled sine function

$$y = \sqrt{b^2 - a^2} \sin \frac{x}{a}, \quad x \in [0, 2\pi a]. \quad \blacksquare$$

The argument leading to the formula for the length of parametrized curve is a rather intuitive and heuristic one. Similar argument is widely used in many applications of the integration. However, the argument is not rigorous. We finish the discussion on arc length by presenting a rigorous argument.

First, similar to the definition of Riemann integral, the arc length of a parametrized curve

$$x = x(t), \quad y = y(t), \quad t \in [a, b]$$

is defined by approximating the curve by straight line segments. Specifically, let

$$P: a = t_0 < t_1 < \cdots < t_n = b$$

be a partition of $[a, b]$. Then we get partition points

$$P_i = (x_i, y_i), \quad x_i = x(t_i), \quad y_i = y(t_i)$$

along the curve. The part of the curve between P_{i-1} and P_i (corresponding to $t \in [t_{i-1}, t_i]$) is approximated by the straight line segment connecting P_{i-1} and P_i . The whole curve is then approximated by all such straight line segments, which has total length

$$L_P = \sum_{i=1}^n \|\overrightarrow{P_{i-1}P_i}\| = \sum_{i=1}^n \sqrt{\Delta x_i^2 + \Delta y_i^2}, \quad \Delta x_i = x_i - x_{i-1}, \quad \Delta y_i = y_i - y_{i-1},$$

where the notation $\|\cdot\|$ denotes the Euclidean length of a vector

$$\|(u, v)\| = \sqrt{u^2 + v^2}$$

and

$$\overrightarrow{P_{i-1}P_i} = (x_i, y_i) - (x_{i-1}, y_{i-1}) = (\Delta x_i, \Delta y_i)$$

is the vector going from P_{i-1} to P_i . We define the arc length of the curve to be the limit

$$\lim_{\|P\| \rightarrow 0} L_P$$

of the total length of the straight line segments.

Now assume $x(t)$ and $y(t)$ are differentiable, with integrable $x'(t)$ and $y'(t)$. Then by the mean value theorem, we have

$$\Delta x_i = x'(t_i^*) \Delta t_i, \quad \Delta y_i = y'(t_i^{**}) \Delta t_i$$

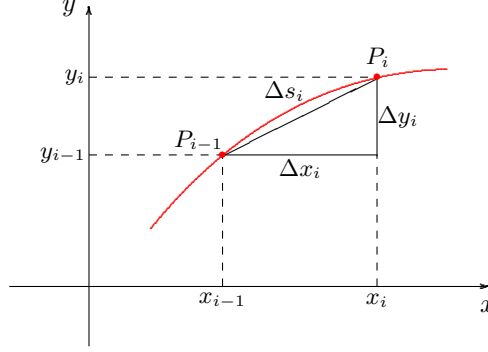


Fig. 7.4 Approximate a curve by straight line segments

for some $t_i^*, t_i^{**} \in [t_{i-1}, t_i]$. Therefore

$$L_P = \sum_{i=1}^n \sqrt{x'(t_i^*)^2 + y'(t_i^{**})^2} \Delta t_i = \sum_{i=1}^n \|(x'(t_i^*), y'(t_i^{**}))\| \Delta t_i,$$

which is almost the same as the Riemann sum

$$S(P, \sqrt{x'^2 + y'^2}) = \sum_{i=1}^n \sqrt{x'(t_i^*)^2 + y'(t_i^*)^2} \Delta t_i = \sum_{i=1}^n \|(x'(t_i^*), y'(t_i^*))\| \Delta t_i.$$

The only difference is that the sample points for x' and y' may not be the same in L_P (by the way, we have chosen the sample points in S to be the same as the ones for x'). The difference can be estimated as follows

$$\begin{aligned} |L_P - S(P, \sqrt{x'^2 + y'^2})| &= \left| \sum_{i=1}^n (\|(x'(t_i^*), y'(t_i^{**}))\| - \|(x'(t_i^*), y'(t_i^*))\|) \Delta t_i \right| \\ &\leq \sum_{i=1}^n \left| \|(x'(t_i^*), y'(t_i^{**}))\| - \|(x'(t_i^*), y'(t_i^*))\| \right| \Delta t_i \\ &\leq \sum_{i=1}^n \|(x'(t_i^*), y'(t_i^{**})) - (x'(t_i^*), y'(t_i^*))\| \Delta t_i \\ &= \sum_{i=1}^n \|(0, y'(t_i^{**}) - y'(t_i^*))\| \Delta t_i \\ &= \sum_{i=1}^n |y'(t_i^{**}) - y'(t_i^*)| \Delta t_i \leq \sum_{i=1}^n \omega_{[t_{i-1}, t_i]}(y') \Delta t_i. \end{aligned}$$

The second inequality follows from the **triangle inequality** for the Euclidean length

$$\|(u_1 + u_2, v_1 + v_2)\| \leq \|(u_1, v_1)\| + \|(u_2, v_2)\|.$$

Since y' is integrable, the sum $\sum_{i=1}^n \omega_{[t_{i-1}, t_i]}(y') \Delta t_i$ approaches 0 as $\|P\|$ approaches 0. Therefore we get

$$\lim_{\|P\| \rightarrow 0} L_P = \lim_{\|P\| \rightarrow 0} S(P, \sqrt{x'^2 + y'^2}) = \int_a^b \sqrt{x'(t)^2 + y'(t)^2} dt.$$

We remark that we have essentially used the triangle inequality and the scaling equality $\|c(u, v)\| = |c| \cdot \|(u, v)\|$ for the Euclidean length. We also essentially used the integrability of x' and y' (continuous differentiability is not necessary). Therefore the argument can be easily extended to high dimensional Euclidean spaces and other non-Euclidean lengths such as $\|(u, v)\| = |u| + |v|$ or $\|(u, v)\| = \max\{|u|, |v|\}$.

7.1.2 Area of Surface of Revolution

Let $f(x)$ be a non-negative continuously differentiable function $[a, b]$. A surface may be produced by revolving the graph of $f(x)$ around the x -axis. We wish to find the area of the surface.

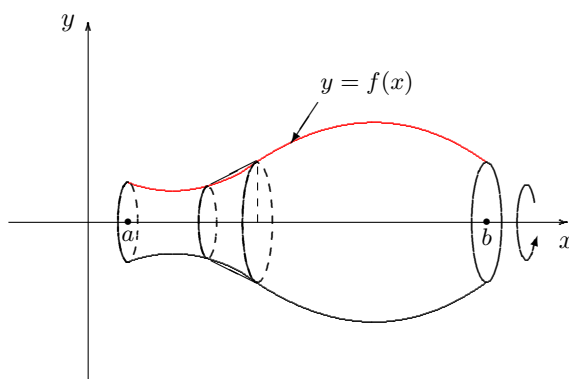


Fig. 7.5 Surface of revolution

Let $A(x)$ be the area of the surface of revolution of $f(x)$ over $[a, x]$. When x is changed by Δx , the area is changed by $\Delta A = A(x + \Delta x) - A(x)$. The absolute value of the area change $|\Delta A|$ is the area of the surface of revolution of $f(x)$ over $[x, x + \Delta x]$. Since $f(x)$ over $[x, x + \Delta x]$ is approximated by the straight line segment connecting $(x, f(x))$ to $(x + \Delta x, f(x + \Delta x))$, the area $|\Delta A|$ is approximated by the area of the surface of revolution of the line segment. The revolution of the line segment can be flattened to become a “circular trapezoid” with top and bottom lengths $2\pi f(x)$ and $2\pi f(x + \Delta x) = 2\pi(f(x) + \Delta f)$, and the distance $\sqrt{\Delta x^2 + (f(x + \Delta x) - f(x))^2} = \sqrt{\Delta x^2 + \Delta f^2}$ between the top and bottom. Therefore the area of the revolution of the line segment is

$$\frac{1}{2}(2\pi f(x) + 2\pi(f(x) + \Delta f))\sqrt{\Delta x^2 + \Delta f^2} = \pi(2f(x) + \Delta f)\sqrt{\Delta x^2 + \Delta f^2}.$$

We get the approximation

$$|\Delta A| \approx \pi(2f(x) + \Delta f)\sqrt{\Delta x^2 + \Delta f^2} \approx \pi(2f(x) + f'(x)\Delta x)\sqrt{1 + f'(x)^2}|\Delta x|.$$

Considering the signs of ΔA and Δx , this implies

$$\frac{dA}{dx} = \lim_{\Delta x \rightarrow 0} \frac{\Delta A}{\Delta x} = \lim_{\Delta x \rightarrow 0} \pi(2f(x) + f'(x)\Delta x)\sqrt{1 + f'(x)^2} = 2\pi f(x)\sqrt{1 + f'(x)^2}.$$

Therefore the area of the revolution is

$$\int_a^b 2\pi f(x)\sqrt{1 + f'(x)^2} dx.$$

The discussion above is as heuristic as the discussion for the arc length. A rigorous argument in terms of the Riemann sum can be similarly carried out and is omitted here.

Example 7.1.8 The sphere of radius R can be obtained by revolving the function $f(x) = \sqrt{R^2 - x^2}$ over $[-R, R]$ around the x -axis. Therefore the area of the sphere is

$$\int_{-R}^R 2\pi\sqrt{R^2 - x^2}\sqrt{1 + \left(\frac{x}{\sqrt{R^2 - x^2}}\right)^2} dx = 2\pi \int_{-R}^R R dx = 4\pi R^2. \quad \blacksquare$$

The key point of the area of the revolution is

$$\int_a^b 2\pi(\text{radius}) ds,$$

where the radius is the radius of the revolution, 2π multiple of the radius is the length of the circle obtained by such revolution, and ds is the arc length. The understanding of the key idea is very important in extending the formula to the other circumstances. For example, the area of the revolution of the graph of $f(x)$ on $[a, b]$ ($0 \leq a < b$) around the y -axis is

$$\int_a^b 2\pi x\sqrt{1 + f'(x)^2} dx.$$

If the curve is given parametrically, then the area of the surface of revolution around the y -axis is

$$\int_a^b 2\pi x(t)\sqrt{x'(t)^2 + y'(t)^2} dt.$$

If the revolution is around the line $x = h$ on the right of the parametrized curve (which means $x(t) \leq h$ for any $t \in [a, b]$), then the area is

$$\int_a^b 2\pi(h - x(t))\sqrt{x'(t)^2 + y'(t)^2} dt.$$

More generally, a surface can be produced by revolving a parametrized curve

$$C: x = x(t), \quad y = y(t), \quad t \in [a, b]$$

around a straight line

$$L: \alpha x + \beta y + \gamma = 0.$$

The condition that C lies on one side of L is the same as $\alpha x(t) + \beta y(t) + \gamma$ is always non-negative or always non-positive. By changing L to $-\alpha x - \beta y - \gamma = 0$ if necessary, we may assume that

$$\alpha x(t) + \beta y(t) + \gamma \geq 0, \text{ for all } t \in [a, b].$$

The radius of the revolution is the distance from a point $P = (x, y)$ on the curve to L , or the distance from P to the point $Q = (\xi, \eta) \in L$, such that PQ is orthogonal to L . The conditions mean two equations

$$\alpha\xi + \beta\eta + \gamma = 0, \quad \frac{x - \xi}{y - \eta} = \frac{\alpha}{\beta}.$$

Solving the system, we get

$$x - \xi = \frac{\alpha}{\alpha^2 + \beta^2}(\alpha x + \beta y + \gamma), \quad y - \eta = \frac{\beta}{\alpha^2 + \beta^2}(\alpha x + \beta y + \gamma).$$

Therefore the radius of revolution is

$$\sqrt{(x - \xi)^2 + (y - \eta)^2} = \frac{\alpha x + \beta y + \gamma}{\sqrt{\alpha^2 + \beta^2}},$$

and the area of the surface of revolution is

$$2\pi \int_a^b \frac{\alpha x(t) + \beta y(t) + \gamma}{\sqrt{\alpha^2 + \beta^2}} ds, \quad ds = \sqrt{x'(t)^2 + y'(t)^2} dt.$$

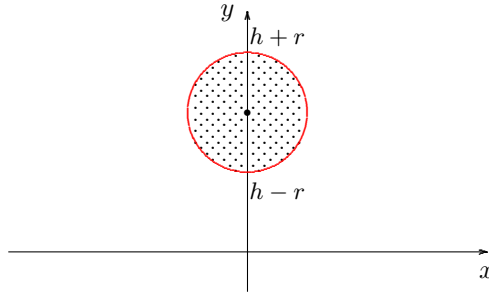


Fig. 7.6 Revolve to get torus.

Example 7.1.9 The torus is obtained by revolving the circle

$$x^2 + (y - h)^2 = r^2, \quad h > r > 0$$

around the x -axis. The circle can be parametrized as

$$x = r \cos t, \quad y = h + r \sin t, \quad t \in [0, 2\pi].$$

Therefore the area of the torus is

$$2\pi \int_0^{2\pi} (h + r \sin t) \sqrt{(-r \sin t)^2 + (0 + r \cos t)^2} dt = 4\pi^2 hr. \quad \blacksquare$$

Example 7.1.10 The revolution of the ellipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \quad a, b > 0$$

around the x -axis is the ellipsoid

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{b^2} = 1.$$

The revolution can be considered as the revolution of the parametrized upper half of the ellipse

$$x = a \cos t, \quad y = b \sin t, \quad t \in [0, \pi].$$

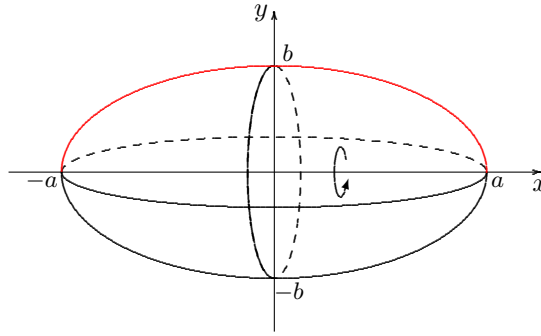


Fig. 7.7 Ellipsoid

The area of the ellipsoid is

$$\begin{aligned} 2\pi \int_0^\pi b \sin t \sqrt{(-a \sin t)^2 + (b \cos t)^2} dt &= 2\pi \int_0^\pi b \sin t \sqrt{a^2 \sin^2 t + b^2 \cos^2 t} dt \\ &= -2\pi \int_0^\pi b \sqrt{a^2(1 - \cos^2 t) + b^2 \cos^2 t} d(\cos t) \\ &= 2\pi b \int_{-1}^1 \sqrt{a^2 + (b^2 - a^2)u^2} du. \end{aligned}$$

By Example 5.4.9, the area is

$$2\pi \frac{a^2 b}{\sqrt{a^2 - b^2}} \arcsin \frac{\sqrt{a^2 - b^2}}{a} + 2\pi b^2$$

when $a > b$. By Example 5.4.11, the area is

$$2\pi \frac{a^2 b}{\sqrt{b^2 - a^2}} \ln \frac{b + \sqrt{b^2 - a^2}}{a} + 2\pi b^2$$

when $a < b$. The case $a = b$ is the sphere, and we already know the area. ■

7.1.3 Volume of Solid of Revolution

For a non-negative function $f(x)$ on $[a, b]$, a solid body may be produced by revolving the region

$$\{(x, y): a \leq x \leq b, 0 \leq y \leq f(x)\}$$

between the graph of $f(x)$ and the x -axis around the x -axis.

Let $V(x)$ be the volume of the part of the solid between a and x . When x is changed by Δx , the volume is changed by $\Delta V = V(x + \Delta x) - V(x)$. The absolute volume of the volume change $|\Delta V|$ is the volume of the part of the solid between x and $x + \Delta x$. Let M and m be the supremum and the infimum of f on $[x, x + \Delta x]$. Then the part of the solid between x and $x + \Delta x$ is sandwiched between two disks with respective radii M and m and with thickness $|\Delta x|$. Therefore

$$\pi m^2 |\Delta x| \leq |\Delta V| \leq \pi M^2 |\Delta x|.$$

Considering the signs of ΔV and Δx , this implies

$$\pi m^2 \leq \frac{\Delta V}{\Delta x} \leq \pi M^2.$$

If $f(x)$ is continuous, then $\lim_{\Delta x \rightarrow 0} m = \lim_{\Delta x \rightarrow 0} M = f(x)$. Therefore by the sandwich rule, we get

$$\frac{dV}{dx} = \lim_{\Delta x \rightarrow 0} \frac{\Delta V}{\Delta x} = \pi [f(x)]^2,$$

and the volume of the revolution is

$$V(b) - V(a) = \int_a^b \pi [f(x)]^2 dx.$$

In fact, the formula also works for integrable $f(x)$.

More generally, for functions $f(x)$ and $g(x)$ on $[a, b]$ satisfying $f(x) \geq g(x) \geq 0$, a solid body may be produced by revolving the region

$$\{(x, y): a \leq x \leq b, g(x) \leq y \leq f(x)\}$$

between the graphs of the two functions around the x -axis. The volume of the solid body is the difference of the two solids of revolution

$$\int_a^b \pi [f(x)]^2 dx - \int_a^b \pi [g(x)]^2 dx = \int_a^b \pi ([f(x)]^2 - [g(x)]^2) dx.$$

Example 7.1.11 To find the volume of the solid ellipsoid in Example 7.1.10, we note that the solid is obtained by revolving the region between

$$y = b\sqrt{1 - \frac{x^2}{a^2}}, \quad x \in [-a, a]$$

and the x -axis. The volume of the ellipsoid solid is

$$\int_{-a}^a \pi y^2 dx = \int_{-a}^a \pi b^2 \left(1 - \frac{x^2}{a^2}\right) dx = \frac{4}{3} \pi a b^2.$$

In particular, the volume of the ball of radius a is $\frac{4}{3} \pi a^3$. ■

Example 7.1.12 The solid torus in Example 7.1.9 is obtained by revolving the region between the upper semicircle $y = h + \sqrt{r^2 - x^2}$ and the lower semicircle $y = h - \sqrt{r^2 - x^2}$. The volume of the torus is

$$\int_{-r}^r \pi \left[(h + \sqrt{r^2 - x^2})^2 dx - (h - \sqrt{r^2 - x^2})^2 \right] dx = \pi \int_{-r}^r 4h\sqrt{r^2 - x^2} dx = 2\pi^2 h r^2. \quad \blacksquare$$

Similar to the area of surface of revolution, the key point of the volume of the solid of revolution is

$$\int_a^b \pi(\text{radius})^2 d\xi,$$

where $\pi(\text{radius})^2$ is the area of the disk obtained by such revolution, and $d\xi$ is the displacement in the direction of the axis of revolution. For example, the volume of the revolution of the graph of an increasing function $f(x)$ on $[a, b]$ ($0 \leq a < b$) around the y -axis is

$$\int_{f(a)}^{f(b)} \pi x^2 dy = \int_a^b \pi x^2 f'(x) dx.$$

In fact, the formula holds as long as $f(a) \leq f(x) \leq f(b)$ for any $x \in [a, b]$ (if the condition is not satisfied, there is a real ambiguity about the meaning of the solid of revolution). If the curve is parametrized and satisfies $x(t) \geq 0$, $y(a) \leq y(t) \leq y(b)$ for any $t \in [a, b]$, then the volume of the revolution around the y -axis is

$$\int_{t=a}^{t=b} \pi x^2 dy = \int_a^b \pi x(t)^2 y'(t) dt.$$

If the curve is parametrized and satisfies $x(t) \leq h$, $y(a) \leq y(t) \leq y(b)$ for any $t \in [a, b]$, then the volume of the revolution around the axis $x = h$ is

$$\int_{t=a}^{t=b} \pi (h - x)^2 dy = \int_a^b \pi (h - x(t))^2 y'(t) dt.$$

More generally, consider the volume of a solid produced by revolving a parametrized curve

$$C: x = x(t), \quad y = y(t), \quad t \in [a, b]$$

around a straight line

$$L: \alpha x + \beta y + \gamma = 0.$$

We further assume the curve essentially proceeds in the positive direction $(\beta, -\alpha)$ of the axis of rotation L

$$\beta x(a) - \alpha y(a) \leq \beta x(t) - \alpha y(t) \leq \beta x(b) - \alpha y(b), \quad \text{for all } t \in [a, b].$$

