AMA1007 / AMA1120
Calculus and Linear Algebra

Department of Applied Mathematics
The Hong Kong Polytechnic University
1st Semester 2022/2023

About the subject

Responsible Staff

Subject Lecturer: **Dr. LEE Heung Wing Joseph, BSc, PhD** 李向榮博士
Contact: via **Zoom Chat Function** (a built-in instant message function in zoom)

Tutor: Mr. YEUNG Hon Keung Angus, BSc, MPhil
Email address: angus.yeung@polyu.edu.hk , Telephone: 2766 6943

Please note that students should be using their official PolyU student account only (email and zoom accounts without changing the original given account name) to contact teaching team members (stating clearly the subject code, student name, and student id) .

Subject Web Page

https://www.polyu.edu.hk/ama/profile/hwlee/ama1007.html

Reading List and References

- *A Short Course in Calculus and Matrices* by **Kwok-Chiu Chung**, McGraw Hill 2008.
- *Calculus. 7th ed.* by **James Stewart**, Brooks/Cole 2012.
- *Thomas' Calculus 12th ed.* by **George B. Thomas Jr., Maurice D. Weir, Joel Hass**, Brooks/Cole 2012.
- *Elementary Linear Algebra 9th ed.* by **Howard Anton, Chris Rorres**, John Wiley and Sons, 2005.

Students must watch all the four **CoCalc** demo videos in full:

https://video.polyu.edu.hk/Panopto/Pages/Viewer.aspx?id=88d760d9-622b-47ed-b43a-ac2201157764

https://video.polyu.edu.hk/Panopto/Pages/Viewer.aspx?id=b7d2c61b-5bb2-4d1c-8af3-ac2200871479

https://video.polyu.edu.hk/Panopto/Pages/Viewer.aspx?id=47b10248-d951-4c23-9bac-ac2300f65c1e

https://video.polyu.edu.hk/Panopto/Pages/Viewer.aspx?id=8e2425b3-8544-4335-9ece-ac2b007e3380

Grading Policy

Continuous Assessment :	Assignments	10%
	Test	30%
Examination :		60%

Midterm Test and Examination rubric

A-/A /A+	80 - 100 (out of 100)
B-/ B /B+	65 - 79 (out of 100)
C-/ C /C+	50 - 64 (out of 100)
D /D+	40 - 49 (out of 100)
F	0 - 39 (out of 100)

Assignments:
There are 5 assignment sets. Solutions with detailed workings and explanations should be submitted by 5pm of the corresponding due dates:
(06 Oct, 13 Oct, 20 Oct, 10 Nov, 24 Nov).
Students should submit their solutions of the assignments via Blackboard.

- Solutions must be scanned into one single clear and readable PDF file using **Microsoft Office Lens**, but
- with file size no bigger than 3MB, and
- the file name must be the student name with surname first, with the covering declaration signed.

Midterm Test:
The Mid-term Test could be scheduled in one of the lecture between Week 10 to Week 12 within normal lecture time. Date and Venue TBA. There are 15 multiple choice questions in the test.

Learning Outcomes

This is a subject to provide students with a solid foundation in Differential and Integral Calculus, and in Matrix Algebra. Upon satisfactory completion of the subject, students are expected to be able to:

- solve problems using the concept of functions and inverse functions
- apply the basic operations of matrices and calculate the determinant
- apply mathematical reasoning to analyse essential features of different mathematical problems such as differentiation and integration
- apply appropriate mathematical techniques to model and solve problems in science and engineering
- extend their knowledge of mathematical techniques and adapt known solutions in different situations

Elementary Functions

Set notations

- A *set* is a collection of *objects*.
- An *element* of a set is an object in the set.
- Objects — lower case; Sets — Upper case.
- $x \in A$ means "x is an element of the set A".
- $x \notin A$ means x is not an element of A.

Sets are described by:

- listing the elements, e.g. $A = \{2, 3, 4, 5\}$.
- stating what special property a typical element x of the set has, e.g. $A = \{x : x$ is an integer and $2 \leq x \leq 5\}$.

Universal and empty sets:

- $\emptyset$ denotes the *empty set*, the set that contains no element (in some other texts, symbol ϕ is used instead).
- Ω denotes the universal set.

Sets sometimes are represented by Venn diagrams. A Venn diagram is an oval drawn on the plane so that all elements of the set are considered to be inside the oval.
In the following, A and B are sets.

- A is a subset of B (written $A \subset B$) if every element of A is an element of B.
- A and B are equal (written $A = B$) if they contain the same elements, i.e. $A \subset B$ and $B \subset A$. The following sets S, T, U are equal.

$$S = \{ 1, 2, 3, 4 \}, \quad T = \{ 2, 4, 3, 1 \}, \quad U = \{ 2, 1, 4, 4, 2, 3 \},$$

- The *intersection* $A \cap B$ is the set $\{ x : x \in A \text{ and } x \in B \}$.
- The *union* $A \cup B$ is the set $\{ x : x \in A \text{ or } x \in B \}$.
- The *relative complement* $A \setminus B$ is the set $\{ x : x \in A \text{ and } x \notin B \}$.
- Absolute complement: $A^c = \Omega \setminus A$, (denoted by $\overline{A}$ in some other texts).
- Disjoint Sets, A and B are disjoint if $A \cap B = \emptyset$.
- Product of 2 sets

$$S \times T = \{(s, t) : s \in S \ \text{ and } \ t \in T\}.$$

Note that the empty set is different from the singleton sets $\{0\}$ or $\{\emptyset\}$.
Of course $\emptyset \in \{\emptyset\}$.
Quick Questions : Is it true that $\emptyset = \{\emptyset\}$?

Note that in some texts the symbol $\subseteq$ is used to denote subset instead of the commonly used symbol $\subset$.

The followings are not part of the formal mathematical language. We only use them in informal occasions.

- For all $\forall$
- there exists $\exists$
- there exists a unique .... $\exists!$
- implies $\Rightarrow$

Example Consider the following statement about the density of real numbers:

$$\forall\ x > 0,\ \exists\ y > 0 \text{ such that } x > y > 0.$$

Laws of Algebra of Sets

$A \cup B = B \cup A$ $A \cap B = B \cap A$	Commutative Laws
$(A \cup B) \cup C = A \cup (B \cup C)$ $(A \cap B) \cap C = A \cap (B \cap C)$	Associative Laws
$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$	Distributive Laws
$A \cup A = A$ $A \cap A = A$	Idempotent Laws
$A \cup \emptyset = A$ $A \cup \Omega = \Omega$ $A \cap \emptyset = \emptyset$ $A \cap \Omega = A$	Identity Laws
$(A^c)^c = A$	Double complementation
$A \cup A^c = \Omega$ $A \cap A^c = \emptyset$	
$\Omega^c = \emptyset$ $\emptyset^c = \Omega$	
$(A \cup B)^c = A^c \cap B^c$ $(A \cap B)^c = A^c \cup B^c$	DeMorgan Laws

Power Set

The symbol $|S|$ denotes the number of elements in the set S. For example, $|\emptyset| = 0$.

$\mathcal{P}(S)$ denotes the **Power set** of the set S, it contains all possible subsets of S.

If $S = \{a, b, c\}$, then $\mathcal{P}(S) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, S\}$.

Quick Question :
If there are n elements in S, how many elements are there in $\mathcal{P}(S)$?
In other words, if $|S| = n$, what is $|\mathcal{P}(S)|$?

Note that $\emptyset \in \mathcal{P}(S)$ and $S \in \mathcal{P}(S)$.

Note also that S is **not** a subset of $\mathcal{P}(S)$. Rather, S is an element of $\mathcal{P}(S)$.
For example, $a \in S$ but $a \notin \mathcal{P}(S)$. Instead, we have $\{a\} \subset S$ and $\{a\} \in \mathcal{P}(S)$.

Real numbers and intervals

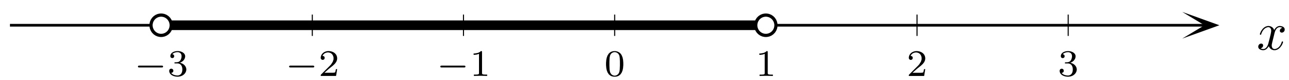
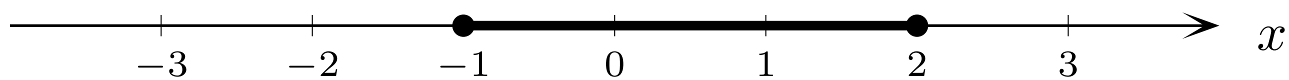
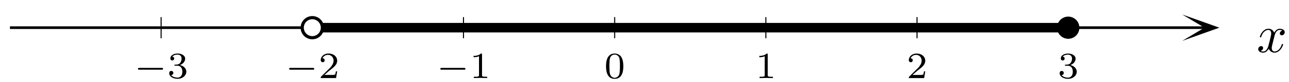
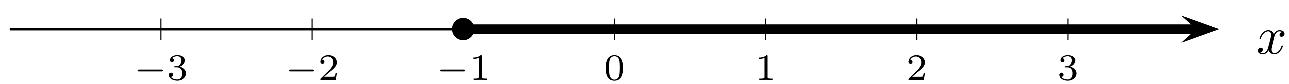
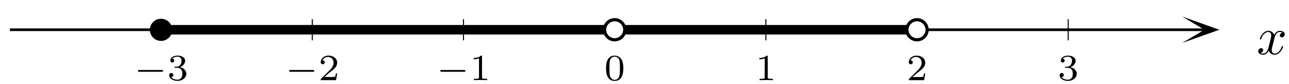
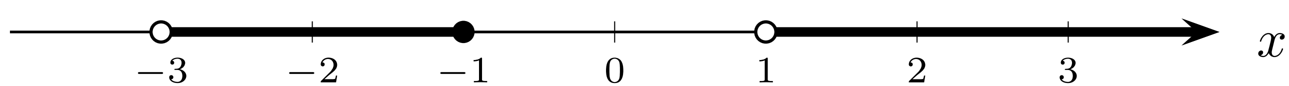
Real numbers are numbers represented as points on a straight line which extends indefinitely on both sides. The set of all real numbers are usually denoted by the symbol $\mathbb{R}$. *Intervals* are subsets of $\mathbb{R}$ described in the following table. The real numbers a and b (with $a < b$) for defining the intervals are the *endpoints* of the intervals.

Notation	Set description	Type
(a, b)	$\{ x \in \mathbb{R} : a < x < b \}$	open
$(a, b]$	$\{ x \in \mathbb{R} : a < x \leq b \}$	half-open
$[a, b)$	$\{ x \in \mathbb{R} : a \leq x < b \}$	half-open
$[a, b]$	$\{ x \in \mathbb{R} : a \leq x \leq b \}$	closed
(a, ∞)	$\{ x \in \mathbb{R} : a < x \}$	open
$[a, \infty)$	$\{ x \in \mathbb{R} : a \leq x \}$	closed
$(-\infty, b)$	$\{ x \in \mathbb{R} : x < b \}$	open
$(-\infty, b]$	$\{ x \in \mathbb{R} : x \leq b \}$	closed
$(-\infty, \infty)$	$\mathbb{R}$	open and closed

Question: Why $\mathbb{R} = (-\infty, \infty)$ could be considered closed?

- Observe that the open interval (a, a) is empty $\emptyset$.
- So $(a, a) = \emptyset$ is open.
- Now observe that the sets $\emptyset$ and $\Omega = \mathbb{R}$ are complementing each other.
- The complement of a closed interval is open, and the complement of an open interval is closed.
- Since $(a, a) = \emptyset$ is open. Therefore, $\mathbb{R}$ could be considered closed.

The following table shows a few examples of subsets of $\mathbb{R}$ as well as their graphs drawn on the x-axis.

Subsets	Diagrams for the subsets
$(-3,\ 1)$	
$[-1,\ 2]$	
$(-2,\ 3]$	
$[-1,\ \infty)$	
$[-3,\ 0) \cup (0, 2)$	
$(-3, -1] \cup (1,\ \infty)$	

.

Absolute values

If x is a real number, the *absolute value* of x is its distance from the origin O. We use the symbol $|x|$ to denote the absolute value. Mathematically,

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

Therefore we have $|3| = 3$, $|-4| = 4$ and $|0| = 0$.

Properties. Let $a, b \in \mathbb{R}$. Then

- $|a - b| =$ distance between a and b on the real line.
- $|ab| = |a||b|$, $\quad |a \pm b| \leq |a| + |b|$ (the triangle inequality).
- $|a| < b$ iff[1] $-b < a < b$. Also, $|a| \leq b$ iff $-b \leq a \leq b$.

[1] iff means "if and only if"

Basic concepts of functions

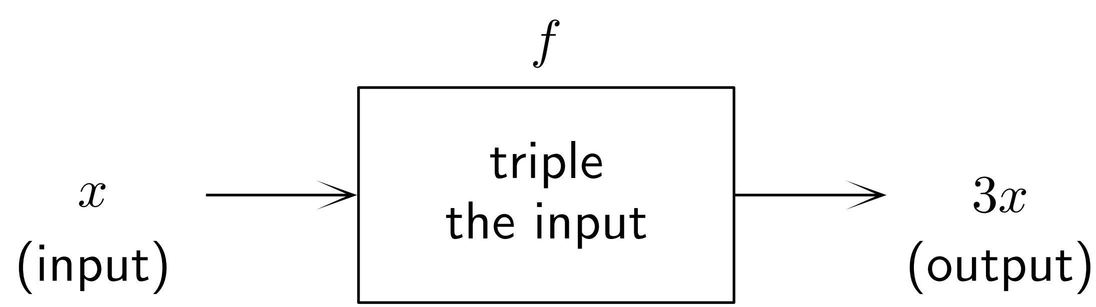
A function is a rule which, when given a number (input), produces a *single* number (output). Consider the function (or the rule) by which the output is three times the input.

input		output
2	$\longmapsto$	6
x	$\longmapsto$	$3x$
t	$\longmapsto$	$3t$
$s-4$	$\longmapsto$	$3(s-4)$.

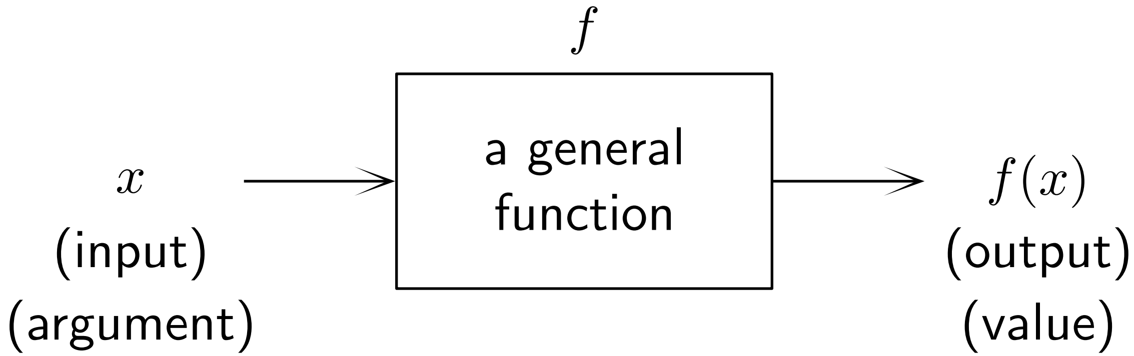
If we denote this function by f, the function can be represented by

$f : x \longmapsto 3x$ or $f(x) = 3x$ or simply $y = 3x$

The above function f can be thought of as a machine that gives an output $3x$ if we input x to it.



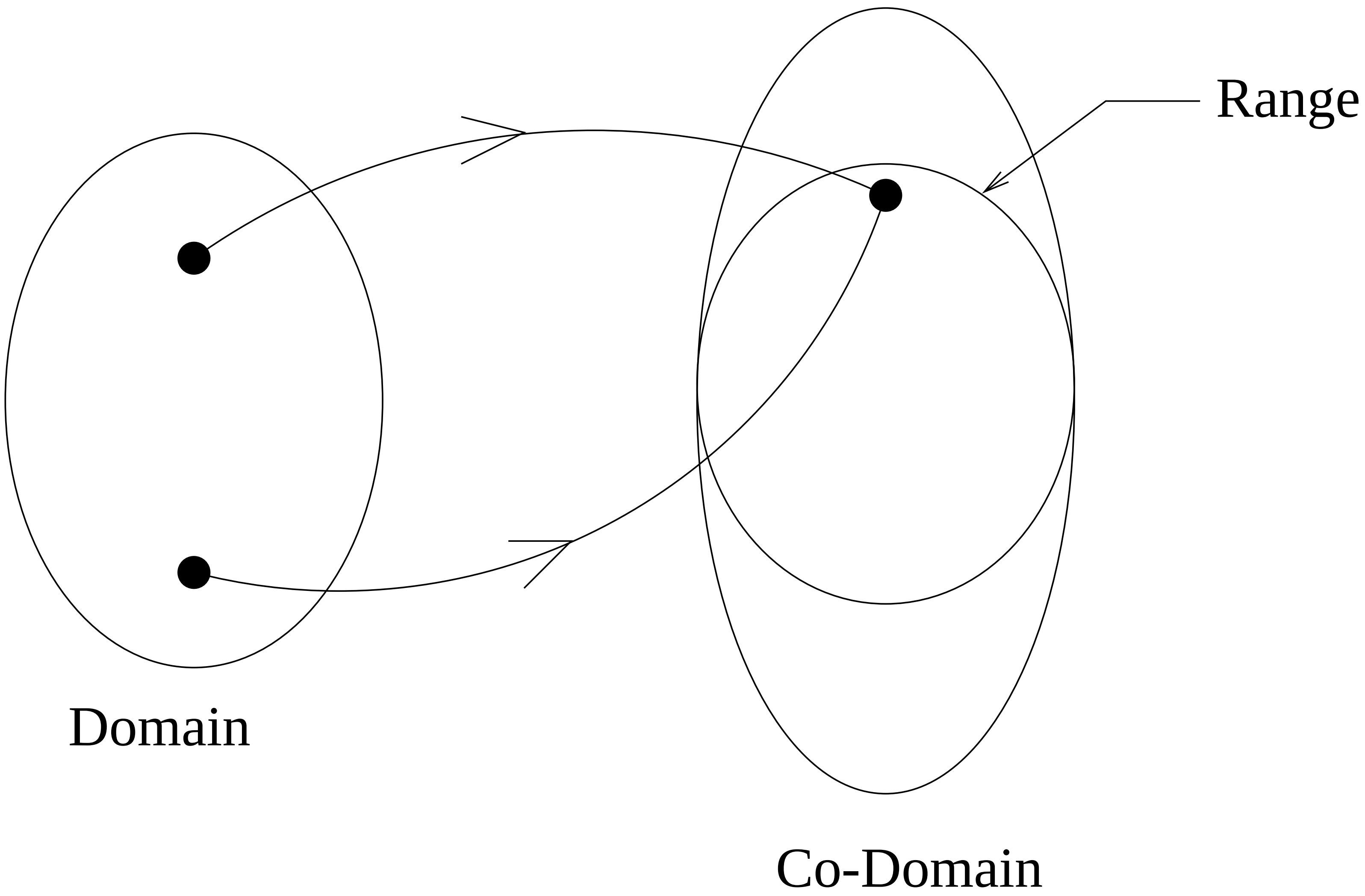
For a general function f, we have



- The input to a function f is called the *argument* and the corresponding output the *value* of the function.
- If the argument is a given number x (so x is given and fixed), the value is denoted by $f(x)$.
- If the argument is a variable number x and $y = f(x)$, then x is called the *independent variable*, y is the *dependent variable*.

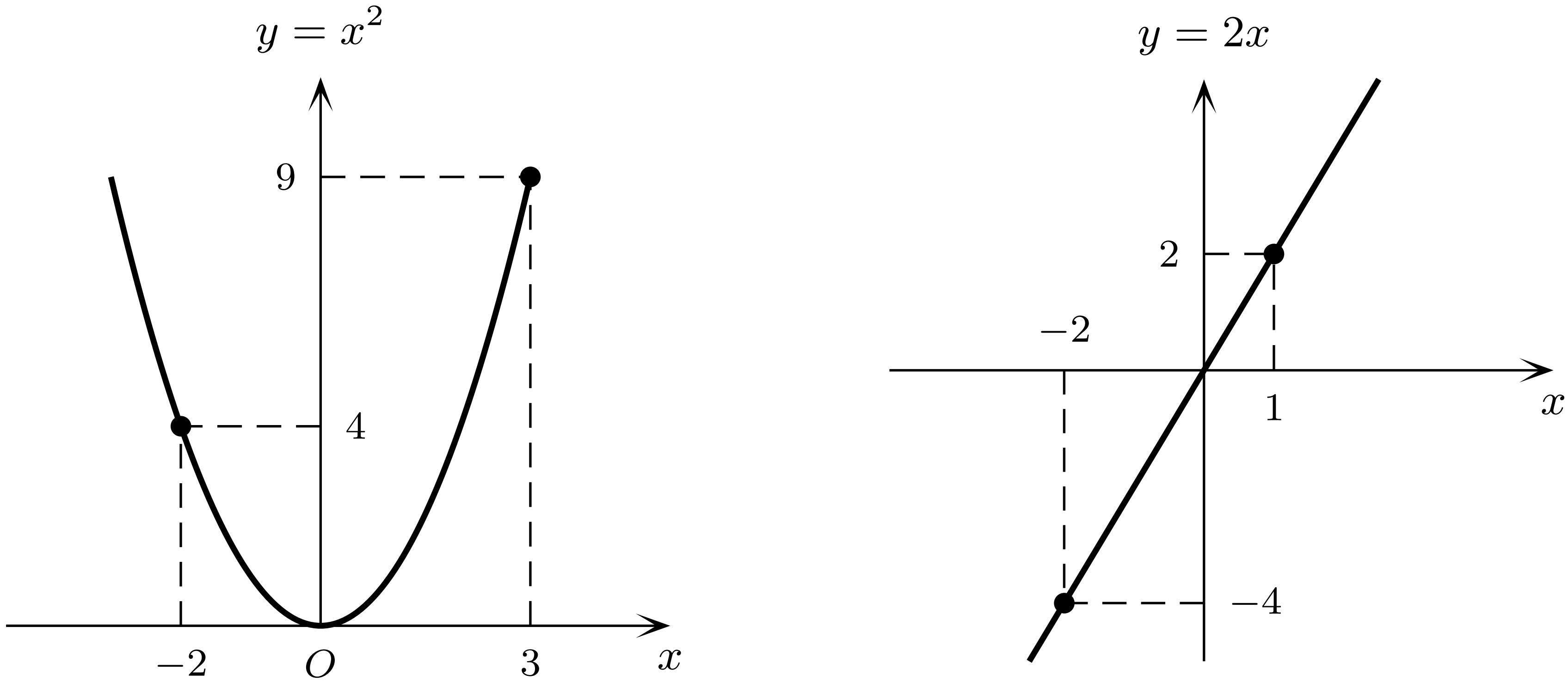
To indicate the symbol (x here) being used as the independent variable, we sometimes denote the function by $f(x)$, rather than just by f, and say that $f(x)$ is a function.

Function: Map



Graph of a function

Given a function $y = f(x)$, we can plot the points (x, y) on the xy-plane so that the function values y are plotted vertically and the x-values horizontally. Figures below show the graphs of two elementary functions.



Graph of $f(x) = x^2$ Graph of $f(x) = 2x$

Figure 3.1: Two examples of graphs of functions.

Domain and range

Consider a given function $y = f(x)$. The set of values that x is allowed to take is called the *domain* of f, written for short as $\mathrm{Dom}\, f$. The domain is sometimes given when a function is defined. For instance, we can define a function f in the following way:

$$f(x) = x + 2, \quad 1 \leq x < 3.$$

This indicates that the domain of the function is the interval $[1, 3)$ so that f is not defined for x lies outside $[1, 3)$. However, if the domain is not explicitly given, it is taken to be the largest set possible. For example, consider the function g defined by

$$g(x) = \sqrt{x - 2},$$

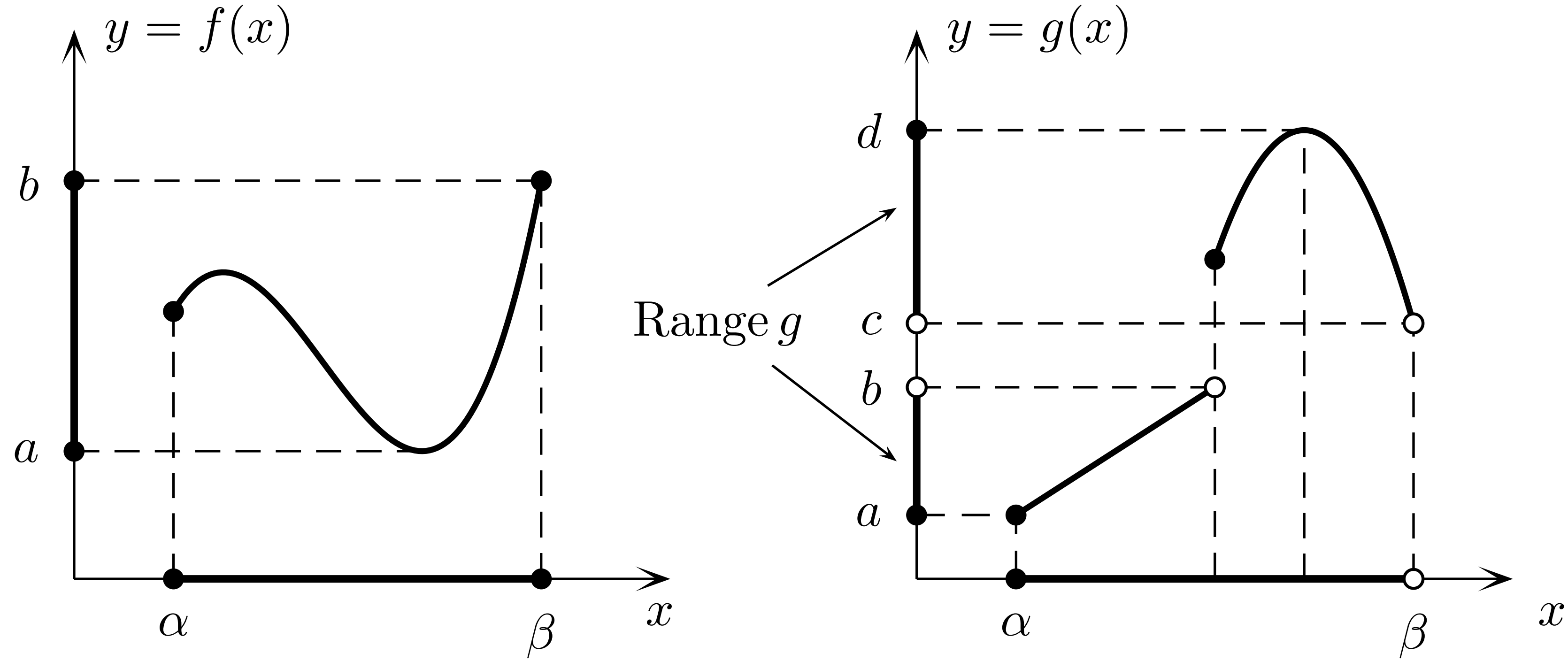
where no domain is explicitly given. However, we understand that the domain of g is $[2, \infty)$ because it is the largest possible set of real numbers x for which $\sqrt{x - 2}$ are real. The domain of

$$F(x) = x^2, \quad -\infty < x < \infty$$

is $\mathbb{R}$. If we restrict the domain of F to $x > 2$ we get a new function G so that

$$G(x) = x^2, \quad x > 2.$$

The set of values that the function f takes on is called the *range* of the function, written for short as Range f. To find $\mathrm{Range}\, f$ we ask the following question: What are the values of $y = f(x)$ where $x \in \mathrm{Dom}\, f$?



$$\mathrm{Dom}\, f = [\alpha, \beta]$$
$$\mathrm{Range}\, f = [a, b]$$

$$\mathrm{Dom}\, g = [\alpha, \beta)$$
$$\mathrm{Range}\, g = [a, b) \cup (c, d]$$

Figure 3.2: The domain and the range of functions.

Example 3.1 Find the domains and ranges of the functions f, g, F and G defined previously.

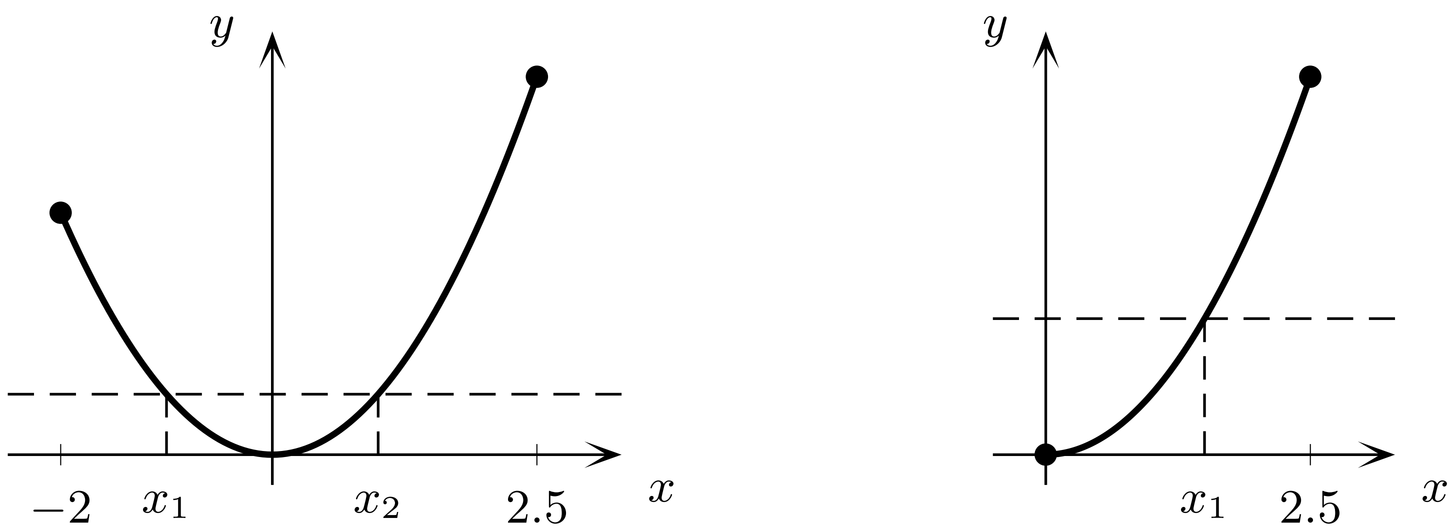
Solution. The answers are shown in the following table:

Function	**Domain**	**Range**
$f(x) = x + 2, \ 1 \le x < 3$	$1 \le x < 3$	$3 \le y < 5$
$g(x) = \sqrt{x - 2}$	$2 \le x < \infty$	$0 \le y < \infty$
$F(x) = x^2, \ x \in \mathbb{R}$	$\mathbb{R}$	$0 \le y < \infty$
$G(x) = x^2, \ x > 2$	$2 < x < \infty$	$4 < y < \infty$

The ranges can be found by considering the graphs of the functions. □

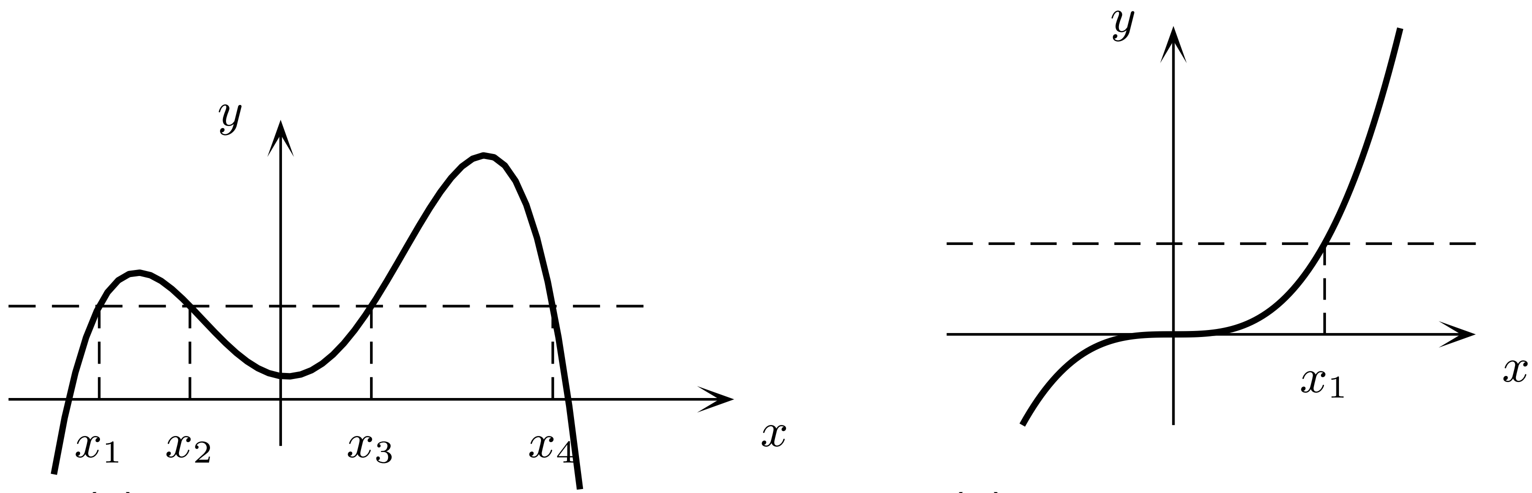
One-to-one functions

Consider the function $f(x) = x^2, \quad -2 \le x \le 2.5$. As -1 and 1 are in the domain and $f(-1) = 1 = f(1)$, we see that two different inputs produce the same output. This is demonstrated in the figure below where there are two distinct numbers x_1, x_2 in the domain with $f(x_1) = f(x_2)$. In this case, we say that the function f is *many-to-one*. A function is *one-to-one* or *injective* if different inputs produce different outputs. If we change the domain of the above function to $0 \le x \le 2.5$, we have a new function $g(x) = x^2$, $0 \le x \le 2.5$. This function g is one-to-one.