The radius of revolution was computed for the area, so that

$$\pi(\text{radius})^2 = \pi \frac{(\alpha x(t) + \beta y(t) + \gamma)^2}{\alpha^2 + \beta^2}.$$

The displacement in the direction of the axis of revolution is the projection of the displacement $(\Delta x, \Delta y)$ along the direction $(\beta, -\alpha)$ of L

$$\Delta \xi = (\Delta x, \Delta y) \cdot \frac{(\beta, -\alpha)}{\sqrt{\beta^2 + \alpha^2}} = \frac{\beta \Delta x - \alpha \Delta y}{\sqrt{\alpha^2 + \beta^2}}.$$

Therefore

$$d\xi = \frac{\beta dx - \alpha dy}{\sqrt{\alpha^2 + \beta^2}} = \frac{\beta x'(t) - \alpha y'(t)}{\sqrt{\alpha^2 + \beta^2}} dt,$$

and the volume of the solid of revolution is

$$\int_a^b \pi \frac{(\alpha x(t) + \beta y(t) + \gamma)^2}{(\sqrt{\alpha^2 + \beta^2})^3} (\beta x'(t) - \alpha y'(t)) dt.$$

Example 7.1.13 Consider the region between

$$y = \cos x, \quad x \in \left[0, \frac{\pi}{2}\right]$$

and the two axes. If we revolve the region around the x -axis, then the solid of revolution has volume

$$\int_0^{\frac{\pi}{2}} \pi y^2 dx = \int_0^{\frac{\pi}{2}} \pi (\cos x)^2 dx = \frac{\pi^2}{4}.$$

If we revolve the region around the y -axis, then the volume is

$$\int_0^1 \pi x^2 dy = - \int_{\frac{\pi}{2}}^0 \pi x^2 \sin x dx = \pi^2 - 2\pi.$$

The integral may be computed by using integration by parts twice. ■

7.1.4 Cavalieri's Principle

The formulae for the area of surface of revolution and the volume of solid of revolution follows from a more general principle. In general, an n -dimensional solid S has n -dimensional size. Here for $n = 1$,

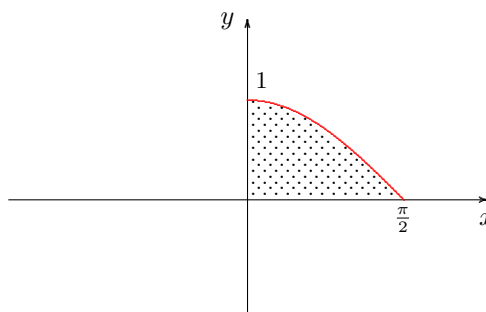


Fig. 7.8 The region bounded by $y = \cos x$ and the axes

the solid is a curve, and the size is the length. For $n = 2$, the solid is a surface or a region on the plane, and the size is the area. For $n \geq 3$, the size is generally called the volume.

To find the size of S , we decompose S into pieces S_t of 1-dimension lower (i.e., S_t has dimension $n - 1$). For $n = 2$, this means that a surface is decomposed into a family of curves. For $n = 3$, this means that a 3-dimensional solid is decomposed into a family of surfaces. The decomposition is **equidistant** if the distance between two nearby pieces does not depend on the location the distance is measured. Then the pieces can really be parametrized by the distance. If t is the distance between the pieces, then the size of the whole solid is

$$\text{Size}(S) = \int_a^b \text{Size}(S_t) dt.$$

Example 7.1.14 Let $f(x) \geq g(x)$, $x \in [a, b]$. To find the area of the region between the graphs of f and g , we decompose the region into vertical straight line intervals $[g(x), f(x)]$ for each $x \in [a, b]$. The family is equidistant, with dx being the distance between nearby intervals. Therefore the area is

$$\int_a^b \text{Size}([g(x), f(x)]) dx = \int_a^b (f(x) - g(x)) dx. \quad \blacksquare$$

Example 7.1.15 Consider the disk of radius R . The disk can be decomposed into circles of radius t , $0 \leq t \leq R$. If all the circles have the same center as the disk, then the family of circles is equidistant, with dt being the distance between nearby circles. Moreover, the size (length) of the circle of radius t is $2\pi t$. Therefore the size (area) of the disk is

$$\int_0^R 2\pi t dt = \pi R^2.$$

The argument also works for a wedge part of the disk, which can be decomposed into circular arcs with centers at the apex of the wedge. Suppose θ is the angle of the wedge. Then the length of the circular arc at radius t is θt . Therefore the area of the wedge is

$$\int_0^R \theta t dt = \frac{\theta}{2} R^2. \quad \blacksquare$$

Example 7.1.16 Suppose S is a solid in the 3-dimensional Euclidean space $\mathbf{R}^3$. Suppose from the viewpoint of the x -axis, the solid lies over the interval $[a, b]$. For each $a \leq x \leq b$, we have the **cross section**

$$S_x = \{(y, z): (x, y, z) \in S\}$$

obtained by intersecting S with the (y, z) -plane perpendicular to the x -axis and passing through x . Then the family of cross sections is equidistant, with the distance between nearby cross sections given by dx . Therefore the volume of the solid S is given by integrating the area of the cross sections

$$\int_a^b \text{Area}(S_x) dx.$$

For the special case that S is obtained from rotating a curve around the x -axis, the area of S_x is πy^2 , and we get the formula for the volume of solid of revolution. ■

Example 7.1.17 Let S is the surface obtained by revolving a non-negative function $y = f(x)$ around the x -axis. Then S is decomposed into the circles S_x obtained by intersecting the (y, z) -plane perpendicular to the x -axis and passing through x . The family of circles is equidistant, and the distance between nearby circles is ds (not the distance dx between the (y, z) -planes). Since the size of the circle is $2\pi y$, the area of the surface of revolution is

$$\int_a^b 2\pi y ds = \int_a^b 2\pi f(x) \sqrt{1 + f'(x)^2} dx.$$

The same argument applies to the surface of revolution around any straight line. ■

Example 7.1.18 For a non-negative function $f(x)$ on $[a, b]$, we have considered the solid body produced by revolving the region

$$R = \{(x, y): a \leq x \leq b, 0 \leq y \leq f(x)\}$$

around the x -axis. If $a \geq 0$, then we may produce another solid body by revolving the region around the y -axis. We try to compute the volume of this solid.

The region R is decomposed into the vertical intervals $[0, f(x)]$ at x for $x \in [a, b]$. The revolution S of R around the y -axis is decomposed into the revolutions S_x of the vertical intervals $[0, f(x)]$ around the y -axis. The family S_x is equidistant, with dx being the distance between nearby pieces. The area of S_x is $2\pi xy = 2\pi x f(x)$. Therefore the volume of the solid of revolution is

$$\int_a^b 2\pi x f(x) dx.$$

This way of deriving the formula for the volume of revolution is called the **cylindrical shell method**. ■

The formula for the volume implies **Cavalieri's Principle**: If two solids are cut by the same families of parallel planes, with the same cross sectional areas, then the two solids have the same volume.

In the subsequent examples, we compute the size of concrete examples.

Example 7.1.19 Let S be the intersection of two circular cylinders of radius 1 around the x -axis and the y -axis. We compute the volume of S by taking the cross sections along the z -axis. For each $-1 \leq z \leq 1$, the cross section at z is a square of side length $2\sqrt{1-z^2}$. By Example 7.1.16, the volume of the solid is

$$\int_{-1}^1 (2\sqrt{1-z^2})^2 dz = 4 \int_{-1}^1 (1-z^2) dz = \frac{16}{3}.$$

Alternatively, we may take the cross section along the x -axis, for any $-1 \leq x \leq 1$, the cross section is the intersection of the disk of radius 1 centered at the origin with the strip $|y| \leq h = \sqrt{1-x^2}$. The area of the section is

$$2xh + \frac{1}{2}4\theta = 2(xh + \theta), \quad 0 \leq \theta \leq \pi, \quad x = \cos \theta.$$

Therefore the volume of the solid is

$$\begin{aligned} \int_{-1}^1 2(xh + \theta) dx &= 4 \int_0^1 (xh + \theta) dx \\ &= -4 \int_0^{\frac{\pi}{2}} (\cos \theta \sin \theta + \theta) d \cos \theta = 4 \int_0^{\frac{\pi}{2}} (\cos \theta \sin^2 \theta d\theta - \theta d \cos \theta) \\ &= 4 \left[\frac{1}{3} \sin^3 \theta \right]_0^{\frac{\pi}{2}} - 4\theta \cos \theta \Big|_0^{\frac{\pi}{2}} + 4 \int_0^{\frac{\pi}{2}} \cos \theta d\theta = \frac{4}{3} - 0 + 4 = \frac{16}{3}. \quad \blacksquare \end{aligned}$$

Example 7.1.20 Let A be a region on a plane. Let v be a point of distance h from the plane. Then by connecting v to all points in A , we get the **pyramid** P of base A and height h .

To compute the volume of P , we put A into the (x, y) -plane and put v on the z -axis, which means $v = (0, 0, h)$ in $\mathbf{R}^3$. Then we consider the cross sections along the z -axis. For each $0 \leq z \leq h$, the cross section P_z is a copy of A shrunk by a factor of $\frac{h-z}{h}$. Therefore the area of the cross section is $\frac{(h-z)^2}{h^2} \text{Area}(A)$, and the volume of the pyramid is

$$\int_0^h \frac{(h-z)^2}{h^2} \text{Area}(A) dz = \frac{1}{3}h \cdot \text{Area}(A). \quad \blacksquare$$

Example 7.1.21 The solid torus in Example 7.1.9 can also be considered as obtained by revolving the region between the left semi-circle $x = -\sqrt{r^2 - (y-h)^2}$ and the right semi-circle $x = \sqrt{r^2 - (y-h)^2}$. In this way, the volume of the torus can also be computed by using the formula in Example 7.1.18, except x and y are exchanged and the region is between two functions. The volume is

$$\begin{aligned} 2\pi \int_{h-r}^{h+r} y \left[(\sqrt{r^2 - (y-h)^2}) - (-\sqrt{r^2 - (y-h)^2}) \right] dy &= 4\pi \int_{-r}^r (h+t) \sqrt{r^2 - t^2} dt \\ &= 4\pi \int_{-r}^r h \sqrt{r^2 - t^2} dt. \end{aligned}$$

The integral is the same as the one in Example 7.1.12. \blacksquare

Example 7.1.22 We redo Example 7.1.13 by using the formula in Example 7.1.18. If we revolve the region around the y -axis, the volume of the solid of revolution is

$$\int_0^{\frac{\pi}{2}} 2\pi xy \, dx = \int_0^{\frac{\pi}{2}} 2\pi x \cos x \, dx = 2\pi \int_0^{\frac{\pi}{2}} x \, d \sin x = 2\pi \left(\frac{\pi}{2} - \int_0^{\frac{\pi}{2}} \sin x \, dx \right) = \pi^2 - 2\pi.$$

If we revolve the region around the x -axis, the volume of the solid of revolution is

$$\int_0^1 2\pi xy \, dy = \int_0^1 \pi x \, d(y^2) = \pi \left(xy^2 \Big|_{(x,y)=(\frac{\pi}{2},0)}^{(x,y)=(0,1)} - \int_{\frac{\pi}{2}}^0 y^2 \, dx \right) = \pi \int_0^{\frac{\pi}{2}} y^2 \, dx.$$

The integral is the same as the one in Example 7.1.13. ■

Example 7.1.23 The unit sphere in $\mathbf{R}^n$

$$S^{n-1} = \{(x_1, x_2, \dots, x_n) : x_1^2 + x_2^2 + \dots + x_n^2 = 1\}$$

is $(n-1)$ -dimensional and is the boundary of the unit ball in $\mathbf{R}^n$

$$B^n = \{(x_1, x_2, \dots, x_n) : x_1^2 + x_2^2 + \dots + x_n^2 \leq 1\}$$

Let A_{n-1} and V_n be the volumes of S^{n-1} and B^n . For example,

$$A_1 = 2\pi, \quad A_2 = 4\pi, \quad V_1 = 2, \quad V_2 = \pi.$$

The unit ball B^n can be decomposed by intersecting with the plane $z_n = t$. The cross section B_t is an $(n-1)$ -dimensional ball of radius $\sqrt{1-t^2}$, and has volume $(\sqrt{1-t^2})^{n-1} V_{n-1}$. The family B_t is equidistant and dt is the distance between nearby cross sections. Therefore

$$V_n = \int_{-1}^1 (\sqrt{1-t^2})^{n-1} V_{n-1} \, dt = 2V_{n-1} \int_0^1 (1-t^2)^{\frac{n-1}{2}} \, dt = 2V_{n-1} I_n = 4V_{n-2} I_n I_{n-1},$$

where I_n is the integral in Example 5.3.13. We have

$$I_n I_{n-1} = \frac{(n-1)(n-3) \cdots}{n(n-2) \cdots} \frac{(n-2)(n-4) \cdots}{(n-1)(n-3) \cdots} \frac{\pi}{2} = \frac{\pi}{2n}.$$

Combined with the values of V_1 and V_2 , we get

$$\begin{aligned} V_{2n} &= \frac{\pi}{n} \frac{\pi}{n-1} \cdots \frac{\pi}{2} V_2 = \frac{\pi^n}{n!}, \\ V_{2n+1} &= \frac{2\pi}{2n+1} \frac{2\pi}{2n-1} \cdots \frac{2\pi}{3} V_1 = \frac{2^{n+1} \pi^n}{(2n+1)!!}. \end{aligned}$$

On the other hand, we may decompose B^n into $(n-1)$ -dimensional spheres S_t of radius t and centered at the origin. The volume of S_t is $t^{n-1} A_{n-1}$. The family S_t is equidistant, and dt is the distance between nearby cross sections. Therefore

$$V_n = \int_0^1 t^{n-1} A_{n-1} \, dt = \frac{1}{n} A_{n-1}.$$

This implies

$$A_{2n-1} = 2nV_{2n} = \frac{2\pi^n}{(n-1)!}, \quad A_{2n} = (2n+1)V_{2n+1} = \frac{2^{n+1}\pi^n}{(2n-1)!!}. \quad \blacksquare$$

Exercises

7.1.1 Compute the arc length for each of the following curves:

1. $y^2 = 2px, \quad x \in [0, a];$
2. $x^2 = 2py, \quad x \in [0, a];$
3. $y = e^x, \quad x \in [0, a];$
4. $y = \ln x, \quad x \in [1, a];$
5. $y = \ln \cos x, \quad x \in [0, a], \quad a < \frac{\pi}{2};$
6. $y = \frac{e^x + e^{-x}}{2}, \quad x \in [-a, a];$
7. $\sqrt{x} + \sqrt{y} = 1, \quad x, y \geq 0;$
8. $y^2 = x^3, \quad x \in [0, a];$
9. $y^4 = x^3, \quad x \in [0, a].$

7.1.2 Find the area of the surface of revolution around the x -axis by each of the following curves:

1. $y^2 = 2px, \quad x \in [0, a];$
2. $x^2 = 2py, \quad x \in [0, a];$
3. $y = e^x, \quad x \in [0, a];$
4. $y = \tan x, \quad x \in [0, a], \quad a < \frac{\pi}{2};$
5. $9x^2 = (y^2 + 2)^3, \quad y \in [0, a];$
6. $y = \frac{e^x + e^{-x}}{2}, \quad x \in [-a, a];$
7. $x^{\frac{3}{2}} + y^{\frac{3}{2}} = 1;$
8. $y^2 = x^3, \quad x \in [0, a].$

7.1.3 Find the volume of the solid of revolution around the x -axis by each of the following curves:

1. $y^2 = 2px, \quad x \in [0, a];$
2. $x^2 = 2py, \quad x \in [0, a];$
3. $y = e^x, \quad x \in [0, a];$
4. $y = \ln x, \quad x \in [1, a];$
5. $y = \sin x, \quad x \in [0, \pi];$
6. $y = \frac{e^x + e^{-x}}{2}, \quad x \in [-a, a];$
7. $x^{\frac{3}{2}} + y^{\frac{3}{2}} = 1;$
8. $y^2 = x^3, \quad x \in [0, a].$

7.1.4 Suppose a circle is rolling along a horizontal line. The track of one point on the circle during the movement is the cycloid. If the circle has radius 1 and the angle of rotation is t , then the cycloid is given by $x = t - \sin t, y = 1 - \cos t$. One period of the cycloid corresponds to $t \in [0, 2\pi]$.

1. Find the length of one period of the cycloid.

2. Find the area of the surface obtained by revolving one period of the cycloid around the x -axis.
3. Find the volume of the solid obtained by revolving the region between one period of the cycloid and the x -axis around the x -axis.

7.1.5 Suppose a line is wrapped around a circle. The track of one point on the line when the line is unwrapped from the circle is the involute of the circle. If the circle has radius 1 and the angle of unwrapping is t , then the curve is given by $x = \cos t + t \sin t$, $y = \sin t - t \cos t$. Let C be part obtained by unwrapping by half of the circle, which corresponds to $t \in [0, \pi]$.

1. Find the length of C .
2. Find the area of the surface obtained by revolving C around the x -axis.
3. Find the volume of the solid obtained by revolving the region between C and the x -axis around the x -axis.

Note that the computation for the third part needs to be split into $t \in [0, \frac{\pi}{2}]$ and $t \in [\frac{\pi}{2}, \pi]$.

7.1.6 Find the volume of the solid of revolution in each of the following problems:

1. region bounded by $y^2 = x$ and $y = x^2$, around x -axis;
2. region below $y = e^{x^2}$, $x \in [0, 1]$, around y -axis;
3. region bounded by $y = x$ and $y = \sqrt{x}$, around $y = 1$;
4. region bounded by $y^2 = x$ and $y = x^2$, around $x = -1$;
5. $x^2 + y^2 = 1$, around $x + y = 2$.

7.1.7 Find the areas of the surfaces obtained by revolving the loop of the curve $3ay^2 = x(a - x)^2$, $a > 0$, around the x - and y -axes.

7.1.8 Find the volume of the solid obtained by cutting the cylinder $x^2 + y^2 \leq 1$ in $\mathbb{R}^3$ by the (x, y) -plane and another plane containing the x -axis.

7.2 POLAR COORDINATES

7.2.1 Curves in Polar Coordinates

The polar coordinates (r, θ) indicates a point on the Euclidean plane with the cartesian coordinates

$$x = r \cos \theta, \quad y = r \sin \theta.$$

Conversely, a point with cartesian coordinates (x, y) has polar coordinates given by

$$r = \sqrt{x^2 + y^2}, \quad (\cos \theta, \sin \theta) = \frac{(x, y)}{\sqrt{x^2 + y^2}}.$$

For a given point on the plane, the angle θ is unique only up to adding a multiple of 2π . Therefore (r, θ) and $(r, \theta + 2\pi)$ represent the same point. Moreover, it is often more convenient to allow r to be negative, and let $(-r, \theta)$ to represent the same point as $(r, \theta + \pi)$. We always need to be aware of the multiple ways of representing the same point in different polar coordinates.

Under the polar coordinates, a curve on the plane can often be divided into pieces, and each piece can be simply parametrized by the angle. Such a parametrization can be denoted as $r = r(\theta)$, called the **polar equation** of the curve. In the usual cartesian coordinates, this means the parametrization

$$x = r(\theta) \cos \theta, \quad y = r(\theta) \sin \theta.$$

In drawing the curves in polar coordinates, it is often useful to apply transformations.

The rotation of (r, θ) by angle α around the origin (in counterclockwise direction) gives $(r, \theta + \alpha)$. The rotation of a curve $r = r(\theta)$ gives the curve $r = r(\theta - \alpha)$. The curve is invariant under the rotation if $r(\theta) = r(\theta - \alpha)$. However, due to the multiple way of representing the same point in different polar coordinates, the curve is also invariant under the rotation if $r(\theta) = r(\theta - \alpha + 2\pi)$, or $r(\theta) = -r(\theta - \alpha + \pi)$, or $r(\theta) = r(\theta - \alpha - 2\pi)$, etc. Basically, the curve is invariant if one of (infinitely many) equations holds.