(a) $f(x) = x^2, \ -2 \le x \le 2.5$ (b) $g(x) = x^2, \ 0 \le x \le 2.5$

Figure 3.3: f is many-to-one and g one-to-one.



(a) A many-to-one function.

(b) A one-to-one function.

Figure 3.4: Many-to-one and one-to-one functions.

Composition of functions

Consider the function $y = 3x^2$. The value y of this function can be obtained in two stages: first we square the input x, then we triple the result.

$$x \longmapsto x^2 \longmapsto 3x^2$$

If g and h are functions defined by

$$g(x) = x^2 \quad \text{and} \quad h(x) = 3x$$

we can write

$$y = 3x^2 = 3g(x) = h(g(x)).$$

The function that is defined in terms of two functions g and h this way is denoted by $h \circ g$. That is,

$$\boxed{h \circ g(x) = h(g(x))}$$

Example 3.2 Consider functions g and h defined by

$$g(x) = x + 1 \quad \text{and} \quad h(x) = x^2 \quad \text{for all } x \in \mathbb{R}.$$

Show that $g \circ h \neq h \circ g$ and $g \circ h(0) = h \circ g(0)$.

Solution. We find that

$$g \circ h(x) = g(h(x)) = g(x^2) = x^2 + 1$$

and

$$h \circ g(x) = h(g(x)) = h(x + 1) = (x + 1)^2.$$

Therefore $g \circ h \neq h \circ g$ and $g \circ h(0) = h \circ g(0)$. □

Inverse functions

For a given function f, suppose that the input x produces the output y, i.e. $y = f(x)$.
We ask: Is there a function g (which depends on the given f) such that

- $\mathrm{Dom}\, g = \mathrm{Range}\, f$, and that
- $g(f(x)) = x$ (i.e. if $y = f(x)$ then $g(y) = x$) for all $x \in \mathrm{Dom}\, f$?

If such a function g exists, we call this the *inverse function* of f and write $g = f^{-1}$.
In this case, we have

$$\mathrm{Dom}\, f = \mathrm{Range}\, f^{-1} \quad \text{and} \quad \mathrm{Dom}\, f^{-1} = \mathrm{Range}\, f.$$

Theorem 3.1 *If the function f is one-to-one, the inverse of f exists so that*

$$f^{-1}(f(x)) = x \quad \textit{for all } x \in \mathrm{Dom}\, f$$

and

$$f(f^{-1}(y)) = y \quad \textit{for all } y \in \mathrm{Dom}\, f^{-1}.$$

Example 3.3 Find the inverse function of $f(x) = \sqrt{x} + 1,\ 0 \leq x \leq 4$.

Solution. Write $y = \sqrt{x} + 1$ and solve x in terms of y. We obtain

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

from which we see that the given f is one-to-one. The range of f is obviously (perhaps from the graph of f) $1 \leq y \leq 3$. Therefore

$$f^{-1}(y) = (y - 1)^2, \quad 1 \leq y \leq 3. \qquad \square$$

If we wish to use x rather than y as the independent valuable, we can replace y by x in the above to get another form of the solution:

$$f^{-1}(x) = (x - 1)^2, \quad 1 \leq x \leq 3.$$

Example 3.4 Consider the function f defined by $f(x) = x^2,\ -3 \le x \le 3$. Show that this function is many-to-one and hence has no inverse.

Solution. If we solve $y = x^2$ for x, we get two results

$$x = \sqrt{y} \quad \text{and} \quad x = -\sqrt{y}.$$

Since $\sqrt{y} \neq -\sqrt{y}$ if $y \neq 0$, we see that there are two different x-values taking the same nonzero y-value. This shows that the function is many-to-one. □

However, if we restrict the domain to say the interval $[0, 3]$, the function f becomes a new function F which is one-to-one and whose range is $[0, 9]$.

Periodic functions

A function $f(x)$ defined on $\mathbb{R}$ is said to be *periodic* if there is a positive constant T (called a *period*) such that

$$f(x+T) = f(x) \quad \text{for all } x \in \mathbb{R}.$$

Clearly, if T is a period of f, so are $2T$, $3T$, etc. Usually, when we say *the period* of a function, we mean the smallest period.

Example 3.5 For the functions shown below, each has a period of 2.

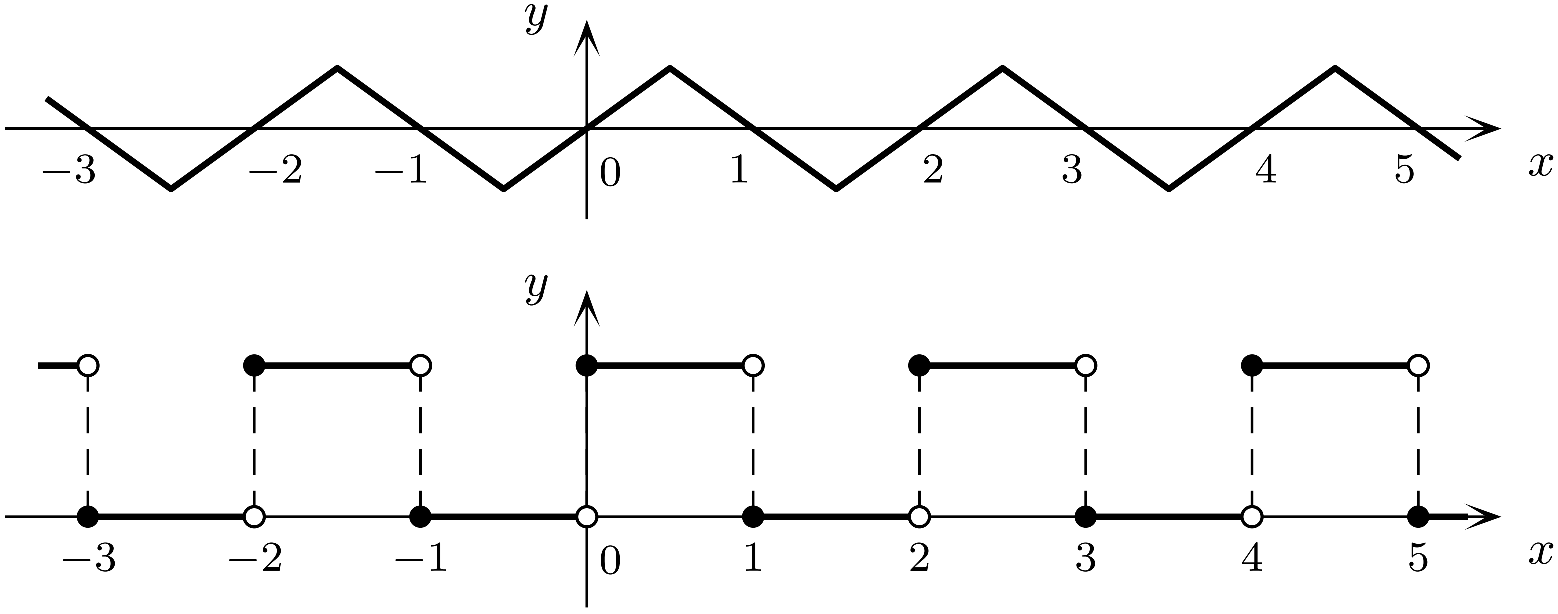
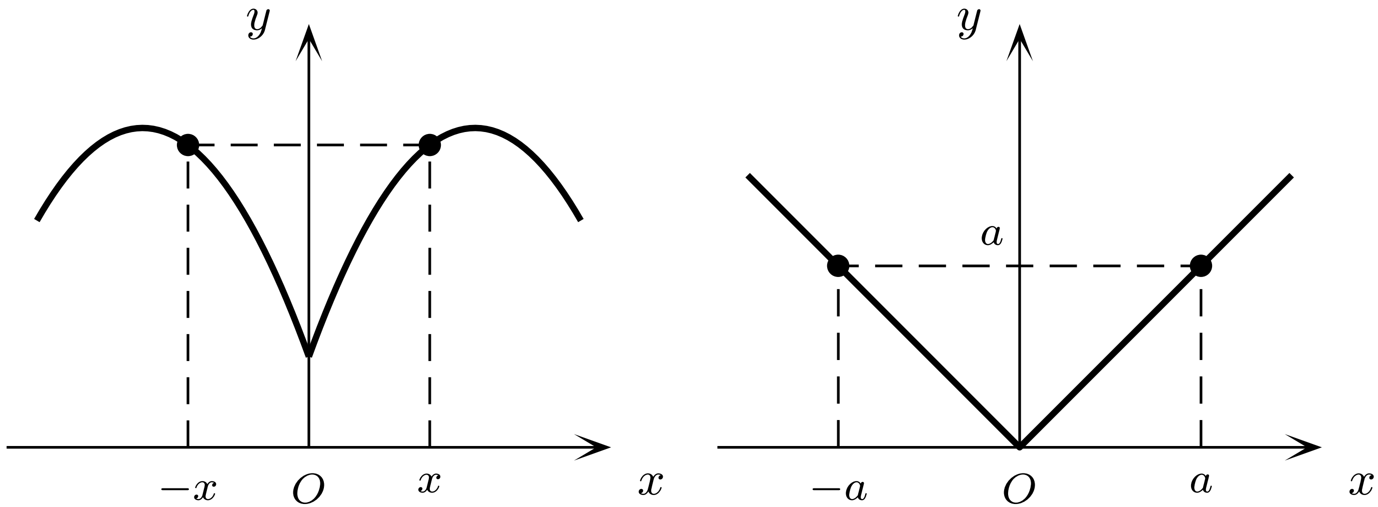


Figure 3.5: Two examples of periodic functions.

Even and odd functions

A function f is an *even* function if $f(-x) = f(x)$ for all $x \in \mathbb{R}$. The graph of an even function is symmetrical about the y-axis.

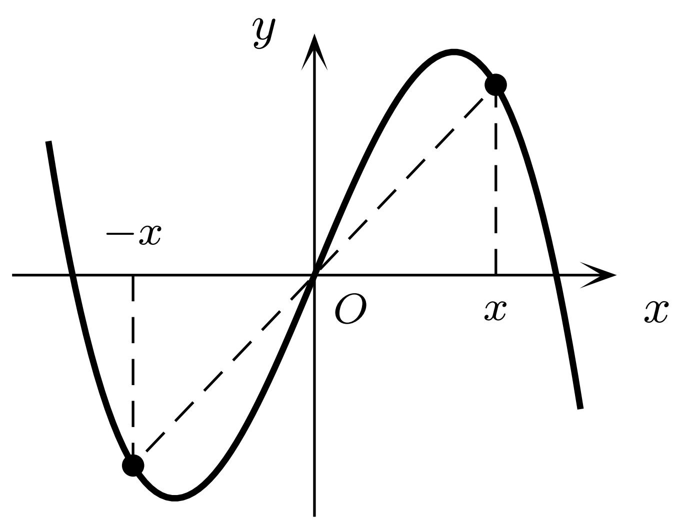


(a) An even function (b) The function $|x|$

Example 3.6 $|x|$, x^{2k} (k is an integer) and $\cos x$ are even functions.

Note that $|x|$ is an even function.

A function f is an *odd* function if $f(-x) = -f(x)$ for all $x \in \mathbb{R}$. The graph of an odd function is symmetrical about the origin.



(c) An odd function

Example 3.7 x^{2k+1} (k is an integer) and $\sin x$ are odd functions.

Polynomials

A polynomial is a function of the form

$$P(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

where $a_0,\ a_1, \ldots,\ a_n$ are given constants (called the *coefficients*) and x is the independent variable. The domain of $P(x)$ is $\mathbb{R}$. If $a_n \neq 0$, n is the *degree* of $P(x)$. We sometimes write $\deg P$ for the degree of $P(x)$.

If all the coefficients $a_0,\ a_1, \ldots,\ a_n$ are zero, the polynomial reduces to *the zero polynomial*. The degree of the zero polynomial is regarded as 0 in this book. A *zero* of $P(x)$ is a root (or a solution) of the equation $P(x) = 0$.

Polynomial	Degree	Name
a_0	0	constant
$a_0 + a_1x,\ (a_1 \neq 0)$	1	linear
$a_0 + a_1x + a_2x^2,\ (a_2 \neq 0)$	2	quadratic
$a_0 + a_1x + a_2x^2 + a_3x^3,\ (a_3 \neq 0)$	3	cubic

The graphs of polynomials are continuous curves.

The following theorems are fundamental.

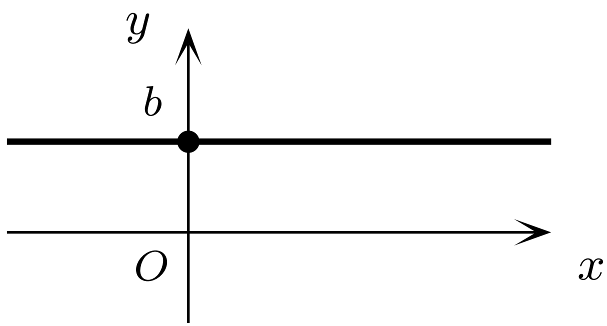
Theorem 3.2 (Remainder theorem) *If we divide a polynomial $P(x)$ by $x - a$, the remainder is $P(a)$.*

Theorem 3.3 (Fundamental theorem of algebra) *If $P(z)$ is a polynomial of degree n (with real or complex coefficients, $n \neq 0$), the equation $P(z) = 0$ has exactly n roots (counting real roots, complex roots and their multiplicities).*

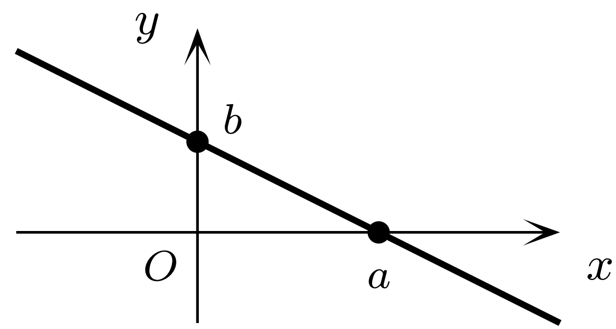
This theorem involves complex numbers[2] and is the rare occasion where complex numbers are mentioned in this book (this set of notes).

---

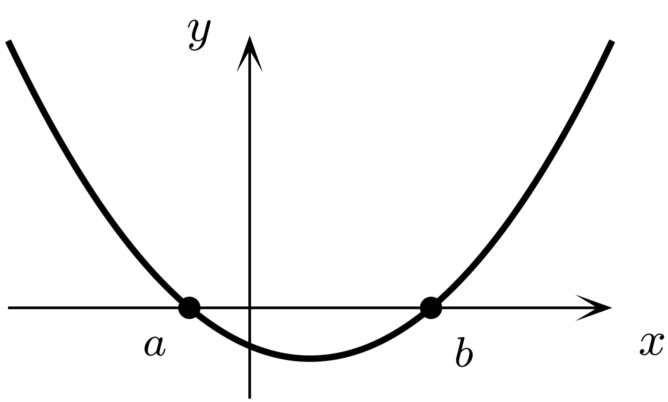
[2]See for example, Weir, et al. *Thomas' Calculus*, 8th edition, 2005



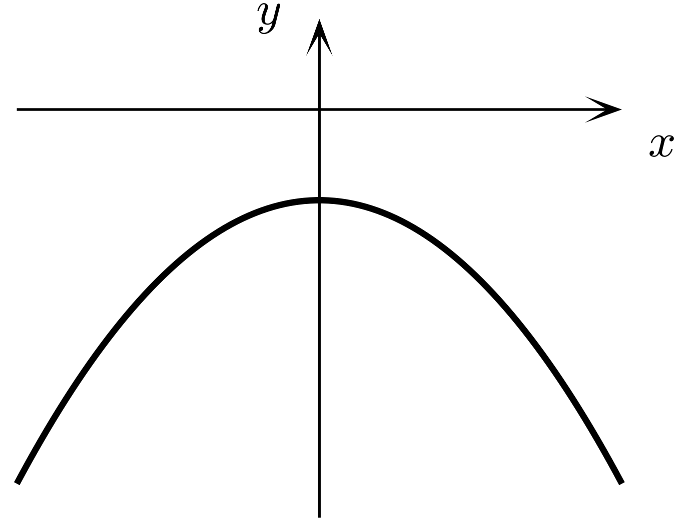
(a) Degree 0: $y = b$



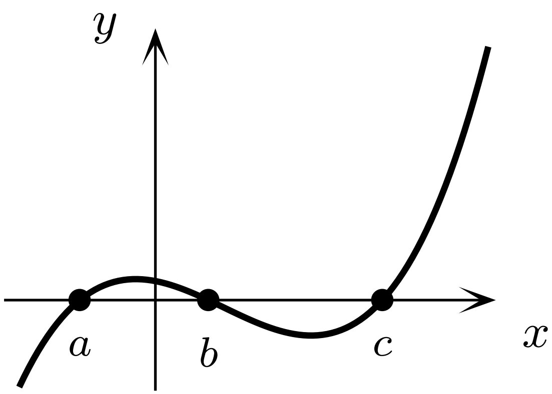
(b) Degree 1: $x/a + y/b = 1$



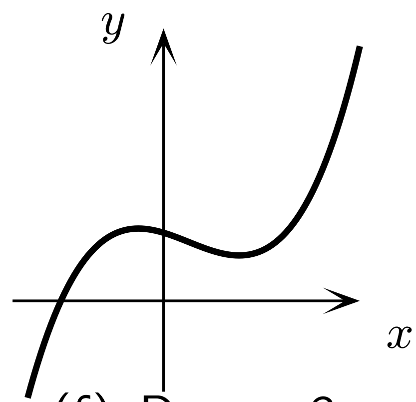
(c) Degree 2: $y = (x - a)(x - b)$



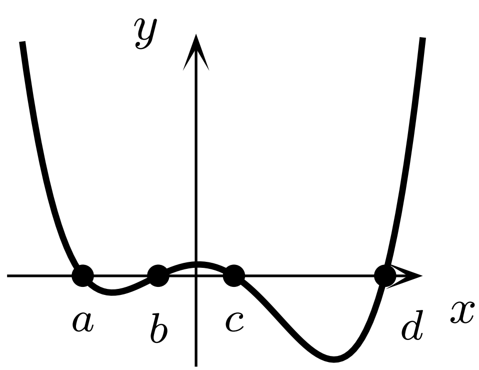
(d) Degree 2: $y = -x^2 - 2$



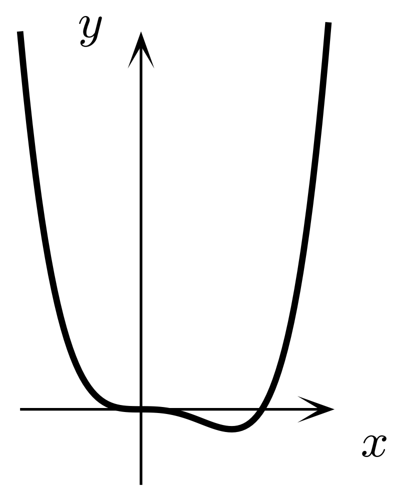
(e) Degree 3:
$y = (x-a)(x-b)(x-c)$



(f) Degree 3:
$y = x^3 - x^2 - x + 2$



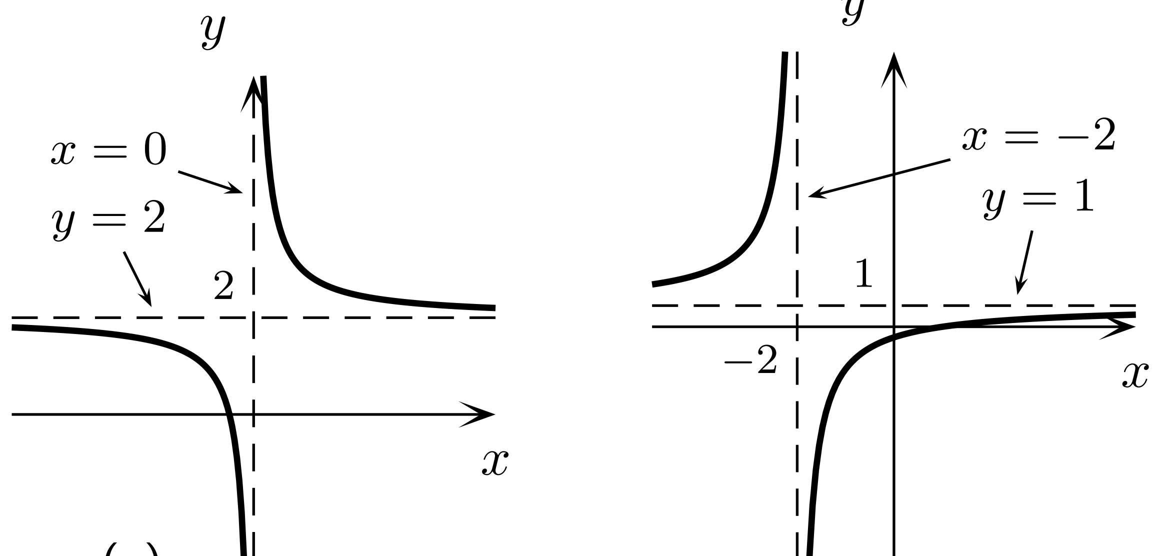
(g) Degree 4:
$y = (x-a)(x-b)(x-c)(x-d)$



(h) Degree 4:
$y = x^4 - x^3$

Rational functions

A *rational function* $f(x)$ is the quotient of two polynomials: $f(x) = \frac{P(x)}{Q(x)}$.
$f(x)$ is not defined when $Q(x) = 0$. The graph of a rational function is formed by continuous curves broken at the zeros of the denominator.

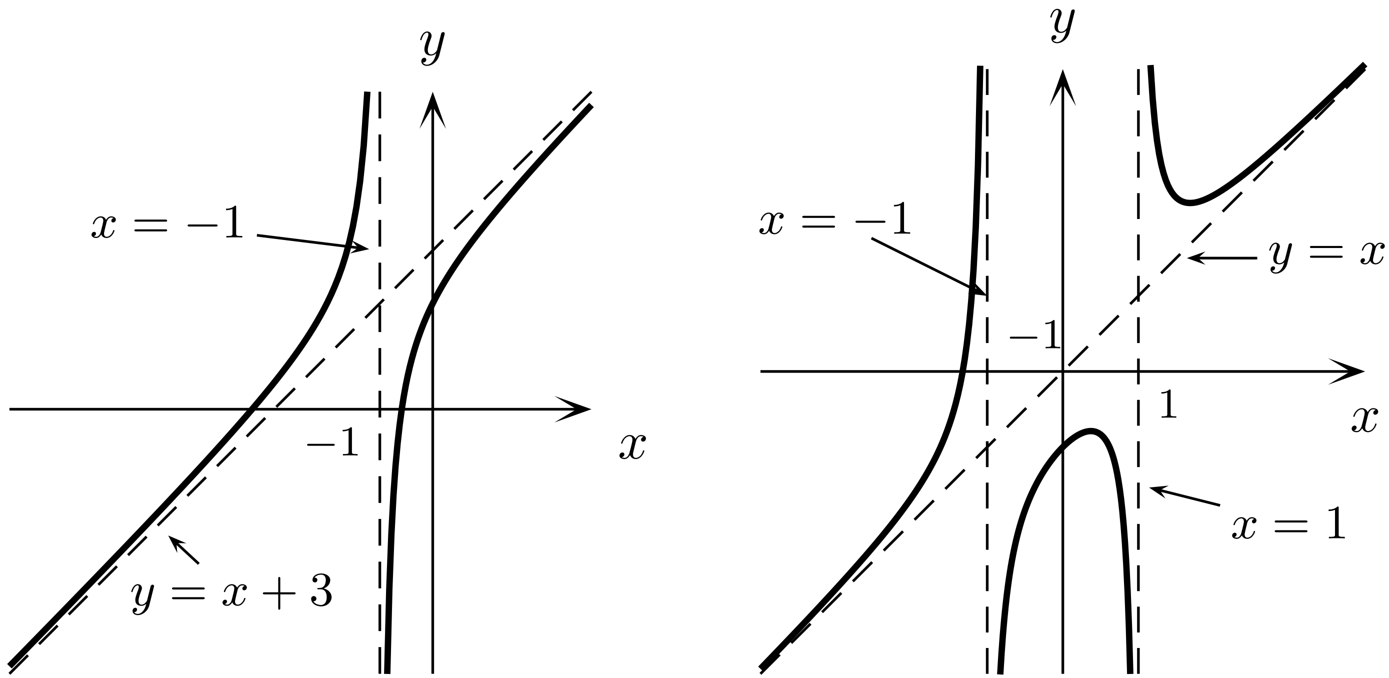


(a) $y = \dfrac{1+2x}{x} = \dfrac{1}{x} + 2$ (b) $y = \dfrac{x-1}{x+2} = 1 - \dfrac{3}{x+2}$

A rational function is *proper* if the degree of the numerator is less than that of the denominator. Otherwise it is *improper*.

. By direct division, we can write a given improper rational function in the form:

$$\begin{pmatrix} \text{An improper} \\ \text{rational function} \end{pmatrix} = \text{a polynomial} + \begin{pmatrix} \text{a proper} \\ \text{rational function} \end{pmatrix} \tag{3.1}$$



(c) $y = x + 3 - \dfrac{1}{x+1}$ (d) $y = x + \dfrac{1}{(x-1)(x+1)}$

Figure 3.6: Examples of rational functions.

See CoCalc
https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page042-CoCalcJupyter.pdf

Example 3.8 Express the improper rational function $\dfrac{x^4 + 4x^3 + 3x^2 - 3}{x^2 + 3x + 2}$ in the form of an improper rational function.

Solution. By long division, we get

$$
\begin{array}{rrrrrrr}
 & x^2 & +x & -2 & & & \\
\cline{2-7}
x^2 + 3x + 2 \,\big) & x^4 & +4x^3 & +3x^2 & +0 & -3 & \\
 & x^4 & +3x^3 & +2x^2 & & & \\
\cline{2-5}
 & & x^3 & +x^2 & +0 & & \\
 & & x^3 & +3x^2 & +2x & & \\
\cline{3-6}
 & & & -2x^2 & -2x & -3 & \\
 & & & -2x^2 & -6x & -4 & \\
\cline{4-6}
 & & & & 4x & +1 &
\end{array}
$$

and hence $$\frac{x^4 + 4x^3 + 3x^2 - 3}{x^2 + 3x + 2} = x^2 + x - 2 + \frac{4x + 1}{x^2 + 3x + 2}. \quad \square$$

Asymptotes of rational functions

We see that the graph of a rational function consists of two or more continuous branches. Each of these branches approaches to a straight line (drawn as a dashed line) as the point on the branch moves towards infinity in a certain direction. Such a straight line is called an *asymptote* of the graph. The equations of the asymptotes of a given rational function can be found using the following theorem.

Theorem 3.4 *Let $P(x)$ and $Q(x)$ be nonzero polynomials having no common factor. Let $f(x) = P(x)/Q(x)$ be a rational function and suppose that $(x - c_1)$, $(x - c_2)$, etc. are factors of $Q(x)$ where c_1, c_2, etc. are distinct real constants.*

- *Then the vertical lines $x = c_1$, $x = c_2$, etc. are asymptotes of the graph of $f(x)$.*
- *Furthermore, if $\deg P \leq \deg Q + 1$ so that $f(x)$ can be resolved in the following special form of (3.1):*

$$f(x) = ax + b + \frac{S(x)}{Q(x)}, \quad \deg(S) < \deg(Q).$$

then the line $y = ax + b$ is also an asymptote of the graph.

Remark 3.1 In the first part of the theorem, the asymptotes are vertical. In the second part of the theorem, if $a \neq 0$, the asymptote is oblique, while if $a = 0$ the asymptote is horizontal. For a proper rational function $(\deg P < \deg Q)$, we have $a = b = 0$ and therefore the x-axis $(y = 0)$ is an asymptote of the graph.

Example 3.9 Find the asymptotes of the rational function $f(x) = \dfrac{x^3 + 2x^2 + 1}{(x-1)(x+2)}$.

Solution. By long division, $f(x) = x + 1 + \dfrac{x+3}{(x-1)(x+2)}$. Therefore the asymptotes are the lines

$$x = 1, \quad x = -2 \quad \text{and} \quad y = x + 1.$$

□

Partial fractions

A proper rational function, with real coefficients, can sometimes be expressed as a sum of two or more proper rational functions, with real coefficients, called *partial fractions*.
For example,

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

In Chapter 5 of the text book, we have to resolve a rational function into partial fractions this way to do integration, an important topic in calculus.
Each factor of the denominator of a given rational function, is associated with a partial fraction or a sum of partial fractions. The rule of association is shown in the table next page for a linear factor and an irreducible quadratic factor (that cannot be factorized into a product of real linear factors).

Rule	Factor of denominator	Form of the partial fractions
1	$ax+b$	$\dfrac{A_1}{ax+b}$
2	$(ax+b)^2$	$\dfrac{A_1}{ax+b}+\dfrac{A_2}{(ax+b)^2}$
3	$(ax+b)^3$	$\dfrac{A_1}{ax+b}+\dfrac{A_2}{(ax+b)^2}+\dfrac{A_3}{(ax+b)^3}$
4	ax^2+bx+c	$\dfrac{A_1x+B_1}{ax^2+bx+c}$
5	$(ax^2+bx+c)^2$	$\dfrac{A_1x+B_1}{ax^2+bx+c}+\dfrac{A_2x+B_2}{(ax^2+bx+c)^2}$

Note: $a, b, c, A_1, A_2, A_3, B_1, B_2$ are real constants, $a \neq 0$.

.
Example 3.10 Resolve $f(x) = \dfrac{x+3}{(x-1)(x-3)}$ into partial fractions.

Solution. First we observe that the given $f(x)$ is a proper rational function. Next we consider each factor of the denominator of $f(x)$. There are two linear factors $x-1$ and $x-3$. By Rule 1 of the table next page, we can assume partial fractions of the forms $\frac{A}{x-1}$ and $\frac{B}{x-3}$ (where A, B are real constants) and get the identity

$$\frac{x+3}{(x-1)(x-3)} \equiv \frac{A}{x-1} + \frac{B}{x-3} \tag{3.2}$$

To find the constants A and B, we remove the denominators and get
$x+3 \equiv A(x-3) + B(x-1)$.
Comparing the coefficient of x and the constant term, we get two equations
$1 = A + B, \quad 3 = -3A - B$. Solving these equations we get $A = -2$, $B = 3$.
Therefore

$$\frac{x+3}{(x-1)(x-3)} \equiv \frac{-2}{x-1} + \frac{3}{x-3} \qquad \square$$

The above method for finding the coefficients A and B is called the *method of undetermined coefficients*.

Example 3.11 Resolve $f(x) = \dfrac{7x+5}{(x+1)^2(x-1)}$ into partial fractions.