Note that the rotation by angle $\alpha = \pi$ is actually the reflection with respect to the origin. Therefore a curve $r = r(\theta)$ is symmetric with respect to the origin if $r(\theta) = r(\theta + n\pi)$ for some odd n or $r(\theta) = -r(\theta + n\pi)$ for some even n .

The reflection of (r, θ) with respect to the direction of angle α gives $(r, 2\alpha - \theta)$. The reflection of a curve $r = r(\theta)$ gives the curve $r = r(2\alpha - \theta)$. The curve is invariant under the reflection if $r(\theta) = r(2\alpha - \theta)$, or one of the many similar equations holds.

The dilation of (r, θ) gives (cr, θ) , where c is the scaling factor. Under the transform, a curve $r = r(\theta)$ becomes $r = cr(\theta)$, which basically expand or shrink (and reflect with respect to the origin if $c < 0$) the curve.

Example 7.2.1 The straight line passing through the origin is given by the polar equation $\theta = \text{constant}$ (this equation does not fit into the form $r = r(\theta)$). In particular, the x -axis is given by $\theta = 0$ and the y -axis is given by $\theta = \frac{\pi}{2}$. Note that the x -axis is also given by $\theta = n\pi$.

For a straight line L not passing through the origin, we draw a straight line segment $L^\perp$ that indicates the shortest distance from the origin to the straight line L . Let the length of $L^\perp$ be a and the angle of L be α . Then the polar equation for L is

$$r = a \sec(\theta - \alpha), \quad -\frac{\pi}{2} + \alpha < \theta < \frac{\pi}{2} + \alpha.$$

For example, the line $x = a$ is given by $r = a \sec \theta$ and the line $y = a$ is given by $r = a \sec\left(\theta - \frac{\pi}{2}\right) = a \csc \theta$. Moreover, for any n ,

$$r = a \sec(\theta - \alpha), \quad n\pi - \frac{\pi}{2} + \alpha < \theta < n\pi + \frac{\pi}{2} + \alpha$$

represents the same straight line. ■

Example 7.2.2 The circle with the origin as the center has polar equation $r = \text{constant}$, where the constant is the radius of the circle.

The circle passing through the origin and with center on the x -axis is given by the polar equation $r = 2a \cos \theta$. More generally, if the circle passes through the origin and the center has angle α , then the polar equation is

$$r = 2a \cos(\theta - \alpha), \quad -\frac{\pi}{2} + \alpha \leq \theta \leq \frac{\pi}{2} + \alpha.$$

For example, the circle passing through the origin and with center on the y -axis is given by $r = 2a \sin \theta$. Note that for any n ,

$$r = 2a \cos(\theta - \alpha), \quad n\pi - \frac{\pi}{2} + \alpha \leq \theta \leq n\pi + \frac{\pi}{2} + \alpha$$

represents the same circle.

Other than the two cases above, the polar equation for a circle is rather complicated. Suppose the center of the circle is at (b, α) and the radius is a , then the circle is given by $b^2 + r^2 - 2br \cos(\theta - \alpha) = a^2$. ■

Example 7.2.3 The equation $r = a + b\theta$ gives a **spiral curve**. The graph can be obtained from the standard spiral curve $r = \theta$ by first rotating by angle $-\frac{a}{b}$ and then dilating by factor b . When you turn the angle, the radius is increasing (when $b > 0$) or decreasing (when $b < 0$) at steady pace.

By $-r(\theta - \pi) = -a - b\theta + b\pi = r(2\alpha - \theta)$ for $\alpha = \frac{\pi}{2} - \frac{a}{b}$, the curve is symmetric with respect to the direction of angle $\frac{\pi}{2} - \frac{a}{b}$. ■

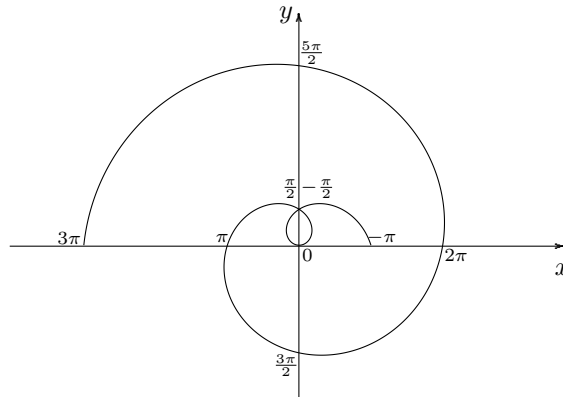
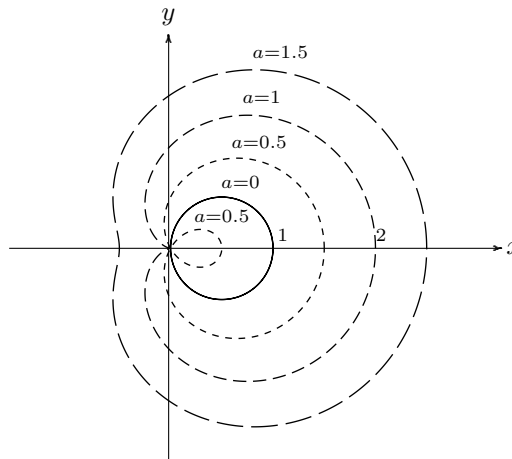


Fig. 7.9 spiral $r = \theta$

Example 7.2.4 The curve $r = 1 + \cos \theta$ is called the **cardioid** because it looks like a heart. The whole curve is traversed once when θ changes from $-\pi$ to π . The cardioid is symmetric with respect to the x -axis because $r = 1 + \cos \theta$ implies $r = 1 + \cos(2 \cdot 0 - \theta)$.

The curve $r = 1 - \cos \theta = 1 + \cos(\theta - \pi)$ is a cardioid obtained by rotating the standard cardioid $r = 1 + \cos \theta$ by angle π , or reflection with respect to the origin.

The curves $r = 1 + \sin \theta$ and $r = 2 + 2 \cos \theta$ are also cardioids, obtained from the standard one by rotating angle $\frac{\pi}{2}$ and by expanding by factor 2.

Fig. 7.10 limaçon $r = a + \cos \theta$

The cardioid belongs to a family of curves $r = a + b \cos \theta$ called **limaçons**. The cardioid corresponds to $a = b$, and the circle corresponds to $a = 0$ or $b = 0$. ■

Example 7.2.5 The curve $r = \cos 2\theta$ satisfies

$$-r\left(\theta + \frac{\pi}{2}\right) = -\cos(2\theta + \pi) = \cos 2\theta = r(\theta).$$

This implies that the curve is invariant after rotating angle $\frac{\pi}{2}$ around the origin. Therefore the curve can be obtained as follows: The part of the curve between the angles $-\frac{\pi}{4}$ and $\frac{\pi}{4}$ looks like a leaf L_1 . By rotating the leaf by angles $\frac{\pi}{2}$, π and $\frac{3\pi}{2}$, we get the second leaf L_2 , the third leaf L_3 and the fourth leaf L_4 . The whole curve is a four-leaved rose.

On close inspection, when θ changes from $-\frac{\pi}{4}$ to $\frac{\pi}{4}$, we get the first leaf L_1 . When θ changes from $\frac{\pi}{4}$ to $\frac{3\pi}{4}$, we get the fourth leaf L_4 (not L_2 , because $r \leq 0$). Next we get the third leaf L_3 for θ between $\frac{3\pi}{4}$ and $\frac{5\pi}{4}$. Finally, we get the second leaf L_2 for θ between $\frac{5\pi}{4}$ and $\frac{7\pi}{4}$.

From the picture, we see the curve is also symmetric with respect to the x -axis. This is analytically justified by

$$r(-\theta) = \cos(-2\theta) = \cos 2\theta = r(\theta).$$

Altogether, the 4-leaved rose has eight symmetries. All symmetries have been verified because they are generated by the rotation of angle $\frac{\pi}{2}$ and the reflection in x -axis.

In general, for even n , the polar equation $r = \cos n\theta$ represents the $2n$ -leaved rose. This is obtained by rotating one leaf (the part between angles $-\frac{\pi}{2n}$ and $\frac{\pi}{2n}$) by the angle $\frac{\pi}{n}$ repeatedly for $2n$ times.

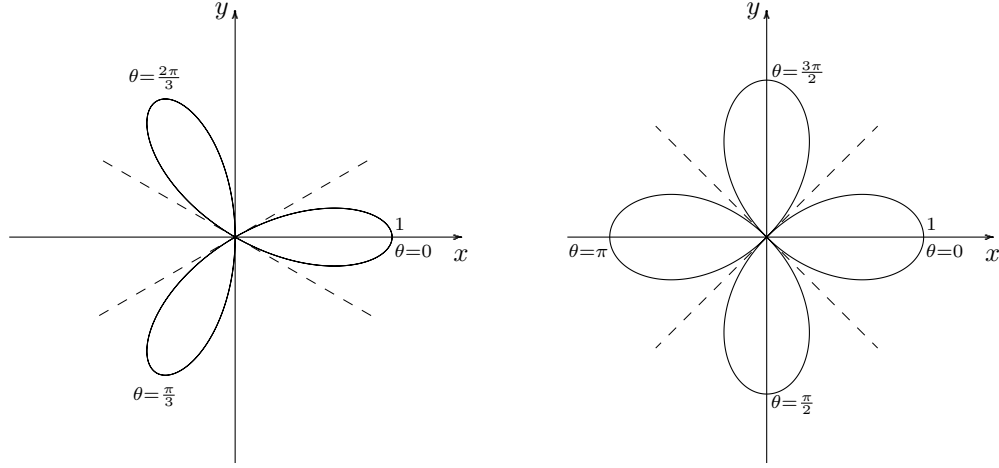


Fig. 7.11 3-leaved and 4-leaved roses

On the other hand, for odd n , the polar equation $r = \cos n\theta$ represents the n -leaved rose. This is obtained by rotating one leaf (the part between angles $-\frac{\pi}{4n}$ and $\frac{\pi}{4n}$) by the angle $\frac{2\pi}{n}$ repeatedly for n times. Although the whole angle 2π is divided into $2n$ equal parts of angle $\frac{\pi}{n}$, the curve appears only in n parts. One can see this by considering the case $n = 3$. Moreover, we note that the case $n = 1$ is the circle of radius $\frac{1}{2}$, which can be considered a one-leaf rose. ■

Example 7.2.6 Let $d, e > 0$. Consider the points on the plane such that the distance from the point to the origin is e multiple of the distance to the straight line $x = d$. Such points are given by the polar equation

$$r = e(d - r \cos \theta),$$

or

$$r = \frac{ed}{1 + e \cos \theta}.$$

In the cartesian coordinates, the curve is given by the equation

$$\sqrt{x^2 + y^2} \left(1 + e \frac{x}{\sqrt{x^2 + y^2}} \right) = ed.$$

This is the same as $x^2 + y^2 = e^2(d - x)^2$. After completing the square, we get

$$(1 - e^2) \left(x + \frac{e^2 d}{1 - e^2} \right)^2 + y^2 = \frac{e^2 d^2}{1 - e^2}.$$

According to the sign of $1 - e^2$, we get three possibilities.

- If $0 < e < 1$, then we have an ellipse of the form $\frac{(x+c)^2}{a^2} + \frac{y^2}{b^2} = 1$, with

$$a = \frac{ed}{1-e^2}, \quad b = \frac{ed}{\sqrt{1-e^2}}.$$

Since $\sqrt{a^2 - b^2} = \frac{e^2 d}{1 - e^2} = c$, we see that the origin is the focus of the ellipse.

- If $e > 1$, then we have (one branch of) a hyperbola of the form $\frac{(x-c)^2}{a^2} - \frac{y^2}{b^2} = 1$, with

$$a = \frac{ed}{e^2 - 1}, \quad b = \frac{ed}{\sqrt{e^2 - 1}}.$$

Since $\sqrt{a^2 + b^2} = \frac{e^2 d}{e^2 - 1} = c$, we see that the origin is the focus of the hyperbola.

- If $e = 1$, then we get a parabola $y^2 = d^2 - 2dx$. The parabola intersects the x -axis at $\frac{d}{2}$. this implies that the origin is the focus of the parabola.

In conclusion, the polar curve $r = \frac{ed}{1 + e \cos \theta}$ is a conic section, and the origin is the **focus** of the curve. The line $x = d$ is the **directrix** and the ratio e is the **eccentricity** of the conic section, and can be recovered from the standard cartesian presentation by

$$e = \frac{c}{a}, \quad d = \frac{b^2}{c}. \quad \blacksquare$$

7.2.2 Length and Area in Polar Coordinates

The length of a curve $r = r(\theta)$, $a \leq \theta \leq b$, under the polar coordinates is

$$L = \int_a^b \sqrt{(r(\theta) \cos \theta)'^2 + (r(\theta) \sin \theta)'^2} d\theta = \int_a^b \sqrt{r^2 + r'^2} d\theta.$$

The polar curve and the origin together also form a wedged region bounded by the curve and two rays $\theta = a$ and $\theta = b$. To find the area of the region, we let $A(\theta)$ be the area of the part between angles a and θ . When the angle θ is changed by $\Delta\theta$, the corresponding area is changed by $\Delta A = A(\theta + \Delta\theta) - A(\theta)$. Let M and m be the supremum and the infimum of $r(\theta)$ on $[\theta, \theta + \Delta\theta]$. Then the part of the wedge between θ and $\theta + \Delta\theta$ is sandwiched between two fans of angle $\Delta\theta$ with respective radii M and m . Since the fan of angle α and radius R has area $\frac{1}{2}R^2\alpha$, we get

$$\frac{1}{2}m^2|\Delta\theta| \leq |\Delta A| \leq \frac{1}{2}M^2|\Delta\theta|.$$

Considering the signs of ΔA and $\Delta\theta$, this implies

$$\frac{1}{2}m^2 \leq \frac{\Delta A}{\Delta\theta} \leq \frac{1}{2}M^2.$$

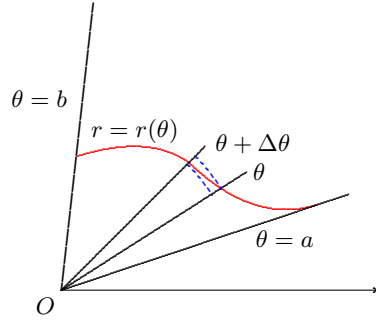


Fig. 7.12 area of polar wedge

If $r(\theta)$ is continuous, then $\lim_{\Delta\theta \rightarrow 0} m = \lim_{\Delta\theta \rightarrow 0} M = r(\theta)$. Therefore by the sandwich rule, we get

$$\frac{dA}{d\theta} = \lim_{\Delta\theta \rightarrow 0} \frac{\Delta A}{\Delta\theta} = \frac{1}{2}[r(\theta)]^2,$$

and the area of the whole wedge is

$$A(b) - A(a) = \frac{1}{2} \int_a^b [r(\theta)]^2 d\theta.$$

In fact, the formula also works for integrable $r(\theta)$.

Example 7.2.7 The length of the cardioid $r = 1 + \cos \theta$ in Example 7.2.4 is

$$\int_{-\pi}^{\pi} \sqrt{(1 + \cos \theta)^2 + (\sin \theta)^2} d\theta = \int_{-\pi}^{\pi} \sqrt{2(1 + \cos \theta)} d\theta = \int_{-\pi}^{\pi} 2 \cos \frac{\theta}{2} d\theta = 8.$$

The area enclosed by the cardioid is

$$\int_{-\pi}^{\pi} \frac{1}{2} (1 + \cos \theta)^2 d\theta = \frac{3}{2} \pi. \quad \blacksquare$$

Example 7.2.8 We would like to find the area of the region inside the cardioid $r = 1 - \cos \theta$ and outside the circle $r = 1$, illustrated in Fig. 7.2.2. The intersections of the two curves can be obtained by solving $1 - \cos \theta = 1$, which shows that the region lies between the angles $\frac{\pi}{2}$ and $\frac{3\pi}{2}$. Therefore the area of the region is

$$\frac{1}{2} \int_{\frac{\pi}{2}}^{\frac{3\pi}{2}} ((1 - \cos \theta)^2 - 1^2) d\theta = \frac{\pi}{4} + 2. \quad \blacksquare$$

Example 7.2.9 The arc length of one leaf of the 3-leaved rose in Fig. 7.11 is

$$\int_{-\frac{\pi}{6}}^{\frac{\pi}{6}} \sqrt{\cos^2 3\theta + 9 \sin^2 3\theta} d\theta = 3 \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sqrt{1 + 8 \sin^2 t} dt = 3 \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sqrt{5 - 4 \cos 2t} dt = \frac{3}{2} \int_0^{2\pi} \sqrt{5 - 4 \cos x} dx.$$

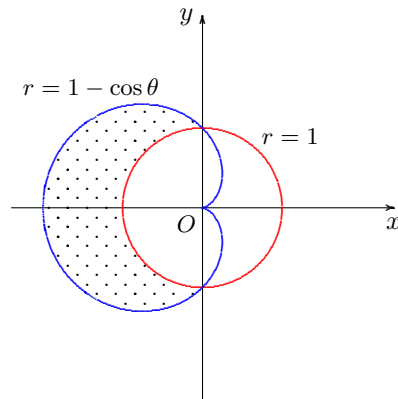


Fig. 7.13 cardioid and circle

This is an elliptic integral that cannot be evaluated in elementary way.

On the other hand, the area of one leaf is

$$\frac{1}{2} \int_{-\frac{\pi}{6}}^{\frac{\pi}{6}} \cos^2 3\theta \, d\theta = \frac{\pi}{12}.$$

■

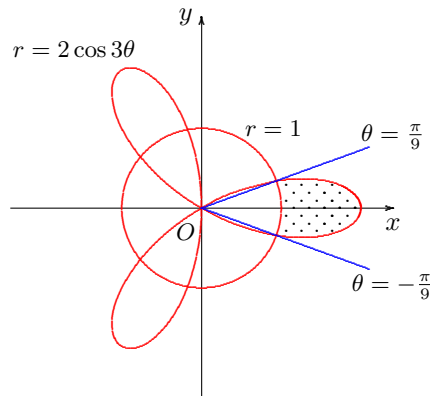


Fig. 7.14 3-leaved rose and circle

Example 7.2.10 We would like to find the area of the region outside the unit circle and inside the 3-leaved rose $r = 2 \cos 3\theta$, illustrated in Fig. 7.14. By solving $2 \cos 3\theta$, we get the angles $-\frac{\pi}{9}$ and $\frac{\pi}{9}$

at the two intersection points. Therefore the region has area

$$\frac{1}{2} \int_{-\frac{\pi}{9}}^{\frac{\pi}{9}} [(2 \cos 3\theta)^2 - 1^2] d\theta = \frac{1}{2\sqrt{3}} + \frac{\pi}{9}. \quad \blacksquare$$

Example 7.2.11 We want 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. As θ increases from 0, curve moves as follows

$$A(\theta = 0) \rightarrow B(\theta = \alpha) \rightarrow O(\theta = \beta) \rightarrow C(\theta = \pi) \rightarrow O(\theta = 2\pi - \beta) \rightarrow \dots,$$

where

$$\alpha = \pi - \arccos \frac{a}{2}, \quad \beta = \pi - \arccos a.$$

The region to be rotated is bounded by the arcs AB , BO , CO (corresponding to angle interval $[\pi, 2\pi - \beta]$, not the arc OC), and the x -axis. The region is the region between the arc AB and the x -axis, subtracting the region between the arc BO and the x -axis, and further subtracting the region between the arc CO and the x -axis. The volume we are interested can be computed by similarly combining the volumes of solids of revolutions of these regions.