Solution. First we observe that the given $f(x)$ is a proper rational function. Next we consider each factor of the denominator of $f(x)$. There are two linear factors $x+1$ (with power 2) and $x-1$. By Rule 1 and Rule 2 of the above table, we can assume partial fractions of the forms $\frac{A}{x+1} + \frac{B}{(x+1)^2}$ and $\frac{C}{x-1}$ (where A, B, C are real constants) and get the identity $\frac{7x+5}{(x+1)^2(x-1)} \equiv \frac{A}{x+1} + \frac{B}{(x+1)^2} + \frac{C}{x-1}$.
Therefore we have $7x+5 \equiv A(x+1)(x-1) + B(x-1) + C(x+1)^2$.
Comparing the coefficient of

$$\begin{aligned} x^2: &\quad 0 = A + C \\ x: &\quad 7 = B + 2C \\ \text{constant term:} &\quad 5 = -A - B + C. \end{aligned}$$

Solving these equations we get $A = -3$, $B = 1$, $C = 3$. Therefore

$$\frac{7x+5}{(x+1)^2(x-1)} \equiv \frac{-3}{x+1} + \frac{1}{(x+1)^2} + \frac{3}{x-1}$$

□

.
Example 3.12 Resolve $f(x) = \dfrac{x-3}{(2x-1)(x^2+1)}$ into partial fractions.

Solution. The denominator has two factors: one is $2x-1$ and the other is x^2+1. By Rule 1 and Rule 4, $f(x)$ has partial fractions in the forms $\frac{A}{2x-1}$ and $\frac{Bx+C}{x^2+1}$ where A, B, C are real constants. Therefore we have the identity

$$\frac{x-3}{(2x-1)(x^2+1)} \equiv \frac{A}{2x-1} + \frac{Bx+C}{x^2+1}$$

and hence $x - 3 \equiv A(x^2+1) + (Bx+C)(2x-1)$. Comparing coefficients of

$$\begin{aligned} x^2: &\qquad 0 = A + 2B \\ x: &\qquad 1 = -B + 2C \\ \text{constant term:} &\qquad -3 = A - C. \end{aligned}$$

Solving the equations, we get $A = -2$, $B = 1$, $C = 1$. Therefore

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

□

Trigonometric functions

Consider the xy-plane in rectangular coordinates such that the scales on both axes are the same. Let P be an arbitrary point on the unit circle (with centre at the origin O and unit radius). If the straight line OP makes an angle θ (in radian) with the positive x-axis and if P has coordinates (x, y), we define the *sine*, *cosine* and *tangent* functions by

$$\cos\theta = x, \quad \sin\theta = y \quad \text{and} \quad \tan\theta = \frac{\sin\theta}{\cos\theta}$$

so that

$$\cos^2\theta + \sin^2\theta = 1.$$

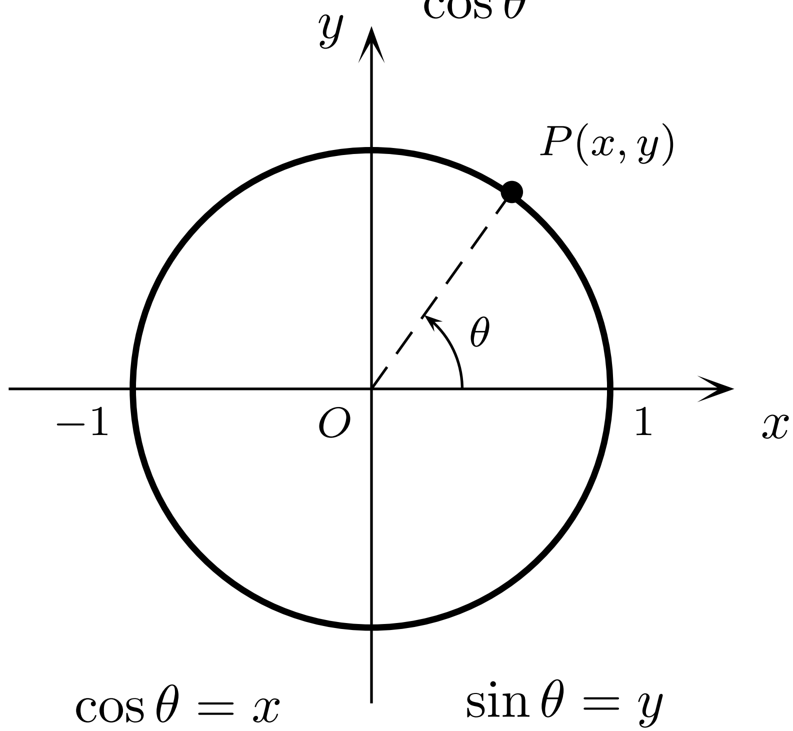


Figure 3.7: Definitions of $\cos\theta$ and $\sin\theta$.

. Both $\cos$ and $\sin$ are continuous periodic functions with a period of 2π. $\cos$ is even while $\sin$ is odd. $\tan x$ is discontinuous at $x = \pm\pi/2, \pm 3\pi/2, \ldots$ where $\cos x = 0$.

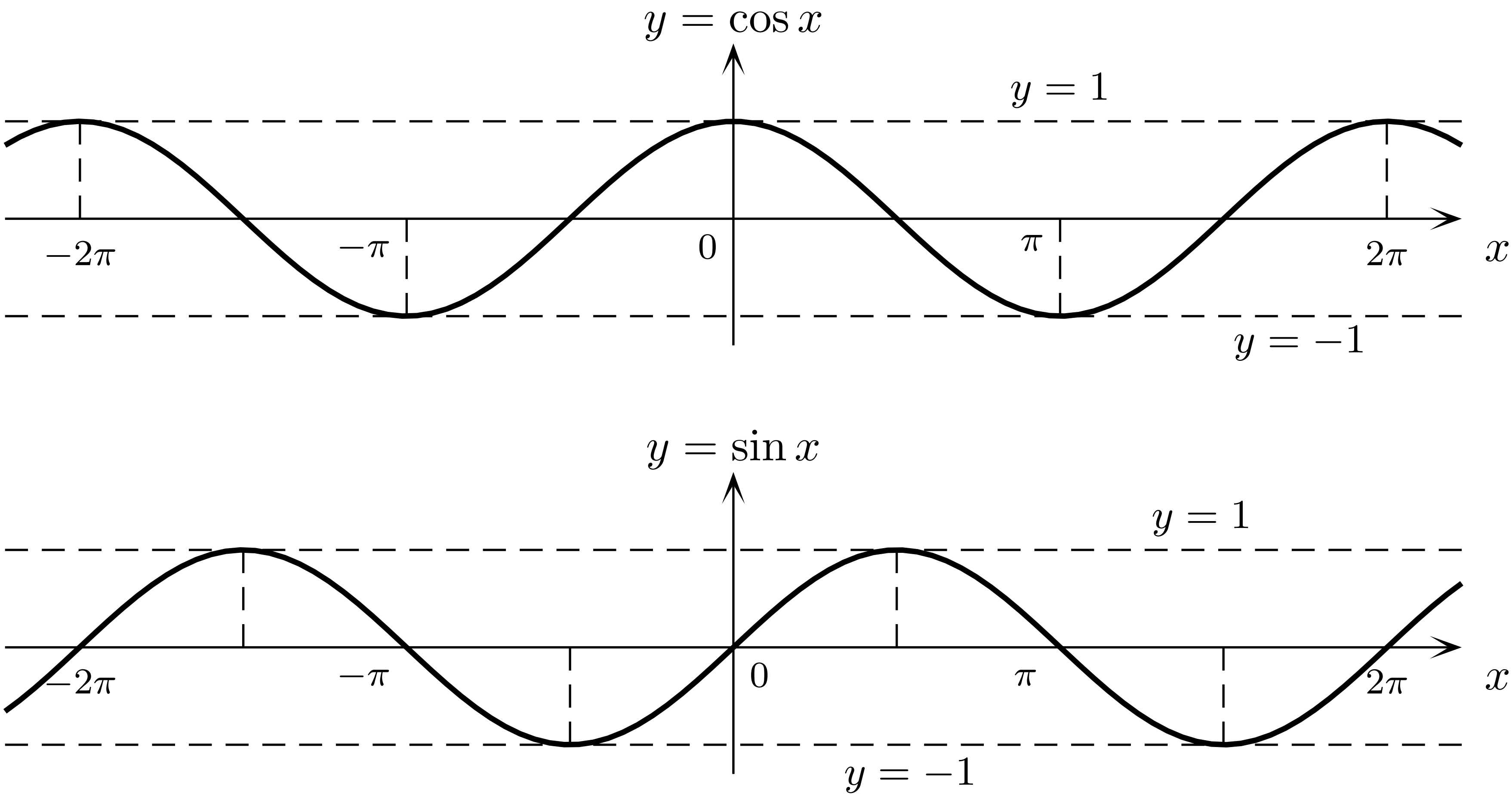


Figure 3.8: The graphs of $\cos x$ and $\sin x$.

.

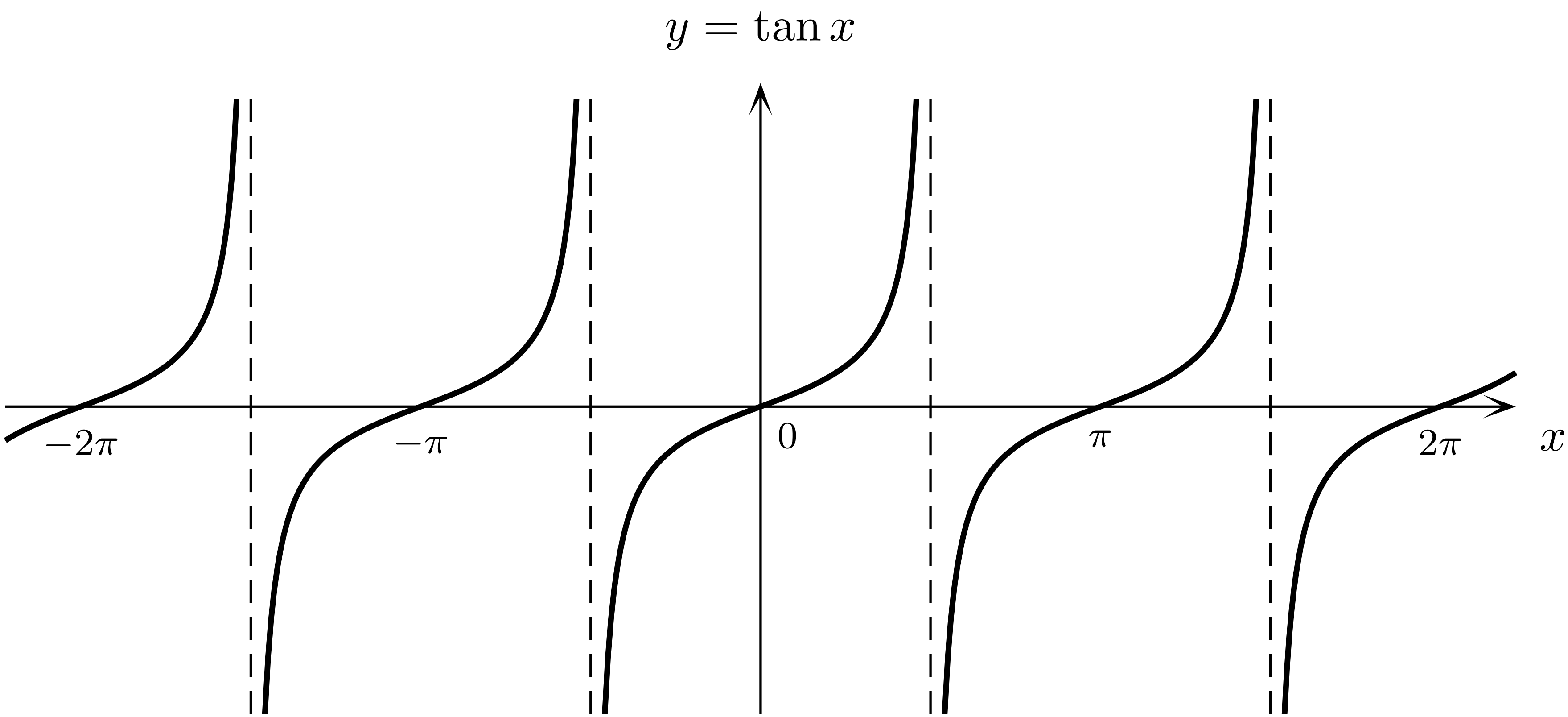


Figure 3.9: The graph of $\tan x$

.

Table of trigonometric formulas

Compound angle formulas

$$\sin(A+B)=\sin A\cos B+\cos A\sin B,\quad \sin(A-B)=\sin A\cos B-\cos A\sin B$$
$$\cos(A+B)=\cos A\cos B-\sin A\sin B,\quad \cos(A-B)=\cos A\cos B+\sin A\sin B$$

$$\tan(A+B)=\frac{\tan A+\tan B}{1-\tan A\tan B},\qquad \tan(A-B)=\frac{\tan A-\tan B}{1+\tan A\tan B}$$

Double angle formulas

$$\sin 2A = 2\sin A\cos A$$
$$\cos 2A = \cos^2 A-\sin^2 A = 1-2\sin^2 A = 2\cos^2 A-1 \qquad \tan 2A=\frac{2\tan A}{1-\tan^2 A}$$

$$\cos^2 A=\frac{1+\cos 2A}{2},\qquad \sin^2 A=\frac{1-\cos 2A}{2}$$

Conversion formulas

$$\sin(x+y)+\sin(x-y)=2\sin x\cos y, \quad \sin(x+y)-\sin(x-y)=2\cos x\sin y$$
$$\cos(x+y)+\cos(x-y)=2\cos x\cos y, \quad \cos(x+y)-\cos(x-y)=-2\sin x\sin y.$$

$$\sin A+\sin B=2\sin\left(\frac{A+B}{2}\right)\cos\left(\frac{A-B}{2}\right)$$

$$\sin A-\sin B=2\cos\left(\frac{A+B}{2}\right)\sin\left(\frac{A-B}{2}\right)$$

$$\cos A+\cos B=2\cos\left(\frac{A+B}{2}\right)\cos\left(\frac{A-B}{2}\right)$$

$$\cos A-\cos B=-2\sin\left(\frac{A+B}{2}\right)\sin\left(\frac{A-B}{2}\right)$$

Inverse trigonometric functions

• **Arcsine.** The function $\sin x$ is many-to-one. However, if we restrict the domain to the interval $[-\pi/2, \pi/2]$, the function becomes one-to-one and its range is $[-1,\ 1]$. With this special domain restriction, the inverse function of $\sin$ exists. It is called the *arcsine* function and is denoted by $\sin^{-1}$ or $\arcsin$. Thus,

$$\boxed{x = \sin^{-1} y \quad \text{iff} \quad y = \sin x \quad \text{and} \quad x \in [-\pi/2, \pi/2].}$$

Note that $\sin(\sin^{-1} y) = y$ for all $-1 \leq y \leq 1$ but $\sin^{-1}(\sin x) = x$ iff $\dfrac{-\pi}{2} \leq x \leq \dfrac{\pi}{2}$.

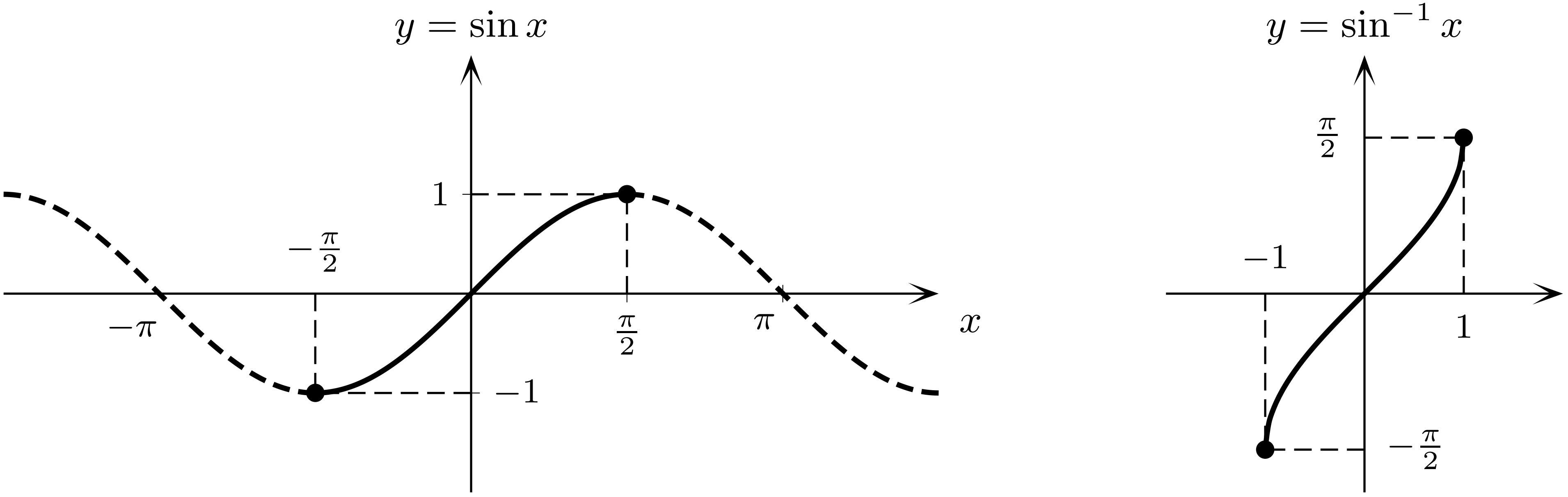


Figure 3.10: Definition of $\sin^{-1} x$.

• **Arccosine.** The function $\cos x$ is many-to-one. However, if we restrict the domain to the interval $[0, \pi]$, the function becomes one-to-one and its range is $[-1,\ 1]$. With this special domain restriction, the inverse function of $\cos$ exists. It is called the *arccosine* function and is denoted by $\cos^{-1}$ or $\arccos$. Thus,

$$\boxed{x = \cos^{-1} y \quad \text{iff} \quad y = \cos x \quad \text{and} \quad x \in [0, \pi].}$$

Note that $\cos(\cos^{-1} y) = y$ for all $-1 \le y \le 1$ but $\cos^{-1}(\cos x) = x$ iff $0 \le x \le \pi$.

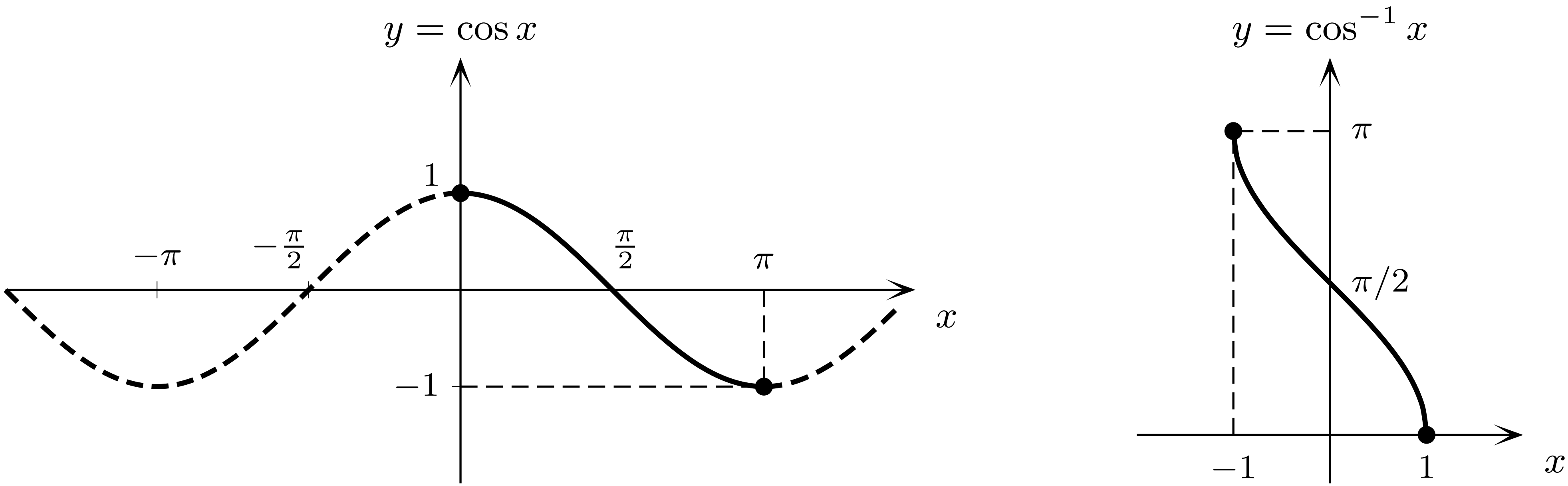


Figure 3.11: Definition of $\cos^{-1} x$.

• **Arctangent.** The function $\tan x$ is many-to-one. However, if we restrict the domain to the interval $(-\pi/2, \pi/2)$, the function becomes one-to-one and its range is $\mathbb{R}$. With this special domain restriction, the inverse function of $\tan$ exists. It is called the *arctangent* function and is denoted by $\tan^{-1}$ or $\arctan$. Thus,

$$\boxed{x = \tan^{-1} y \quad \text{iff} \quad y = \tan x \quad \text{and} \quad x \in (-\pi/2, \pi/2).}$$

Note that $\tan(\tan^{-1} y) = y$ for all $-\infty < y < \infty$ but $\tan^{-1}(\tan x) = x$ iff $\frac{-\pi}{2} < x < \frac{\pi}{2}$.

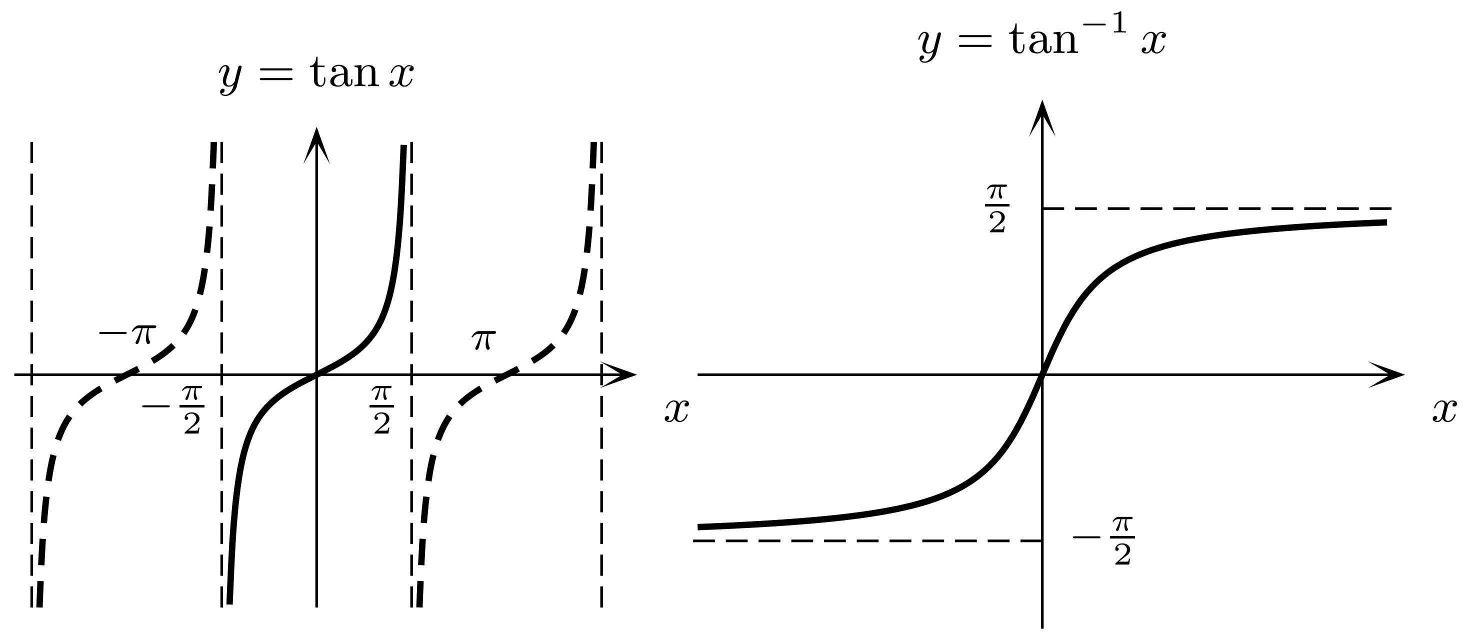


Figure 3.12: Definition of $\tan^{-1} x$.

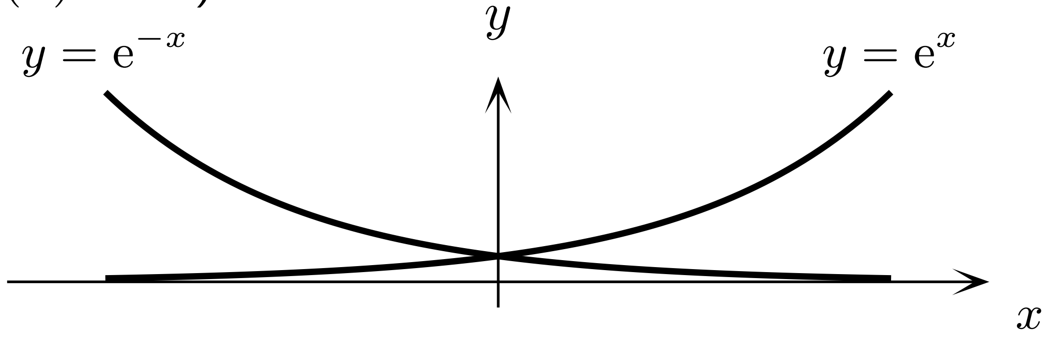
Exponential functions

The function $y = a^x$ is called an *exponential function*. The number a is the *base* and x the *exponent* (or *index*, or *power*). In order that $y = a^x$ takes on real values for all real x, we must assume $a > 0$.

Law of indices

$a^m a^n = a^{m+n}$	$a^m / a^n = a^{m-n}$	$(a^m)^n = a^{mn}$
$a^0 = 1$	$a^{-1} = 1/a$	$a^{-m} = 1/a^m$

If the base is the number e $= 2.718281828459\cdots$, the exponential function is denoted by $\exp$ ($\exp(x) \equiv \mathrm{e}^x$).



The graph of $y = \mathrm{e}^x$ shows the exponential growth while that of $y = \mathrm{e}^{-x}$ shows the exponential decay.

Logarithmic functions

Let a be a positive constant and consider the exponential function $y = a^x$. This function is one-to-one and its range is $(0, \infty)$. The inverse function of this exponential function is called a *logarithmic* function and is defined by

$$x = \log_a y \ \ (0 < y < \infty) \quad \text{iff} \quad y = a^x \ \ (-\infty < x < \infty)$$

The number $\log_a y$, where $y > 0$, is called the *logarithm* of y to the *base* a.

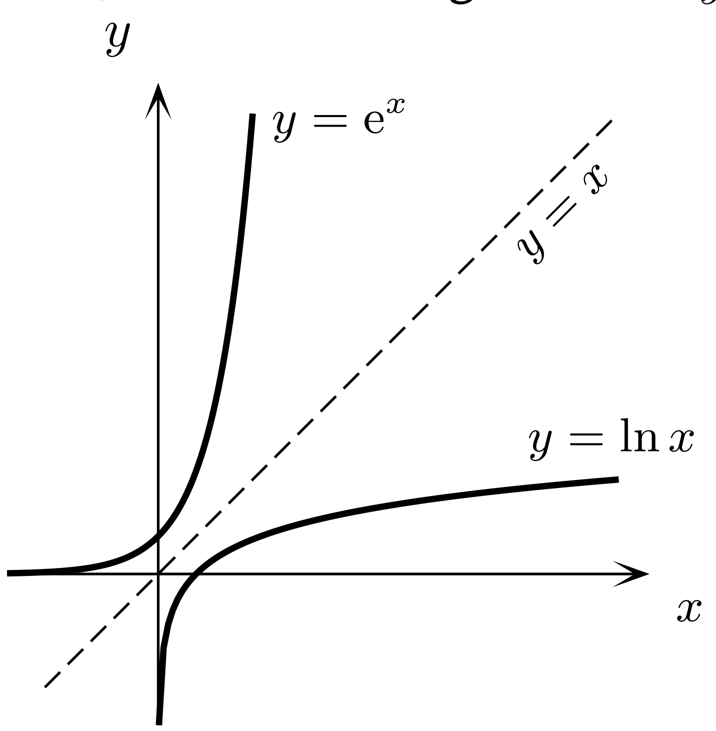


Figure 3.13: Graphs of $\exp x$ and $\ln x$.

.

Rules of logarithm Let $a,\ b,\ x,\ y$ be positive real numbers.	
$\log_a(xy) = \log_a x + \log_a y$	$\log_a 1 = 0$
$\log_a(x/y) = \log_a x - \log_a y$	$\log_a x^m = m \log_a x$
$\log_a x = \dfrac{\log_b x}{\log_b a}$	where m is real

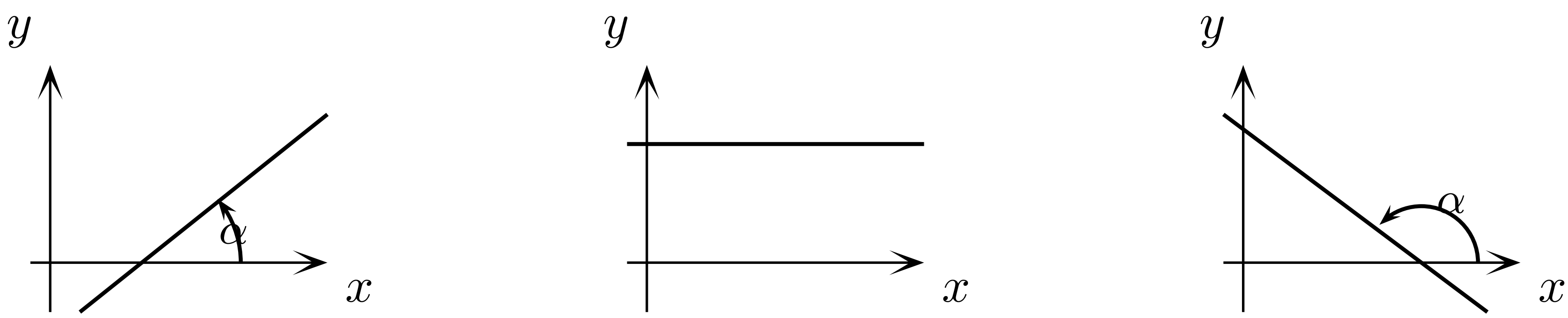
If the base is e $= 2.718281828\cdots$, the logarithm function is denoted by $\ln$ or $\log$. Thus

$$\boxed{y = \ln x \ \ (0 < x < \infty) \quad \text{iff} \quad x = \mathrm{e}^y \ \ (-\infty < y < \infty).}$$

Slope of a straight line

Consider any straight line on the xy-plane. If the line is *not* parallel to the x-axis, the *angle of inclination* is the angle between the line and the positive x-axis. If the line is parallel to the x-axis, the angle of inclination is 0.
If the angle of inclination of a straight line is α and $\alpha \neq \pi/2$, then the *slope* of the straight line is the real number

$$\boxed{\text{slope } = \tan\alpha.}$$



Inclined line	Horizontal line.	Inclined line
$0 < \alpha < \pi/2$	Angle of incl. $= 0$	$\pi/2 < \alpha < \pi$
Slope ($= \tan\alpha$) is positive	Slope $= \tan 0 = 0$	Slope ($= \tan\alpha$) is negative

The slope gives a measure of the steepness of the straight line. Also the sign of the slope tells us in which direction the straight line is running. Using the following theorem, we can find the slope of a non-vertical line based on the coordinates of two distinct points on the line.

Theorem 3.5 *Let (x_0, y_0) and (x_1, y_1) be two points on a straight line with $x_0 \neq x_1$, then the slope m of the straight line is given by the formula $m = (y_1 - y_0)/(x_1 - x_0)$.*

Example 3.13 Find the slope the line containing the points $(3, -2)$ and $(-4, 1)$.

Solution. The slope is $(1 - (-2))/(-4 - 3) = -3/7$. □

Limits and Continuity

One-sided limits

Consider a function $f(x)$ which is defined in $\mathbb{R}$ so that

$$f(x) = \begin{cases} x^2 & \text{if} \quad x < 2, \\ g(x) & \text{otherwise} \end{cases} \tag{4.1}$$

where $g(x)$ is a function which is not important (in fact not considered) in the following discussion.

x	0.0	1.0	1.90	1.99	1.999	1.9999
y	0.0	1.0	3.61	3.96	3.996	3.9996

The table shows that the values of $f(x)$ approaches the number 4 as x increases and approaches 2. This number 4 is called the limit of $f(x)$ as x approaches 2 from the left.