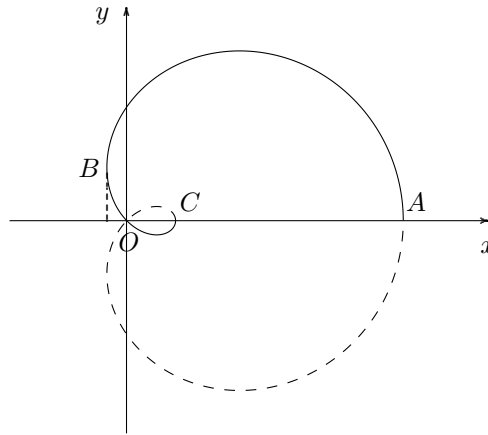


Fig. 7.15 revolving limaçon around x -axis

The volume of the solid of revolution is given by

$$\int \pi y^2 dx = \int \pi (a + \cos \theta)^2 \sin^2 \theta d[(a + \cos \theta) \cos \theta].$$

Let us consider

$$\int_{\theta=0}^{\theta=\pi} \pi y^2 dx = \int_0^{\alpha} \pi y^2 dx + \int_{\alpha}^{\beta} \pi y^2 dx + \int_{\beta}^{\pi} \pi y^2 dx.$$

Since x is decreasing along the arc AB , the integral $\int_0^{\alpha} \pi y^2 dx$ is the negative of volume of the solid of revolution of the region between AB and the x -axis. Since x is decreasing along the arcs BO and

OC , the integrals $\int_{\alpha}^{\beta} \pi y^2 dx$ and $\int_{\beta}^{\pi} \pi y^2 dx$ are the (positive) volumes of the solids of revolution of the regions between BO , OC and the x -axis. Moreover, by the symmetry, $\int_{\beta}^{\pi} \pi y^2 dx$ is also the volume of the solid of revolution of the region between CO and the x -axis. We conclude that the volume we are interested in is

$$\begin{aligned}
 - \int_{\theta=0}^{\theta=\pi} \pi y^2 d\theta &= - \int_0^{\pi} \pi (a + \cos \theta)^2 \sin^2 \theta d[(a + \cos \theta) \cos \theta] \\
 &= - \int_0^{\pi} \pi (a + \cos \theta)^2 (1 - \cos^2 \theta) (a + 2 \cos \theta) d(\cos \theta) \\
 &= - \int_1^{-1} \pi (a + t)^2 (1 - t^2) (a + 2t) dt \\
 &= \int_{-1}^1 \pi (a^2 + 2at + t^2) (a + 2t) (1 - t^2) dt \\
 &= \int_{-1}^1 \pi (a(a^2 + t^2) + 4at^2) (1 - t^2) dt \\
 &= 2\pi a \left(a^2 + \frac{1}{3}(5 - a^2) - 1 \right) = \frac{4}{3}\pi a(a^2 + 1).
 \end{aligned}$$

■

Exercises

7.2.1 Sketch the region for each of the following problems in polar coordinates:

1. $r > 2$, $0 < \theta < \pi$;
2. $1 \leq r \leq 2$, $1 \leq \theta \leq 2$;
3. $r \geq 0$, $\theta^2 > 1$;
4. $r^2 > 1$, $\frac{\pi}{2} < \theta < \pi$.

7.2.2 Find the distance between the two points in polar coordinates.

7.2.3 Sketch the graph and find the cartesian equation:

1. $r = -\cos \theta$;
2. $r = \epsilon + \cos \theta$, $\epsilon > 0$ small;
3. $r = -\epsilon + \cos \theta$, $\epsilon > 0$ small;
4. $r = \cos \theta + \sin \theta$;
5. $r = \sqrt{2} + \cos \theta + \sin \theta$;
6. $r = -\theta$;
7. $r = \theta^2$;
8. $r\theta = 1$;
9. $r = \sin 2\theta$;
10. $r = \sin 3\theta$.

7.2.4 Sketch the graphs of $r = \cos \frac{\theta}{2}$, $r = \cos \frac{2\theta}{3}$ and $r = \sin \frac{\theta}{2}$. Then explain the graph of $r = \cos \lambda\theta$ for a rational coefficient λ .

7.2.5 Find the length for each of the following curves:

1. $r = \sqrt{2} + \cos \theta + \sin \theta$;
2. $r = \sin \theta$, $0 \leq \theta \leq a$, where $0 < a \leq \pi$;
3. $r = \theta$, $0 \leq \theta \leq \pi$;
4. $r = \theta^2$, $0 \leq \theta \leq a$, where $0 < a \leq 2\pi$.

7.2.6 Find the area for each of the following problems:

1. region enclosed by $r = \theta$, $0 \leq \theta \leq \pi$;
2. region enclosed by the cardioid $r = 1 + \cos \theta$ in the first quadrant;
3. one leaf in the 8-leaved rose $r = \sin 4\theta$;
4. one leaf in the 7-leaved rose $r = \cos 7\theta$;
5. region enclosed by $r^2 = \sin 2\theta$;
6. region inside $r = a \sin \theta$ and outside $r = b \cos \theta$;
7. region inside $r = 2 + \cos \theta$ and outside $r = 3 \cos \theta$.

7.3 PHYSICAL APPLICATIONS

7.3.1 Center of Mass

Consider a particle of mass m lying at the coordinate x on a horizontal line. The **moment** of the particle with respect to the coordinate ξ on the line is $m(x - \xi)$. A system of n particles with masses $m_1, m_2, \dots, m_n$ at $x_1, x_2, \dots, x_n$ has the total moment $\sum m_i(x_i - \xi)$ with respect to ξ . The **center of the mass** $\bar{x}$ for the system is the point such that the total moment with respect to the point is zero (i.e., the system is **balanced** with respect to the point). By solving $\sum m_i(x_i - \xi) = 0$ for ξ , we get the formula

$$\bar{x} = \frac{\sum m_i x_i}{\sum m_i}$$

for the center of mass.

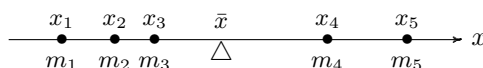


Fig. 7.16 center of mass for a discrete system

The center of mass has the distribution property. Suppose a discrete system is divided into a system S of $m_1, m_2, \dots, m_n$ at $x_1, x_2, \dots, x_n$ and another system S' of $m'_1, m'_2, \dots, m'_{n'}$ at $x'_1, x'_2, \dots, x'_{n'}$. Then S and S' have masses

$$m_S = \sum m_i, \quad m_{S'} = \sum m'_{i'},$$

and centers of mass

$$\bar{x}_S = \frac{\sum m_i x_i}{\sum m_i}, \quad \bar{x}_{S'} = \frac{\sum m'_{i'} x'_{i'}}{\sum m'_{i'}}.$$

Then the center of mass of the whole system is

$$\bar{x}_{S \cup S'} = \frac{\sum m_i x_i + \sum m'_{i'} x'_{i'}}{\sum m_i + \sum m'_{i'}} = \frac{m_S \bar{x}_S + m_{S'} \bar{x}_{S'}}{m_S + m_{S'}}.$$

The right side is the center of mass for a synthetic system of two particles of masses $m_S, m_{S'}$ at $\bar{x}_S$ and $\bar{x}_{S'}$.

The distribution property can be extended to the division of the system into more parts. Suppose a system S is divided into disjoint systems $S_1, S_2, \dots, S_k$. Suppose each S_i has total mass m_{S_i} and center of mass $\bar{x}_{S_i}$. Then the center of mass for S is the same as the center of mass of the synthetic system of masses $m_{S_1}, m_{S_2}, \dots, m_{S_k}$ at $\bar{x}_{S_1}, \bar{x}_{S_2}, \dots, \bar{x}_{S_k}$.

Now consider a rod lying horizontally. We attach a coordinate system to the rod and assume the rod extends from $x = a$ to $x = b$. The rod has a continuously distributed mass given by the **density function** $\rho(x)$. The total mass of the rod is $\int_a^b \rho(x) dx$.

To find the moment of the rod with respect to ξ , we take a partition P of the rod (or the interval $[a, b]$). Each part $[x_{i-1}, x_i]$ of the rod has mass approximated by

$$\Delta m_i \approx \rho(x_i^*) \Delta x_i, \quad x_i^* \in [x_{i-1}, x_i]$$

and the moment with respect to ξ approximated by

$$(x_i^* - \xi) \Delta m_i \approx (x_i^* - \xi) \rho(x_i^*) \Delta x_i.$$

The total moment is then approximated by

$$\sum (x_i^* - \xi) \Delta m_i \approx \sum (x_i^* - \xi) \rho(x_i^*) \Delta x_i,$$

which is the Riemann sum of the function $(x - \xi)\rho(x)$ on $[a, b]$. By taking the limit as $\|P\| \rightarrow 0$, we get the moment $\int_a^b (x - \xi)\rho(x) dx$. By solving $\int_a^b (x - \xi)\rho(x) dx = 0$, we see that the rod is balanced with respect to ξ if ξ is the **center of mass**

$$\bar{x} = \frac{\int_a^b x \rho(x) dx}{\int_a^b \rho(x) dx}.$$

The center of mass in the continuous setting also has the distribution property. For example, suppose $[a, b]$ is divided into two parts $[a, c]$ and $[c, b]$. Then masses and centers of mass of the parts are

$$m_{[a,c]} = \int_a^c \rho(x) \, dx, \quad m_{[c,b]} = \int_c^b \rho(x) \, dx, \quad \bar{x}_{[a,c]} = \frac{\int_a^c x\rho(x) \, dx}{\int_a^c \rho(x) \, dx}, \quad \bar{x}_{[c,b]} = \frac{\int_c^b x\rho(x) \, dx}{\int_c^b \rho(x) \, dx}.$$

By $\int_a^b = \int_a^c + \int_c^b$, we get

$$\bar{x}_{[a,b]} = \frac{m_{[a,c]}\bar{x}_{[a,c]} + m_{[c,b]}\bar{x}_{[c,b]}}{m_{[a,c]} + m_{[c,b]}}.$$

Example 7.3.1 For a rod of constant density $\rho(x) = \rho$, the center of mass is

$$\bar{x} = \frac{\int_a^b x\rho \, dx}{\int_a^b \rho \, dx} = \frac{\rho \frac{b^2 - a^2}{2}}{\rho(b - a)} = \frac{a + b}{2},$$

which is the middle point of the rod. If the mass is linearly increasing, then $\rho(x) = \lambda + \mu x$, and the center of mass is

$$\bar{x} = \frac{\int_a^b x(\lambda + \mu x) \, dx}{\int_a^b (\lambda + \mu x) \, dx} = \frac{\lambda \frac{b^2 - a^2}{2} + \mu \frac{b^3 - a^3}{3}}{\lambda(b - a) + \mu \frac{b^2 - a^2}{2}} = \frac{\lambda \frac{a + b}{2} + \mu \frac{a^2 + ab + b^2}{3}}{\lambda + \mu \frac{a + b}{2}}. \quad \blacksquare$$

The discussion of the center of mass can be extended to higher dimensions. For example, we may consider a discrete system in the Euclidean plane, consisting of n masses $m_1, m_2, \dots, m_n$ at $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. With respect to a point (ξ, η) in the plane, we have the total moments

$$\sum m_i(x_i - \xi), \quad \sum m_i(y_i - \eta)$$

in the x - as well as y -directions. The point (ξ, η) is the center of mass when both moments vanish. This gives us the center of mass $(\bar{x}, \bar{y})$ for the system given by

$$\bar{x} = \frac{\sum m_i x_i}{\sum m_i}, \quad \bar{y} = \frac{\sum m_i y_i}{\sum m_i}.$$

Similar formula may be derived for a discrete system in the 3-dimensional Euclidean space. Moreover, the distribution property also extends to (each coordinate of) higher dimension.

Now we consider a continuous system of mass distributed along a curve. Suppose the wire lies in the Euclidean plane and is parametrized as $x = x(t)$, $y = y(t)$ for $t \in [a, b]$. Suppose the curve has

density function $\rho(t)$. Then for a partition P of the wire (or the interval $[a, b]$ for the parameter), we have the approximate mass

$$\Delta m_i \approx \rho(t_i^*) \Delta s_i, \quad t_i^* \in [t_{i-1}, t_i]$$

for each segment and the approximate total moments with respect to (ξ, η) in x - and y -directions

$$\sum (x(t_i^*) - \xi) \rho(t_i^*) \Delta s_i, \quad \sum (y(t_i^*) - \eta) \rho(t_i^*) \Delta s_i.$$

As $\|P\| \rightarrow 0$, we get the total moments

$$\begin{aligned} \int_a^b (x - \xi) \rho \, ds &= \int_a^b (x(t) - \xi) \rho(t) \sqrt{x'(t)^2 + y'(t)^2} \, dt, \\ \int_a^b (y - \eta) \rho \, ds &= \int_a^b (y(t) - \eta) \rho(t) \sqrt{x'(t)^2 + y'(t)^2} \, dt. \end{aligned}$$

By letting both moments vanish and solving for ξ and η , we get the center of mass $(\bar{x}, \bar{y})$ for the system given by

$$\begin{aligned} \bar{x} &= \frac{\int_a^b x \rho \, ds}{\int_a^b \rho \, ds} = \frac{\int_a^b x(t) \rho(t) \sqrt{x'(t)^2 + y'(t)^2} \, dt}{\int_a^b \rho(t) \sqrt{x'(t)^2 + y'(t)^2} \, dt}, \\ \bar{y} &= \frac{\int_a^b y \rho \, ds}{\int_a^b \rho \, ds} = \frac{\int_a^b y(t) \rho(t) \sqrt{x'(t)^2 + y'(t)^2} \, dt}{\int_a^b \rho(t) \sqrt{x'(t)^2 + y'(t)^2} \, dt}. \end{aligned}$$

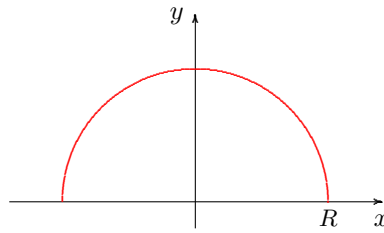


Fig. 7.17 semi-circle

Example 7.3.2 Consider a semi-circular wire of constant density. We have

$$x = R \cos \theta, \quad y = R \sin \theta, \quad 0 \leq \theta \leq \pi,$$

and $\rho(\theta) = \rho$ is a constant. Then $ds = R d\theta$ and the center of mass is

$$\bar{x} = \frac{\int_0^\pi R \cos \theta \rho R d\theta}{\int_0^\pi \rho R d\theta} = 0, \quad \bar{y} = \frac{\int_0^\pi R \sin \theta \rho R d\theta}{\int_0^\pi \rho R d\theta} = R \frac{2}{\pi}. \quad \blacksquare$$

Next we study the center of mass for a flat plate of *constant density*. Suppose the flat plate is the region between functions $f(x)$ and $g(x)$ for $x \in [a, b]$.

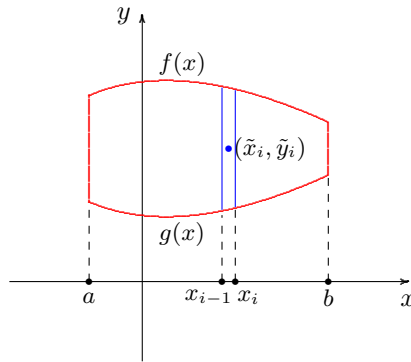


Fig. 7.18 flat plate

Let P be a partition of the interval $[a, b]$. The mass of the strip over $[x_{i-1}, x_i]$ is approximately

$$\Delta m_i \approx \rho(f(x_i^*) - g(x_i^*))\Delta x_i, \quad x_i^* = [x_{i-1}, x_i].$$

Since the density is constant, by Example 7.3.1, the center of the mass for the strip is approximately given by

$$\bar{x}_i \approx x_i^*, \quad \bar{y}_i \approx \frac{1}{2}(f(x_i^*) + g(x_i^*)).$$

Since the whole plate is made up of all such strips, by the distribution property, the center of mass for the whole plate is approximated by

$$\begin{aligned} \bar{x} &\approx \frac{\sum \bar{x}_i \Delta m_i}{\sum \Delta m_i} \approx \frac{\sum \rho x_i^* (f(x_i^*) - g(x_i^*)) \Delta x_i}{\sum \rho (f(x_i^*) - g(x_i^*)) \Delta x_i}, \\ \bar{y} &\approx \frac{\sum \bar{y}_i \Delta m_i}{\sum \Delta m_i} \approx \frac{\frac{1}{2} \sum \rho (f(x_i^*)^2 - g(x_i^*)^2) \Delta x_i}{\sum \rho (f(x_i^*) - g(x_i^*)) \Delta x_i}. \end{aligned}$$

Taking the limit as $\|P\| \rightarrow 0$, we get

$$\bar{x} = \frac{\int_a^b x(f(x) - g(x)) dx}{\int_a^b (f(x) - g(x)) dx}, \quad \bar{y} = \frac{\frac{1}{2} \int_a^b (f(x)^2 - g(x)^2) dx}{\int_a^b (f(x) - g(x)) dx}.$$

Example 7.3.3 Consider the semi-disk of constant density. The disk lies between $f(x) = \sqrt{R^2 - x^2}$ and $g(x) = 0$ over $x \in [-R, R]$. To compute the center of mass, we note that

$$\int_{-R}^R (f(x) - g(x)) \, dx = \frac{\pi}{2} R^2$$

is the area of the semi-disk, and

$$\begin{aligned} \int_{-R}^R (f(x) - g(x)) \, dx &= \int_{-R}^R x\sqrt{R^2 - x^2} \, dx = 0, \\ \frac{1}{2} \int_{-R}^R ([f(x)]^2 - [g(x)]^2) \, dx &= \frac{1}{2} \int_{-R}^R (R^2 - x^2) \, dx = \frac{2}{3} R^3. \end{aligned}$$

Therefore the center of mass is given by

$$\bar{x} = 0, \quad \bar{y} = \frac{4}{3\pi} R. \quad \blacksquare$$

7.3.2 Work

Suppose a constant force is applied to an object, so that the object moves a distance d in the direction of the force. Then the **work** done by the force is

$$W = Fd$$

is the product of the force and the displacement.

What if the force varies? Suppose the object moves from a to b along the x -axis, and the force is $F(t)$ when the object is located at t . Then we partition the movement into small increments. In other words, we consider a partition P of the interval $[a, b]$. The work done by the force in moving the object from x_{i-1} to x_i is approximately $F(x_i^*)\Delta x_i$. Therefore the total work is approximately $\sum F(x_i^*)\Delta x_i$. When $\|P\| \rightarrow 0$, we get the work

$$W = \int_a^b F(x) \, dx$$

done by the force in moving the object from a to b .

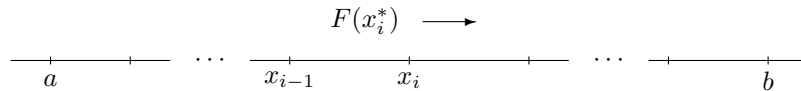


Fig. 7.19 work done by a variable force

Example 7.3.4 Suppose one end of spring is fixed and the other end is attached to an object. When the spring is neither stretched nor compressed, the spring exercises no force on the object and the object is in the natural position, which we denote as the origin.