Intuitive definition

Definition 4.1 Let a be a point on the real axis such that $f(x)$ is defined when x is on the left of a ($x < a$) and near to a and let L be a real number. If the values of $f(x)$ approaches L as x increases and approaches a, we call L *the limit of* $f(x)$ *as* x *approaches* a *from the left* and write

$$\lim_{x \to a^-} f(x) = L.$$

Remark 4.1 The symbol $x \to a^-$ represents "x approaches a from the left" which means "x is getting closer and closer to a though it is always on the left of a". The number L, usually dependent on $f(x)$ and a, is also called *the left-hand limit of* $f(x)$ *at* a. The value $f(a)$, whether it is defined or not, plays no part in the definition of the left-hand limit at a.

Graphical demonstration

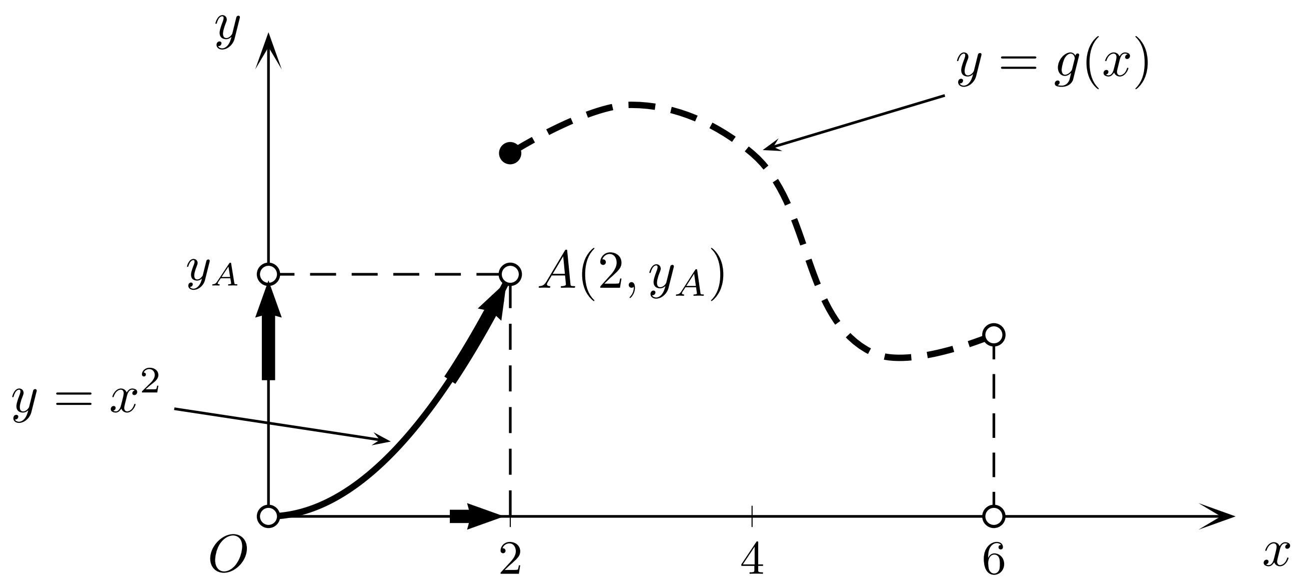


Figure 4.1: The graph of $f(x)$ in $0 < x < 6$.

On an interval to the left of 2, say $(0, 2)$, the graph of the function $f(x)$ is a portion of the parabola $y = x^2$ as shown in Fig. 4.1.
On the graph, we see that as x moves towards 2 from the left the points on the parabola moves towards the point A and hence y moves towards the y-coordinate y_A of A. Since the curve OA here is part of the parabola $y = x^2$, we see that $y_A = 2^2 = 4$ and hence $\lim_{x \to 2^-} f(x) = 4$.
Note that the value of $f(2)$ and the function $g(x)$ appearing in (4.1) do not play any part in the definition of the left-hand limit at $x = 2$.

Right-hand limit

Analogous to the left-hand limit, we define the right-hand limit as follows:

Definition 4.2 Let a be a point on the real axis such that $f(x)$ is defined when x is on the right of a $(x > a)$ and near to a and let R be a real number. If the values of $f(x)$ approaches R as x decreases and approaches a, we call R *the limit of* $f(x)$ *as* x *approaches* a *from the right* and write

$$\lim_{x \to a^+} f(x) = R.$$

Remark 4.2 The symbol $x \to a^+$ represents "x approaches a from the right." The number R, usually dependent on $f(x)$ and a, is also called *the right-hand limit of* $f(x)$ *at* a. The value $f(a)$, whether it is defined or not, plays no part in the definition of the right-hand limit at a.

Examples

Example 4.1 Let $F(x)$ be the piecewise-defined function defined by the graph shown in Fig. 4.2. Find the left-hand limits and the right-hand limits of $F(x)$ at the points $x = 1$, 2, 3, 4, 5 and 6.

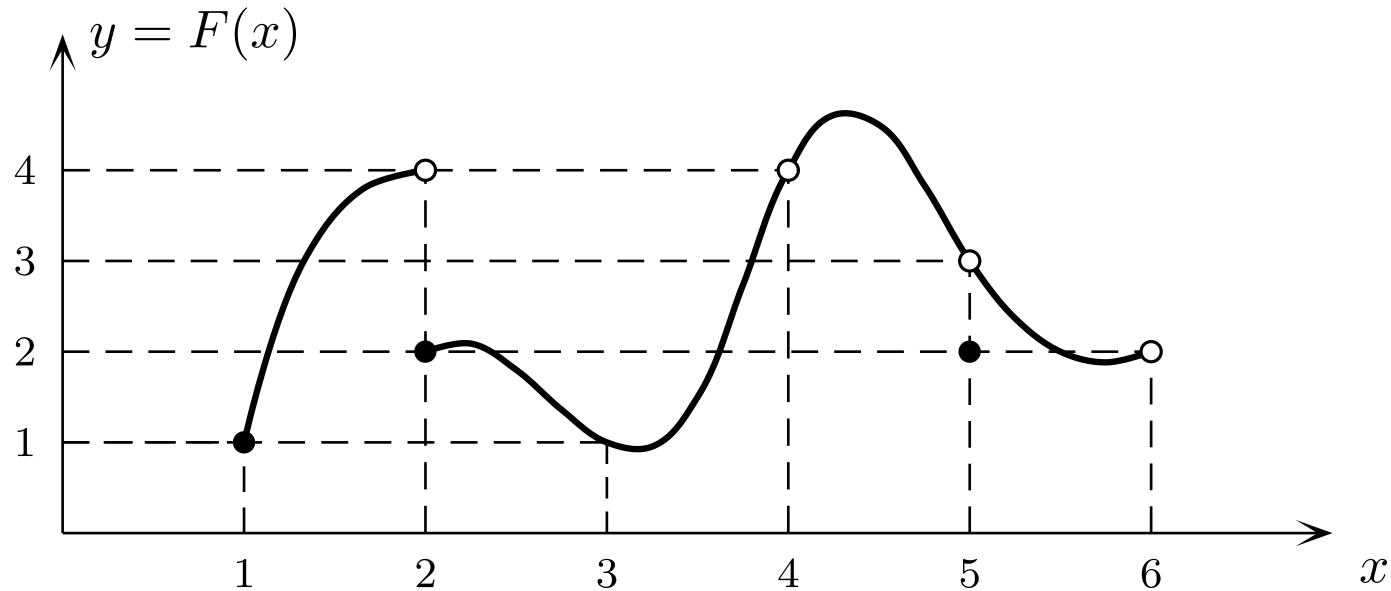


Figure 4.2: The graph of $F(x)$ defined on $[1, 4) \cup (4, 6)$.

Solution. From the graph we see that

$$\lim_{x\to 1^-} F(x) \text{ is not defined,} \qquad \lim_{x\to 1^+} F(x) = 1,$$

$$\lim_{x\to 2^-} F(x) = 4, \qquad \lim_{x\to 2^+} F(x) = 2,$$

$$\lim_{x\to 3^-} F(x) = 1, \qquad \lim_{x\to 3^+} F(x) = 1,$$

$$\lim_{x\to 4^-} F(x) = 4, \qquad \lim_{x\to 4^+} F(x) = 4,$$

$$\lim_{x\to 5^-} F(x) = 3, \qquad \lim_{x\to 5^+} F(x) = 3,$$

$$\lim_{x\to 6^-} F(x) = 2, \qquad \lim_{x\to 6^+} F(x) \text{ is not defined.}$$

□

Example 4.2 Let $f(x)$ be the function on $(0,4)$ defined by

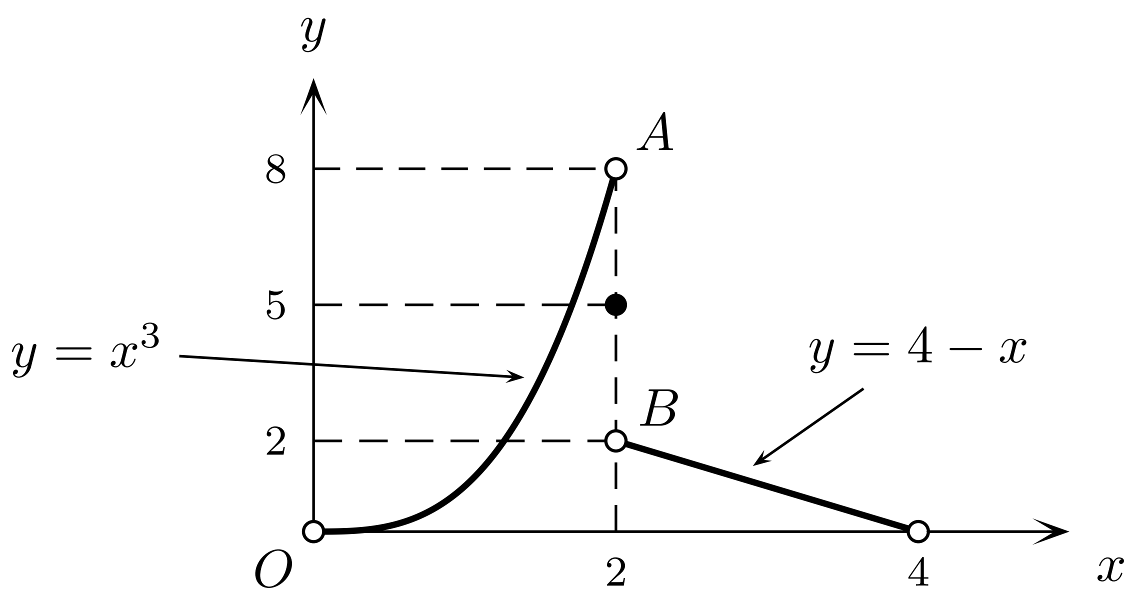
$$f(x) = \begin{cases} x^3 & \text{if } 0 < x < 2, \\ 5 & \text{if } x = 2, \\ 4 - x & \text{if } 2 < x < 4. \end{cases}$$

Sketch the graph of this function and find the limits $\lim_{x\to 2^-} f(x)$ and $\lim_{x\to 2^+} f(x)$.

See CoCalc

https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page070-CoCalcJupyter.pdf

Solution.



From the graph we see that

$$\lim_{x \to 2^-} f(x) = \text{the } y\text{-coordinate of } A = \left[x^3\right]_{x=2} = 8.$$

$$\lim_{x \to 2^+} f(x) = \text{the } y\text{-coordinate of } B = \left[4 - x\right]_{x=2} = 2.$$

□

Note that, without referring to the graph, we can simply work out the limits as follows:

$$\lim_{x \to 2^-} f(x) = \left[x^3\right]_{x=2} = 8, \quad \text{and} \quad \lim_{x \to 2^+} f(x) = \left[4 - x\right]_{x=2} = 2.$$

Limits of functions

Let

$$L = \lim_{x \to a^-} F(x) \quad \text{and} \quad R = \lim_{x \to a^+} F(x).$$

Using the results of Example 4.1 (page 69)

- At $a = 1$: $L \neq R$ since L is not defined.
- At $a = 2$: $L \neq R$.
- At $a = 3$: $L = R = f(a)$.
- At $a = 4$: $L = R \neq f(a)$ since $f(a)$ is not defined.
- At $a = 5$: $L = R \neq f(a)$.
- At $a = 6$: $L \neq R$ since R is not defined.

These results shows that for the equality and inequality of L and R at a general point a, there are three possibilities:

Three possibilities at a

1.	$L \neq R$.	a is an endpoint of the domain of $f(x)$, or there is a vertical gap (or jump) at a.
2.	$L = R \neq f(a)$.	There is no vertical gap but there is a hole on the graph. $f(a)$ may or may not be defined in this case.
3.	$L = R = f(a)$.	There is neither a vertical gap nor a hole on the graph.

When Case **2** or Case **3** in the table occurs we call the common values of L and R simply the limit of the function.

Definition of Limit of a function

Definition 4.3 (Limit of a function) If the one-sided limits $\lim_{x \to a^-} f(x)$ and $\lim_{x \to a^+} f(x)$ exist and both are equal to L, we say that *the limit of* $f(x)$ *as* x *approaches* a *is* L and write

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

If either of the one-sided limits does not exist or, if they exist but are not equal, we say that *the limit of* $f(x)$ *as* x *approaches* a *does not exist.*

Examples

Example 4.3 Consider again the piecewise-defined function $F(x)$ defined by the graph in Fig. 4.2. Find the limits of $F(x)$ (if exist) at the points $x = 1$, 2, 3, 4, 5 and 6.

Solution.

$\lim_{x \to 1} F(x)$ does not exist.

$\lim_{x \to 2} F(x)$ does not exist.

$\lim_{x \to 3} F(x) = 1.$

$\lim_{x \to 4} F(x) = 4.$

$\lim_{x \to 5} F(x) = 3.$

$\lim_{x \to 6} F(x)$ does not exist. □

Example 4.4 Fig. 4.3 shows the graph of $f(x) = (\sin x)/x$. Based on the graph, we see that the limit $\lim_{x \to 0} f(x)$ exists. Find this limit by evaluating the values of $f(x)$ near $x = 0$.

See CoCalc
https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page076-CoCalcJupyter.pdf

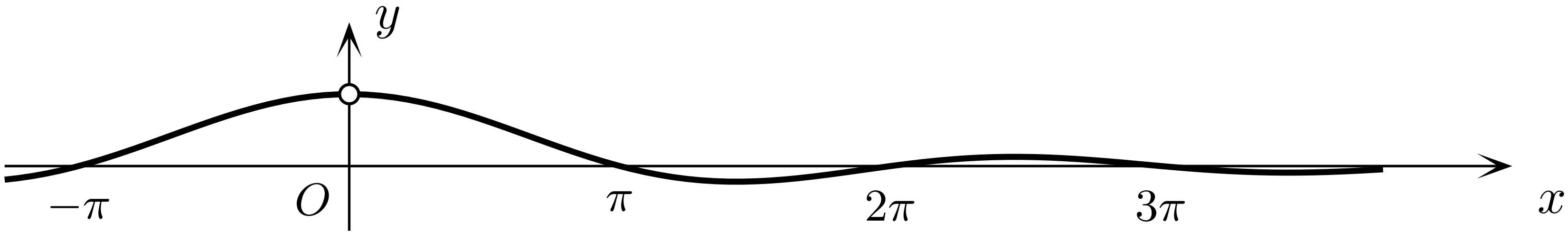


Figure 4.3: The graph of $(\sin x)/x$.

Solution. The values of $f(x) = (\sin x)/x$ are computed by a desk calculator at some (numerically) small values of x approaching 0. table:

x	± 0.100	± 0.080	± 0.060	± 0.040	± 0.020	± 0.010	± 0.008
y	0.9983	0.9989	0.9994	0.9997	0.9999	1.0000	1.0000

From the table, we see that $\lim_{x \to 0} \frac{\sin x}{x} = 1$. □

Example 4.5 Fig. 4.4 shows the graph of $f(x) = x\sin(1/x)$ near $x = 0$. Based on the graph, find $\lim_{x \to 0}[x\sin(1/x)]$ if it exists.

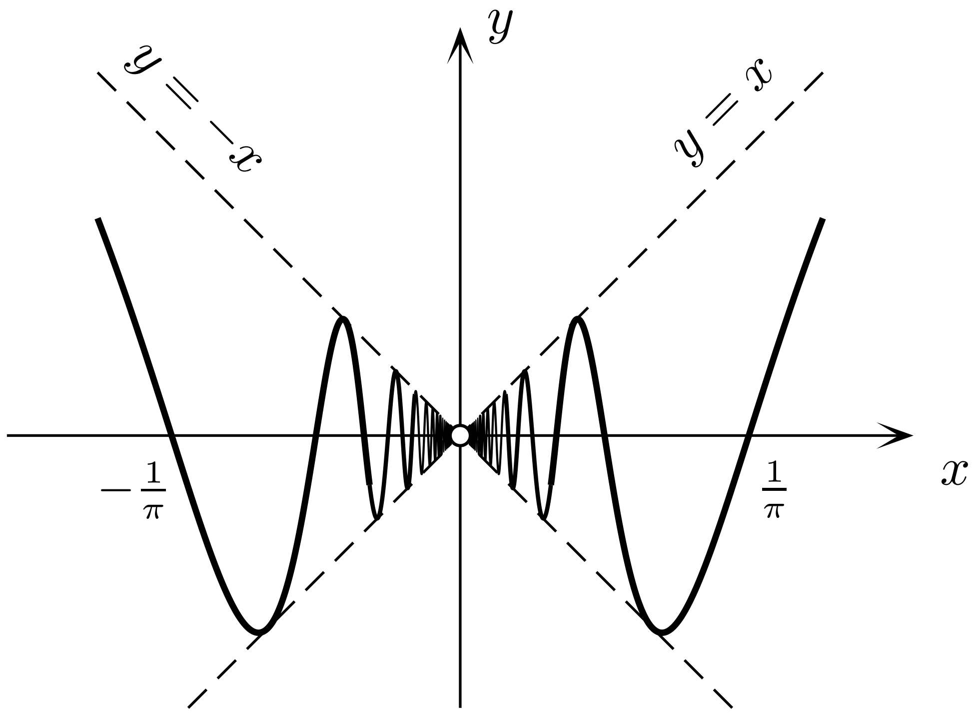


Figure 4.4: The graph of $x\sin(1/x)$ near $x = 0$.

Solution. From the graph, we see that $\lim_{x \to 0} x\sin(1/x) = 0$. □

Example 4.6 Fig. 4.5 shows the graph of $f(x) = \sin(1/x)$ near $x = 0$. Based on the graph, find $\lim_{x\to 0} \sin(1/x)$ if it exists.

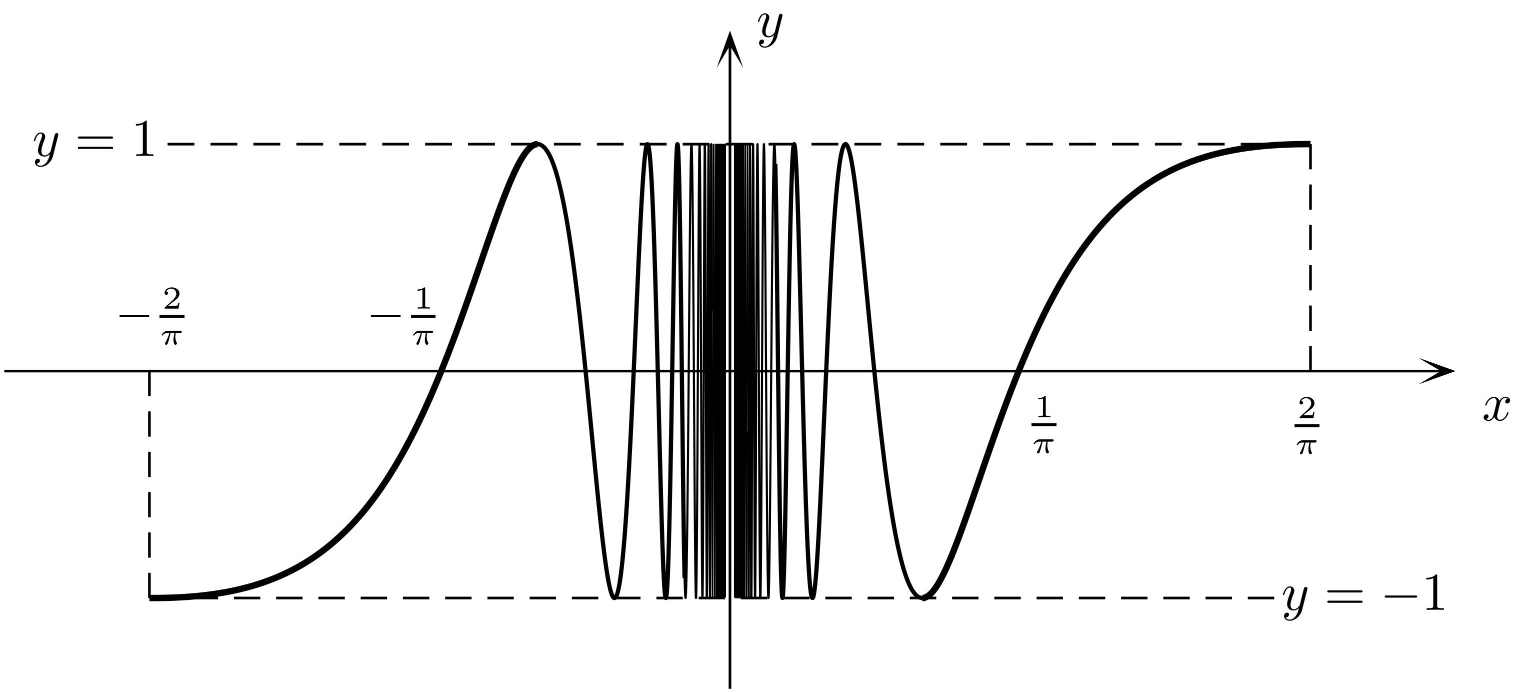


Figure 4.5: The graph of $\sin(1/x)$ near $x = 0$.

Solution. From the graph, we see that $\lim_{x\to 0} \sin(1/x)$ does not exist. □

Limit theorems

The following theorems are important as they help us find limits of functions derived by algebraic operations on elementary functions.

Theorem 4.1 *Let n be a positive integer and k a constant. Assume that the limits $\lim_{x\to a} f(x)$ and $\lim_{x\to a} g(x)$ exist. Then*

1. $\lim_{x\to a} k = k$

2. $\lim_{x\to a} x = a$

3. $\lim_{x\to a} kf(x) = k \lim_{x\to a} f(x)$

4. $\lim_{x\to a} [f(x) + g(x)] = \lim_{x\to a} f(x) + \lim_{x\to a} g(x)$

5. $\lim_{x\to a} [f(x)g(x)] = \lim_{x\to a} f(x) \lim_{x\to a} g(x)$

6. $\lim_{x\to a} [f(x)/g(x)] = \lim_{x\to a} f(x) \Big/ \lim_{x\to a} g(x)$ *if* $\lim_{x\to a} g(x) \neq 0$

7. $\lim_{x\to a} [f(x)]^n = \left[\lim_{x\to a} f(x)\right]^n$

8. $\lim_{x\to a} \sqrt[n]{f(x)} = \sqrt[n]{\lim_{x\to a} f(x)}$ *(assume $f(x) \geq 0$ near $x = a$ if n is even.)*

Theorem 4.2 (Composite function) *If* $\lim_{x \to a} f(x) = A$ *and* $\lim_{u \to A} g(u) = B$ *then* $\lim_{x \to a} g(f(x)) = B$.

Theorem 4.3 (Squeeze Theorem) *Let* $f(x)$, $g(x)$, $h(x)$ *be functions such that* $f(x) \le g(x) \le h(x)$ *for all* x *near* a, *except possibly at* a *itself. If*

$$\lim_{x \to a} f(x) = \lim_{x \to a} h(x) = L$$

then $\lim_{x \to a} g(x) = L$.

Definition 4.4 Let S be a subset of the domain of a function $f(x)$. We say that $f(x)$ is *bounded on* S if there is a constant K such that

$$|f(x)| \leq K \quad \text{for all } x \in S.$$

In particular, if S is an open interval (p, q) where $p < a < q$, we say that $f(x)$ is *bounded near* a.

The composite functions of the form $\sin F(x)$ and $\cos F(x)$ are bounded functions as $|\sin F(x)| \leq 1$ and $|\cos F(x)| \leq 1$. These functions are bounded on the domain of $F(x)$.

The following theorem follows directly from Theorem 4.3.

Theorem 4.4 *If $g(x)$ is bounded near a, except possibly at a itself, and if* $\lim_{x\to a} f(x) = 0$ *then* $\lim_{x\to a} f(x)g(x) = 0$.

Remark 4.3 Obviously, the above theorems are true if we replace the limits by the left-hand limits (or by the right-hand limits).

Continuity of functions

We know from the table on page 73 that at a general point a there are three possibilities. If the third case in the table is true, i.e. if on the graph there is no gap or hole at a, we say that the function is continuous at a.

Definition 4.5 (Continuity at a point) If $\lim_{x \to a} f(x) = f(a)$, we say that $f(x)$ *is continuous at* a.

Definition 4.6 (Discontinuity at a point) We say that $f(x)$ is discontinuous (i.e. not continuous) at a if any one of the following holds:

1. $f(a)$ is not defined

2. $\lim_{x \to a} f(x) \neq f(a)$

3. $\lim_{x \to a} f(x)$ does not exist

If the function is defined only on one side of a, then the associated one-sided limit is used instead of two-sided limits in the above definitions.

Properties of continuity

Theorem 4.5 (Properties of continuity) *Let n be a positive integer and k a constant. If $f(x)$ and $g(x)$ are continuous at a then the following functions are also continuous at a:*

1. (Scalar multiple) $kf(x)$ *where k is a constant.*
2. (Sum and difference) $f(x)+g(x)$ *and* $f(x)-g(x)$.
3. (Product) $f(x)g(x)$.
4. (Quotient) $f(x)/g(x)$ *if* $g(a)\neq 0$.
5. (Power) $\big[f(x)\big]^n$.
6. (Root) $\sqrt[n]{f(x)}$ *(assume $f(a)\geq 0$ if n is even.)*

Theorem 4.6 *If $f(x)$ is continuous at $x=a$ and $g(u)$ is continuous at $u=f(a)$ then $g(f(x))$ is continuous at $x=a$.*

Continuity on an interval

Definition 4.7 (Continuity on an interval) Let J be an interval. If $f(x)$ is continuous at every point of J, we say that $f(x)$ *is continuous on* J, otherwise $f(x)$ is *not continuous on* J.

Continuity on an interval means that the graph of the function is a continuous or one-piece curve, i.e. the graph can be drawn without lifting the pencil from the paper.

Example 4.7 Consider again the function $F(x)$ defined by Fig. 4.2. Find all the x-values at which the function is discontinuous. Find the intervals on which the function is continuous.

Solution. The function is not continuous at $x = 2$ (where there is a gap), $x = 4$ (where there is a hole), $x = 5$ (where there is a hole) and at all points outside the interval $[1, 6)$ (where $F(x)$ is not defined). The function is continuous at every point in the intervals $[1, 2)$, $[2, 4)$, $(4, 5)$ and $(5, 6)$. □

Continuous elementary functions

Elementary functions continuous on the interval J

Function $f(x)$	Interval J
Polynomials, $\exp kx$ ($k =$ const.)	$\mathbb{R}$
$\sin kx$, $\cos kx$ ($k =$ const.)	$\mathbb{R}$
$\tan kx$ ($k =$ const., $k \neq 0$)	$\ldots, \left(\frac{-3\pi}{2k}, \frac{-\pi}{2k}\right), \left(\frac{-\pi}{2k}, \frac{\pi}{2k}\right), \left(\frac{\pi}{2k}, \frac{3\pi}{2k}\right), \ldots$
$\ln kx$ ($k =$ const., $k > 0$)	$(0, \infty)$
Rational functions $P(x)/Q(x)$	$(-\infty, x_0), (x_0, x_1), (x_1, x_2), \ldots, (x_n, \infty)$ where $x_0, x_1, \ldots, x_n$ are distinct zeros of $Q(x)$.

The above table can be further extended by Theorem 4.5 to include many other functions generated from elementary functions by algebraic operations.

Examples of functions continuous on the interval J

$f(x)$	J	Remarks
$x^3 - 2x^2 + x - 2$	$\mathbb{R}$	Polynomial
$5 \sin 3x$	$\mathbb{R}$	Scalar multiple a of trigonometric function
$\exp 2x$	$\mathbb{R}$	Exponential function
$\sin 2x + \exp 3x$	$\mathbb{R}$	Sum of continuous functions
$e^x \cos 3x$	$\mathbb{R}$	Product of continuous functions
$\cos(x^2 + x + 1)$	$\mathbb{R}$	Composition of continuous functions
$\sin \ln x$	$(0, \infty)$	Composition of continuous functions
$\sqrt{\ln 2x}$	$[1/2, \infty)$	Square-root of a continuous function
$\dfrac{\cos x}{(x-1)(x-2)}$	$(-\infty, 1)$, $(1, 2)$, $(2, \infty)$	Quotient of continuous functions

Methods of finding limits: Substitution

In the previous section, we see that for function $f(x)$ continuous at a, we have the formula:

$$\lim_{x \to a} f(x) = f(a), \tag{4.2}$$

i.e. the limit can be found *by substitution*. If $f(x)$ is generated from elementary functions by algebraic operations and if $f(a)$ is defined then $f(x)$ is continuous at a and hence the substitution formula (4.2) works.

Example 4.8 By substitution, we get

1. $\lim_{x\to 2}(x^3-4x^2+x+5)=2^3-4\times 2^2+2+5=-1.$
2. $\lim_{x\to 2}\dfrac{x+1}{x^2-1}=\dfrac{2+1}{2^2-1}=1.$
3. $\lim_{x\to 0}\dfrac{\mathrm{e}^x}{1+\mathrm{e}^{2x}}=\dfrac{\mathrm{e}^0}{1+\mathrm{e}^0}=\dfrac{1}{1+1}=\dfrac{1}{2}.$
4. $\lim_{x\to 2}\sin^2 3x=\sin^2(3\times 2)=\sin^2 6.$
5. $\lim_{x\to 0}[(x^2-2)\cos 3x]=(0-2)\cos 0=-2.$
6. $\lim_{x\to 1}\dfrac{\sin x}{1+x^2}=\dfrac{\sin 1}{2}.$
7. $\lim_{t\to 2}\tan(3t^2-2)=\tan(3\times 2^2-2)=\tan 10.$
8. $\lim_{y\to 1}\cos\exp(y^2+1)=\cos\exp(1^1+1)=\cos\mathrm{e}^2.$
9. $\lim_{u\to\pi/2}\sqrt{u-\sin u}=\sqrt{\pi/2-\sin(\pi/2)}=\sqrt{\pi/2-1}.$

Example 4.9 The method of substitution fails in the following cases:

1. $\lim_{x\to 1} \dfrac{1}{x^2-1} = \dfrac{1}{0}$ which is undefined.

2. $\lim_{x\to \pi/2} \tan x = \tan(\pi/2)$ which is undefined.

3. $\lim_{x\to 0} \dfrac{\cos x}{1-\cos x} = \dfrac{1}{0}$ which is undefined.

4. $\lim_{x\to 1} \ln(x^2-3) = \ln(-2)$ which is undefined.

5. $\lim_{x\to 0} \dfrac{\sin x}{x} = \dfrac{0}{0}$ which is undefined.

6. $\lim_{x\to 0} x\sin\dfrac{1}{x} = 0\cdot\sin(1/0)$ which is undefined.

Cancellation of factors

Example 4.10 Find $\lim_{x \to 2} \dfrac{x^2 - 4}{x - 2}$.

Solution. On substitution, we get $0/0$. The difficulty can be removed by observing that we do not have to consider $x = 2$ for the limit. Since

$$\frac{x^2 - 4}{x - 2} = x + 2 \quad \text{for } x \neq 2,$$

we have

$$\lim_{x \to 2} \frac{x^2 - 4}{x - 2} = \lim_{x \to 2} (x + 2) = 2 + 2 = 4.$$

□

Example 4.11 Let n be a positive integer. Find $\lim_{x \to a} \dfrac{x^n - a^n}{x - a}$.

Solution. Using the factorization

$$x^n - a^n = (x-a)(x^{n-1} + x^{n-2}a + \cdots + xa^{n-2} + a^{n-1})$$

we get on substitution

$$\begin{aligned} \lim_{x \to a} \frac{x^n - a^n}{x - a} &= \lim_{x \to a}(x^{n-1} + x^{n-2}a + \cdots + xa^{n-2} + a^{n-1}) \\ &= a^{n-1} + a^{n-1} + \cdots + a^{n-1} + a^{n-1} \quad (n \text{ terms}) \\ &= na^{n-1}. \end{aligned}$$

□

A trigonometric formula

Theorem 4.7 *If θ is in radians, then*

$$\boxed{\lim_{\theta \to 0} \frac{\sin\theta}{\theta} = 1.} \tag{4.3}$$

Remark 4.4 The above limit formula holds for θ in radians, not in degrees. Unless otherwise stated, the unit for angles in this book is radian.

Example 4.12 Find $\lim_{x\to 1} \frac{\sin(x^3+2x-3)}{x^3+2x-3}$.

Solution. Put $u = x^3+2x-3$. Then $\lim_{x\to 1} u = 1+2-3 = 0$. Applying Theorem 4.2 and Theorem 4.7 above, we get

$$\lim_{x\to 1} \frac{\sin(x^3+2x-3)}{x^3+2x-3} = \lim_{u\to 0} \frac{\sin u}{u} = 1. \qquad \square$$

Example 4.13 Find $\lim_{x\to 0}\dfrac{1-\cos x}{x}$.

Solution. Using the formula $1-\cos 2\theta = 2\sin^2\theta$ and Theorem 4.7, we get

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

Example 4.14 Find $\lim_{x \to 0} \sqrt{\frac{7x + \sin 2x}{x}}$.

Solution. By Theorem 4.1 and Theorem 4.7, we get

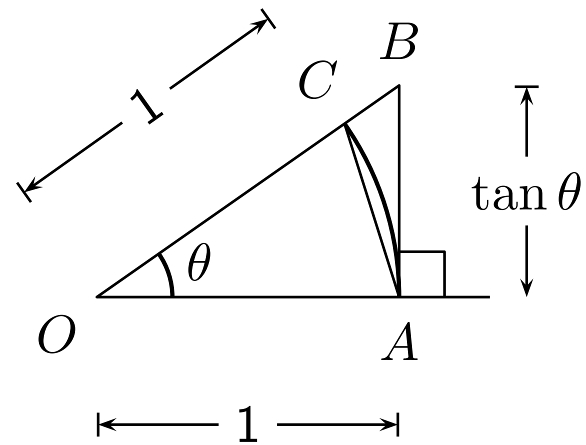
$$\lim_{x \to 0} \sqrt{\frac{7x + \sin 2x}{x}} = \sqrt{\lim_{x \to 0} \frac{7x + \sin 2x}{x}}$$

$$= \sqrt{\lim_{x \to 0} \left(7 + \frac{\sin 2x}{x}\right)} = \sqrt{7 + 2 \times 1} = 3.$$

□

Proof

First consider the case when $\theta > 0$. Since θ is approaching 0, we can assume that $0 < \theta < \pi/2$ and construct a right-angled triangle OAB with $OA = 1$ and $\angle BOA = \theta$ as shown in the diagram. The arc $\overset{\frown}{AC}$ is part of the circle centred at O with unit radius.



Comparing areas, we have $\triangle OAC <$ sector $OAC < \triangle OAB$, i.e. $\frac{1}{2}\sin\theta < \frac{1}{2}\theta < \frac{1}{2}\tan\theta$. Therefore, $1 < \frac{\theta}{\sin\theta} < \frac{1}{\cos\theta}$ and hence, for $0 < \theta < \pi/2$,

$$1 > \frac{\sin\theta}{\theta} > \cos\theta. \tag{4.4}$$

Since $(\sin\theta)/\theta$ and $\cos\theta$ are even functions, the inequalities (4.4) are true also for $-\pi/2 < \theta < 0$.

As $\lim_{\theta \to 0} \cos\theta = 1$, we get by Theorem 4.3

$$\lim_{\theta \to 0^+} \frac{\sin\theta}{\theta} = 1. \tag{4.5}$$

For the case $\theta < 0$, we write $\phi = -\theta > 0$ and get

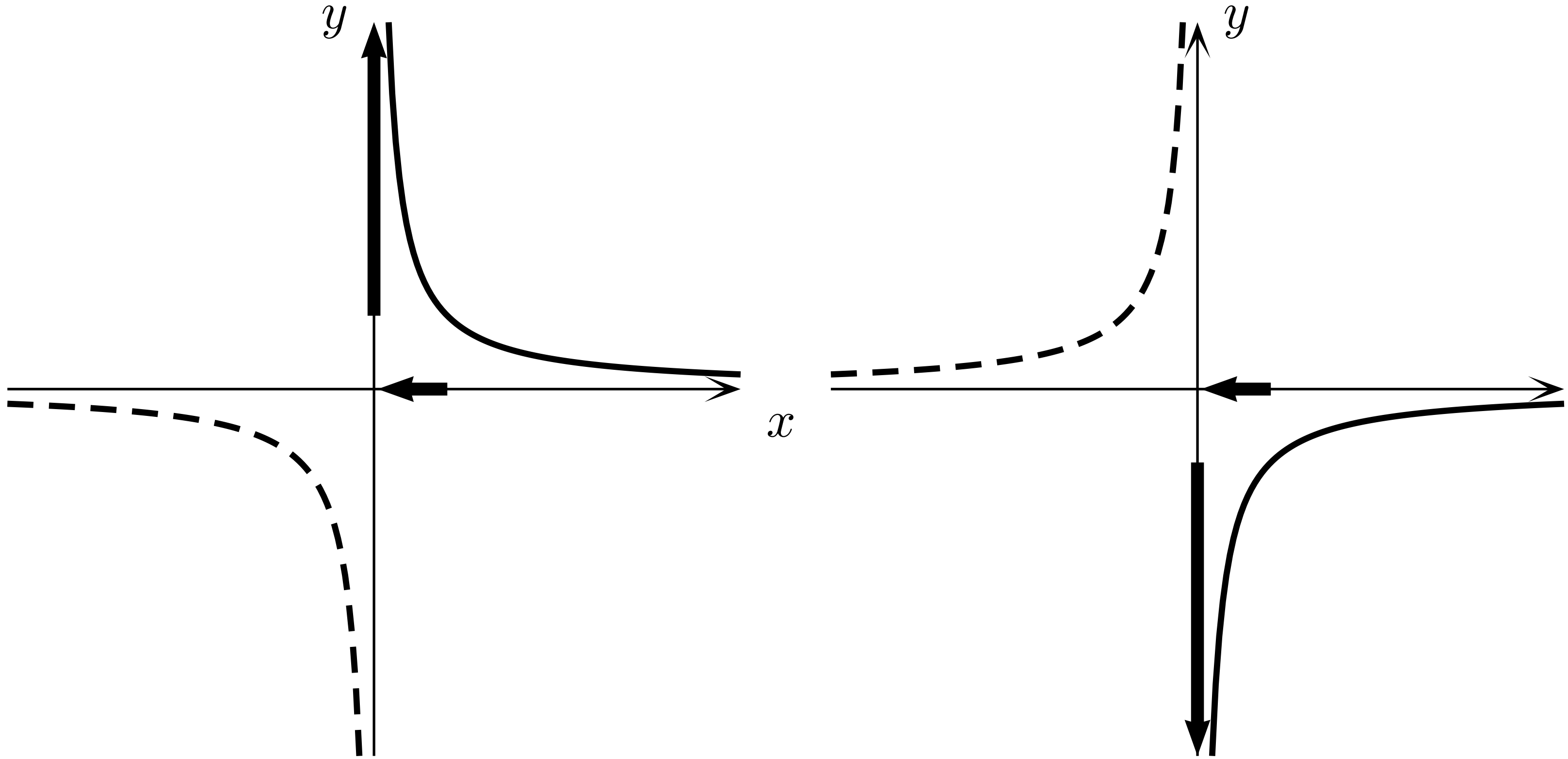
$$\lim_{\theta \to 0^-} \frac{\sin\theta}{\theta} = \lim_{\phi \to 0^+} \frac{\sin\phi}{\phi} = 1. \tag{4.6}$$

Combining (4.5) and (4.6), we get the required result (4.3).

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

Infinite limits

Consider the function $f(x) = 1/x$ near $x = 0$ (see Fig. 4.6(a)). If $x > 0$ anthe value of $f(x)$ becomes very big and can be made as big as we please (as indicated by the thick black vertical arrow).
To describe this property of $f(x)$ we introduce the following definitions.



(a) The graph of $y = 1/x$. (b) The graph of $y = -1/x$.

Figure 4.6: Functions tending to ∞ and $-\infty$ as $x \to 0^+$.

Definitions

Definition 4.8 If the value of $f(x)$ can be made bigger than any prescribed positive and large number by taking $x > a$ and close enough to a, we say that $f(x)$ *approaches to infinity as* x *approaches* a *from the right* and we write

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

The situation in Fig 4.6(b) motivates the next definition.

Definition 4.9 If the value of $f(x)$ can be made smaller than any prescribed number (usually negative and numerically large) by taking $x > a$ and close enough to a, we say that $f(x)$ *approaches to negative infinity as* x *approaches* a *from the right* and we write

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

Clearly, by considering x approaching from the left, i.e. $x < a$ instead of $x > a$ in the above two definitions, we can define the two *infinite limits from the left*:

$$\lim_{x \to a^-} f(x) = \infty \quad \text{and} \quad \lim_{x \to a^-} f(x) = -\infty.$$

Definition 4.10

- If $\lim_{x\to a^-} f(x) = \lim_{x\to a^+} f(x) = \infty$, we write $\lim_{x\to a} f(x) = \infty$.
- If $\lim_{x\to a^-} f(x) = \lim_{x\to a^+} f(x) = -\infty$, we write $\lim_{x\to a} f(x) = -\infty$.

The symbol ∞ stands for "infinity". It is not a real number and is used as a "quantity" greater than any real number. In the same way, $-\infty$ is a "quantity" smaller than any real number. Infinity or negative infinity is always associated with limits such as those defined above.

Formulas

By actually computing $1/x$ using small values of x or by observing the graph of $1/x$ in Fig 4.6(a), we see that

$$\lim_{x\to 0^+} \frac{1}{x} = \infty \quad \text{and} \quad \lim_{x\to 0^-} \frac{1}{x} = -\infty.$$

On the other hand, direct substitution gives

$$\lim_{x\to 0^+} \frac{1}{x} = \frac{1}{0^+} \quad \text{and} \quad \lim_{x\to 0^-} \frac{1}{x} = \frac{1}{0^-}$$

which are undefined. Therefore, we can write

$$\frac{1}{0^+} = \infty \quad \text{and} \quad \frac{1}{0^-} = -\infty \tag{4.7}$$

where 0^- means a function approaching 0 from the left and 0^+ means from the right. Although ∞ is not a number, the above formulas in (4.7) are sometimes useful for finding limits. These, together with other useful formulas about infinity are listed in the following table.

Useful formulas involving infinity

(k is a constant, m and n are positive integers)

$$\infty + k = \infty, \qquad (-\infty) + k = -\infty. \tag{4.8}$$

$$\infty + \infty = \infty, \qquad (-\infty) - \infty = -\infty, \qquad \infty - \infty = \text{indeterminate}. \tag{4.9}$$

$$k \cdot \infty = \infty, \qquad k \cdot (-\infty) = -\infty \qquad \text{if } k > 0. \tag{4.10}$$

$$k \cdot \infty = -\infty, \qquad k \cdot (-\infty) = \infty \qquad \text{if } k < 0. \tag{4.11}$$

$$\infty \cdot \infty = \infty, \qquad \infty \cdot (-\infty) = -\infty, \qquad (-\infty) \cdot (-\infty) = \infty. \tag{4.12}$$

$$k \div \infty = 0, \qquad k \div (-\infty) = 0, \qquad \infty \div \infty = \text{indeterminate}. \tag{4.13}$$

$$1 \div 0^+ = \infty, \qquad 1 \div 0^- = -\infty, \qquad 0 \cdot \infty = \text{indeterminate}. \tag{4.14}$$

$$\sqrt[m]{\infty} = \infty, \qquad \sqrt[n]{-\infty} = -\infty \qquad \text{if } n \text{ is odd}. \tag{4.15}$$

Remark 4.5 In (4.8) the formula "$\infty + k = \infty$" means that if $\lim f(x) = \infty$ then $\lim(f(x) + k) = \infty$. In (4.9) the formula "$\infty + \infty = \infty$" means that if $\lim f(x) = \infty$ and $\lim g(x) = \infty$ then $\lim[f(x) + g(x)] = \infty$. Also the formula "$\infty - \infty =$ indeterminate" means that we cannot draw any conclusion on $\lim[f(x) - g(x)]$ if we only know that $\lim f(x) = \infty$ and $\lim g(x) = \infty$. Other formulas in the above table are to be interpreted similarly.

Based on the algebraic operations on infinity given in the above table, Theorem 4.1 (page 79) is also true if $\lim f(x)$ and $\lim g(x)$ are infinite.

Examples

Example 4.15 Is it true that $\lim_{x \to 0} \frac{1}{x} = \infty$?

Solution. No, because $\lim_{x \to 0^-} \frac{1}{x} = -\infty$ and $\lim_{x \to 0^+} \frac{1}{x} = \infty$ and they are not the same. □

Example 4.16 Find $\lim_{x\to 0} \frac{1}{x^2}$.

Solution. On substitution, $\lim_{x\to 0^-} \frac{1}{x^2} = \frac{1}{0^+} = \infty$ and $\lim_{x\to 0^+} \frac{1}{x^2} = \frac{1}{0^+} = \infty$.
Therefore $\lim_{x\to 0} \frac{1}{x^2} = \infty$. □

Example 4.17 Find $\lim_{x\to 1^-} \dfrac{x+2}{(x-1)(x+1)}$ and $\lim_{x\to 1^+} \dfrac{x+2}{(x-1)(x+1)}$.

Solution. For x near to 1, the factor $x-1$ is negative if $x<1$, positive if $x>1$. Therefore we have on substitution,

$$\lim_{x\to 1^-} \frac{x+2}{(x-1)(x+1)} = \frac{3}{(0^-)(2)} = -\infty \text{ and } \lim_{x\to 1^+} \frac{x+2}{(x-1)(x+1)} = \frac{3}{(0^+)(2)} = \infty.\square$$

See CoCalc
https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page109-CoCalcJupyter.pdf

Example 4.18 Find $\lim\limits_{x\to 1^-} \dfrac{1}{x^2-1}$ and $\lim\limits_{x\to 1^+} \dfrac{1}{x^2-1}$.

Solution. Similar to the previous example,

$$\lim_{x\to 1^-} \frac{1}{x^2-1} = \frac{1}{0^-} = -\infty \quad \text{and} \quad \lim_{x\to 1^+} \frac{1}{x^2-1} = \frac{1}{0^+} = \infty. \qquad \square$$

Example 4.19 Find $\lim_{x\to\pi/2^-} \tan x$ and $\lim_{x\to\pi/2^+} \tan x$.

Solution. Using $\tan x = (\sin x)/\cos x$ and the fact that

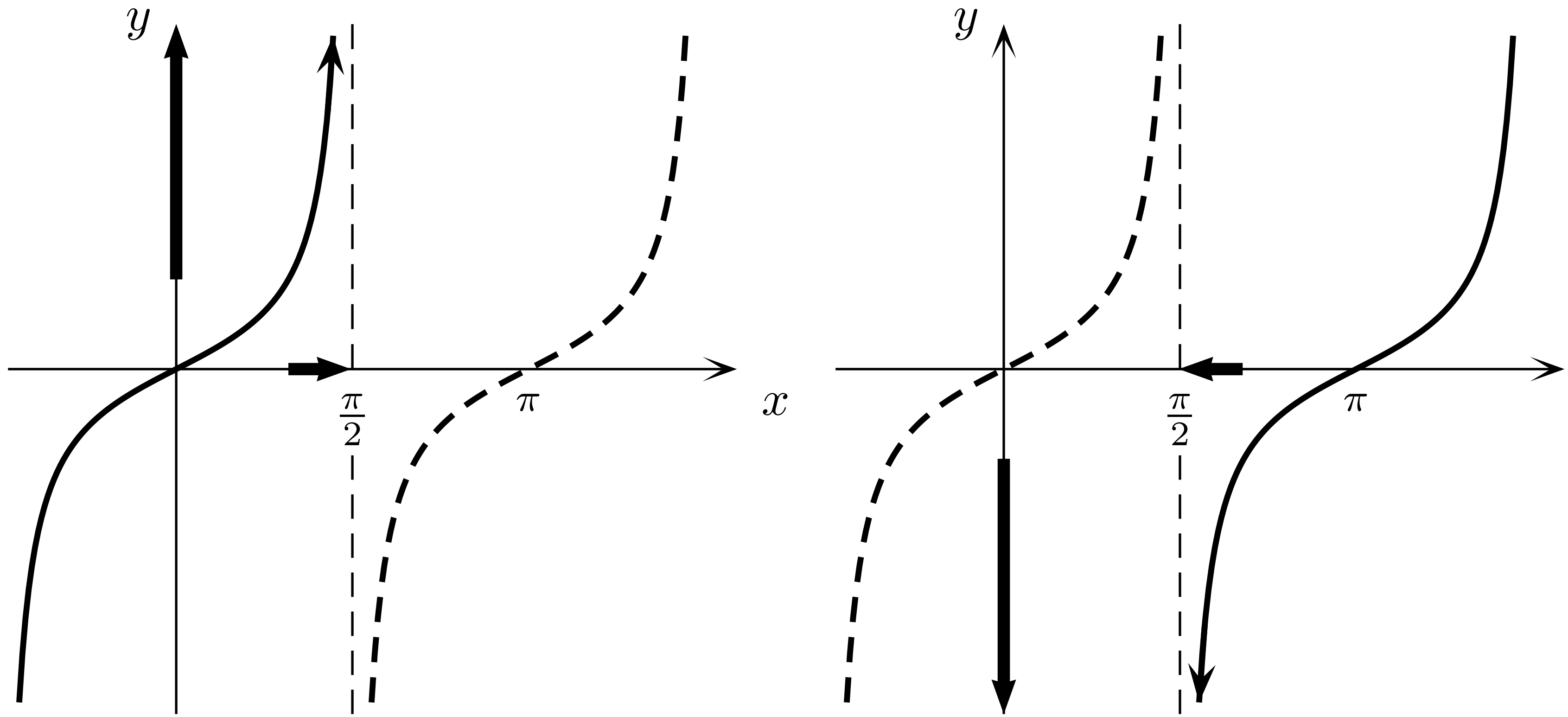
$$\cos x \quad \text{is near to } 0 \text{ and } \begin{cases} \text{positive} & \text{if } x < \pi/2 \text{ and near to } \pi/2, \\ \text{negative} & \text{if } x > \pi/2 \text{ and near to } \pi/2 \end{cases}$$

we get

$$\lim_{x\to\pi/2^-} \tan x = \frac{1}{0^+} = +\infty \quad \text{and} \quad \lim_{x\to\pi/2^+} \tan x = \frac{1}{0^-} = -\infty. \qquad \square$$

The results can also be seen from the graph of $\tan x$ directly (Fig. 4.7).

Remark 4.6 The formula (4.7) is *not* saying that $1/0 = \pm\infty$ is true. In fact there are examples of functions $f(x)$ which approach to 0 but the limits of their reciprocals $1/f(x)$ do not exist and are neither ∞ nor $-\infty$.



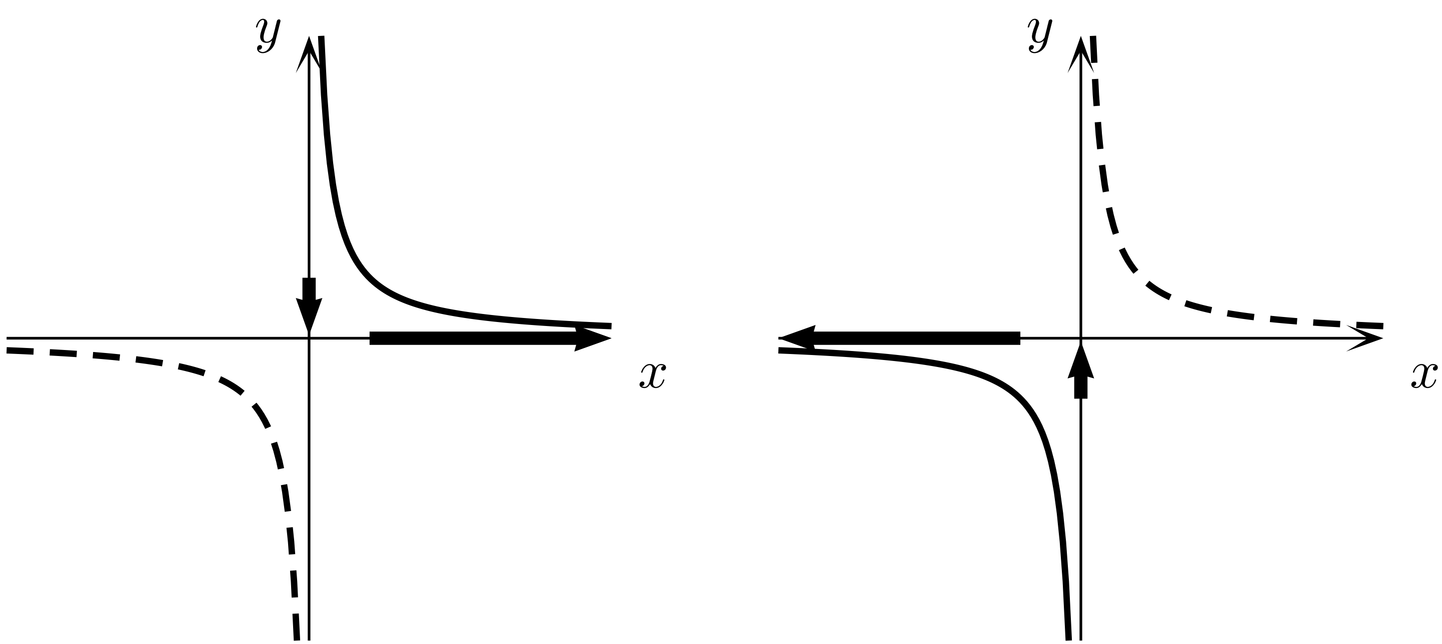
(a) $\tan x \to \infty$ as $x \to \pi/2^-$. (b) $\tan x \to -\infty$ as $x \to \pi/2^+$.

Figure 4.7: Infinite limits of $\tan x$ at $x = \pi/2$.

Limits at infinity

Consider again the function $f(x) = 1/x$. As x increases without bound ($x \to \infty$) or decreases without bound ($x \to -\infty$), the value of $f(x)$ approaches to 0. These facts can be seen from the graph of $1/x$ (see Fig. 4.8) and are stated mathematically as

$$\lim_{x\to\infty} \frac{1}{x} = 0 \quad \text{and} \quad \lim_{x\to-\infty} \frac{1}{x} = 0.$$



(a) $y \to 0$ as $x \to \infty$.   (a) $y \to 0$ as $x \to -\infty$.

Figure 4.8: Function $1/x$ tending to 0 as x tends to ∞ and $-\infty$.

Definitions

Definition 4.11 Let A and B be real numbers.

1. We write $\lim_{x\to\infty} f(x) = A$ if $f(x)$ approaches A as x increases without bound.

2. We write $\lim_{x\to-\infty} f(x) = B$ if $f(x)$ approaches B as x decreases without bound.

Similar definitions are for

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

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

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

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

The above six types of limits are called limits at infinity. By the formulas in the table on page 105 we see that Theorem 4.1 (page 79) is true also for limits at infinity.

Examples

Example 4.20 By inspecting the graph of $y = \tan^{-1} x$ we see that

$$\lim_{x \to -\infty} \tan^{-1} x = -\pi/2 \quad \text{and} \quad \lim_{x \to \infty} \tan^{-1} x = \pi/2.$$

Example 4.21 Find $\lim_{x\to\infty}(3x-2)$ and $\lim_{x\to-\infty}(3x-2)$.

Solution. Using formulas (4.10) and (4.8) on page 105, we get

$$\lim_{x\to\infty}(3x-2) = 3\cdot\infty - 2 = \infty - 2 = \infty.$$

$$\lim_{x\to-\infty}(3x-2) = 3\cdot(-\infty) - 2 = -\infty - 2 = -\infty.$$

□

Example 4.22 Find $\lim_{x\to\infty}(-2x^3+x^2+4x-3)$ and $\lim_{x\to-\infty}(-2x^3+x^2+4x-3)$.

Solution.

$$\begin{aligned}
\lim_{x\to\infty}(-2x^3+3x^2+4x-5) &= \lim_{x\to\infty} x^3\Big(-2+\frac{3}{x}+\frac{4}{x^2}-\frac{5}{x^3}\Big)\\
&= \infty(-2+0+0-0) = -\infty.\\
\lim_{x\to-\infty}(-2x^3+3x^2+4x-5) &= \lim_{x\to-\infty} x^3\Big(-2+\frac{3}{x}+\frac{4}{x^2}-\frac{5}{x^3}\Big)\\
&= -\infty(-2+0+0-0) = \infty.
\end{aligned}$$

□

Polynomials at infinity

From these examples, we obtain the rule:

Theorem 4.8 *If $P(x)$ is a polynomial of degree n with positive leading coefficient:*

$$P(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0, \quad a_n > 0,$$

then

$$\lim_{x \to \infty} P(x) = \infty \quad \textit{and} \quad \lim_{x \to -\infty} P(x) = \begin{cases} \infty & \textit{if } n \textit{ is even,} \\ -\infty & \textit{if } n \textit{ is odd.} \end{cases}$$

Rational functions at infinity

In the following examples, we find limits of rational functions at infinity. For this, we first divide the numerator and denominator by the highest-degree term of the denominator.

Examples

Example 4.23 Find $\lim_{x\to\infty} \dfrac{4x-1}{x^2+1}$ and $\lim_{x\to-\infty} \dfrac{4x-1}{x^2+1}$.

Solution.

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

Similarly we get

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

□

Example 4.24 Find $\lim_{x\to\infty} \frac{2x^2+3x-1}{4x^2+1}$ and $\lim_{x\to-\infty} \frac{2x^2+3x-1}{4x^2+1}$.

Solution.

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

Clearly we get the same answer for the second limit. □

See CoCalc
https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page121-CoCalcJupyter.pdf

Example 4.25 Find $\lim_{x\to\infty} \frac{-3x^3+4x^2-1}{2x^2+x}$ and $\lim_{x\to-\infty} \frac{-3x^3+4x^2-1}{2x^2+x}$.

Solution.

$$\lim_{x\to\infty} \frac{-3x^3+4x^2-1}{2x^2+x} = \lim_{x\to\infty} \frac{-3x+4-\frac{1}{x^2}}{2+\frac{1}{x}} = \frac{-\infty+4-0}{2+0} = -\infty.$$

$$\lim_{x\to-\infty} \frac{-3x^3+4x^2-1}{2x^2+x} = \lim_{x\to-\infty} \frac{-3x+4-\frac{1}{x^2}}{2+\frac{1}{x}} = \frac{-3(-\infty)+4-0}{2+0} = \infty. \quad \square$$

Limits of rational functions

From the above examples, we obtain the following rules for limits of rational functions $P(x)/Q(x)$ at infinity. Here we assume that $P(x)$ and $Q(x)$ are polynomials:

$$P(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0, \quad a_n \neq 0,$$

$$Q(x) = b_m x^m + b_{m-1} x^{m-1} + \cdots + b_1 x + b_0, \quad b_m \neq 0.$$

Theorem 4.9 *Let* $n = \deg P(x)$ *and* $m = \deg Q(x)$ *so that* $a_n \neq 0$ *and* $b_m \neq 0$.

1. *If* $n < m$, *then both limits* $\lim_{x \to \pm\infty} P(x)/Q(x) = 0$.

2. *If* $n = m$, *then both limits* $\lim_{x \to \pm\infty} P(x)/Q(x) = a_n/b_m$.

3. *If* $n > m$, *then the limits* $\lim_{x \to \pm\infty} P(x)/Q(x) = -\infty$ *or* $+\infty$ *depending on the signs of the ratio* a_n/b_m, *and whether* $n - m$ *is even or odd.*

Further examples

Example 4.26 Find $\lim\limits_{x\to\infty} \dfrac{2x+3}{\sqrt{x^2+4}}$.

Solution.

$$\lim_{x\to\infty} \frac{2x+3}{\sqrt{x^2+4}} = \lim_{x\to\infty} \frac{2+\dfrac{3}{x}}{\sqrt{1+\dfrac{4}{x^2}}} = \frac{2+0}{\sqrt{1+0}} = 2. \qquad \square$$

Example 4.27 Find $\lim_{x\to\infty} \frac{x+1}{\sqrt[3]{x^2+1}}$.

Solution.

$$\begin{aligned}\lim_{x\to\infty} \frac{x+1}{\sqrt[3]{x^2+1}} &= \lim_{x\to\infty} \frac{(x+1) \div x^{2/3}}{(\sqrt[3]{x^2+1}) \div x^{2/3}} \\ &= \lim_{x\to\infty} \frac{x^{1/3} + x^{-2/3}}{\sqrt[3]{1+x^{-2}}} \\ &= \frac{\infty + 0}{1+0} = \infty.\end{aligned}$$

□

Example 4.28 Find $\lim_{x\to\infty}\left(\sqrt{x^2+3}-x\right)$.

Solution.

$$\lim_{x\to\infty}\left(\sqrt{x^2+3}-x\right)=\lim_{x\to\infty}\frac{(x^2+3)-x^2}{\sqrt{x^2+3}+x}=\lim_{x\to\infty}\frac{3}{\sqrt{x^2+3}+x}=\frac{3}{\infty}=0.\ \square$$

See CoCalc

https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page126-CoCalcJupyter.pdf

Limits of functions
- a more rigorous approach

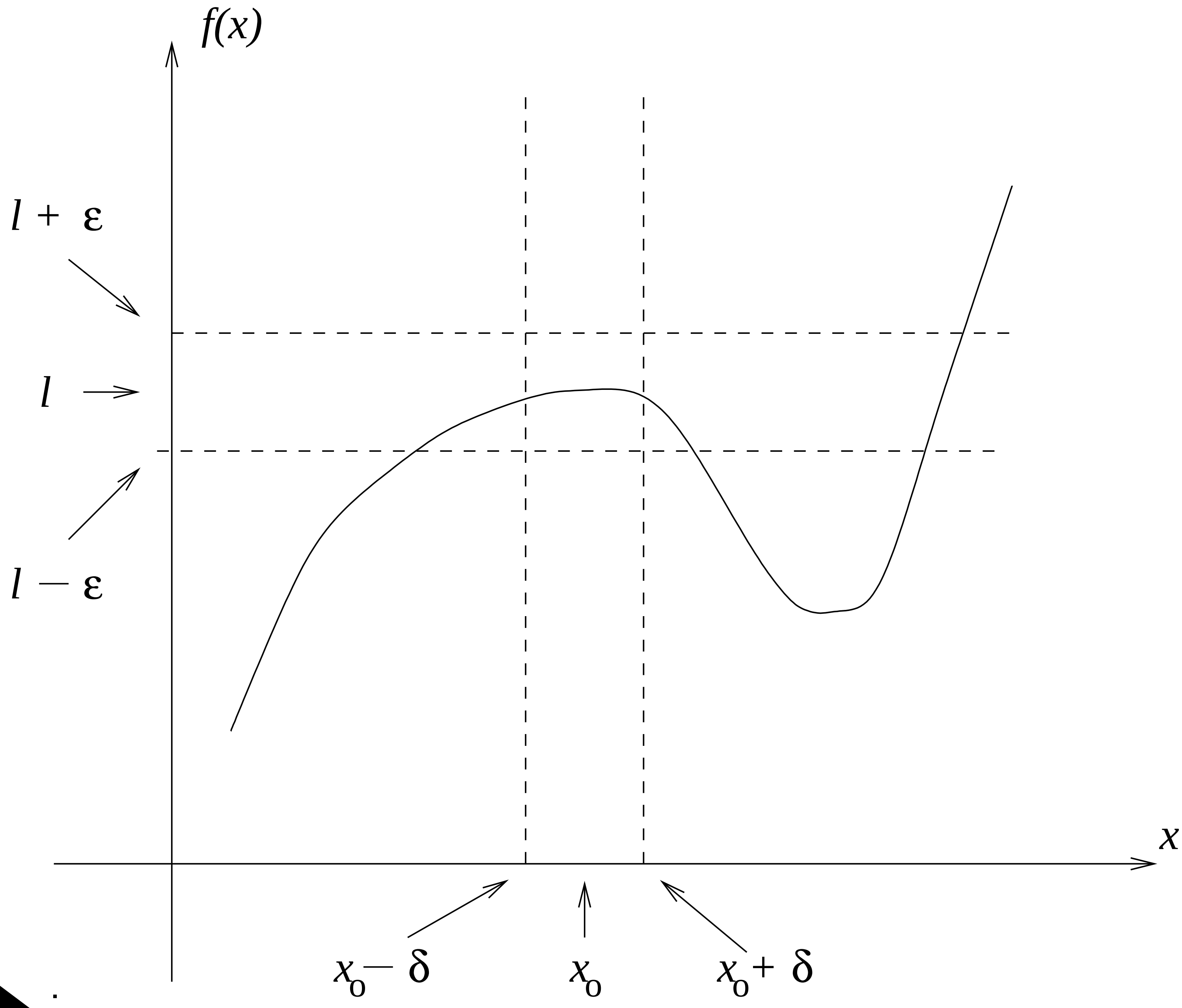
Definition
The set of all points x such that $|x - x_0| < \delta$ is called a δ neighborhood of the point x_0. The set of all points x such that $0 < |x - x_0| < \delta$ in which $x = x_0$ is excluded, is called a deleted δ neighborhood of x_0.