Hooke's law says that the force exercised by the spring is proportional to the distance the spring is stretched. In other words, when the object is at distance x from its natural position, the spring exercises force

$$F(x) = kx$$

on the object, where k is the **spring constant**. 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

$$\int_0^a kx \, dx = \frac{k}{2}a^2. \quad \blacksquare$$

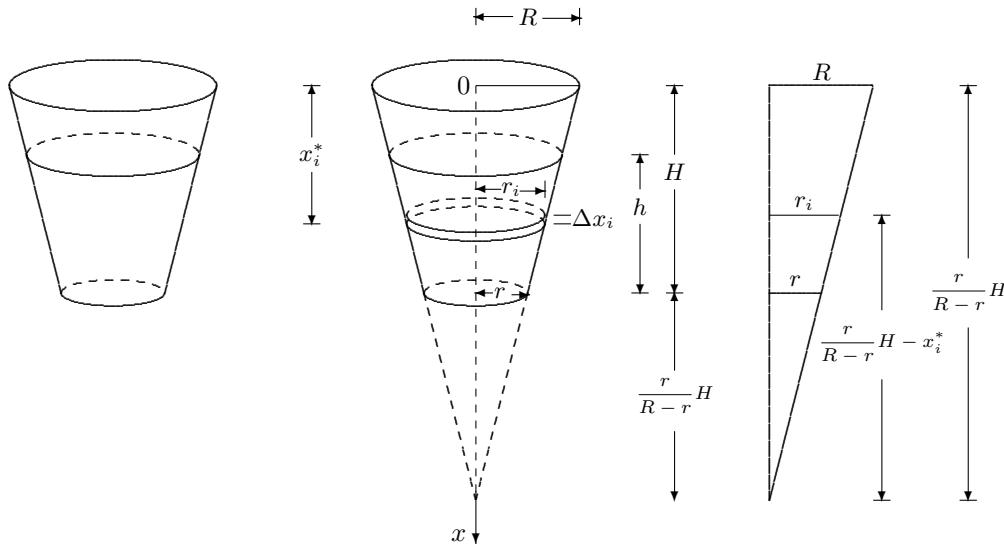


Fig. 7.20 a frustum container

Example 7.3.5 A frustum of a cone of height H , top radius R and base radius r contains some liquid of density ρ . Suppose that the depth of the liquid in the frustum is h . How much work does it take to pump the liquid to the top of the container?

Let x be the distance to the top of the container. Then the liquid extends from $x = H - h$ to $x = H$. The radius $r(x)$ of the disk at depth x satisfies

$$\frac{r(x) - r}{R - r} = \frac{H - x}{H}.$$

Therefore

$$r(x) = (R - r) \frac{H - x}{H} + r.$$

Let P be a partition of the interval $[H - h, H]$. Then the thin layer corresponding to $[x_{i-1}, x_i]$ has approximate volume

$$\Delta V_i \approx \pi r(x_i^*)^2 \Delta x_i,$$

and the work needed to pump the layer to the top of the container is approximately (g is the gravitational constant)

$$\Delta W_i \approx (g\rho \Delta V_i)x_i^* = \pi g\rho x_i^* r(x_i^*)^2 \Delta x_i.$$

Therefore the total work needed to pump all the liquid to the top of container is approximately

$$\sum \Delta W_i = \pi g\rho \sum x_i^* r(x_i^*)^2 \Delta x_i.$$

Taking the limit as $\|P\| \rightarrow 0$, we get the total work

$$\begin{aligned} W &= \pi g\rho \int_{H-h}^H x[r(x)]^2 dx = \frac{\pi g\rho}{H^2} \int_{H-h}^H x[(R-r)(H-x) + rH]^2 dx \\ &= \pi g\rho H^2 R^2 \left(a^2 b + \frac{1}{2} a(2a - 3b)b^2 + \frac{1}{3} (1-a)(1-3a)b^3 - \frac{1}{4} (1-a)^2 b^4 \right), \end{aligned}$$

where $a = \frac{r}{R}$ and $b = \frac{h}{H}$. ■

Exercises

7.3.1 Find the center of mass of the parabola $y = x^2$, $x \in [0, 2]$, of constant density.

7.3.2 Find the center of mass of a triangle (not including interior) of constant density and with vertices at $(-1, 0)$, $(0, \sqrt{15})$ and $(7, 0)$.

7.3.3 Find the center of mass of the triangular flat plate with constant density and the vertices at $(0, 0)$, $(a, 0)$, $(0, b)$. Then use this to find the center of mass of a general triangular flat plate with constant density.

7.3.4 Find the center of mass of the region bounded by the given curves:

1. $y = x$, $y = x^2$;

3. $y = \frac{1}{x}$, $y = 0$, $x \in [0, 1]$;

2. $y = \sin x$, $y = \cos x$, $x \in \left[\frac{\pi}{4}, \frac{3\pi}{4}\right]$;

4. $y = e^x$, $y = 0$, $x \in [0, 1]$.

7.3.5 Suppose a flat plate is divided into four parts of equal weight. If the centers of mass for the first two pieces are (a, b) and (c, d) , what is the center of mass for the remaining two pieces together?

7.3.6 A spring has natural length a . If force F is needed to stretch the spring to length b , how much work is needed in order to stretch the spring from the natural length to the length b ?

7.3.7 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?

7.3.8 A cable of mass m with length l hangs vertically from the top of a building. How much work is required to lift the whole cable to the top of the building?

7.3.9 Newton's law of gravitation says that two bodies with masses m_1 and m_2 attract each other with a force

$$F = \frac{gm_1m_2}{d^2},$$

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?

7.4 SUMMARY

Applications

- **Arc Length**

For a given curve

$$C : x = x(t), \quad y = y(t) \quad (a \leq t \leq b),$$

the arc length is

$$L = \int_a^b \sqrt{x'(t)^2 + y'(t)^2} dt.$$

In particular, for $y = f(x)$, $a \leq x \leq b$, the arc length is

$$L = \int_a^b \sqrt{1 + f'(x)^2} dx.$$

- **Area of Surface of Revolution**

Let $f(x) \geq 0$ be a continuous function defined over the interval $[a, b]$. Then the surface area of revolution for the graph of $f(x)$ is

$$S = 2\pi \int_a^b f(x) \sqrt{1 + f'(x)^2} dx.$$

- **Volume of Revolution**

Let $f(x) \geq 0$ be a continuous function defined over the interval $[a, b]$. Then the volume of revolution for the graph of $f(x)$ is

$$V = \pi \int_a^b [f(x)]^2 dx.$$

- **Arc Length of a Polar Curve**

For a polar curve $r = r(\theta)$, $a \leq \theta \leq b$, the arc length is

$$L = \int_a^b \sqrt{[r(\theta)]^2 + [r'(\theta)]^2} \, d\theta.$$

- **Area Bounded by a Polar Curve**

For a wedged region bounded by a polar curve $r = r(\theta)$ and two rays $\theta = a$ and $\theta = b$, its area is

$$A = \int_a^b \frac{1}{2} [r(\theta)]^2 \, d\theta.$$

- **Center of Mass**

– 1D:

If the density function is $\rho(x)$, then the center of mass is located at

$$\bar{x} = \frac{\int_a^b x \rho(x) \, dx}{\int_a^b \rho(x) \, dx}.$$

– 2D:

If the mass is evenly distributed on the plate that is bounded by two curves $y = f_1(x)$ and $y = f_2(x)$ between $x = a$ and $x = b$, then the center of mass $(\bar{x}, \bar{y})$ is

$$\begin{aligned} \bar{x} &= \frac{\int_a^b x [f_2(x) - f_1(x)] \, dx}{\int_a^b [f_2(x) - f_1(x)] \, dx}, \\ \bar{y} &= \frac{\frac{1}{2} \int_a^b \{ [f_2(x)]^2 - [f_1(x)]^2 \} \, dx}{\int_a^b [f_2(x) - f_1(x)] \, dx}. \end{aligned}$$

8

Infinite Series

8.1 SERIES

We first study series of numbers.

8.1.1 Sum of Series

A **series** (or **infinite series**) is an infinite sum

$$\sum_{n=1}^{\infty} a_n = a_1 + a_2 + a_3 + \cdots.$$

The following are some examples:

$$\begin{aligned}\sum_{n=1}^{\infty} \frac{1}{n^p} &= 1 + \frac{1}{2^p} + \frac{1}{3^p} + \cdots + \frac{1}{n^p} + \cdots; \\ \sum_{n=1}^{\infty} \frac{1}{n(n+1)} &= \frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \frac{1}{3 \cdot 4} + \cdots + \frac{1}{n(n+1)} + \cdots; \\ \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n!} &= 1 - \frac{1}{2} + \frac{1}{3} - \cdots + \frac{(-1)^{n+1}}{n!} + \cdots; \\ \sum_{n=0}^{\infty} \frac{1}{n!} &= 1 + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} + \cdots;\end{aligned}$$

$$\sum_{n=0}^{\infty} r^n = 1 + r + r^2 + r^3 + \cdots + r^n + \cdots.$$

Similar to sequences, series do not have to start at $n = 1$. Sometimes it is more convenient to start at $n = 0$ or some other numbers.

Finite sums have definite values. To define the value of an infinite sum, we introduce the **partial sum** of a series

$$s_n = \sum_{i=1}^n a_i = a_1 + a_2 + \cdots + a_n.$$

Definition 8.1.1 A series $\sum_{n=1}^{\infty} a_n$ **converges** (or is **summable**) if the sequence of partial sums converges. Moreover, the **sum** of the series is

$$\sum_{n=1}^{\infty} a_n = \lim_{n \rightarrow \infty} s_n.$$

A series **diverges** if it does not converge.

Like sequence, if a series is modified by finitely many terms, then the new partial sum s'_n is related to the old one s_n by $s'_n = s_n + C$ for a constant C . This implies that the convergence of a series is not changed by modifying finitely many terms.

Example 8.1.1 The **geometric series** is

$$\sum_{n=0}^{\infty} r^n = 1 + r + r^2 + r^3 + \cdots + r^n + \cdots.$$

The partial sum $s_n = 1 + r + r^2 + r^3 + \cdots + r^n$ satisfies

$$rs_n = r + r^2 + r^3 + \cdots + r^{n+1} = s_n - 1 + r^{n+1}.$$

Therefore

$$s_n = \frac{1 - r^{n+1}}{1 - r}$$

for $r \neq 1$, and the series converges if and only if $|r| < 1$

$$\sum_{n=0}^{\infty} r^n = 1 + r + r^2 + r^3 + \cdots + r^n + \cdots = \lim_{n \rightarrow \infty} \frac{1 - r^{n+1}}{1 - r} = \frac{1}{1 - r}. \quad \blacksquare$$

Example 8.1.2 Example 2.1.28 tells us that the series

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

converges. Example 2.1.29 tells us that the **harmonic series**

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

diverges. ■

Example 8.1.3 The series

$$\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = \frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \cdots + \frac{1}{n(n+1)} + \cdots$$

has the partial sum

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

Therefore

$$\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = \lim_{n \rightarrow \infty} \left(1 - \frac{1}{n+1}\right) = 1. \quad \blacksquare$$

If series $\sum a_n$ and $\sum b_n$ converge, then by the arithmetic properties of limits, the series $\sum(a_n + b_n)$ and $\sum ca_n$ converge, and

$$\sum(a_n + b_n) = \sum a_n + \sum b_n, \quad \sum ca_n = c \sum a_n.$$

The following is another property of convergent series.

Theorem 8.1.2 *If $\sum a_n$ converges, then $\lim_{n \rightarrow \infty} a_n = 0$.*

Proof. If the series converges, then $\lim_{n \rightarrow \infty} s_n = l$ converges, and we get

$$\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} (s_n - s_{n-1}) = \lim_{n \rightarrow \infty} s_n - \lim_{n \rightarrow \infty} s_{n-1} = \lim_{n \rightarrow \infty} s_n - \lim_{n \rightarrow \infty} s_n = 0. \quad \blacksquare$$

The theorem immediately implies that the series such as $\sum 1$, $\sum (-1)^n$, $\sum \frac{n}{n+1}$ do not converge. The following is a more sophisticated example.

Example 8.1.4 The series $\sum \sin na$ converges when a is a multiple of π . If a is not a multiple of π , we will show that there are infinitely many natural numbers n_k such that $|\sin n_k a| \geq \epsilon$ for some $\epsilon > 0$. This implies that $\sin na$ does not converge to 0. By Theorem 8.1.2, the series $\sum \sin na$ diverges.

Let $0 < a < \pi$. Then $[a_k, b_k] = \left[\frac{(2k+1)\pi - a}{2}, \frac{(2k+1)\pi + a}{2}\right]$ are disjoint intervals of length a . Moreover, for $\epsilon = \sin \frac{\pi - a}{2} = \cos \frac{a}{2} > 0$, we have $|\sin t| \geq \epsilon$ on $[a_k, b_k]$. Since the length of $[a_k, b_k]$

is a , there is $n_k a \in [a_k, b_k]$ for some natural number n_k . Since the intervals are disjoint, n_k are distinct. Therefore we have infinitely many n_k such that $|\sin n_k a| \geq \epsilon$.

For general a , we have $a = q\pi + b$ for an integer q and $0 \leq b < \pi$. Then for a not to be a multiple of π is the same as $0 < b < \pi$. We also have $\sin na = (-1)^{nq} \sin nb$. The argument above also shows that $\sum (-1)^{nq} \sin nb$ diverges. ■

However, there are also plenty of divergent series $\sum a_n$ satisfying $\lim_{n \rightarrow \infty} a_n = 0$, such as the harmonic series $\sum \frac{1}{n}$. Therefore more refined criteria are needed. For example, if $a_n \geq 0$, then the partial sums are nondecreasing. By Proposition 2.1.7, we have the following criterion.

Theorem 8.1.3 *If all terms in a series are non-negative, then the series converges if and only if the partial sums are bounded.*

In general we may apply the Cauchy criterion to the convergence of the partial sum. Since

$$s_n - s_{m-1} = a_m + a_{m+1} + \cdots + a_n, \quad n \geq m,$$

The Cauchy criterion for the convergence of series takes the following form.

Theorem 8.1.4 *A series $\sum a_n$ converges if and only if for any $\epsilon > 0$, there is N , such that*

$$n \geq m > N \implies |a_m + a_{m+1} + \cdots + a_n| < \epsilon.$$

The special case $n = m + 1$ gives exactly Lemma 8.1.2. Moreover, such Cauchy criterion has already been used in Examples 2.1.28 and 2.1.29.

8.1.2 Comparison Test

Similar to improper integrals, we may compare the convergence of series.

Theorem 8.1.5 (Comparison Test) *If $|a_n| \leq b_n$ for sufficiently large n , and $\sum b_n$ converges, then $\sum a_n$ also converges.*

Note that by Theorem 8.1.3, the convergence of $\sum b_n$ is equivalent to the partial sums being bounded.

Proof. The convergence of $\sum b_n$ tells us that the series satisfies the Cauchy criterion: For any $\epsilon > 0$, there is N , such that

$$n > m > N \implies |b_m + b_{m+1} + \cdots + b_n| < \epsilon.$$

Since

$$|a_m + a_{m+1} + \cdots + a_n| \leq |a_m| + |a_{m+1}| + \cdots + |a_n| \leq b_m + b_{m+1} + \cdots + b_n.$$

We see that the series $\sum a_n$ also satisfies the Cauchy criterion. ■

Example 8.1.5 Consider the series $\sum a^{-n^2 + \frac{1}{n}}$. By Lemma 8.1.2, for the series to converge, we must have $\lim_{n \rightarrow \infty} a^{-n^2 + \frac{1}{n}} = 0$. This implies that we must have $a > 1$.

Since $-n^2 + \frac{1}{n} < -n$ for large n (actually $n > 1$ is enough), for $a > 1$ we have

$$0 < a^{-n^2 + \frac{1}{n}} < a^{-n}.$$

By Example 8.1.1, the series $\sum a^{-n} = \sum \left(\frac{1}{a}\right)^n$ converges. Therefore by the comparison test, $\sum a^{-n^2 + \frac{1}{n}}$ also converges. We conclude that $\sum a^{-n^2 + \frac{1}{n}}$ converges if and only if $a > 1$. ■

Example 8.1.6 The series $\sum \frac{n + \sin n}{n^3 + n + 2}$ satisfies

$$\lim_{n \rightarrow \infty} \frac{\frac{n + \sin n}{n^3 + n + 2}}{\frac{1}{n^2}} = 1.$$

This implies that

$$\left| \frac{n + \sin n}{n^3 + n + 2} \right| \leq \frac{2}{n^2}$$

for sufficiently large n . Since we know the series $\sum \frac{1}{n^2}$ converges, we conclude that $\sum \frac{n + \sin n}{n^3 + n + 2}$ also converges. ■

The argument in the example above is quite useful. Suppose $a_n, b_n > 0$ and the limit $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = l$ converges. Then $0 < a_n < (l + 1)b_n$ for sufficiently large n . The comparison test tells us that the convergence of $\sum b_n$ implies the convergence of $\sum a_n$. Note that if $l \neq 0$, then the limit $\lim_{n \rightarrow \infty} \frac{b_n}{a_n} = \frac{1}{l}$ also converges, and we conclude that $\sum a_n$ converges if and only if $\sum b_n$ converges.

We already fully understand the convergence of the geometric series in Example 8.1.1. By specifically comparing with the geometric series, we get the following useful tests.

Theorem 8.1.6 (Root Test) Suppose the limit

$$\lim_{n \rightarrow \infty} \sqrt[n]{|a_n|} = r$$

converges. If $r < 1$, then $\sum a_n$ converges. If $r > 1$, then $\sum a_n$ diverges.

Theorem 8.1.7 (Ratio Test) Suppose the limit

$$\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| = r$$

converges. If $r < 1$, then $\sum a_n$ converges. If $r > 1$, then $\sum a_n$ diverges.

In case $r = 1$, the test is not conclusive, and other methods have to be used to determine the convergence.

Proof. If $\lim_{n \rightarrow \infty} \sqrt[n]{|a_n|} > 1$, then $\sqrt[n]{|a_n|} > 1$ for sufficiently large n . This implies $|a_n| > 1$ for sufficiently large n . By Lemma 8.1.2, the series $\sum a_n$ must diverge.

Suppose $\lim_{n \rightarrow \infty} \sqrt[n]{|a_n|} = r < 1$ and fix some R satisfying $r < R < 1$. Then $\sqrt[n]{|a_n|} < R$ for sufficiently large n . This implies that $|a_n| < R^n$ for sufficiently large n . By the convergence of $\sum R^n$ and the comparison test, the series $\sum a_n$ must converge.

If $\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| > 1$, then $|a_{n+1}| > |a_n|$ for sufficiently large n . This implies that a_n does not converge to 0 and the series $\sum a_n$ must diverge.

Suppose $\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| = r < 1$ and fix some R satisfying $r < R < 1$. Then $\left| \frac{a_{n+1}}{a_n} \right| < R$ for sufficiently large n , say $n \geq N$. This implies

$$|a_n| = |a_N| \left| \frac{a_{N+1}}{a_N} \right| \left| \frac{a_{N+2}}{a_{N+1}} \right| \cdots \left| \frac{a_n}{a_{n-1}} \right| \leq |a_N| R^{n-N} = cR^n, \quad n \geq N$$

for a constant $c = \frac{|a_N|}{R^N}$. By the convergence of $\sum R^n$ and the comparison test, the series $\sum a_n$ must converge. ■

Example 8.1.7 Consider the series $\sum n^a b^n$. We have

$$\lim_{n \rightarrow \infty} \sqrt[n]{|n^a b^n|} = |b|.$$

Therefore the series converges when $|b| < 1$ and diverges when $|b| > 1$.

In case $b = 1$, the series is $\sum n^a$. If $a \geq -1$, then $n^a \geq \frac{1}{n}$. By the divergence of $\sum \frac{1}{n}$ and the comparison test, the series $\sum n^a$ diverges. If $a \leq -2$, then $n^a \leq \frac{1}{n^2}$. By the convergence of $\sum \frac{1}{n^2}$ and the comparison test, the series $\sum n^a$ converges. In fact, in Example 8.1.10, we will see that $\sum n^a$ also converges for all $a < -1$. The case $b = -1$ will be discussed in Example 8.1.13. ■

Example 8.1.8 Consider the series $\sum \frac{n!}{n^n} a^n$. We have

$$\lim_{n \rightarrow \infty} \left| \frac{\frac{(n+1)!}{(n+1)^{n+1}} a^{n+1}}{\frac{n!}{n^n} a^n} \right| = \lim_{n \rightarrow \infty} \frac{(n+1)}{(n+1)^{n+1}} n^n |a| = \lim_{n \rightarrow \infty} \frac{|a|}{\left(\frac{n+1}{n}\right)^n} = \frac{|a|}{e}.$$

Therefore the series converges when $|a| < e$ and diverges when $|a| > e$. Moreover, by $\left(\frac{n+1}{n}\right)^n < e$, we know the quotient is > 1 when $|a| = e$. Therefore the terms in the series are nondecreasing and cannot converge to 0. We conclude the series also diverges for $|a| = e$. ■

Upper Limit

Note that the root and ratio tests assume that the limits converge. In case the limits do not converge, we may still use the spirit of the tests. By carefully examining the proof, we note that for

the convergence, all we need is that $\sqrt[n]{|a_n|} < R$ or $\left| \frac{a_{n+1}}{a_n} \right| < R$ for sufficiently large n and a constant $R < 1$. The condition can be reformulated in terms of the **upper limit**.