Definition
The number l is the limit of $f(x)$ as x approaches x_0 denoted by

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

if $\forall\ \epsilon > 0,\ \exists\ \delta > 0$ such that $|f(x) - l| < \epsilon$ whenever $0 < |x - x_0| < \delta$.

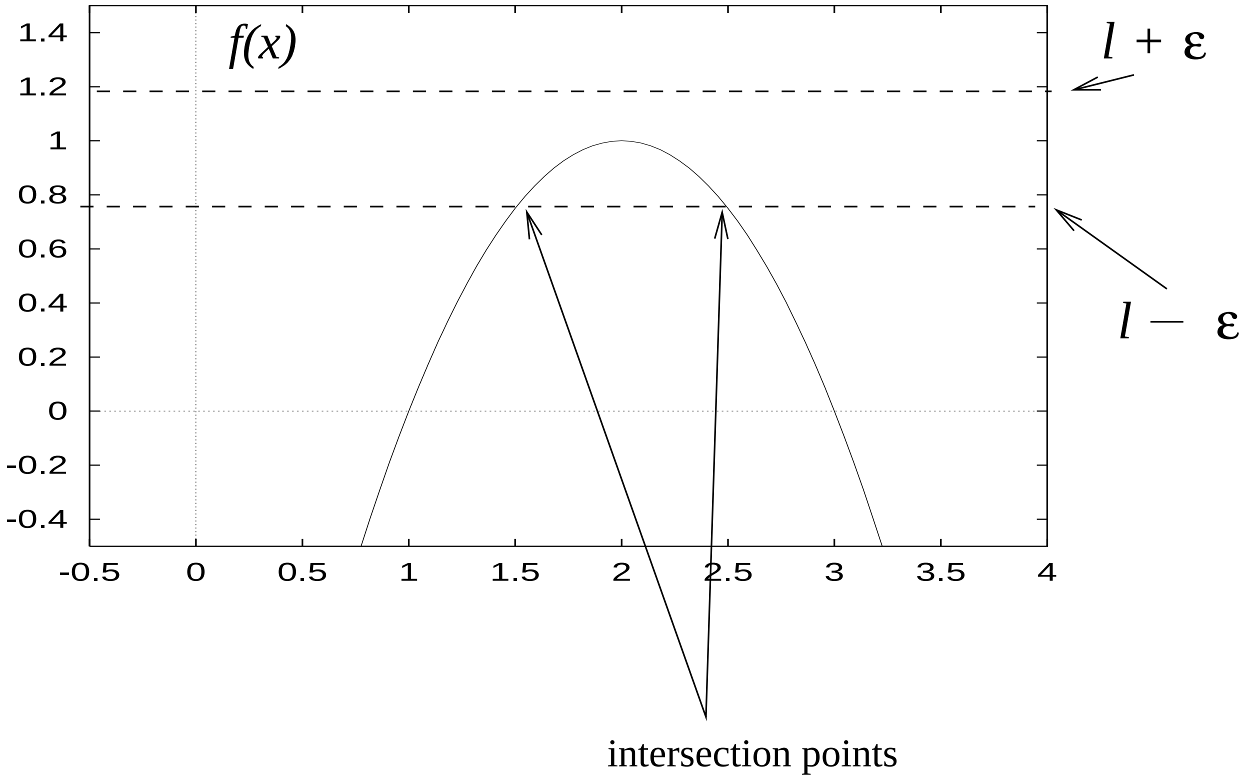
This definition simply says that for any positive number ϵ (however small) we can find some positive number δ (usually depending on ϵ) such that whenever x in the deleted δ neighbourhood of x_0, $x \in (x_0 - \delta\, ,\, x_0 + \delta)\backslash\{x_0\}$, then $f(x) \in (l - \epsilon\, ,\, l + \epsilon)$.
Note that $f(x_0)$ may not equals the limit value l according to the definition.


f(x)
l + ε
l
l − ε
x
x₀− δ
x₀
x₀+ δ

Example : Consider the function

$$f(x) = -x^2 + 4x - 3$$

We are going to show that $\lim_{x \to 2} f(x) = 1$ by definition.



Note that no matter how small we choose the number ϵ , the intersection points indicated in the diagram can be obtained by solving

$$-x^2 + 4x - 3 = 1 - \epsilon$$
$$x^2 - 4x + 4 - \epsilon = 0,$$

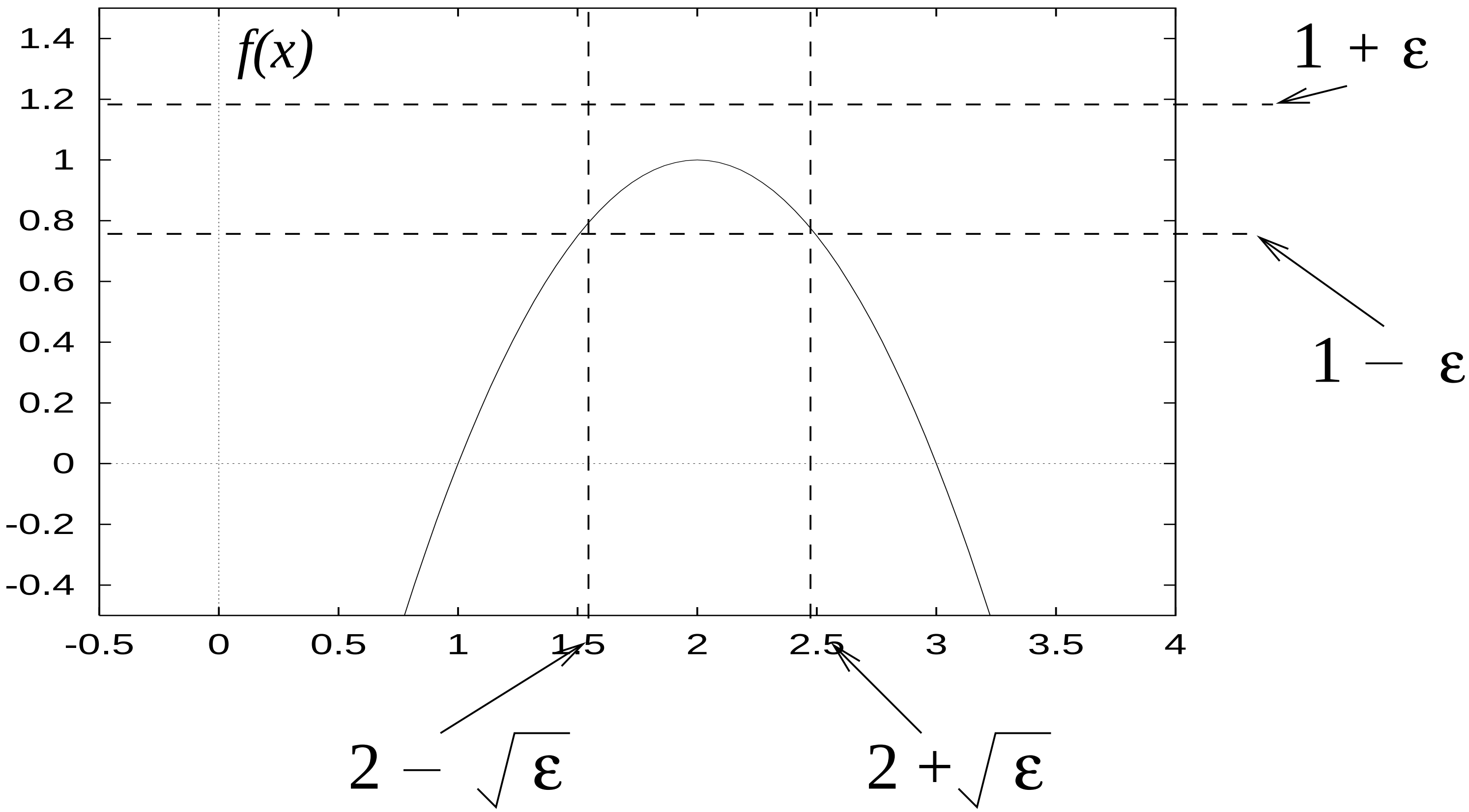
thus, the intersection points are given by

$$\frac{4 \pm \sqrt{16 - 4(4 - \epsilon)}}{2} = 2 \pm \sqrt{\epsilon}.$$

Let $\delta = \sqrt{\epsilon}$. Hence, no matter how small the number ϵ is, we are going to have

$$|f(x) - 1| < \epsilon$$

whenever $0 < |x - 2| < \sqrt{\epsilon} = \delta$. Therefore, by definition, $\lim_{x \to 2} f(x) = 1$.


f(x)
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.5
0
0.5
1
1.5
2
2.5
3
3.5
4
1 + ε
1 − ε
2 − √ε
2 + √ε

Right and Left Hand Limits

We call l^+ the **right hand limit** of $f(x)$ at x_0 if $\forall\ \epsilon > 0,\ \exists\ \delta > 0$ such that $|f(x) - l^+| < \epsilon$ whenever $0 < |x - x_0| < \delta$ **and** $x > x_0$. We write it as

$$\lim_{x \to x_0^+} f(x) = l^+ = f(x_0^+).$$

Similarly, the **Left hand limit** can be defined with the alternate condition $x < x_0$ the same way as the above.
Example : Consider the function

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

What is $\lim\limits_{x \to 0^+} f(x) = f(0^+)$?
What is $\lim\limits_{x \to 0^-} f(x) = f(0^-)$?
In this example, both $f(0^+)$ and $f(0^-)$ exist, but $\lim\limits_{x \to 0} f(x)$ does not.
See CoCalc
https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page132-CoCalcJupyter.pdf

Differentiation

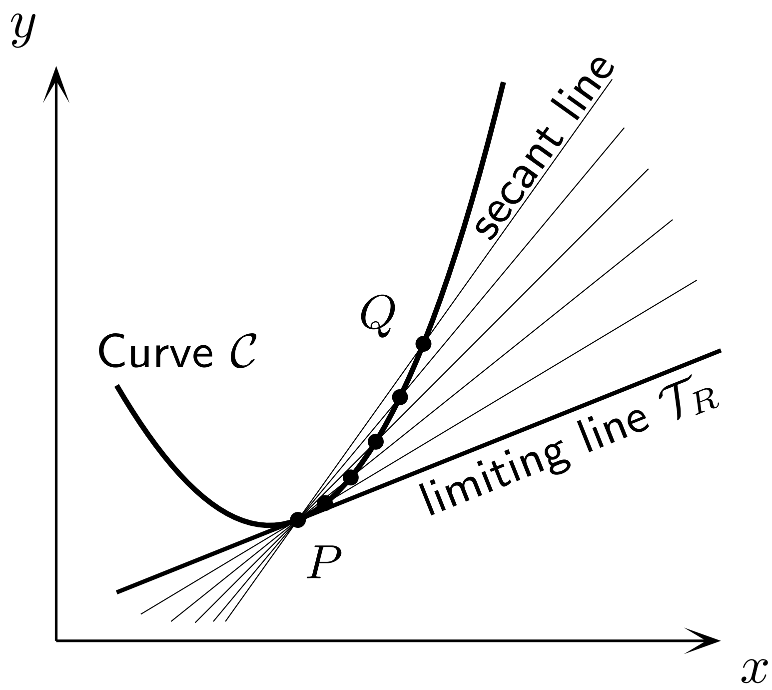
Secant line and Tangent line

Definition 5.1

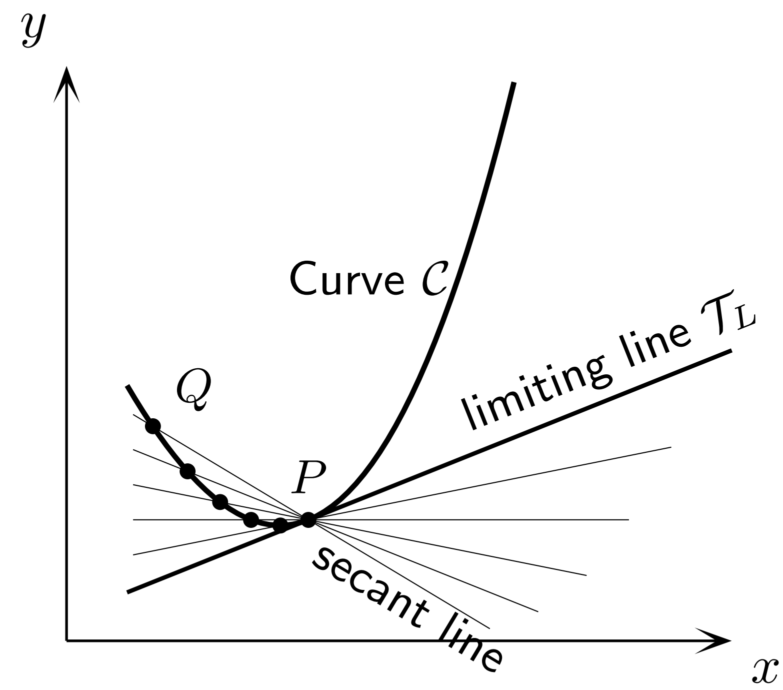
(a) Let P and Q be two distinct points on a curve $\mathcal{C}$. The straight line which passes through P and Q is called a *secant line* (or simply a *secant*) of the curve $\mathcal{C}$.

(b) If P is fixed and we allow Q to move along the curve towards P from both sides of P, and if the secants PQ approaches to the same limiting straight line, we call this limiting straight line the *tangent line* (or simply the *tangent*) to the curve at P.

Tangent to a curve at a point



(a) $Q \to P$ from the right along $\mathcal{C}$



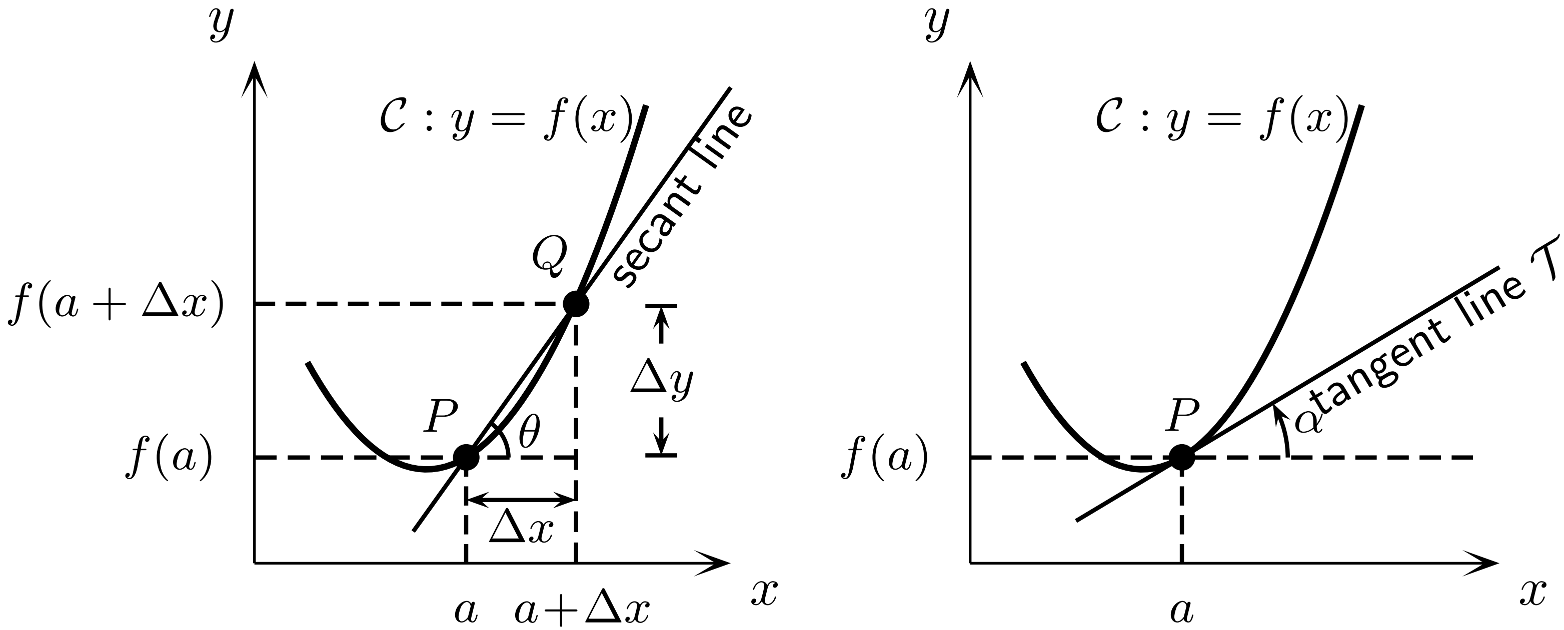
(b) $Q \to P$ from the left along $\mathcal{C}$

Figure 5.1: Tangent line as the limit of secant lines.

In Fig. 5.1, we see that as Q approaches P from the left and from the right, the limiting lines $\mathcal{T}_R$ and $\mathcal{T}_L$ so obtained are the same straight line and therefore $\mathcal{T}_R (= \mathcal{T}_L)$ is the tangent line to the curve at P.

Slope of a curve at a point

The slope is a quantity which measures the steepness of the straight line.



(a) The slope of the secant line PQ is $\tan\theta = \Delta y/\Delta x$

(b) The slope of the tangent line is $\tan\alpha = \lim_{\Delta x \to 0} \Delta y/\Delta x$

Figure 5.2: The slope of a curve as a limit.

Definition 5.2 Let P be a point on a curve $\mathcal{C}$ and let $\mathcal{T}$ be the tangent to $\mathcal{C}$ at P. Then the *slope* of the curve $\mathcal{C}$ at P is the slope of $\mathcal{T}$, if $\mathcal{T}$ is not vertical. (See Fig. 5.2(b).)

Let $\mathcal{C}$ be a given curve whose equation is $y = f(x)$ and let $P(a, f(a))$ be a point on $\mathcal{C}$. Let Q be another point on $\mathcal{C}$ with coordinates $(a + \Delta x, f(a + \Delta x))$ (Fig. 5.2(a)). Here Δx is called an *increment in* x. The point $x = a + \Delta x$ is on the left or on the right of the point $x = a$ according as Δx is negative or positive. In Fig. 5.2(a), Q is on the right of P and therefore Δx is positive. The *increment in* y is defined by

$$\Delta y = f(a + \Delta x) - f(a). \tag{5.1}$$

As $Q \to P$ from both sides along $\mathcal{C}$, $\Delta x \to 0$ from both sides. Consequently, if the curve $\mathcal{C}$ has a tangent line at P, we have:

$$\begin{aligned}
\text{secant line } PQ \quad &\longrightarrow \quad \text{tangent line at } P \\
\theta \quad &\longrightarrow \quad \alpha \\
\frac{f(a+\Delta x) - f(a)}{\Delta x} = \frac{\Delta y}{\Delta x} = \tan\theta \quad &\longrightarrow \quad \tan\alpha \\
&\qquad\quad = \text{the slope of the tangent line at } P \\
&\qquad\quad = \text{the slope of the curve at } P.
\end{aligned}$$

The above limiting values can be seen intuitively from Fig. 5.2 though it shows only the case when $\Delta x \to 0^+$. The result is stated as follows:

Theorem 5.1 *The slope of the curve $\mathcal{C} : y = f(x)$ at $x = a$ is equal to the limit*

$$\lim_{\Delta x \to 0} \frac{f(a + \Delta x) - f(a)}{\Delta x} \quad \textit{or} \quad \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} \tag{5.2}$$

where Δy is defined by (5.1).

Examples

Example 5.1 Use the formula (5.2) to find the slope of the curve $y = x^2$ at $x = a$.

Solution. Let $f(x) = x^2$. Then the required slope is equal to

$$\lim_{\Delta x \to 0} \frac{(a + \Delta x)^2 - a^2}{\Delta x} = \lim_{\Delta x \to 0} (2a + \Delta x) = 2a. \qquad \square$$

Example 5.2 Show that the absolute value function $f(x) = |x|$ has no slope at $x = 0$.

Solution. $f(x) = x$ if $x > 0$; $f(x) = -x$ if $x < 0$ and $f(0) = 0$. Therefore, for $\Delta x > 0$, we have $f(0 + \Delta x) = f(\Delta x) = \Delta x$ and hence

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

However, for $\Delta x < 0$, we have $f(0 + \Delta x) = f(\Delta x) = -\Delta x$ and hence

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

Since the left-hand and right-hand limits are not equal, $f(x)$ has no slope at $x = 0$. □

Derivative of a function

The slope of the graph of $y = f(x)$ at $x = a$ is the limit (5.2). This is a number dependent on the function $f(x)$ and on the constant a. If the limit (5.2) exists, we say that the function $f(x)$ is *differentiable* at $x = a$.
If we consider $f(x)$ as a given function and replace a by the variable x, then the limit (5.2), if exists, becomes a function of x called the *derivative* of the function $f(x)$. The derivative of $f(x)$, which is defined by the formula

$$\lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} \quad \text{or} \quad \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} \tag{5.3}$$

where $\Delta y = f(x + \Delta x) - f(x)$, is denoted by the symbols

$$f'(x) \quad \text{or} \quad y' \quad \text{or} \quad \frac{\mathrm{d}y}{\mathrm{d}x} \quad \text{or} \quad \frac{\mathrm{d}f}{\mathrm{d}x}.$$

Note that all these symbols represent the same function of x.

Differentiation

We often say that we *differentiate* the function $f(x)$ to get its derivative $f'(x)$, and *differentiation* means the process of getting the derivative $f'(x)$ from $f(x)$. We also say that the function $f(x)$ is *differentiable* on an interval J if $f'(x)$ exists at every x in J.

The value of the derivative at a particular point $x = a$ is denoted by

$$f'(a) \quad \text{or} \quad y'(a) \quad \text{or} \quad \left.\frac{\mathrm{d}y}{\mathrm{d}x}\right|_{x=a} \quad \text{or} \quad \left.\frac{\mathrm{d}f}{\mathrm{d}x}\right|_{x=a}.$$

This value is the slope of the curve $y = f(x)$ at $x = a$.

Summary

We summarize the above definitions and notations in the following table:

	Derivative of $f(x)$	Derivative of $f(x)$ at $x = a$
Definition:	$\lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$	$\lim_{\Delta x \to 0} \frac{f(a+\Delta x) - f(a)}{\Delta x}$
Nature:	A function of x	A real number
Notation:	$f'(x)$ or y' or $\frac{\mathrm{d}y}{\mathrm{d}x}$ or $\frac{\mathrm{d}f}{\mathrm{d}x}$	$f'(a)$ or $y'(a)$ or $\left.\frac{\mathrm{d}y}{\mathrm{d}x}\right\|_{x=a}$ or $\left.\frac{\mathrm{d}f}{\mathrm{d}x}\right\|_{x=a}$
Geometric meaning:	The slope of the curve $y = f(x)$ at a general point x.	The slope of the curve $y = f(x)$ at a particular point where $x = a$. This is equal to $\tan\alpha$ in Fig. 5.2.

Examples

Example 5.3 Use the formula (5.3) to differentiate $y = x^2$.

Solution. Let $f(x) = x^2$. Then

$$y' = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{(x + \Delta x)^2 - x^2}{\Delta x}$$

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

□

Note that the above process of getting the derivative is exactly the same as that in Example 5.1 except that we have replaced the constant a by the variable x. In the forthcoming section, we shall list formulas of derivatives in a table so that we may use the formulas to get derivatives directly without spending time in evaluating limits.

Differentiability implies continuity

Using (5.2), we can prove that:

Theorem 5.2 *If* $f'(a)$ *exists then the function* $f(x)$ *is continuous at* $x = a$.

The converse of the theorem is not true. The function $f(x) = |x|$ gives a counter-example. This $f(x)$ is continuous at $x = 0$ but $f(x)$ is not differentiable at $x = 0$.

Rate of change

Suppose an object moves along a straight line. Its distance from a certain fixed point O on the line at time t is given by $y = F(t)$. Over the time interval $[t_0,\, t_0 + \Delta t]$, the object covers a distance equal to $\Delta y = F(t_0 + \Delta t) - F(t_0)$. The *difference quotient*

$$\frac{\Delta y}{\Delta t} = \frac{F(t_0 + \Delta t) - F(t_0)}{\Delta t}$$

is the average *velocity* of the object over the time interval $[t_0,\, t_0 + \Delta t]$.
As a result, the derivative $f'(t_0)$ is simply the *instantaneous velocity* of the object at the instant t_0.
More generally, for $y = f(x)$, the *difference quotient*

$$\frac{f(b) - f(a)}{b - a}$$

is called the *average rate of change of* y *with respect to* x *over the interval* $[a, b]$, and the derivative $f'(a)$ at $x = a$ is called the *rate of change of* y *with respect to* x *at* $x = a$.

Differentiation by the first principle

In addition to Example 5.3, we give more examples to show how differentiation formulas can be derived from the first principle, i.e. from the definition (5.3) of derivatives.

Examples

Example 5.4 Let $y = f(x) = C$, a constant. Show that $\dfrac{\mathrm{d}y}{\mathrm{d}x} = 0$ for every x.

Proof. Since

$$\Delta y = f(x + \Delta x) - f(x) = C - C = 0,$$

we get

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{0}{\Delta x} = 0. \qquad \square$$

Example 5.5 Let $y = f(x) = x^n$, where n is a positive integer. Show from the first principle that $\frac{\mathrm{d}y}{\mathrm{d}x} = nx^{n-1}$.

Proof. Let x be fixed and Δx approach to 0 but not equal to 0. Write $z = x + \Delta x$. Then $\Delta x \to 0$ means $z \to x$. As $\Delta x \neq 0$, we have

$$\frac{\Delta y}{\Delta x} = \frac{(x + \Delta x)^n - x^n}{\Delta x} = \frac{z^n - x^n}{z - x}$$

$$= z^{n-1} + z^{n-2}x + z^{n-3}x^2 + \cdots + zx^{n-2} + x^{n-1} \quad (n \text{ terms})$$

Therefore,

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{z \to x} (z^{n-1} + z^{n-2}x + z^{n-3}x^2 + \cdots + x^{n-1}) = nx^{n-1}.$$

□

Example 5.6 Let $y = 1/x$. Show from the first principle that $y' = -1/x^2$.

Proof.

$$y' = \lim_{\Delta x \to 0} \frac{\dfrac{1}{x + \Delta x} - \dfrac{1}{x}}{\Delta x}$$

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

□

Example 5.7 Let $y = \sin x$. Show from the first principle that $y' = \cos x$.

Proof. Using the identity $\sin A - \sin B = 2\cos\left(\frac{A+B}{2}\right)\sin\left(\frac{A-B}{2}\right)$, we have, for any fixed x and any $\Delta x \neq 0$,

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

Since

$$\lim_{\Delta x\to 0}\frac{\sin\frac{\Delta x}{2}}{\frac{\Delta x}{2}} = 1 \quad \text{and} \quad \lim_{\Delta x \to 0}\cos\left(x+\frac{\Delta x}{2}\right) = \cos x,$$

we have

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \lim_{\Delta x\to 0}\frac{\sin(x+\Delta x)-\sin x}{\Delta x} = \cos x.$$

□

Table of differentiation formulas

	$f(x)$	$f'(x)$	
1.	constant	0	
2.	x^n	nx^{n-1}	$n =$ real constant
3.	$\sin x$	$\cos x$	
4.	$\cos x$	$-\sin x$	
5.	$\tan x$	$\sec^2 x$	
6.	$\cot x$	$-\csc^2 x$	
7.	$\sec x$	$\sec x \tan x$	
8.	$\csc x$	$-\csc x \cot x$	
9.	e^x	e^x	$\mathrm{e} = 2.718281828\cdots$
10.	a^x	$a^x \ln a$	$a > 0$, real constant
11.	$\ln x$	$1/x$	$x > 0$
12.	$\log_a x$	$(\log_a \mathrm{e})/x$	$a > 0$, real constant
13.	$\sin^{-1} x$	$1/\sqrt{1-x^2}$	
14.	$\cos^{-1} x$	$-1/\sqrt{1-x^2}$	
15.	$\tan^{-1} x$	$1/(1+x^2)$	

Examples

Example 5.8 If $y = 1/x^3$, find y'.

Solution. Since we can write $y = x^{-3}$, we get, using formula 2 with $n = -3$,

$$y' = (-3)x^{-3-1} = -3x^{-4}. \qquad \square$$

Example 5.9 If $y = \sqrt[3]{x}$, find y'.

Solution. Since we can write $y = x^{1/3}$, we get , using formula 2 with $n = 1/3$,

$$y' = \frac{1}{3}x^{1/3-1} = \frac{1}{3x^{2/3}}.$$

□

Example 5.10 The point $P(\pi/4, 1)$ lies on the curve $y = \tan x$. Find the slope of the curve at this point.

Solution. Using formula 5, we have $y' = \sec^2 x$. Therefore at $x = \pi/4$, the slope of the curve $y = \tan x$ is given by

$$m = \sec^2(\pi/4) = 2.$$

□

Example 5.11 The point $A(2, 8)$ lies on the curve $y = x^3$. Find the equation of the tangent line at A.

Solution. Using formula 2 with $n = 3$, we have $y' = 3x^2$. Therefore at $x = 2$, the slope of the curve $y = x^3$ is $m = 3 \cdot 2^2 = 12$. Hence the equation of the tangent line through $A(2, 8)$ is

$$y - 8 = 12(x - 2) \quad \text{or} \quad y = 12x - 16. \qquad \square$$

Basic rules of differentiation

(Scalar multiplication) $y = kf(x)$, k **is a constant :** $\dfrac{\mathrm{d}y}{\mathrm{d}x} = k\,\dfrac{\mathrm{d}f}{\mathrm{d}x} = kf'(x)$

(Sum) $y = f(x) + g(x)$**:** $\dfrac{\mathrm{d}y}{\mathrm{d}x} = \dfrac{\mathrm{d}f}{\mathrm{d}x} + \dfrac{\mathrm{d}g}{\mathrm{d}x} = f'(x) + g'(x)$

(Difference) $y = f(x) - g(x)$**:** $\dfrac{\mathrm{d}y}{\mathrm{d}x} = \dfrac{\mathrm{d}f}{\mathrm{d}x} - \dfrac{\mathrm{d}g}{\mathrm{d}x} = f'(x) - g'(x)$

(Product) $y = f(x)g(x)$**:** $\dfrac{\mathrm{d}y}{\mathrm{d}x} = f(x)\dfrac{\mathrm{d}g}{\mathrm{d}x} + g(x)\dfrac{\mathrm{d}f}{\mathrm{d}x} = f(x)g'(x) + g(x)f'(x)$

(Quotient) $y = \dfrac{f(x)}{g(x)}$**:** $\dfrac{\mathrm{d}y}{\mathrm{d}x} = \dfrac{g(x)f'(x) - f(x)g'(x)}{[g(x)]^2}$

(Composite function) $y = f(u)$ & $u = g(x)$**:** $\dfrac{\mathrm{d}y}{\mathrm{d}x} = \dfrac{\mathrm{d}y}{\mathrm{d}u} \cdot \dfrac{\mathrm{d}u}{\mathrm{d}x} = f'(u)g'(x)$

(Inverse function) $y = f(x) \& x = f^{-1}(y)$**:** $\dfrac{\mathrm{d}x}{\mathrm{d}y} = \dfrac{1}{\dfrac{\mathrm{d}y}{\mathrm{d}x}} = \dfrac{1}{f'(x)}$ if $f'(x) \neq 0$

Examples: Constant multiplication, sums and differences

Example 5.12 If $y = 4x^3$, find y'.

Solution. $y' = 4 \cdot 3x^2 = 12x^2$. □

Example 5.13 If $y = x^3 + x^2 - 2x + 3$, find y'.