Given a sequence a_n and a number l , we ask the following question: How many terms in the sequence are above l ? Imagine that we start with a very high l and then gradually lower l . In the process, more and more terms will be above l . If a_n is bounded, then for sufficiently high l , there is no $a_n > l$, and for sufficiently low l , every $a_n > l$. The transition from “no $a_n > l$ ” to “every $a_n > l$ ” has to pass through a unique critical moment λ satisfying

- If $l > \lambda$, then only finitely many $a_n > l$.
- If $l < \lambda$, then infinitely many $a_n > l$.

The upper limit is this critical moment

$$\overline{\lim}_{n \rightarrow \infty} x_n = \lambda.$$

The sequence a_n has no upper bound if and only if there are always infinitely many $a_n > l$ for any l . In this case $\overline{\lim}_{n \rightarrow \infty} x_n = +\infty$.

The sequence a_n diverges to $-\infty$ if and only if there are only finitely many $a_n > l$ for any l . In this case $\overline{\lim}_{n \rightarrow \infty} a_n = -\infty$.

The upper limit has the following properties:

1. If a sequence converges, then the upper limit is the limit.
2. In general, a sequence may have (perhaps infinitely) many convergent subsequences. The upper limit is the maximum of the limits of convergent subsequences.
3. If $c \geq 0$, then $\overline{\lim}_{n \rightarrow \infty} (c \cdot a_n) = c \cdot \overline{\lim}_{n \rightarrow \infty} a_n$.
4. The upper limit of a_n is the limit if the monotonic sequence $\bar{a}_n = \sup\{x_k : k \geq n\}$.

The second property is the reason for the terminology “upper limit”.

Theorem 8.1.8 (Root Test) *If $\overline{\lim}_{n \rightarrow \infty} \sqrt[n]{|a_n|} < 1$, then $\sum a_n$ converges. If $\overline{\lim}_{n \rightarrow \infty} \sqrt[n]{|a_n|} > 1$, then $\sum a_n$ diverges.*

Proof. Suppose $1 < \overline{\lim}_{n \rightarrow \infty} \sqrt[n]{|a_n|}$. By the second property in the definition of the upper limit, we have $\sqrt[n]{|a_n|} > 1$ for infinitely many n . This implies $|a_n| > 1$ for infinitely many n , and further implies that a_n does not converge to 0. Therefore the series $\sum a_n$ diverges.

Suppose $1 > \overline{\lim}_{n \rightarrow \infty} \sqrt[n]{|a_n|}$. Then fix some $R < 1$ satisfying $R > \overline{\lim}_{n \rightarrow \infty} \sqrt[n]{|a_n|}$. By the first property in the definition of the upper limit, there are only finitely many n such that $\sqrt[n]{|a_n|} > R$. Let N be the largest such n . Then $\sqrt[n]{|a_n|} \leq R$ for all $n > N$. This implies $|a_n| < R^n$ for $n > N$. By the convergence $\sum R^n$ and the comparison test, the series $\sum a_n$ converges. ■

Example 8.1.9 Consider the series $\sum a^{(-1)^n n}$. We have $\sqrt[n]{a^{(-1)^n n}} = a$ for even n and $\sqrt[n]{a^{(-1)^n n}} = a^{-1}$ for odd n . Therefore $\overline{\lim}_{n \rightarrow \infty} \sqrt[n]{a^{(-1)^n n}} = \max\{a, a^{-1}\}$ (note that in the sequence $\sqrt[n]{a^{(-1)^n n}}$, both a and a^{-1} appear infinitely many times). In particular, $\overline{\lim}_{n \rightarrow \infty} \sqrt[n]{a^{(-1)^n n}} > 1$ when $a > 1$, and the series diverges. For $a = 1$, the series also obviously diverges. Therefore the series diverges for all a . ■

The ratio test can be extended in a similar way. For the convergence, we need the upper limit of the ratio. For the divergence, however, we need the **lower limit**. The lower limit $\lim_{n \rightarrow \infty} a_n$ is the critical moment μ satisfying

- If $l < \mu$, then only finitely many $a_n < l$.
- If $l > \mu$, then infinitely many $a_n < l$.

Theorem 8.1.9 (Ratio Test) If $\overline{\lim}_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| < 1$, then $\sum a_n$ converges. If $\overline{\lim}_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| > 1$, then $\sum a_n$ diverges.

8.1.3 Integral Comparison Test

The convergence of series can also be tested by comparing with the convergence of improper integrals.

Theorem 8.1.10 (Integral Test) Suppose $f(x)$ is defined on $[a, +\infty)$ and is nonincreasing. Then the series $\sum f(n)$ converges if and only if the improper integral $\int_a^{+\infty} f(x) dx$ converges.

Proof. Since $f(x)$ is nonincreasing, we have $\lim_{x \rightarrow +\infty} f(x) = l$, where l is either bounded l or $l = -\infty$.

In the case $l \neq 0$, both $\sum f(n)$ and $\int_a^{+\infty} f(x) dx$ diverge (the divergence of the improper integral can be seen by the Cauchy criterion). Therefore we may assume $\lim_{x \rightarrow +\infty} f(x) = 0$ in the subsequent discussion. In particular, since $f(x)$ is nonincreasing, the assumption implies that $f(x) \geq 0$. Then the convergence of $\sum f(n)$ and $\int_a^{+\infty} f(x) dx$ is equivalent to the boundedness of the partial sum s_n and the “partial” integral $\int_a^n f(x) dx$.

Without loss of generality, we assume $a = 1$. Since $f(x)$ is nonincreasing, we get

$$f(n) \geq \int_n^{n+1} f(x) dx \geq f(n+1),$$

so that the partial sum s_n of $\sum f(n)$ satisfies

$$s_n = f(1) + f(2) + \cdots + f(n) \geq \int_1^{n+1} f(x) dx \geq f(2) + f(3) + \cdots + f(n+1) = s_{n+1} - f(1).$$

Therefore the sequence s_n is bounded if and only if the sequence $\int_1^n f(x) dx$ is bounded. ■

Example 8.1.10 For $p > 0$, the series $\sum \frac{1}{n^p}$ converges if and only if the improper integral $\int_1^{+\infty} \frac{dx}{x^p}$ converges. By Example 6.3.4, this happens (i.e., the series converges) if and only if $p > 1$.

We also note that $\sum \frac{1}{n^p}$ diverges for $p \leq 0$ because the individual term $\frac{1}{n^p}$ does not converge to 0.

Similarly, by comparing with the improper integral in Example 6.3.7, the series $\sum \frac{1}{n(\log n)^p}$ converges if and only if $p > 1$. ■

Example 8.1.11 Consider the series $\sum \frac{\log n}{n^p}$. If $p \leq 1$, then $\frac{\log n}{n^p} \geq \frac{1}{n^p} \geq \frac{1}{n}$. By the divergence of $\sum \frac{1}{n}$ and the comparison test, we see that the series $\sum \frac{\log n}{n^p}$ diverges.

If $p > 1$, then choose q satisfying $p > q > 1$. Since $p - q > 1$, we have

$$\lim_{n \rightarrow \infty} \frac{\frac{\log n}{n^p}}{\frac{1}{n^q}} = \lim_{n \rightarrow \infty} \frac{\log n}{n^{p-q}} = 0.$$

Then by the convergence of $\sum \frac{1}{n^q}$ and the comparison test, we conclude that $\sum \frac{\log n}{n^p}$ converges. ■

8.1.4 Absolute and Conditional Convergence

A series $\sum a_n$ **absolutely converges** if the corresponding absolute value series $\sum |a_n|$ converges. The comparison test Theorem 8.1.5 tells us the following

- If $|a_n| \leq b_n$ and $\sum b_n$ converges, then $\sum a_n$ absolutely converges.
- If $\sum a_n$ absolutely converges, then $\sum a_n$ converges.

The first statement can be obtained by taking a_n to be $|a_n|$ in the comparison test. The first statement can be obtained by taking b_n to be $|a_n|$ in the comparison test.

This leaves the possibility that a series $\sum a_n$ may converge, but the corresponding absolute value series $\sum |a_n|$ may diverge. In this case, we say $\sum a_n$ **converge conditionally**.

Example 8.1.12 Consider the series

$$\sum \frac{(-1)^{n+1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots + \frac{1}{2n-1} - \frac{1}{2n} + \cdots$$

The corresponding absolute value series is the divergent harmonic series $\sum \frac{1}{n}$.

Let s_n be the partial sum of $\sum \frac{(-1)^{n+1}}{n}$. Then s_{2n} is the partial sum of the series $\sum \left(\frac{1}{2n-1} - \frac{1}{2n} \right)$.

By

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

and the convergence of $\sum \frac{1}{n^2}$, we conclude that the series $\sum \left(\frac{1}{2n-1} - \frac{1}{2n} \right)$ converges. In other words, the subsequence s_{2n} converges. Moreover, by $s_{2n-1} = s_{2n} + \frac{1}{2n}$, the subsequence s_{2n-1} also converges, and we have

$$\lim_{n \rightarrow \infty} s_{2n-1} = \lim_{n \rightarrow \infty} s_{2n}.$$

Therefore $\lim_{n \rightarrow \infty} s_n$ converges and the series $\sum \frac{(-1)^{n+1}}{n}$ (conditionally) converges. ■

Note that the conclusion of the comparison test is always absolute convergence. Therefore the comparison test cannot be used, at least directly, to show the convergence of conditionally convergent series. This is parallel to the convergence of improper integrals, for which we also have the Dirichlet and Abel tests. Before extending such tests to series, we state a simple criterion for the convergence.

With exactly the same reasoning, we can generalize the above example to the following result, due to Leibniz.

Theorem 8.1.11 (Leibniz Test) *If a_n is decreasing and $\lim_{n \rightarrow \infty} a_n = 0$, then the series*

$$\sum_{n=1}^{\infty} (-1)^{n+1} a_n = a_1 - a_2 + a_3 - \cdots$$

converges.

For $a_n > 0$, the terms in the series $\sum (-1)^{n+1} a_n$ takes alternating signs. Such series are called **alternating series**.

Proof. The proof is the generalization of the example above. The even partial sum s_{2n} is the partial sum of the series $\sum (a_{2n-1} - a_{2n})$. Since a_n is decreasing, the terms in the series $\sum (a_{2n-1} - a_{2n})$ are non-negative, and the convergence is the same as the partial sum being bounded above. By

$$s_{2n} = a_1 - (a_2 - a_3) - (a_4 - a_5) - \cdots - (a_{2n-2} - a_{2n-1}) - a_{2n} \leq a_1,$$

we see that the series $\sum (a_{2n-1} - a_{2n})$ converges, and $\lim_{n \rightarrow \infty} s_{2n}$ converges. Moreover, by $\lim_{n \rightarrow \infty} a_n = 0$, we also have

$$\lim_{n \rightarrow \infty} s_{2n+1} = \lim_{n \rightarrow \infty} (s_{2n} + a_{2n+1}) = \lim_{n \rightarrow \infty} s_{2n}.$$

Therefore $\lim_{n \rightarrow \infty} s_n$ converges. ■

The idea used in the proof above also gives us an estimation for the **remainder**

$$R_n = \sum_{i=1}^{+\infty} (-1)^{i+1} a_i - s_n = \sum_{i=n+1}^{+\infty} (-1)^{i+1} a_i = (-1)^n a_{n+1} + (-1)^{n+1} a_{n+2} + (-1)^{n+2} a_{n+3} + \cdots$$

of the partial sum. We have

$$R_{2n} = a_{2n+1} - a_{2n+2} + a_{2n+3} - a_{2n+4} + \cdots = (a_{2n+1} - a_{2n+2}) + (a_{2n+3} - a_{2n+4}) + \cdots \geq 0,$$

and

$$R_{2n} = a_{2n+1} - (a_{2n+2} - a_{2n+3}) - (a_{2n+4} - a_{2n+5}) - \cdots \leq a_{2n+1}.$$

Therefore $0 \leq R_{2n} \leq a_{2n+1}$. By similar argument, we have $0 \geq R_{2n-1} \geq -a_{2n}$. In general, we have

$$|R_n| \leq a_{n+1}.$$

Example 8.1.13 By the Leibniz test, the series

$$\sum \frac{(-1)^{n+1}}{n^p} = 1 - \frac{1}{2^p} + \frac{1}{3^p} - \frac{1}{4^p} + \cdots + \frac{(-1)^{n+1}}{n^p} + \cdots$$

converges for any $p > 0$. However, By Example 8.1.10, we also know the corresponding absolute value series $\sum \frac{1}{n^p}$ converges if and only if $p > 1$. Therefore the alternating series converges conditionally for $0 < p \leq 1$. ■

Example 8.1.14 Consider the alternating series $\sum (-1)^n \frac{n^2 + a}{n^3 + b}$. The corresponding absolute value series is comparable to the harmonic series and therefore diverges. Then we try to use the Leibniz test by showing that $f(x) = \frac{x^2 + a}{x^3 + b}$ is decreasing. By considering the sign of

$$f'(x) = \frac{-x^4 - 3ax^2 + 2xb}{(x^3 + b)^2},$$

the function indeed decreases for sufficiently large x . Therefore the Leibniz test can be used, and the series $\sum (-1)^n \frac{n^2 + a}{n^3 + b}$ converges. ■

The Leibniz test is too simple to deal with more sophisticated series. In this case, we may use the analogue of the Dirichlet or Abel test for improper integrals.

Theorem 8.1.12 (Dirichlet Test) Suppose the partial sum of $\sum a_n$ is bounded. Suppose b_n is monotonic and $\lim_{n \rightarrow \infty} b_n = 0$. Then $\sum a_n b_n$ converges.

Theorem 8.1.13 (Abel Test) Suppose $\sum a_n$ converges. Suppose b_n is monotonic and bounded. Then $\sum a_n b_n$ converges.

Example 8.1.15 Consider the series $\sum \frac{\sin na}{n}$. 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

$$|s_n| \leq \frac{1}{\left|\sin \frac{a}{2}\right|}$$

in case a is not a multiple of π . On the other hand, if a is a multiple of π , then the terms in the series are zero, and the series still converges.

On the other hand, the divergence of $\int_1^{+\infty} \frac{|\sin x|}{x} dx$ in Example 6.3.10 suggests that the absolute value series $\sum \frac{|\sin na|}{n}$ should diverge in case a is not a multiple of π . However, the earlier argument cannot be applied here because $\frac{|\sin ax|}{x}$ is not a decreasing function. Using the argument in Example 8.1.4, for $0 < a < \pi$, we find $n_k a \in [a_k, b_k] = \left[\frac{(2k+1)\pi - a}{2}, \frac{(2k+1)\pi + a}{2} \right]$ satisfying $|\sin n_k a| \geq \epsilon = \cos \frac{a}{2}$. Then by $n_k \leq \frac{b_k}{a} < (k+1)\frac{\pi}{a}$, we have

$$\frac{|\sin n_k a|}{n_k} \geq \frac{a\epsilon}{(k+1)\pi}.$$

The diverges of $\sum_k \frac{1}{k+1}$ implies the divergence of $\sum_k \frac{|\sin n_k a|}{n_k}$, which further implies the divergence of $\sum_n \frac{|\sin na|}{n}$.

For general a that is not a multiple of π , we have $a = q\pi + b$ for an integer q and $0 < b < \pi$. Then $\sum_n \frac{|\sin na|}{n} = \sum_n \frac{|\sin nb|}{n}$ still diverges.

We conclude that $\sum \frac{\sin na}{n}$ converges conditionally when a is not a multiple of π . ■

8.1.5 Rearrangement

In a finite sum, the order of the terms does not affect the value of sum. For example, we have $a + b + c + d = c + b + a + d$. However, we need to be careful in rearranging orders in an infinite sum.

Example 8.1.16 By Leibniz test, we know the alternating series $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots$ converges. We also know the sum $s > 0$. Now we rearrange the terms so that one positive term is followed by two negative terms

$$1 - \frac{1}{2} - \frac{1}{4} + \frac{1}{3} - \frac{1}{6} - \frac{1}{8} + \cdots$$

The new series has sum

$$\left(1 - \frac{1}{2}\right) - \frac{1}{4} + \left(\frac{1}{3} - \frac{1}{6}\right) - \frac{1}{8} + \cdots = \frac{1}{2} - \frac{1}{4} + \frac{1}{6} - \frac{1}{8} + \cdots = \frac{1}{2}s.$$

We get a different sum after rearrangement. ■

The reason for getting different sum after rearrangement is because the original series converges conditionally.

Theorem 8.1.14 *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 series is s .*

The question of order arises when we deal with the product of two series. Motivated by the product of finite sums, we expect to have

$$\left(\sum_{i=1}^{\infty} a_i\right) \left(\sum_{j=1}^{\infty} b_j\right) = \sum_{i,j=1}^{\infty} a_i b_j.$$

However, the terms on the right have two indices and there are many natural ways of ordering them. In light of Theorem 8.1.14, we do not need to be worried about the many possible choices of the order under the absolute convergence condition.

Theorem 8.1.15 *If $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ converge absolutely, then $\sum_{i,j=1}^{\infty} a_i b_j$ also converge absolutely in any arrangement of the order, and with sum*

$$\sum_{i,j=1}^{\infty} a_i b_j = \left(\sum_{n=1}^{\infty} a_n\right) \left(\sum_{n=1}^{\infty} b_n\right).$$

Example 8.1.17 We will see in Example 8.2.1 that $\sum_{n=0}^{\infty} \frac{x^n}{n!}$ absolutely converges to e^x . Then

$$\begin{aligned} e^x e^y &= \left(\sum_{n=0}^{\infty} \frac{x^n}{n!}\right) \left(\sum_{n=0}^{\infty} \frac{y^n}{n!}\right) = \sum_{i,j=0}^{\infty} \frac{x^i y^j}{i! j!} = \sum_{n=0}^{\infty} \sum_{i+j=n} \frac{x^i y^j}{i! j!} \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{i=0}^n \frac{n!}{i!(n-i)!} x^i y^{n-i} = \sum_{n=0}^{\infty} \frac{1}{n!} (x+y)^n = e^{x+y}. \end{aligned}$$

In the second to the last equality, we used the binomial expansion. ■

Exercises

8.1.1 For each of the following series, show that it converges and find its sum:

1. $\frac{2-3}{5^2} + \frac{2^2+3^2}{5^3} + \cdots + \frac{2^n + (-1)^n 3^n}{5^{n+1}} + \cdots;$
2. $1 + 3r + 5r^2 + \cdots + (2n+1)r^n + \cdots;$
3. $\frac{1}{a(a+d)} + \frac{1}{(a+d)(a+2d)} + \cdots + \frac{1}{(a+nd)(a+(n+1)d)} + \cdots;$

$$4. \frac{1}{a(a+d)(a+2d)} + \frac{1}{(a+d)(a+2d)(a+3d)} + \cdots + \frac{1}{(a+nd)(a+(n+1)d)(a+(n+2)d)} + \cdots;$$

$$5. r \sin \theta + r^2 \sin 2\theta + \cdots + r^n \sin n\theta + \cdots;$$

$$6. r \cos \theta + r^2 \cos 2\theta + \cdots + r^n \cos n\theta + \cdots.$$

8.1.2 Prove that if $a_n \geq 0$ and $\sum a_n$ converges, then $\sum a_n^2$ also converges. Moreover, construct an example for which $\sum a_n^2$ converges but $\sum a_n$ diverges.

8.1.3 Prove that if $\sum a_n^2$ converges, then $\sum \frac{a_n}{n}$ converges.

8.1.4 Prove 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.

8.1.5 Prove that if a_n is bounded, then $\sum a_n r^n$ converges for $|r| \leq 1$.

8.1.6 Determine the convergence for each of the following series:

$$1. 1 + \frac{1}{2} - \frac{1}{3} + \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \cdots;$$

$$11. \sum \left(\frac{\log n}{n} \right)^n;$$

$$2. \frac{4}{2} + \frac{4 \cdot 7}{2 \cdot 6} + \frac{4 \cdot 7 \cdot 10}{2 \cdot 6 \cdot 10} + \cdots;$$

$$12. \sum \frac{2 + (-1)^n}{2^n};$$

$$3. \sum \frac{x^n}{n!};$$

$$13. \sum \frac{n^3(a + (-1)^n)^n}{b^n}, a > 1, b > 0;$$

$$4. \sum \frac{(n!)^2}{(2n)!};$$

$$14. \sum \frac{\sqrt{n+a} - \sqrt{n+b}}{n^p};$$

$$5. \sum \frac{a^n n^n}{n!};$$

$$15. \sum \left(\frac{1}{\sqrt{n}} - \sqrt{\log \frac{n+1}{n}} \right);$$

$$6. \sum \frac{n^2}{\left(2 + \frac{1}{n}\right)^n};$$

$$16. \sum \left(\cos \frac{a}{n} \right)^{n^2};$$

$$7. \sum \frac{n^{n+\frac{1}{n}}}{\left(1 + \frac{1}{n}\right)^n};$$

$$17. \sum \frac{n^{2n}}{(n+a)^{n+b}(n+b)^{n+a}}, a, b > 0;$$

$$18. \sum \frac{1}{n^p} \sin \frac{1}{n};$$

$$8. \sum \frac{1}{\sqrt[n]{\log n}};$$

$$19. \sum (\sqrt{n+1} - \sqrt{n})^p \log \frac{n-1}{n+1};$$

$$9. \sum \frac{n^{n-1}}{(2n^2 + n + 1)^{\frac{n+1}{2}}};$$

$$20. \sum \left(e - \left(1 + \frac{1}{n} \right)^n \right)^p;$$

$$21. \sum n^{\log x};$$

$$10. \sum \left(1 + \frac{1}{n} \right)^{-n^2};$$

$$22. \sum \frac{1}{n^p (\log n)^q};$$

23. $\sum \frac{1}{n \log n \log \log n}.$

8.1.7 Suppose $\sum a_n$ converges and $\lim_{n \rightarrow \infty} \frac{b_n}{a_n} = 1$. Can you conclude that $\sum b_n$ also converges?

8.1.8 Determine absolute or conditional convergence for each of the following series:

1. $\sum (-1)^n \frac{n^p + a}{n^q + b};$
2. $\sum (-1)^n \frac{(\log n)^p}{n^q};$
3. $\sum (-1)^n \sin \frac{a}{n^p};$
4. $\sum \frac{(-1)^n \sin na}{n^p}.$

8.2 POWER SERIES

8.2.1 Taylor Series

A function is smooth if it has derivatives of arbitrary order. If $f(x)$ is smooth at a , then the high order approximations of the function gives us the **Taylor series**

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!} (x-a)^n = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!} (x-a)^2 + \frac{f'''(a)}{3!} (x-a)^3 + \cdots + \frac{f^{(n)}(a)}{n!} (x-a)^n + \cdots.$$

The partial sum of the series is $T_n(x)$, the n -th order **Taylor expansion** of $f(x)$ at a .

We expect $T_n(x)$ to be better and better approximation of $f(x)$ as n become bigger. In other words, we wish to have $\lim_{n \rightarrow \infty} T_n(x) = f(x)$, at least for those x near a . This means exactly that the Taylor series converges to the function

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!} (x-a)^2 + \frac{f'''(a)}{3!} (x-a)^3 + \cdots + \frac{f^{(n)}(a)}{n!} (x-a)^n + \cdots.$$

To study whether this is indeed the case, we introduce the **remainder**

$$R_n(x) = f(x) - T_n(x).$$

Theorem 3.3.2 tells us that if $f(x)$ has n -th order derivative at a , then $\lim_{x \rightarrow a} \frac{R_n(x)}{(x-a)^n} = 0$. Under additional assumption, the proof of the theorem, which appeared in Example 4.2.4, can also be used to prove a more detailed formula for the remainder.

Theorem 8.2.1 Suppose $f(x)$ has $(n+1)$ -st order derivative between a and x . Then the remainder of the n -th order Taylor expansion is given by

$$R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!} (x-a)^{n+1}$$

for some c between a and x .

Proof. We only prove the case $n = 2$. Similar to the argument in Example 4.2.4, by repeatedly using Cauchy's Means Value Theorem, we get

$$\begin{aligned}\frac{R_2(x)}{(x-a)^3} &= \frac{R_2(x) - R_2(a)}{(x-a)^3 - (a-a)^3} = \frac{R'_2(c_1)}{3(c_1-a)^2} \\ &= \frac{R'_2(c_1) - R'_2(a)}{3[(c_1-a)^2 - (a-a)^2]} = \frac{R''_2(c_2)}{3 \cdot 2(c_2-a)} \\ &= \frac{R''_2(c_2) - R''_2(a)}{3 \cdot 2(c_2-a)} = \frac{R'''_2(c_3)}{3 \cdot 2 \cdot 1},\end{aligned}$$

where c_1 is between a and x , c_2 is between a and c_1 , and c_3 is between a and c_2 . Since $R'''_2 = f'''$, we get

$$R_2(x) = \frac{f'''(c_3)}{3!}(x-a)^3. \quad \blacksquare$$

The formula for the remainder allows us to show that the same function is indeed the sum of its Taylor series.

Example 8.2.1 Example 3.3.8 contains the Taylor series for the power function $(1+x)^p$, the exponential function e^x , then logarithmic function $\ln(1+x)$, and the trigonometric functions $\sin x$ and $\cos x$ expanded at $a = 0$. For $f(x) = e^x$, the remainder satisfies

$$|R_n(x)| = \frac{e^c}{(n+1)!}|x|^{n+1} \leq e^{|x|} \frac{|x|^{n+1}}{(n+1)!}.$$

Since the right side converges to 0 for each x , we conclude that

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \cdots,$$

for all x . Similarly, for $f(x) = \sin x$ and $f(x) = \cos x$, we have $|f^{(n+1)}(c)| \leq 1$ because the high order derivatives are either $\pm \sin x$ or $\pm \cos x$. Therefore the corresponding remainder satisfies

$$|R_n(x)| = \frac{|f^{(n+1)}(c)|}{(n+1)!}|x|^{n+1} \leq \frac{|x|^{n+1}}{(n+1)!},$$

and we also get

$$\begin{aligned}\sin x &= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots + (-1)^{n+1} \frac{x^{2n-1}}{(2n-1)!} + \cdots, \\ \cos x &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots + (-1)^{n+1} \frac{x^{2n}}{(2n)!} + \cdots,\end{aligned}$$

for all x .

In applying the remainder formula to $f(x) = \ln(1+x)$, we get $f^{(n+1)}(x) = \frac{(-1)^{n-1}(n-1)!}{(1+x)^n}$ and

$$|R_n(x)| = \frac{1}{(n+1)!} \frac{n!}{|1+c|^{n+1}} |x|^{n+1} = \frac{|x|^{n+1}}{(n+1)|1+c|^{n+1}}$$

Knowing only that c lies between 0 and x , we are sure that

$$\ln x = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots + (-1)^{n+1} \frac{x^n}{n!} + \cdots$$

for $|x| < \frac{1}{2}$. In a later example, we will use other method to show that the equality holds for all $|x| < 1$. ■

Example 8.2.2 We try to use $e = e^1 = \sum_{n=0}^{\infty} \frac{1}{n!}$ to compute the value of e accurate up to the 6-th digit. The remainder satisfies

$$|R_n(1)| \leq \frac{e^c}{(n+1)!} 1^{n+1} \leq \frac{e}{(n+1)!}.$$

Since $\frac{e}{10!} < 10^{-6}$, we find

$$e \approx 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{9!} \approx 2.718285$$

is accurate up to the 6-th digit. ■

Example 8.2.3 In Example 3.3.5, we used the second order approximation to compute $\sqrt{3.96}$ and $\sqrt{4.05}$. The error can be estimated by the remainder formula

$$R_2(x) = \frac{1}{16\sqrt{c^5}}(x-4)^3.$$

Specifically, for $\sqrt{3.96}$, we have $3.96 < c < 4$ and

$$|\sqrt{3.96} - 1.989975| \leq \frac{1}{16\sqrt{c^5}} |3.96 - 4|^3 \approx \frac{0.04^3}{16 \cdot 2^5} = 0.000000125. \quad \blacksquare$$

Example 8.2.4 Using the Taylor expansion and the remainder of e^x , we have

$$e^{-x^2} = 1 - \frac{x^2}{1!} + \frac{x^4}{2!} - \frac{x^6}{3!} + \frac{x^8}{4!} + R_4(-x^2),$$

where

$$|R_4(-x^2)| \leq \frac{|-x^2|^5}{5!} = \frac{x^{10}}{5!}.$$

Thus we get an approximate value of the integral of e^{-x^2}

$$\int_0^1 e^{-x^2} dx \approx 1 - \frac{1}{3 \cdot 1!} + \frac{1}{5 \cdot 2!} - \frac{1}{7 \cdot 3!} + \frac{1}{9 \cdot 4!} \approx 0.7475.$$

The error may be estimated by

$$\left| \int_0^1 e^{-x^2} dx - 0.7475 \right| \leq \int_0^1 |R_4(-x^2)| dx \leq \int_0^1 \frac{x^{10}}{5!} dx = \frac{1}{11 \cdot 5!} < 0.001. \quad \blacksquare$$

8.2.2 Power Series

A **power series** is a series of the form

$$\sum_{n=0}^{\infty} a_n x^n.$$

Taylor series are power series, up to substituting x by $x - a$.

Example 8.2.5 Example 8.1.1 tells us that the power series

$$\sum_{n=0}^{\infty} x^n = 1 + x + x^2 + x^3 + \cdots + x^n + \cdots$$

converges for $|x| < 1$ and diverges for $|x| \geq 1$.

Example 8.1.8 tells us that the power series

$$\sum_{n=1}^{\infty} \frac{n!}{n^n} x^n = x + \frac{1}{2}x^2 + \frac{2}{9}x^3 + \cdots + \frac{n!}{n^n}x^n + \cdots$$

converges for $|x| < e$ and diverges for $|x| > e$.

Example 8.2.1 tells us that the power series

$$\begin{aligned} \sum_{n=0}^{\infty} \frac{x^n}{n!} &= 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \cdots, \\ \sum_{n=1}^{\infty} (-1)^{n+1} \frac{x^{2n-1}}{(2n-1)!} &= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots + (-1)^{n+1} \frac{x^{2n-1}}{(2n-1)!} + \cdots, \\ \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!} &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots + (-1)^{n+1} \frac{x^{2n}}{(2n)!} + \cdots, \end{aligned}$$

converge for all x . ■

Lemma 8.2.2 For any power series $\sum a_n x^n$, there is $R \geq 0$, such that the power series converges absolutely for $|x| < R$ and diverges for $|x| > R$.

The number R is called **radius of convergence** of the 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 .

Proof. The lemma is a consequence of the following claim: If $\sum a_n r^n$ converges, then $\sum a_n x^n$ absolutely converges for $|x| < |r|$. The radius of convergence is then the supremum of $|r|$ such that $\sum a_n r^n$ converges.

Suppose $\sum a_n r^n$ converges. Then we have $|a_n r^n| < M$ for a constant M and all n . For $|x| < |r|$, this implies that

$$|a_n x^n| = |a_n r^n| \cdot \left| \frac{x}{r} \right|^n \leq M \left| \frac{x}{r} \right|^n.$$

Since $\left| \frac{x}{r} \right| < 1$, by Example 8.1.1, the series $\sum \left| \frac{x}{r} \right|^n$ converges. Then by the comparison test, the series $\sum a_n x^n$ absolutely converges. ■

If $\lim_{n \rightarrow \infty} \sqrt[n]{|a_n|}$ converges, then

$$\lim_{n \rightarrow \infty} \sqrt[n]{|a_n x^n|} = |x| \cdot \lim_{n \rightarrow \infty} \sqrt[n]{|a_n|}.$$

By the root test Theorem 8.1.6, we see that the power series converges if

$$|x| \cdot \lim_{n \rightarrow \infty} \sqrt[n]{|a_n|} < 1,$$

and the power series diverges if

$$|x| \cdot \lim_{n \rightarrow \infty} \sqrt[n]{|a_n|} > 1.$$

This gives us the formula

$$R = \frac{1}{\lim_{n \rightarrow \infty} \sqrt[n]{|a_n|}}$$

for the radius of convergence. By the similar argument, if $\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right|$ converges, then

$$R = \frac{1}{\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right|}.$$

In general, by

$$\overline{\lim}_{n \rightarrow \infty} \sqrt[n]{|a_n x^n|} = |x| \cdot \overline{\lim}_{n \rightarrow \infty} \sqrt[n]{|a_n|},$$

and Theorem 8.1.8, we always have

$$R = \frac{1}{\overline{\lim}_{n \rightarrow \infty} \sqrt[n]{|a_n|}}.$$

Example 8.2.6 By

$$\lim_{n \rightarrow \infty} \sqrt[n]{1} = \lim_{n \rightarrow \infty} \sqrt[n]{\left| \frac{(-1)^{n+1}}{n} \right|} = 1,$$

the radius of convergence for the geometrical series and the Taylor series for $\ln(1+x)$ is 1.

By

$$\lim_{n \rightarrow \infty} \frac{\frac{1}{(n+1)!}}{\frac{1}{n!}} = \lim_{n \rightarrow \infty} \frac{1}{n+1} = 0,$$

the radius of convergence for the Taylor series of e^x is $+\infty$. In other words, the Taylor series converges for all x . Similar computation shows that the radius of convergence for the Taylor series of $\sin x$ and $\cos x$ is also $+\infty$. ■

Example 8.2.7 For any a , we have $\lim_{n \rightarrow \infty} \sqrt[n]{n^p} = 1$. Therefore the radius of convergence for the power series $\sum n^p x^n$ is 1. The example already appeared in Example 8.1.7.

Example 8.2.8 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

$$\lim \left| \frac{\frac{(-1)^{n+1}}{2^{2n+2} ((n+1)!)^2}}{\frac{(-1)^n}{2^{2n} (n!)^2}} \right| = \lim \frac{1}{4(n+1)^2} = 0,$$

the later series converges for all x . Therefore the Bessel function is defined for all x .

8.2.3 Operation of Power Series

By suitable substitution and arithmetic combination, new power series may be constructed from the known power series.

Example 8.2.9 From Example 8.1.1, we know $\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$ for $|x| < 1$. Then we get

$$\begin{aligned} \frac{1}{1+x} &= 1 - x + x^2 - x^3 + \cdots + (-1)^n x^n + \cdots, \quad |x| < 1; \\ \frac{x}{1+x} &= x - x^2 + x^3 - x^4 + \cdots + (-1)^{n+1} x^n + \cdots, \quad |x| < 1; \\ \frac{1}{x} &= 1 - (x-1) + (x-1)^2 - (x-1)^3 + \cdots + (-1)^n (x-1)^n + \cdots, \quad 0 \leq x \leq 2; \\ \frac{1}{1+x^2} &= 1 - x^2 + x^4 - x^6 + \cdots + (-1)^n x^{2n} + \cdots, \quad |x| < 1; \\ \frac{1}{2-x} &= \frac{1}{2} + \frac{1}{2^2} x + \frac{1}{2^3} x^2 + \frac{1}{2^4} x^3 + \cdots + \frac{1}{2^{n+1}} x^n + \cdots, \quad |x| < 2. \end{aligned}$$

■

Because power series converge absolutely within the radius of convergence, by Theorem 8.1.15, we can multiply two power series together within the common radius of convergence.

Theorem 8.2.3 Suppose $\sum a_n x^n$ and $\sum b_n x^n$ have radii of convergence R and R' . Then

$$\left(\sum a_n x^n\right) \left(\sum b_n x^n\right) = \sum c_n x^n, \quad c_n = a_0 b_n + a_1 b_{n-1} + \cdots + a_n b_0$$

for $|x| < \min\{R, R'\}$.

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 of power x^n .

The power series can also be differentiated or integrated term by term within the radius of convergence.

Theorem 8.2.4 Suppose

$$f(x) = \sum_{n=0}^{\infty} a_n x^n = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots + a_n x^n + \cdots$$

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

$$\int_0^x f(t) dt = \sum_{n=0}^{\infty} \int_0^x a_n t^n dt = a_0 x + \frac{a_1}{2} x^2 + \frac{a_2}{3} x^3 + \frac{a_3}{4} x^4 + \cdots + \frac{a_n}{n+1} x^{n+1} + \cdots$$

for $|x| < R$.

Example 8.2.10 Taking the derivative of $\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$, we get

$$\begin{aligned} \frac{1}{(1-x)^2} &= 1 + 2x + 3x^2 + 4x^3 + \cdots + n x^{n-1} + \cdots, \\ \frac{2}{(1-x)^3} &= 2 \cdot 1 + 3 \cdot 2x + 4 \cdot 3x^2 + 5 \cdot 4x^3 + \cdots + n(n-1)x^{n-2} + \cdots. \end{aligned}$$

Therefore

$$\begin{aligned} &1^2 x + 2^2 x^2 + 3^2 x^3 + \cdots + n^2 x^n + \cdots \\ &= \sum_{n=1}^{\infty} n^2 x^n = x \sum_{n=1}^{\infty} n x^{n-1} + x^2 \sum_{n=2}^{\infty} n(n-1) x^{n-2} \\ &= x \frac{1}{(1-x)^2} + x^2 \frac{2}{(1-x)^3} = \frac{x(1+x)}{(1-x)^3}. \end{aligned}$$

If we integrate instead, we get

$$\ln(1-x) = -\int_0^x \frac{dx}{1-x} = -x - \frac{x^2}{2} - \frac{x^3}{3} - \cdots - \frac{x^n}{n!} - \cdots$$

for $|x| < 1$. Substituting $-x$ for x , we get the Taylor expansion of $\ln(1+x)$

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots + (-1)^{n+1} \frac{x^n}{n!} + \cdots$$

for $|x| < 1$. Note that in Example 8.2.1, by estimating the remainder, we were able to prove the equality rigorously only for $|x| < \frac{1}{2}$. Here we get the equality for all x within the radius of convergence by using term wise integration. ■

Example 8.2.11 By integrating $\frac{1}{1+x^2} = \sum_{n=0}^{\infty} (-1)^n x^{2n}$, we get the Taylor expansion of $\arctan x$

$$\arctan x = \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} x^{2n+1} = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots + \frac{(-1)^n}{2n+1} x^{2n+1} + \cdots$$

for $|x| < 1$. ■

Exercises

8.2.1 Determine the radius of convergence for each of the following power series:

1. $\sum \frac{x^n}{n^p}$;
2. $\sum \frac{(-1)^n}{n} (x-1)^n$;
3. $\sum \left(\frac{a^n}{n} + \frac{b^n}{n^2} \right) x^n$;
4. $\sum \frac{x^n}{a^n + b^n}$;
5. $\sum \frac{(n!)^2}{(2n)!} x^n$;
6. $\sum \frac{(2n)!!}{(2n+1)!!} x^n$;
7. $\sum \left(1 + \frac{1}{n} \right)^{n^2} x^n$;
8. $\sum \frac{\left(1 + 2 \cos \frac{n}{4} \pi \right)^n}{\ln n} x^n$;
9. $\sum a^{n^2} x^n$;
10. $\sum a^{n^2} x^{n^2}$.

8.2.2 Find Taylor expansion and determine the radius of convergence:

1. $\frac{1}{(x-1)(x-2)}$, at 0;
2. $\sqrt{x}$, at $x=2$;
3. $\sin x^2$, at 0;
4. $\sin^2 x$, at 0;
5. $\sin x$, at $\frac{\pi}{2}$;
6. $\sin 2x$, at $\frac{\pi}{2}$;
7. $\arcsin x$, at 0;
8. $\int_0^x \frac{\sin t}{t} dt$, at 0.