Solution. $y' = 3x^2 + 2x - 2$. □

Example 5.14 If $y = 5x^3 + x + 4/x$, find y'.

Solution. $y' = 5 \cdot 3x^2 + 1 - 4/x^2 = 15x^2 + 1 - 4/x^2$. □

Example 5.15 If $y = \dfrac{3}{x^2} - \dfrac{2}{x}$, find y'.

Solution. $y' = \dfrac{3 \cdot (-2)}{x^3} - \dfrac{2 \cdot (-1)}{x^2} = \dfrac{2}{x^2} - \dfrac{6}{x^3}$ □

Example 5.16 If $y = x^n + bx^3 + c$ where n, b, c are constants. Find y'.

Solution. $y' = nx^{n-1} + 3bx^2$. □

Example 5.17 If $y = 2\cos x + \sin x$, find y'.

Solution. $y' = -2\sin x + \cos x$. $\square$

Example 5.18 If $y = 2\mathrm{e}^x + 3\sin x$, find y'.

Solution. $y' = 2\mathrm{e}^x + 3\cos x$. □

Example 5.19 If $y = 4x + \ln x$, find y'.

Solution. $y' = 4 + 1/x$. □

Example 5.20 If $y = 3x^2 - 2\tan x$, find y'.

Solution. $y' = 6x - 2\sec^2 x$. □

Example 5.21 If $y = 2x^3 + 3e^x - \sin x$, find y'.

Solution. $y' = 6x^2 + 3e^x - \cos x$. □

Example 5.22 If $y = 1 + x + x^2 + x^3 + x^4 + x^5$, find y'.

Solution. $y' = 1 + 2x + 3x^2 + 4x^3 + 5x^4$. □

Examples: Products and quotients

Example 5.23 If $y = x^3 \sin x$, find y'.

Solution.

$$y' = x^3 \frac{\mathrm{d}}{\mathrm{d}x} \sin x + \sin x \frac{\mathrm{d}}{\mathrm{d}x} x^3 = x^3 \cos x + 3x^2 \sin x.$$

□

Example 5.24 If $y = x^2 \mathrm{e}^x$, find y'.

Solution.

$$y' = x^2 \frac{\mathrm{d}}{\mathrm{d}x} \mathrm{e}^x + \mathrm{e}^x \frac{\mathrm{d}}{\mathrm{d}x} x^2 = x^2 \mathrm{e}^x + 2x \mathrm{e}^x = \mathrm{e}^x x(x+2).$$

□

Example 5.25 If $y = (x^2 + x - 2)\mathrm{e}^x$, find y'.

Solution.

$$
\begin{aligned}
y' &= (x^2 + x - 2)\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{e}^x + \mathrm{e}^x\frac{\mathrm{d}}{\mathrm{d}x}(x^2 + x - 2) \\
&= (x^2 + x - 2)\mathrm{e}^x + \mathrm{e}^x(2x + 1) = \mathrm{e}^x(x^2 + 3x - 1).
\end{aligned}
$$

□

Example 5.26 If $y = (x^2 + 3x - 2)\cos x$, find y'.

Solution.

$$\begin{aligned} y' &= (x^2 + 3x - 2)\frac{\mathrm{d}}{\mathrm{d}x}\cos x + \cos x \frac{\mathrm{d}}{\mathrm{d}x}(x^2 + 3x - 2) \\ &= (x^2 + 3x - 2)(-\sin x) + (\cos x)(2x + 3) \\ &= -(x^2 + 3x - 2)\sin x + (2x + 3)\cos x. \end{aligned}$$

□

Example 5.27 If $y = \dfrac{x^2}{\sin x}$, find y'.

Solution.

$$y' = \frac{(\sin x)\dfrac{\mathrm{d}}{\mathrm{d}x}(x^2) - x^2\dfrac{\mathrm{d}}{\mathrm{d}x}\sin x}{[\sin x]^2} = \frac{2x\sin x - x^2\cos x}{\sin^2 x}.$$

$\square$

See CoCalc
https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page173-CoCalcJupyter.pdf

Example 5.28 If $y = \dfrac{x^2+3x+2}{x^2+2}$, find y'.

Solution.

$$y' = \frac{(x^2+2)\dfrac{\mathrm{d}}{\mathrm{d}x}(x^2+3x+2) - (x^2+3x+2)\dfrac{\mathrm{d}}{\mathrm{d}x}(x^2+2)}{(x^2+2)^2}$$

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

□

Example 5.29 If $y = \dfrac{\mathrm{e}^x}{\cos x}$, find y'.

Solution.

$$y' = \frac{(\cos x)\dfrac{\mathrm{d}}{\mathrm{d}x}(\mathrm{e}^x) - \mathrm{e}^x\dfrac{\mathrm{d}}{\mathrm{d}x}\cos x}{\cos^2 x} = \frac{\mathrm{e}^x(\cos x + \sin x)}{\cos^2 x}.$$

□

Examples: Chain rule

Example 5.30 If $y = \sin 3x$, find y'.

Solution. Let $y = \sin u$ and $u = 3x$. Then

$$y' = \frac{\mathrm{d}y}{\mathrm{d}u} \cdot \frac{\mathrm{d}u}{\mathrm{d}x} = (\cos u)(3) = 3\cos 3x.$$

$\square$

Example 5.31 If $y = \cos x^3$, find y'.

Solution. Let $y = \cos u$ and $u = x^3$. Then

$$y' = \frac{\mathrm{d}y}{\mathrm{d}u} \cdot \frac{\mathrm{d}u}{\mathrm{d}x} = (-\sin u)(3x^2) = -3x^2 \sin x^3.$$

$\square$

Example 5.32 If $y = (x^2 - 3x + 2)^6$, find y'.

Solution. Let $y = u^6$ and $u = x^2 - 3x + 2$. Then

$$y' = \frac{\mathrm{d}y}{\mathrm{d}u} \cdot \frac{\mathrm{d}u}{\mathrm{d}x} = (6u^5)(2x - 3) = 6(2x - 3)(x^2 - 3x + 2)^5.$$

□

Example 5.33 If $y = \mathrm{e}^{-4x^2+3x-1}$, find y'.

Solution. Let $y = \mathrm{e}^u$ and $u = -4x^2 + 3x - 1$. Then

$$y' = (\mathrm{e}^u)(-8x+3) = (-8x+3)\exp(-4x^2+3x-1).$$

□

Example 5.34 If $y = \ln(2x^2 + 1)$, find y'.

Solution. Let $y = \ln u$ and $u = 2x^2 + 1$. Then

$$y' = \frac{1}{u} \cdot (4x) = \frac{4x}{2x^2 + 1}.$$

□

Example 5.35 Let $y = \sqrt{x^2 + a^2}$ where a is a nonzero constant. Find y'.

Solution. Let $y = u^{1/2}$ and $u = x^2 + a^2$. Then

$$y' = \frac{1}{2}u^{-1/2}(2x) = \frac{x}{\sqrt{x^2 + a^2}}.$$

□

Example 5.36 If $y = \dfrac{1}{x^2 + 4x + 2}$, find y'.

Solution. Let $y = u^{-1}$ and $u = x^2 + 4x + 2$. Then

$$y' = -u^{-2}(2x + 4) = -\frac{2(x + 2)}{(x^2 + 4x + 2)^2}.$$

□

Example 5.37 If $y = \sin \ln x^2$, find y'.

Solution. Let $y = \sin u$, $u = \ln v$ and $v = x^2$. Then

$$y' = \frac{\mathrm{d}y}{\mathrm{d}u} \cdot \frac{\mathrm{d}u}{\mathrm{d}v} \cdot \frac{\mathrm{d}v}{\mathrm{d}x} = (\cos u)\frac{1}{v}(2x) = \frac{2(\cos \ln x^2)}{x}.$$

$\square$

Example 5.38 If $y = (\sin x^2)\exp(\cos x)$, find y'.

Solution.

$$y' = \exp(\cos x)\frac{\mathrm{d}}{\mathrm{d}x}(\sin x^2) + (\sin x^2)\frac{\mathrm{d}}{\mathrm{d}x}\exp(\cos x)$$

$$= 2x\exp(\cos x)(\cos x^2) - (\sin x)\,(\sin x^2)\,\exp(\cos x).$$

□

See CoCalc
https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page184-CoCalcJupyter.pdf

Example 5.39 If $y = \ln\left(\dfrac{2x+3}{x+2}\right)$, find y'.

Solution.

$$y' = \frac{x+2}{2x+3} \cdot \frac{2(x+2) - (2x+3)}{(x+2)^2}$$

$$= \frac{1}{(2x+3)(x+2)}.$$

□

Example 5.40 Let f and g be two differentiable functions. If $y = f(x^2)g(3x+2)$, find y' in terms of f, g, f' and g'.

Solution.

$$\begin{aligned} y' &= f(x^2)\frac{\mathrm{d}}{\mathrm{d}x}g(3x+2) + g(3x+2)\frac{\mathrm{d}}{\mathrm{d}x}f(x^2) \\ &= 3\,f(x^2)\,g'(3x+2) + 2x\,g(3x+2)\,f'(x^2). \end{aligned}$$

□

Examples: Inverse functions

Example 5.41 Let $y = \sin x^2$, $0 < x < 1$. Find $\dfrac{\mathrm{d}x}{\mathrm{d}y}$ in terms of x.

Solution.

$$\frac{\mathrm{d}x}{\mathrm{d}y} = 1\Big/\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{1}{2x\cos x^2}.$$

□

Example 5.42 Show that $\dfrac{\mathrm{d}}{\mathrm{d}x}\left(\sin^{-1}x\right) = \dfrac{1}{\sqrt{1-x^2}}$ for $|\,x\,| < 1.$

Proof. If $y = f(x) = \sin x$, $-\pi/2 < x < \pi/2$, then $x = g(y) = \sin^{-1} y$. Since $\cos x > 0$ whenever $-\pi/2 < x < \pi/2$, we have

$$\frac{\mathrm{d}x}{\mathrm{d}y} \quad \text{or} \quad g'(y) = \frac{1}{f'(x)} = \frac{1}{\cos x} = \frac{1}{\sqrt{1-\sin^2 x}} = \frac{1}{\sqrt{1-y^2}},$$

for every $y \in (-1,\ 1)$. Changing the dummy variable y to x, we have

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\sin^{-1}x\right) = \frac{1}{\sqrt{1-x^2}} \quad \text{for } |\,x\,| < 1.$$

$\square$

Remark 5.1 Similarly, we can establish the formula

$$\frac{\mathrm{d}}{\mathrm{d}x}(\cos^{-1}x) = -\frac{1}{\sqrt{1-x^2}} \quad \text{for } |\,x\,| < 1.$$

Implicit functions

In the xy-plane, the unit circle with centre at the origin O can be represented by the equation

$$x^2 + y^2 = 1. \tag{5.4}$$

We see that the circle describes two functions of x. One of these can be represented by the upper half of the circle while the other by the lower half (see Fig. 5.3).

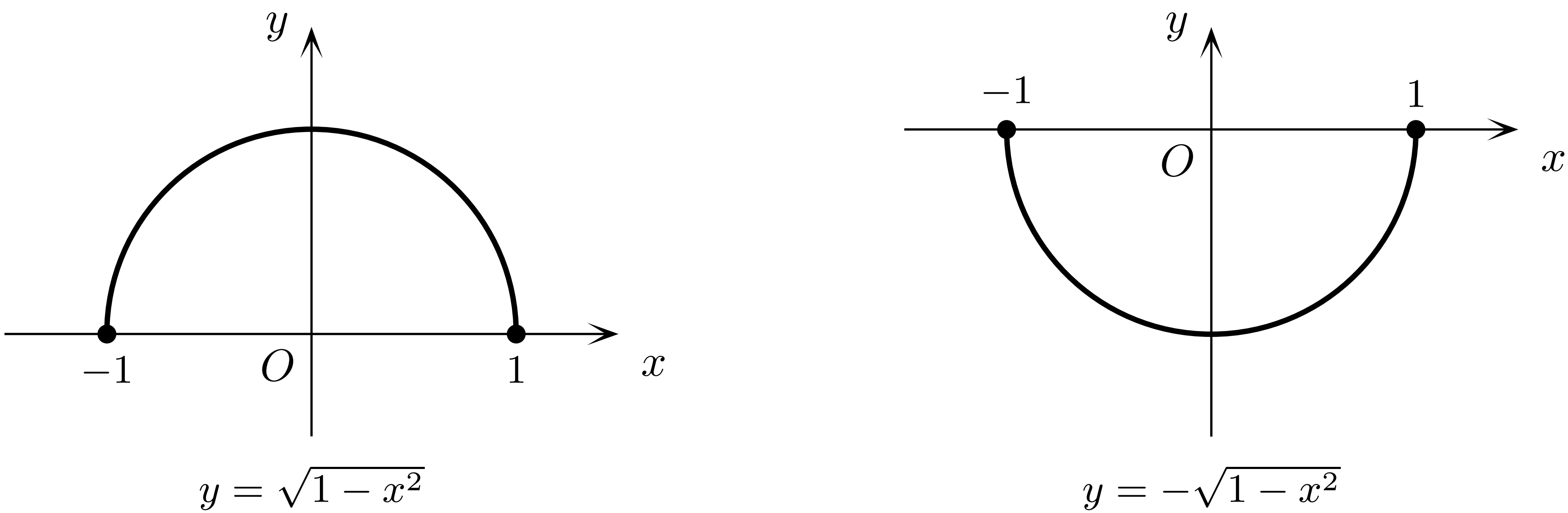


Figure 5.3: Two functions defined by the equation $x^2 + y^2 = 1$.

Each of these functions is called an *implicit function* defined by the equation (5.4). In fact, the functions in this special case can be defined explicitly by:

$$y = \sqrt{1 - x^2}, \quad -1 \leq x \leq 1$$

and

$$y = -\sqrt{1 - x^2}, \quad -1 \leq x \leq 1.$$

However, not all implicit functions can be expressed in explicit forms. For example, the equation

$$y + \sin(xy) = 2\pi \tag{5.5}$$

defines y as one or more functions of x implicitly, but not explicitly.

Examples

Example 5.43 Find y' if y is the function defined implicitly by the equation (5.5). Show that the point $P(1,\ 2\pi)$ lies on the curve defined by (5.5) and find the slope of the curve at P.
See CoCalc
https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page191-CoCalcJupyter.pdf

Solution. Regarding y as a function of x, we can differentiate both sides of (5.5) with respect to x:

$$\frac{\mathrm{d}}{\mathrm{d}x}(y + \sin xy) = \frac{\mathrm{d}}{\mathrm{d}x}(2\pi).$$

It follows that $y' + (\cos xy)(xy' + y) = 0$. Solving for y', we get

$$y' = \frac{-y\cos xy}{1 + x\cos xy}.$$

At $x = 1, y = 2\pi$, the LHS of (5.5) is $2\pi + \sin 2\pi$ which is equal to the RHS. Therefore the point $P(1,\ 2\pi)$ lies on the curve defined by (5.5). At this point, $x = 1$ and $y = 2\pi$. Therefore the slope of the curve at P is

$$y' = \frac{-y\cos xy}{1 + x\cos xy} = \frac{-2\pi\cos(2\pi)}{1 + \cos(2\pi)} = -\pi.$$

□

Higher derivatives

If $y = f(x)$ has a derivative $f'(x)$ and if $f'(x)$ also has a derivative, this derivative of the derivative of $f(x)$ is called the *second order derivative* of $f(x)$. The second order derivative is denoted by

$$y'' \quad \text{or} \quad f''(x) \quad \text{or} \quad \frac{\mathrm{d}^2 y}{\mathrm{d}x^2}.$$

If we differentiate the second order derivative, we get the *third order derivative* denoted by:

$$y''' \quad \text{or} \quad f'''(x) \quad \text{or} \quad \frac{\mathrm{d}^3 y}{\mathrm{d}x^3}.$$

In this way, the nth order derivative is defined and is denoted by

$$y^{(n)} \quad \text{or} \quad f^{(n)}(x) \quad \text{or} \quad \frac{\mathrm{d}^n y}{\mathrm{d}x^n}.$$

For convenience, we also define

$$y^{(0)} = f^{(0)}(x) = f(x)$$

so that $y^{(n)}$ or $f^{(n)}(x)$ is defined for $n = 0, 1, 2, 3, \ldots$.

Examples

Example 5.44 Let $y = x^3 - 4\ln x$. Find y', y'' and y'''.

Solution. $y' = 3x^2 - 4/x, \quad y'' = 6x + 4/x^2$ and $y''' = 6 - 8/x^3$. □

A useful rule for differentiating a product n times is:

Leibniz's rule

For differentiable functions $u(x)$ and $v(x)$,

$$(uv)^{(n)} = \sum_{k=0}^{n} \binom{n}{k} u^{(n-k)} v^{(k)}$$

where $\binom{n}{k}$ denotes the coefficient of t^k in the binomial expansion of $(1+t)^n$.

The formula for the binomial coefficient is

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad \text{for } n = 1, 2, 3, \ldots; \;\; k = 0, 1, 2, \ldots, n$$

and the rules for $n = 1, 2, 3, 4$ are:

$$\begin{aligned}
(uv)' &= u'v + uv' \\
(uv)'' &= u''v + 2u'v' + uv'' \\
(uv)''' &= u'''v + 3u''v' + 3u'v'' + uv''' \\
(uv)^{(4)} &= u^{(4)}v + 4u'''v' + 6u''v'' + 4u'v''' + uv^{(4)}
\end{aligned}$$

Examples

Example 5.45 Find y'' if $y = x^3 \sin 2x$.

Solution.

$$\begin{aligned} y'' &= x^3(-4\sin 2x) + 2(3x^2)(2\cos 2x) + 6x\sin 2x \\ &= (-4x^3 + 6x)\sin 2x + 12x^2\cos 2x. \end{aligned}$$

□

Example 5.46 Let $y = x^2 \mathrm{e}^{2x}$. Find $y^{(n)}$ for $n \geq 0$.

Solution.
Let $u = x^2$ and $v = \mathrm{e}^{2x}$. Then $u^{(k)} = 0$ for $k \geq 3$. By Leibniz's rule,

$$\begin{aligned} y^{(n)} &= x^2(2^n \mathrm{e}^{2x}) + n(2x)(2^{n-1}\mathrm{e}^{2x}) + \frac{n(n-1)}{2}(2)(2^{n-2}\mathrm{e}^{2x}) + 0 + \cdots \\ &= 2^{n-2}\mathrm{e}^{2x}[4x^2 + 4nx + n(n-1)] \end{aligned} \tag{5.6}$$

for $n \geq 2$. Furthermore, direct differentiation gives

$$y = x^2\mathrm{e}^{2x} \quad \text{and} \quad y' = \mathrm{e}^{2x}(2x + 2x^2)$$

showing that (5.6) is also true for $n = 0$ and 1. □

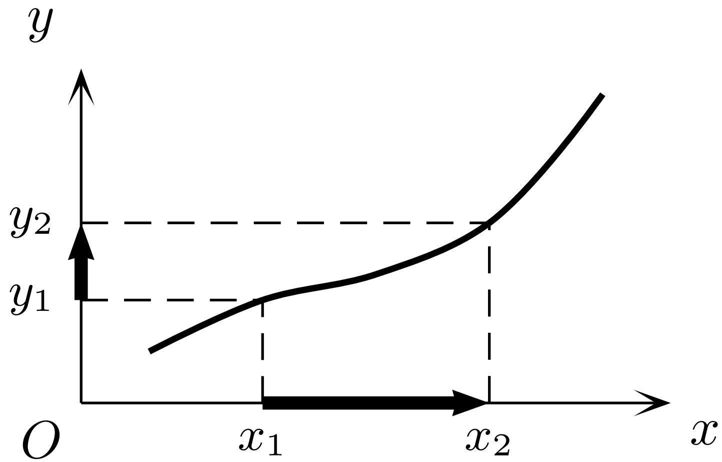
See CoCalc
https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page196-CoCalcJupyter.pdf

Applications of Differentiation

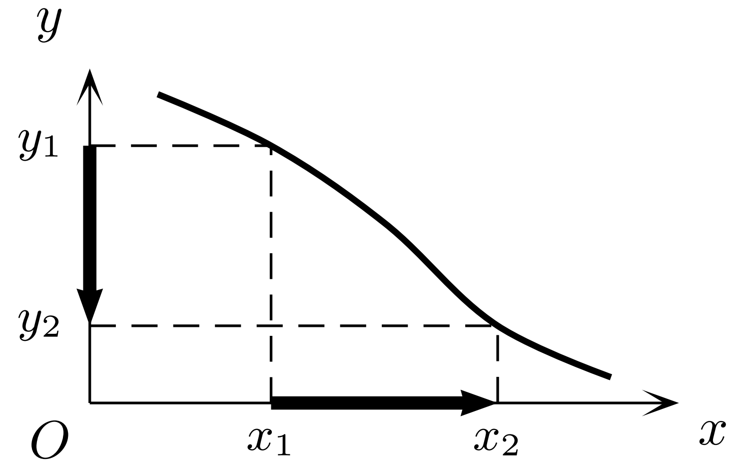
Increasing and decreasing functions

Definition 6.1 Let J be an open interval.

- A function $y = f(x)$ is *increasing* on J if y increases as x increases on J.
- The function is *decreasing* on J if y decreases as x increases on J.



An increasing function.



An decreasing function.

Figure 6.1: The meaning of an increasing function and a decreasing function.

Example 6.1 Show that $y = x^2$ is increasing in the interval $J = (0, \infty)$

Solution. For any two distinct numbers x_1 and x_2 in $J = (0, \infty)$, we have

$$\frac{x_1^2 - x_2^2}{x_1 - x_2} = x_1 + x_2 > 0.$$

Therefore $x_1^2 < x_2^2$ whenever $x_1 < x_2$. Hence $y = x^2$ is increasing in J. □

If a function is differentiable, we can determine its monotonicity by considering the sign of its derivative.

Theorem 6.1 *Let $f(x)$ be differentiable on an open interval J. Then*

- $f'(x) > 0$ *on* $J \implies f$ *is increasing on* J.
- $f'(x) < 0$ *on* $J \implies f$ *is decreasing on* J.

Examples

Example 6.2 Use the above theorem to show that the function $y = x^3$ is increasing on the interval $J_1 = (-\infty, 0)$ and on the interval $J_2 = (0, \infty)$.

Solution. Since $y' = 3x^2 > 0$ on J_1 and on J_2, the above theorem asserts that the function $y = x^3$ is increasing on either interval. □

In fact using directly the definition of increasing function we can show that the function $y = x^3$ is increasing over the entire real line.

Example 6.3 Determine the open intervals in which the function $f(x) = x^4 + 4x$ is increasing or decreasing.

Solution. $f'(x) = 4x^3 + 4 = 4(x^3 + 1)$.

When $x < -1$, $f' < 0$. $\therefore f$ is decreasing in the interval $(-\infty, -1)$.

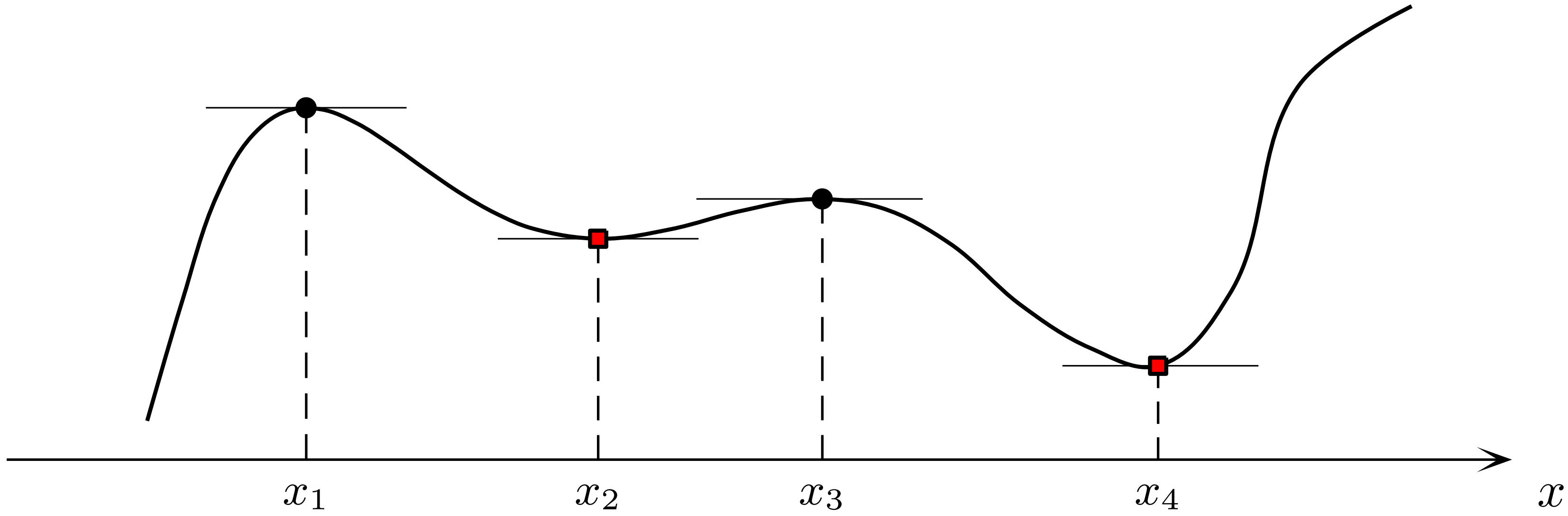
When $x > -1$, $f' > 0$. $\therefore f$ is increasing in the interval $(-1, \infty)$. $\square$

Local maxima and minima

Definition 6.2 Consider $y = f(x)$ and $a \in \operatorname{Dom} f$.

- $f(x)$ has a *local* (or *relative*) *maximum* at a, and $f(a)$ is a *local* (or *relative*) *maximum* if $f(a) \geq f(x)$ for all x in an open interval containing a.
- The definition for a *local* (or *relative*) *minimum* is similar.
- We say that $f(x)$ has a *local extremum* at a if it has either a local maximum or a local minimum at a.

The local extrema of a differentiable function are illustrated in Fig. 6.2. From the diagram, we see that the local extrema occur when the tangents to the graph are horizontal, i.e. when the graph has zero slope.



• Local maxima at x_1 and x_3. ■ Local minima at x_2 and x_4.

Figure 6.2: The meaning of local extrema.

Stationary point

Definition 6.3 If $f'(a) = 0$, we say that $x = a$ is a *stationary point*.

In Fig. 6.2, x_1, x_2, x_3 and x_4 are stationary points.

Theorem 6.2 (A necessary condition for local extrema) *If $f(x)$ is differentiable in an open interval J and if it has a local extremum at $x = a$ in J, then $x = a$ must be a stationary point, i.e. $f'(a) = 0$.*

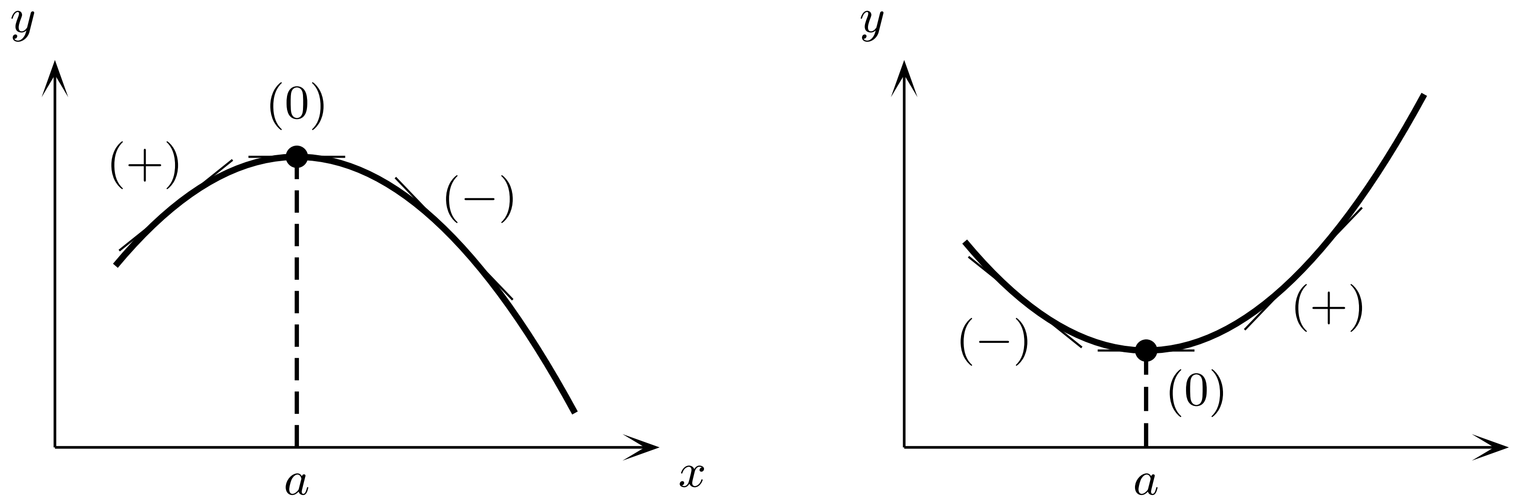
The theorem tells us that for differentiable functions, the local extrema can be found by considering only the stationary points. Having found the stationary points, we can test them one by one for local maxima or local minima using the tests introduced in the following sections.

First derivative test

Assume $f(x)$ is differentiable in an interval J containing a and $f'(a) = 0$.

If $f'(x)$ changes sign from positive to negative as x increases through $x = a$ then $f(x)$ has a local maximum at a.

If $f'(x)$ changes sign from negative to positive as x increases through $x = a$ then $f(x)$ has a local minimum at $x = a$.



(a) Local maximum is attained at a.
Slope changes sign from $+$ to $-$.

(b) Local minimum is attained at a.
Slope changes sign from $-$ to $+$.

Figure 6.3: Slope changes sign through a local extremum.

Examples

Example 6.4 Find the local extrema of $y = x^3 - 9x^2 + 24x + 5$.

Solution.
$$y' = 3x^2 - 18x + 24 = 3(x-2)(x-4).$$

$$\therefore \quad y' = 0 \text{ when } x = 2, 4.$$

We can find the signs of y' in the intervals separated by $x = 2$ and $x = 4$ as shown in the table:

x	$x < 2$	$x = 2$	$2 < x < 4$	$x = 4$	$4 < x$
y'	$+$	0	$-$	0	$+$

Since y' changes sign from positive to negative as x increases through $x = 2$, we conclude that y has a local maximum at $x = 2$. Similarly, y has a local minimum at $x = 4$.

By direct substitution, we get

$$\text{local maximum} = 2^3 - 9 \cdot 2^2 + 24 \times 2 + 5 = 25 \text{ attained at } x = 2.$$

$$\text{local minimum} = 4^3 - 9 \times 4^2 + 24 \times 4 + 5 = 21 \text{ attained at } x = 4.$$

□

Second derivative test

Assume $f(x)$ is twice differentiable at $x = a$. The second derivative test is:

- $f'(a) = 0$ and $f''(a) < 0 \implies$ Local maximum is attained at $x = a$.
- $f'(a) = 0$ and $f''(a) > 0 \implies$ Local minimum is attained at $x = a$.
- $f'(a) = 0$ and $f''(a) = 0 \implies$ No conclusion can be made.

Examples

Example 6.5 Find the local maximum and local minimum of

$$y = x^3 - 9x^2 + 24x + 5$$

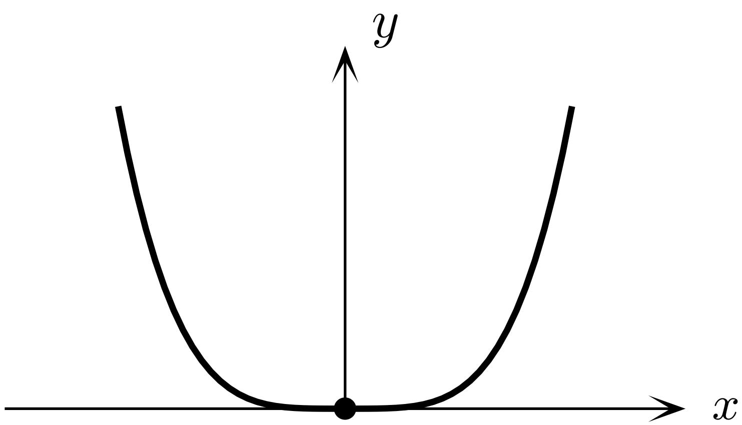
by the second derivative test. (Compare with the previous example.)

Solution. $y' = 3x^2 - 18x + 24 = 3(x-2)(x-4)$.