8.2.3 Find the sum of series:

$$1. \sum_{n=1}^{\infty} (-1)^n \frac{x^{2n+1}}{n!};$$

$$3. \sum_{n=2}^{\infty} \frac{(x-1)^n}{n(n-1)};$$

$$2. \sum_{n=1}^{\infty} n^3 x^n;$$

$$4. \sum_{n=1}^{\infty} \frac{x^n}{2^n(2n-1)!}.$$

8.2.4 Use the product of power series to verify the identity $\sin 2x = 2 \sin x \cos x$.

8.2.5 Suppose $\sum_{n=0}^{\infty} a_n x^n$ is the Taylor expansion of $f(x)$. Find the Taylor expansion of $\frac{f(x)}{1+x}$.

8.2.6 Prove that the function $f(x) = \sum_{n=0}^{\infty} \frac{x^n}{(n!)^2}$ satisfies $xf'' + f' - f = 0$.

8.3 FOURIER SERIES

8.3.1 Fourier Coefficients

A function $f(x)$ is **periodic** if $f(x+p) = f(x)$ for some constant p , called the **period**. The trigonometric functions $\sin x$ and $\cos x$ have period 2π , and $\tan x$ has period π .

If $f(x)$ and $g(x)$ are periodic with the same period p , then $f(x) + g(x)$ and $cf(x)$ is also periodic with period p . Moreover, if $f(x)$ has period p , then $f(cx)$ has period $\frac{p}{c}$.

A periodic function is determined by its value on one interval of period length. For example, the periodic function of period 1 and satisfying $f(x) = x$ on $[0, 1)$ is given by

$$f(x) = x - n, \quad n \leq x < n+1, n \in \mathbb{N}.$$

The Taylor series can be considered as building up functions based on the power functions x^n . For the periodic functions, it is often more natural and useful to build up functions based on the trigonometric functions. Specifically, we wish to write a periodic function $f(x)$ of period 2π as

$$f(x) = a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx) = a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + \cdots.$$

For a periodic function with other period, we may multiply a suitable constant to the variable to convert the period to 2π .

Given a periodic function $f(x)$ of period 2π , how do we construct the trigonometric function series as above? We note that the trigonometric functions are “orthogonal” in the sense that

$$\int_0^{2\pi} \cos mx \cos nx \, dx = \begin{cases} 0, & \text{if } m \neq n, \\ \pi, & \text{if } m = n \neq 0, \\ 2\pi, & \text{if } m = n = 0; \end{cases}$$

$$\begin{aligned}\int_0^{2\pi} \sin mx \sin nx \, dx &= \begin{cases} 0, & \text{if } m \neq n, \\ \pi, & \text{if } m = n \neq 0; \end{cases} \\ \int_0^{2\pi} \sin mx \cos nx \, dx &= 0.\end{aligned}$$

Then the equalities above imply

$$\begin{aligned}\int_0^{2\pi} f(x) \cos nx \, dx &= a_0 \int_0^{2\pi} \cos nx \, dx + \sum_{k=1}^{\infty} \left(a_k \int_0^{2\pi} \cos kx \cos nx \, dx + b_k \int_0^{2\pi} \sin kx \cos nx \, dx \right) \\ &= \begin{cases} \pi a_n, & \text{if } n \neq 0, \\ 2\pi a_0, & \text{if } n = 0; \end{cases} \\ \int_0^{2\pi} f(x) \sin nx \, dx &= a_0 \int_0^{2\pi} \sin nx \, dx + \sum_{m=1}^{\infty} \left(a_m \int_0^{2\pi} \cos mx \sin nx \, dx + b_m \int_0^{2\pi} \sin mx \sin nx \, dx \right) \\ &= \pi b_n, \text{ if } n \neq 0.\end{aligned}$$

Strictly speaking, the exchange of the integral and the infinite sum needs to be justified in the computation above. But at the computation formally suggests the following definition.

Definition 8.3.1 *The Fourier series of a periodic function $f(x)$ of period 2π is*

$$f(x) \sim a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx),$$

where

$$\begin{aligned}a_0 &= \frac{1}{2\pi} \int_0^{2\pi} f(x) \, dx, \\ a_n &= \frac{1}{\pi} \int_0^{2\pi} f(x) \cos nx \, dx, \quad n \neq 0, \\ b_n &= \frac{1}{\pi} \int_0^{2\pi} f(x) \sin nx \, dx, \quad n \neq 0.\end{aligned}$$

are the **Fourier coefficients**.

The notation $\sim$ only indicates the relation between the function and the series. The equality of the values is yet to be established.

Note that since the functions have period 2π , the integral on $[0, 2\pi]$ can be replaced by the integral on any interval of length 2π .

If $f(x)$ has period $p = 2L$, then $f\left(\frac{Lx}{\pi}\right)$ has period 2π , and the Fourier series of $f\left(\frac{Lx}{\pi}\right)$ becomes the Fourier series of $f(x)$

$$f(x) \sim a_0 + \sum_{n=1}^{\infty} \left(a_n \cos \frac{n\pi x}{L} + b_n \sin \frac{n\pi x}{L} \right),$$

where

$$\begin{aligned} a_0 &= \frac{1}{2\pi} \int_0^{2\pi} f\left(\frac{Lx}{\pi}\right) dx = \frac{1}{2L} \int_0^{2L} f(x) dx, \\ a_n &= \frac{1}{\pi} \int_0^{2\pi} f\left(\frac{Lx}{\pi}\right) \cos nx \, dx = \frac{1}{L} \int_0^{2L} f(x) \cos \frac{n\pi x}{L} dx, \quad n \neq 0, \\ b_n &= \frac{1}{\pi} \int_0^{2\pi} f\left(\frac{Lx}{\pi}\right) \sin nx \, dx = \frac{1}{L} \int_0^{2L} f(x) \sin \frac{n\pi x}{L} dx, \quad n \neq 0. \end{aligned}$$

Example 8.3.1 Let $0 < a < 2\pi$. The function

$$f(x) = \begin{cases} 0, & \text{if } 2n\pi \leq x < 2n\pi + a, \\ 1, & \text{if } 2n\pi + a \leq x < 2(n+1)\pi, \end{cases}$$

is the periodic function of period 2π that uniquely extends the step function

$$f(x) = \begin{cases} 1, & \text{if } 0 \leq x < a, \\ 0, & \text{if } a \leq x < 2\pi, \end{cases}$$

on $[0, 2\pi)$. The Fourier coefficients are

$$a_0 = \frac{1}{2\pi} \int_0^a 1 \, dx = \frac{a}{2\pi}, \quad a_n = \frac{1}{\pi} \int_0^a \cos nx \, dx = \frac{\sin na}{n\pi}, \quad b_n = \frac{1}{\pi} \int_0^a \sin nx \, dx = \frac{1 - \cos na}{n\pi}.$$

The Fourier series is

$$f(x) \sim \frac{a}{2\pi} + \sum_{n=1}^{\infty} \frac{1}{n\pi} (\sin na \cos nx + (1 - \cos na) \sin nx). \quad \blacksquare$$

Example 8.3.2 We start with function x on $[0, 1)$. The function extends to an even periodic function of period 2

$$f(x) = |x - 2n|, \quad 2n - 1 < x < 2n + 1.$$

The value of $f(2n + 1)$ does not affect the Fourier coefficients, which are

$$\begin{aligned} a_0 &= \frac{1}{2} \int_{-1}^1 |x| \, dx = \int_0^1 x \, dx = \frac{1}{2}, \\ a_n &= \int_{-1}^1 |x| \cos n\pi x \, dx = 2 \int_0^1 x \cos n\pi x \, dx = \frac{2(1 - (-1)^n)}{n^2\pi^2}, \\ b_n &= \int_{-1}^1 |x| \sin n\pi x \, dx = 0. \end{aligned}$$

The coefficient $b_n = 0$ because $|x| \sin n\pi x$ is an odd function. Thus we get

$$|x| \sim \frac{1}{2} + \sum \frac{4}{(2n+1)^2\pi^2} \cos(2n+1)\pi x, \quad |x| < 1.$$

We may also extend x on $[0, 1)$ to an odd periodic function of period 2

$$f(x) = x - 2n, \quad 2n - 1 < x < 2n + 1.$$

The Fourier coefficients $a_n = 0$ for the odd function, and

$$b_n = \int_{-1}^1 x \sin n\pi x \, dx = 2 \int_0^1 x \sin n\pi x \, dx = \frac{(-1)^{n+1}2}{n\pi}.$$

Thus the Fourier series is

$$x \sim \sum \frac{(-1)^{n+1}2}{n\pi} \sin n\pi x. \quad \blacksquare$$

Example 8.3.3 The periodic function $f(x)$ of period 1 and satisfying $f(x) = x^2$ on $[0, 1)$ has the Fourier coefficients (note that $L = \frac{1}{2}$)

$$\begin{aligned} a_0 &= \int_0^1 x^2 \, dx = \frac{1}{3}, \\ a_n &= 2 \int_0^1 x^2 \cos 2n\pi x \, dx = \frac{1}{n^2\pi^2}, \\ b_n &= 2 \int_0^1 x^2 \sin 2n\pi x \, dx = -\frac{1}{n\pi}. \end{aligned}$$

Thus the Fourier series is

$$x^2 \sim \frac{1}{3} + \sum \frac{\cos 2n\pi x - n\pi \sin 2n\pi x}{n^2\pi^2}, \quad 0 < x < 1. \quad \blacksquare$$

8.3.2 Convergence of Fourier Series

Although the Fourier coefficients are computed based on the assumption that the function is the value of the Fourier series, the equality between the function and the series is yet to be rigorously established. The following gives the value of the Fourier series for reasonably good functions.

Theorem 8.3.2 Suppose $f(x)$ is a periodic integrable function. Suppose $f(x)$ has left limit $f(a^-)$ and right limit $f(a^+)$ at a , and there are $M, \delta > 0$, such that

$$0 < t < \delta \implies |f(a+t) - f(a^+)| \leq Mt, \quad |f(a-t) - f(a^-)| \leq Mt.$$

Then the Fourier series of $f(x)$ converges to $\frac{f(a^+) + f(a^-)}{2}$ at $x = a$.

The condition is satisfied when $f(x)$ has left and right derivatives at a . In particular, the Fourier series converges to $f(a)$ when $f(x)$ is differentiable at a .

Example 8.3.4 In Example 8.3.2, we computed the Fourier series of the odd function x on $(-1, 1)$. Since the function is differentiable on $(-1, 1)$, we have

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}2}{n} \sin n\pi x = x, \quad |x| < 1.$$

Near $x = 1$, the periodic function of period 2 is

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

It has left and right limits

$$f(1^-) = \lim_{x \rightarrow 1^-} x = 1, \quad f(1^+) = \lim_{x \rightarrow 1^+} f(x - 2) = \lim_{x \rightarrow 1^+} (x - 2) = -1,$$

and is left and right differentiable

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

Therefore we conclude that the value of the Fourier series at $x = 1$ is

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1} 2}{n\pi} \sin n\pi = \frac{1 + (-1)}{2} = 0.$$

Indeed this is true because $\sin n\pi = 0$ for all n .

Taking the value of the Fourier series at $x = \frac{1}{2}$ and $x = \frac{1}{4}$, we get

$$\begin{aligned} \frac{1}{2} &= \sum_{n=1}^{\infty} \frac{(-1)^{n+1} 2}{n\pi} \sin \frac{n\pi}{2} = \frac{2}{\pi} \sum_{n=0}^{\infty} \frac{1}{2n+1} \sin \frac{(2n+1)\pi}{2} = \frac{2}{\pi} \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} \\ &= \frac{2}{\pi} \left(1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots \right); \\ \frac{1}{4} &= \sum_{n=1}^{\infty} \frac{(-1)^{n+1} 2}{n\pi} \sin \frac{n\pi}{4} \\ &= \frac{2}{\pi} \sum_{n=0}^{\infty} \left(\frac{1}{4n+1} \sin \frac{(4n+1)\pi}{4} - \frac{1}{4n+2} \sin \frac{(4n+2)\pi}{4} + \frac{1}{4n+3} \sin \frac{(4n+3)\pi}{4} \right) \\ &= \frac{2}{\pi} \sum_{n=0}^{\infty} (-1)^n \left(\frac{1}{(4n+1)\sqrt{2}} - \frac{1}{4n+2} + \frac{1}{(4n+3)\sqrt{2}} \right) \\ &= \frac{2}{\pi\sqrt{2}} \left(1 + \frac{1}{3} - \frac{1}{5} - \frac{1}{7} + \cdots \right) - \frac{2}{\pi} \frac{1}{2} \left(1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots \right). \end{aligned}$$

This tells us

$$\begin{aligned} \frac{\pi}{4} &= \sum_{n=0}^{\infty} (-1)^n \frac{1}{2n+1} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \cdots, \\ \frac{\pi}{2\sqrt{2}} &= \sum_{n=0}^{\infty} (-1)^n \left(\frac{1}{4n+1} + \frac{1}{4n+3} \right) = 1 + \frac{1}{3} - \frac{1}{5} - \frac{1}{7} + \frac{1}{9} + \frac{1}{11} - \cdots \end{aligned}$$

■

Example 8.3.5 The periodic function of period 1 in Example 8.3.3 satisfies the condition of the convergence theorem for the Fourier series. By $f(0^+) = 0$ and $f(0^-) = f(1^-) = 1$, we have

$$\frac{1}{2} = \frac{1}{3} + \sum_{n=1}^{\infty} \frac{\cos 2n\pi - n\pi \sin 2n\pi}{n^2\pi^2} = \frac{1}{3} + \sum_{n=1}^{\infty} \frac{1}{n^2\pi^2}.$$

Therefore

$$\frac{\pi}{6} = \sum_{n=1}^{\infty} \frac{1}{n^2}. \quad \blacksquare$$

Exercises

8.3.1 The periodic functions are given in one interval of period length. Compute Fourier series.

1. $\sin^2 x$ on $(-\pi, \pi)$;
2. $|x|$ on $(-L, L)$;
3. x^2 on $(-1, 1)$;
4. $x \sin x$ on $(-\pi, \pi)$;
5. $\sin x$ on $(0, \pi)$;
6. $\cos ax$ on $(-\pi, \pi)$;
7. $f(x) = \begin{cases} ax, & \text{if } x \in (-1, 0], \\ bx, & \text{if } x \in [0, 1) \end{cases}$ on $(-1, 1)$.

8.3.2 Find the relation between the Fourier coefficients of $f(x)$, $f'(x)$ and $\int_0^x f(t)dt$. Then use the Fourier series of x on $(-1, 1)$ to get the Fourier series of x^2 and x^3 on $(-1, 1)$.

8.3.3 Extend x^2 on $(0, \pi)$ to an even periodic function $f(x)$ and an odd periodic function $g(x)$, both with period 2π . Use the Fourier series of $f(x)$ and $g(x)$ to compute the sum of series

$$\sum_{n=1}^{\infty} \frac{1}{n^2}, \quad \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^2}, \quad \sum_{n=1}^{\infty} \frac{1}{(2n-1)^2}.$$

8.4 SUMMARY

Definitions

- The **partial sum** of a series $\sum_{n=1}^{\infty} a_n$ is $s_n = \sum_{i=1}^n a_i$. The series **converges** if the sequence s_n converges, and has **sum** $\sum_{n=1}^{\infty} a_n = \lim_{n \rightarrow \infty} s_n$.
- A series $\sum a_n$ **absolutely converges** if $\sum |a_n|$ converges. The series **conditionally converges** if $\sum a_n$ converges and $\sum |a_n|$ diverges.

- A **power series** is a series of the form $\sum_{n=0}^{\infty} a_n x^n$.
- The **Taylor series** of a smooth function $f(x)$ at $x = a$ is $\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!} (x - a)^n$.
- The **Fourier series** of a periodic function of period $2L$ is

$$f(x) \sim a_0 + \sum_{n=1}^{\infty} \left(a_n \cos \frac{n\pi x}{L} + b_n \sin \frac{n\pi x}{L} \right),$$

with **Fourier coefficients**

$$a_0 = \frac{1}{2L} \int_0^{2L} f(x) dx, \quad a_n = \frac{1}{L} \int_0^{2L} f(x) \cos \frac{n\pi x}{L} dx, \quad b_n = \frac{1}{L} \int_0^{2L} f(x) \sin \frac{n\pi x}{L} dx.$$

- The **upper limit** $\overline{\lim}_{n \rightarrow \infty} a_n$ of a sequence a_n is the unique λ satisfying (1) If $l > \lambda$, then finitely many $a_n > l$; (2) If $l < \lambda$, then infinitely many $a_n > l$.

Theorems

- If $\sum a_n$ converges, then $\lim_{n \rightarrow \infty} a_n = 0$.
- A series with non-negative terms converges if and only if the partial sums are bounded.
- Cauchy Criterion: A series $\sum a_n$ converges if and only if for any $\epsilon > 0$, there is N , such that $n \geq m > N$ implies $|a_m + a_{m+1} + \cdots + a_n| < \epsilon$.
- Comparison Test: If $|a_n| \leq b_n$ and $\sum b_n$ converges, then $\sum a_n$ converges.
- Absolute convergence implies convergence.
- Root Test: If $\lim_{n \rightarrow \infty} \sqrt[n]{|a_n|} < 1$, then $\sum a_n$ converges. If $\lim_{n \rightarrow \infty} \sqrt[n]{|a_n|} > 1$, then $\sum a_n$ diverges.
- Ratio Test: If $\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| < 1$, then $\sum a_n$ converges. If $\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| > 1$, then $\sum a_n$ diverges.
- Integral Test: If $f(x)$ is nonincreasing and $\lim_{x \rightarrow +\infty} f(x) = 0$, then $\sum f(n)$ converges if and only if the improper integral $\int_a^{+\infty} f(x) dx$ converges.
- Leibniz Test: If a_n is decreasing and $\lim_{n \rightarrow \infty} a_n = 0$, then the **alternating series** $\sum (-1)^{n+1} a_n$ converges.
- Dirichlet Test: If the partial sum of $\sum a_n$ is bounded, b_n is monotonic and $\lim_{n \rightarrow \infty} b_n = 0$, then $\sum a_n b_n$ converges.

- Abel Test: If $\sum a_n$ converges, b_n is monotonic and bounded, then $\sum a_n b_n$ converges.
- The sum of an absolutely convergent series is independent of the order of the terms. The sum of a conditionally convergent series can be any number after rearranging the order of the terms.
- Product Formula: If $\sum_n a_n$ and $\sum_n b_n$ absolutely converge, then an arrangement of the double series $\sum_{m,n} a_m b_n$ absolutely converge, and the sum $\sum_{m,n} a_m b_n = (\sum_n a_n)(\sum_m b_m)$.
- Remainder of Taylor Expansion: Suppose $f(x)$ has $(n+1)$ st order derivative between a and x , and $T_n(x)$ is the n -th partial sum of the Taylor series, then the **remainder** $R_n(x) = f(x) - T_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}$ for some c between a and x .
- Radius of Convergence: For any power series $\sum a_n x^n$, there is a number $R \geq 0$ (possibly $+\infty$), such that the series converges for $|x| < R$ and diverges for $|x| > R$. The **radius of convergence** is given by $R = \frac{1}{\lim_{n \rightarrow \infty} \sqrt[n]{|a_n|}}$.
- Product of Power Series: $(\sum a_n x^n)(\sum b_n x^n) = \sum c_n x^n$, where $c_n = a_0 b_n + a_1 b_{n-1} + \cdots + a_n b_0$.
- Power series can be differentiated and integrated term by term within the radius of convergence.
- Convergence of Fourier Series: The Fourier series of $f(x)$ converges to $f(a)$ at $x = a$ if the function is differentiable at a . Moreover, it converges to $\frac{f(a^+) + f(a^-)}{2}$ at $x = a$ if $f(x)$ is left and right differentiable at a .