$\therefore \quad y'' = 6x - 18 = 6(x-3)$ and $y' = 0$ when $x = 2,\ 4$.
At $x = 2$, $y'' < 0$. $\therefore$ y has a local maximum at $x = 2$ and the local maximum is 25.
At $x = 4$, $y'' > 0$. $\therefore$ y has a local minimum at $x = 4$ and the local minimum is 21.
□

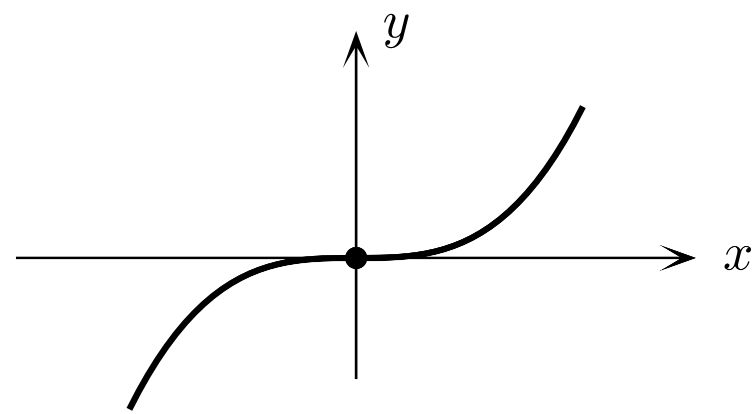
Remark

Fig. 6.4 illustrates that the test fails if we know only $f'(a) = f''(a) = 0$. Indeed, consider the following three cases: (i) $f(x) = x^4$ (ii) $f(x) = -x^4$ (iii) $f(x) = x^3$. In all these cases, we have $f'(0) = f''(0) = 0$. However, at $x = 0$, $f(x)$ has a local minimum in (i), a local maximum in (ii) and has no local extremum at all in (iii).



$f(x) = x^4$, $f'(0) = f''(0) = 0$.

A local minimum at $x = 0$.



$f(x) = x^3$, $f'(0) = f''(0) = 0$.

Neither a local minimum

nor a local maximum at $x = 0$.

Figure 6.4: Failing cases of the second derivative test.

Global maxima and minima

In practical problems, we often require to find the actual maximum or minimum value rather than the local maxima or minima.

Definition 6.4 Let a be a point in the domain J of a function $f(x)$.

- If $f(a) \geq f(x)$ for all x in J, we say that $f(a)$ is the *maximum value* of $f(x)$ on J and that $f(x)$ has its *maximum attained at* $x = a$. The maximum value is also called the *global maximum* or the *absolute maximum*, or simply the *maximum*.
- We have similar definitions for *minimum value* on J.
- The *extremum* means the maximum value or the minimum value.
- The function $f(x)$, whose extrema are being sought, is called the *objective function*.

The method of locating global extrema varies and is dependent on the objective function $f(x)$ and on the type of its domain. In the following sections, we consider the cases where (i) J is a closed and bounded interval of the form $[a, b]$, and (ii) J is an interval of any other types.

Global extrema on $[a, b]$

Suppose that the domain of a continuous function $f(x)$ is a closed and bounded interval $J = [a, b]$. The global extrema must exist in J. Furthermore, if $f(x)$ is differentiable[3] on $[a, b]$, an extremum must be attained at an endpoint or at a stationary point of J.

Theorem 6.3 *Let $f(x)$ be differentiable on its domain $J = [a, b]$ and $c \in J$. If $f(c)$ is an extremum of $f(x)$, then c is an endpoint of J or $f'(c) = 0$.*

Hence for differentiable functions on $[a, b]$, the extrema can be obtained by finding and comparing the function values at the endpoints and the stationary points.

[3] Differentiability at an endpoint of J means the associated one-sided limit of (5.3) exists.

Examples

Example 6.6 Find the extrema of the function $y = x^3 - 6x^2 + 9x,\ 0 \le x \le 5$.

Solution. Solving $y' = 3x^2 - 12x + 9 = 3(x-1)(x-3) = 0$, we get $x = 1$ or $x = 3$. Therefore the function y has stationary points at $x = 1$ and $x = 3$. We compute the y-values at the end-points and the stationary points:

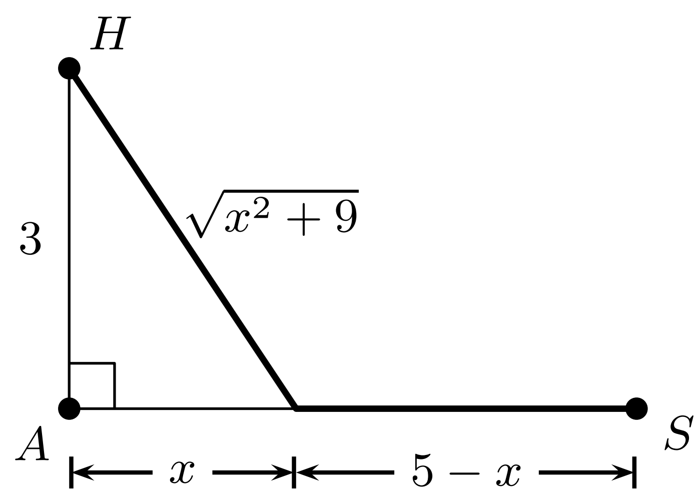
	x	y
Endpoint :	0	0
Endpoint :	5	20
Stationary point:	1	4
Stationary point:	3	0

Comparing y-values in the table, we see that

$$\max y = 20 \text{ attained at } x = 5; \quad \min y = 0 \text{ attained at } x = 0 \text{ and at } x = 3. \quad \square$$

Example 6.7 A lighthouse H is in the sea 3 km from a point A of a straight coastline which is perpendicular to the line joining A to H. There is a store S located 5 km down the coast from A. The lighthouse keeper can row his boat at 4 km per hour and he can walk at 6 km per hour. To what point of the shore should he row so as to reach the store in shortest time? What is this shortest time?

Solution.



If the lighthouse keeper lands x km from A, he must row for $\sqrt{x^2+9}$ km and walk for $(5-x)$ km. The total time required on this route is

$$T(x) = \frac{\sqrt{x^2+9}}{4} + \frac{5-x}{6}, \quad 0 \le x \le 5.$$

To minimize the time T, we differentiate T to get

$$T'(x) = \frac{x}{4\sqrt{x^2+9}} - \frac{1}{6}.$$

Solving $T'(x) = 0$ in $0 \le x \le 5$, we get $x = 6/\sqrt{5}$. Comparing the time T:

	x	T
Endpoint :	0	1.583
Endpoint :	5	1.458
Stationary point :	$6/\sqrt{5}$	1.392

we see that the shortest time is 1.392 hours when $x = 6/\sqrt{5}$. □

Global extrema on an interval of other types

Suppose that the domain of a differentiable function $f(x)$ is an interval J not of the form $[a, b]$. This includes the cases like $J = (a, \infty)$, $J = (-\infty, b]$ and $J = (-\infty, \infty)$. In each of these cases, the existence of a global extremum is not guaranteed. This can be seen from Fig. 6.5.

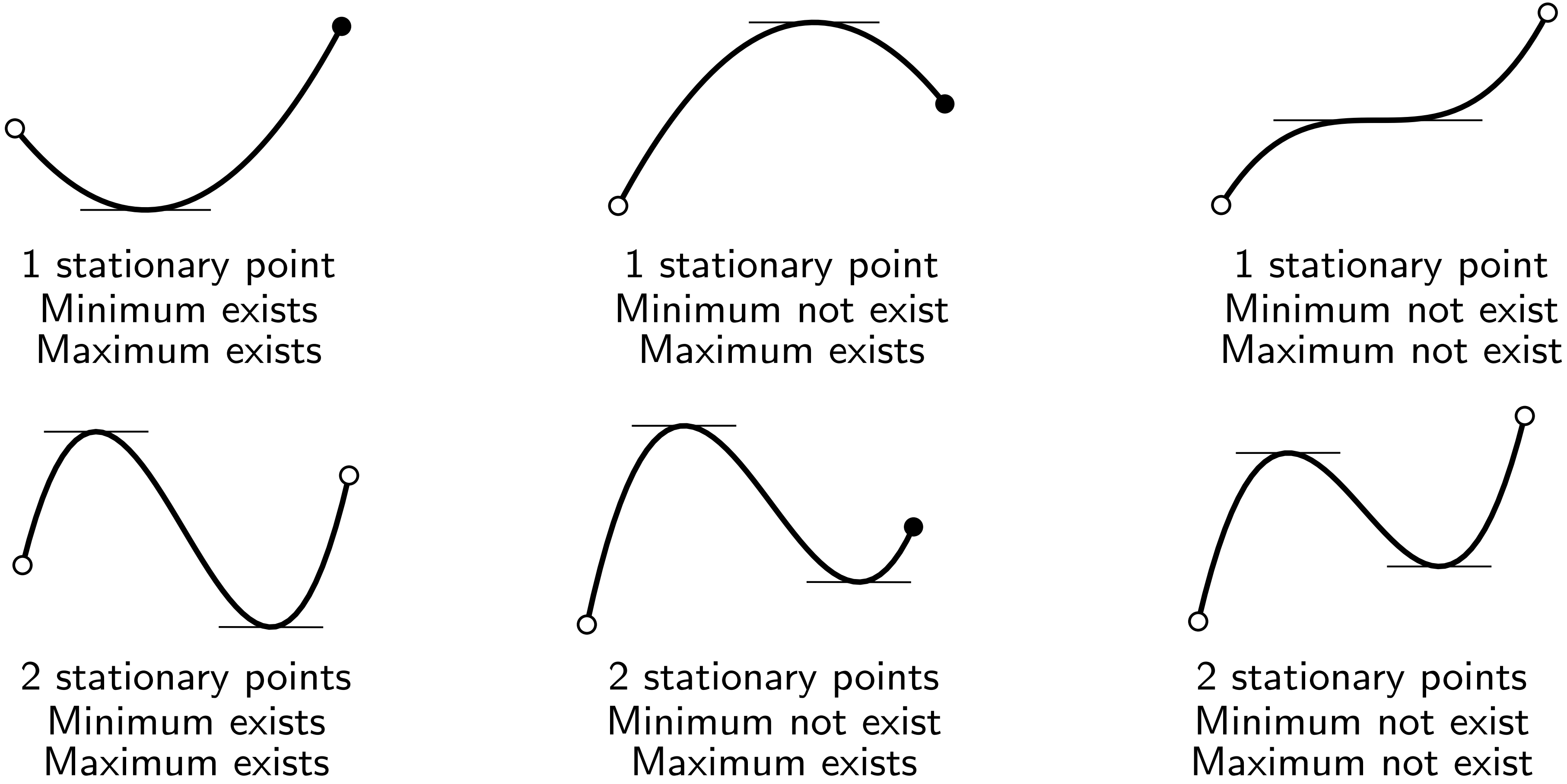


Figure 6.5: Global extrema may or may not exist.

If $f(x)$ is differentiable on the interval J, we can get the global extrema by simply comparing values of $f(x)$ at the stationary points and at the endpoints a and b. Though the values at a and b may not be defined, they can be replaced by their limiting values.
For instance, if $f(x)$ is not defined at the endpoint b of J where b is finite or $b = \infty$, we can use $\lim_{x \to b^-} f(x)$ or $\lim_{x \to \infty} f(x)$ to replace $f(b)$ in the comparison. If the largest value in the comparison table is attained in the domain J, this value is the global maximum. If the largest value is not attained in J, the global maximum does not exists. Similar conclusions are for the global minimum.

Example 6.8 The profit function of producing and selling x units of a kind of product is given by $P(x) = -x^3 + 9x^2 - 15x$. Assuming that x is a continuous variable and $0 \leq x < \infty$, find the maximum profit.

Solution. Being a polynomial, P is differentiable in $J = [0, \infty)$. We differentiate P to get $P' = -3x^2 + 18x - 15 = -3(x-1)(x-5)$. Solving $P' = 0$, we get the stationary points $x = 1$ and $x = 5$. We compute the limiting values of P at infinity and the P-values at $x = 0$ and at the stationary points:

	x	$P(x)$
Endpoint :	0	$P(0) = 0$
Endpoint :	∞	$\lim_{x\to\infty} P(x) = -\infty$
Stationary point :	1	$P(1) = -7$
Stationary point :	5	$P(5) = 25$

By comparison we see that the maximum profit is 25 attained at $x = 5$.

□

Case: exactly one stationary point

When there is only one stationary point and the domain is an interval, there is a simple rule to get the global extremum.

Theorem 6.4 *Let $f(x)$ be a function defined on an interval J. Assume that $f(x)$ is differentiable on J and has exactly one stationary point a in J. Then,*

- *$f(a)$ is a local minimum $\Longrightarrow$ $f(a)$ is the global minimum.*
- *$f(a)$ is a local maximum $\Longrightarrow$ $f(a)$ is the global maximum.*

The theorem is obviously true for the case $J = (a, b)$ as a global extremum must be a local one. It is true for an interval J of any type and the proof is omitted here. In many practical problems, the objective functions satisfy the assumption of the theorem. For such a problem, the global extremum is the same as the local extremum if the objective function has only one stationary point in J.

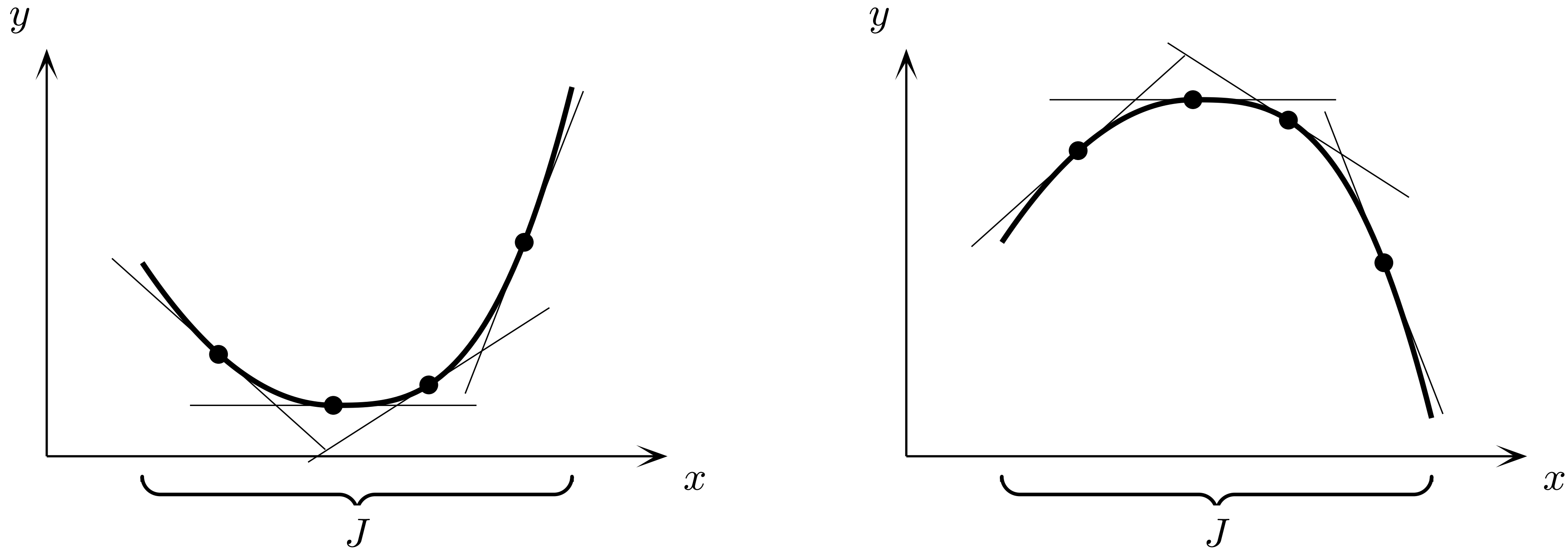
Example 6.9 Find the smallest sum of two positive numbers if their product is 16.

Solution. Let x and y be positive numbers satisfying $xy = 16$. Let $S = x + y$ with $x > 0$ and $y > 0$. Since $y = 16/x$, we have to minimize $S(x) = x + 16/x$ over the domain $J = (0, \infty)$.
We first differentiate S with respect to x and get $S' = 1 - 16/x^2$ and $S'' = 32/x^3$.
Solving $S' = 0$, we have $x = \pm 4$. We reject the solution $x = -4$ as it is not in the domain J. Therefore there is only one stationary point $x = 4$.
Since $S'(4) = 0$ and $S''(4) > 0$, S has a local minimum at the point $x = 4$. By Theorem 6.4, S has a global minimum at the point $x = 4$ and the minimum value is $4 + 4 = 8$. □

Concavity and inflection points

Definition 6.5 Let $f(x)$ be differentiable on an open interval J.

- $f(x)$ is said to be *concave up* on J if for any point a in J, the graph of $f(x)$ near $x = a$ lies above its tangent at a. Similarly,
- $f(x)$ is said to be *concave down* on J if its graph lies below its tangents.



(a) Concave up on J: Curve above its tangents.

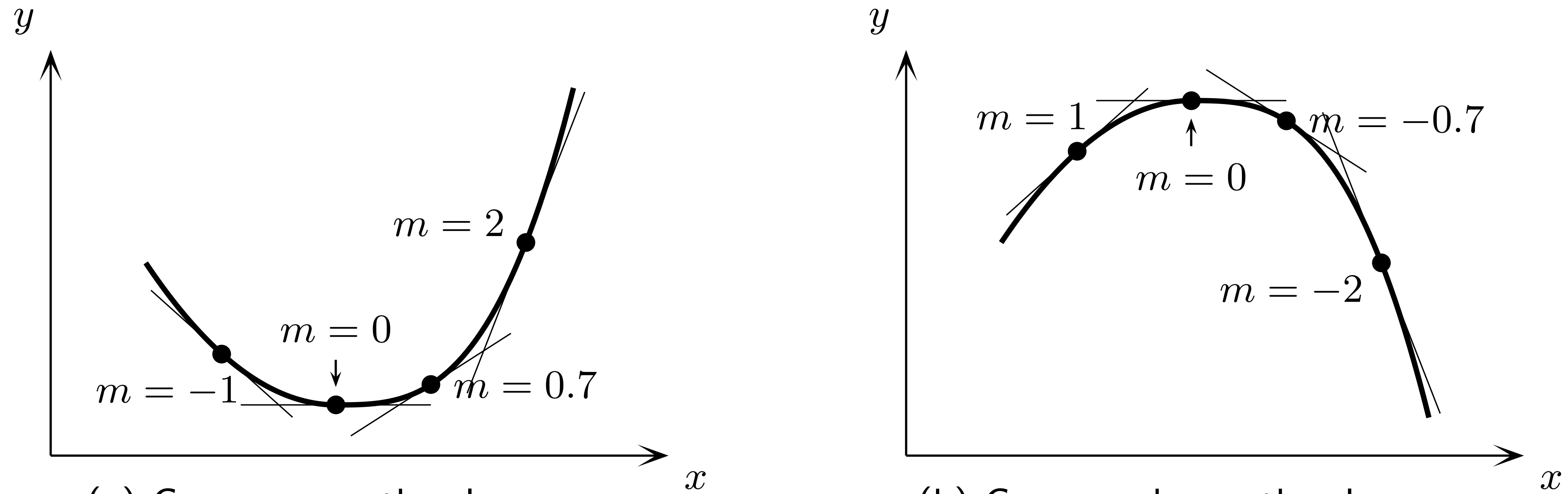
(b) Concave down on J: Curve below its tangents.

Figure 6.6: Meaning of concave up and concave down over an interval.

Concavity

Theorem 6.5 *Let J be an open interval.*

- *$f(x)$ is concave up on J iff $f'(x)$ is increasing on J.*
- *$f(x)$ is concave down on J iff $f'(x)$ is decreasing on J.*



(a) Concave up: the slope m is increasing.

(b) Concave down: the slope m is decreasing.

Figure 6.7: Concavity of $f(x)$ as monotonicity of $f'(x)$.

Furthermore, if $f''(x)$ exists, we have the following theorem.

Theorem 6.6 *Let $f(x)$ be twice differentiable on an open interval J.*

- $f(x)$ *is concave up on* J *iff* $f''(x) > 0$ *on* J.
- $f(x)$ *is concave down on* J *iff* $f''(x) < 0$ *on* J.

Summary

We summarize the above definitions and theorems in the following table.

	$f(x)$ is concave up on J	$f(x)$ is concave down on J
Definition:	Graph above its tangents on J	Graph below its tangents on J
Theorem 6.5:	$f'(x)$ is increasing on J	$f'(x)$ is decreasing on J
Theorem 6.6:	$f''(x) > 0$ on J	$f''(x) < 0$ on J

Example

Example 6.10 Find the open intervals in which the function $f(x) = x^3 + 3x^2 - 4x + 2$ is concave up. Also find the open intervals in which $f(x)$ is concave down.

Solution. $f(x) = x^3 + 3x^2 - 4x + 2$. $\therefore$ $f'(x) = 3x^2 + 6x - 4$ and hence $f''(x) = 6x + 6$. In the interval $(-1, \infty)$, we have $f''(x) > 0$ and $f(x)$ is concave up. In the interval $(-\infty, -1)$, we have $f''(x) < 0$ and $f(x)$ is concave down. $\square$

Inflection points

Definition 6.6 Let $f(x)$ be differentiable on an open interval J and a be a point in J. We say that $(a, f(a))$ is an *inflection point* of $f(x)$ if $f(x)$ is concave up on one side and concave down on the other side of a.

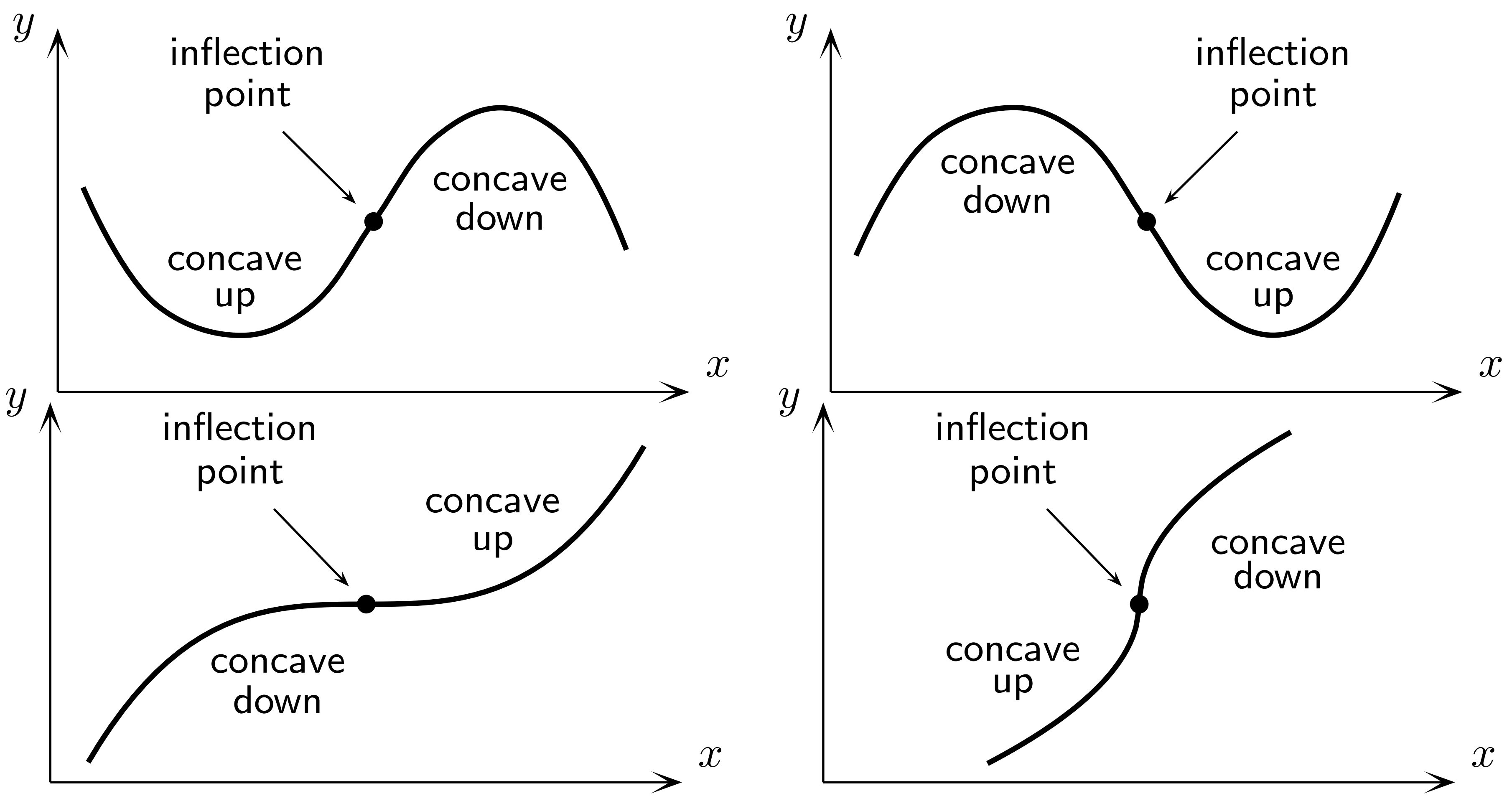


Figure 6.8: Meaning of inflection point.

By definition, if $(a, f(a))$ is an inflection point of the function $f(x)$, then $f''(x)$ is positive on one side of a and $f''(x) < 0$ on the other side. Therefore if $f''(x)$ is continuous, we have $f''(a) = 0$.
Conversely, the condition $f''(a) = 0$ is not sufficient to deduce that $(a, f(a))$ is an inflection point. Take for example the smooth function $f(x) = x^4$. Clearly this $f(x)$ has no inflection point but $f''(0) = 0$.
However, the following theorem gives a sufficient condition to identify an inflection point. This theorem is useful when the third order derivative can be found without difficulty.

Theorem 6.7 *Suppose that $f'''(x)$ exists over an open interval containing a. If $f''(a) = 0$ but $f'''(a) \neq 0$ then $(a, f(a))$ is an inflection point of $f(x)$.*

Example

Example 6.11 Find the inflection points (if any) of the function

$$f(x) = x^3 + 3x^2 - 4x + 2.$$

Solution. Let $y = x^3 + 3x^2 - 4x + 2$. As in the previous example, we have $y' = 3x^2 + 6x - 4$ and $y'' = 6x + 6$. To find the inflection point, we solve for x the equation $y'' = 0$, i.e. $6x + 6 = 0$. The only solution is $x = -1$ at which $y = 8$. Since $y''' = 6 \neq 0$, the function $f(x)$ has an inflection point $(-1, 8)$. □

Another example Using CoCalc to find inflection points of a rational function

https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/rational-function-inflection-points.pdf

Curve sketching
The general procedure

1. Find the domain of the function if it is not given. Observe if there is any symmetry. If $f(x)$ is even $(f(-x) = f(x))$, the graph is symmetrical about the y-axis. If it is odd $(f(-x) = -f(x))$, the graph is symmetrical about the origin.

2. Find the first derivative $f'(x)$. Hence obtain the points where $f'(x)$ is undefined and the stationary points where $f'(x) = 0$. Find also the local extrema and the intervals of monotonicity. Find the points on the graph at the stationary points.

3. Find the second derivative $f''(x)$ (if not too difficult), hence obtain the intervals of concavity and the inflection points (if any).

4. Summarize the results of No. 2 to 3 in a table.

5. Find the asymptotes (if any).

6. If necessary, find additional points that help to fix the position of the graph.

7. If necessary, find the limiting value of $f(x)$ like $\lim_{x\to\infty} f(x)$.

8. Based on No. 4, 5, 6 and 7, select good x-range and y-range to include the main characteristics of the graph. Sketch the graph.

Examples of curve sketching

Example 6.12 Sketch the curve $y = x^3 - 3x$.

Solution

1. The domain is $\mathbb{R}$. The function is odd as $(-x)^3 - 3(-x) = -(x^3 - 3x)$.
2. First derivative: $y' = 3x^2 - 3 = 3(x^2 - 1)$. Therefore there are two stationary points $x = -1$ and $x = 1$. These stationary points subdivide the real axis into three intervals:

$$(-\infty, -1), \quad (-1, 1), \quad (1, \infty)$$

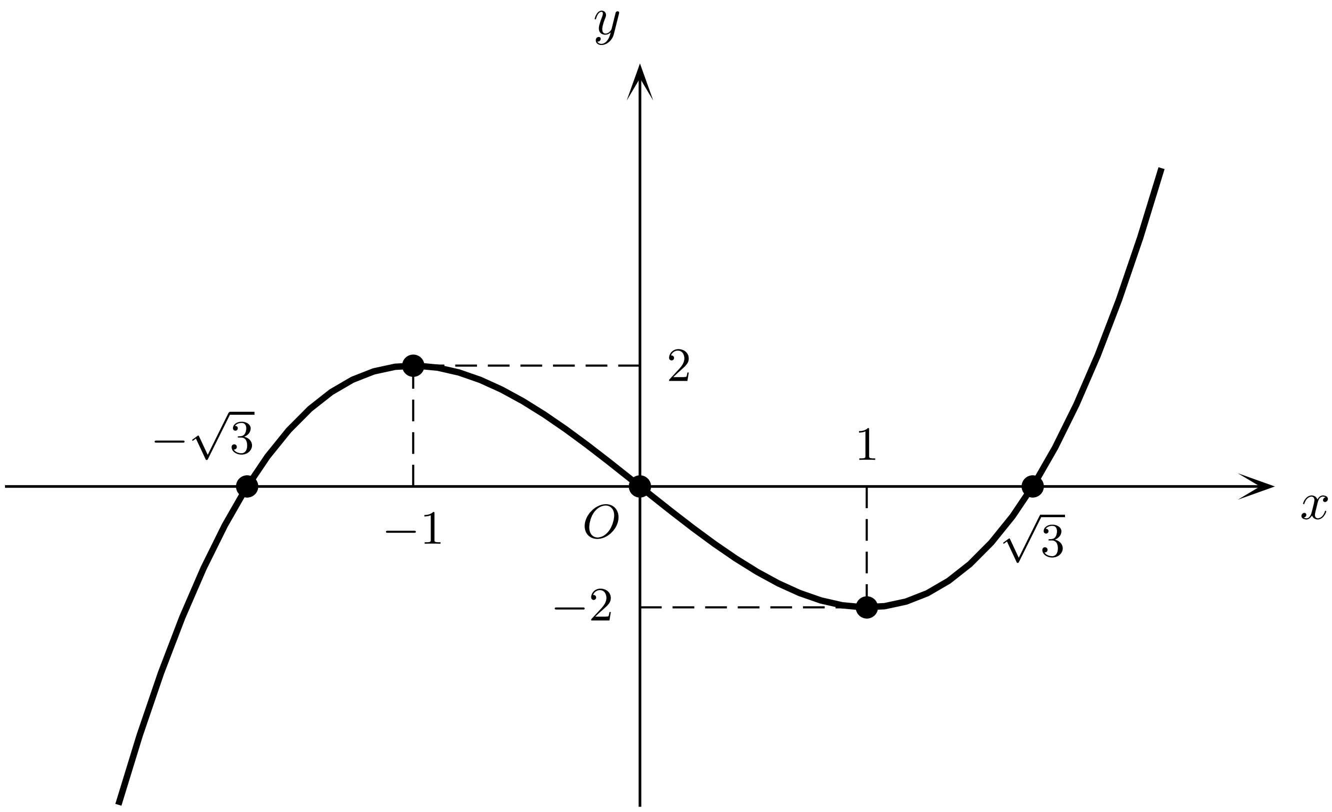
   We get: $y' > 0$ (y increasing) in $(-\infty, -1)$ and in $(1, \infty)$; $y' < 0$ (y decreasing) in $(-1, 1)$. By the first derivative test, the function has a local maximum at $x = -1$ and has a local minimum at $x = 1$. The extremum points on the graph are $(-1, 2)$ and $(1, -2)$.
3. Second derivative: $y'' = 6x$. Therefore, $y'' < 0$ (y concave down) if $x < 0$; $y'' > 0$ (y concave up) if $x > 0$. It follows that the origin $(0, 0)$ is an inflection point.

4. No. 3 and 4 are summarized as (see footnote[4] for the meanings of symbols):

x	$x < -1$	-1	$-1 < x < 0$	0	$0 < x < 1$	1	$1 < x$
y'	$+$	0	$-$	$-$	$-$	0	$+$
y''	$-$	$-$	$-$	0	$+$	$+$	$+$
y	$\nearrow$	loc. max	$\searrow$	$\searrow$	$\searrow$	loc. min	$\nearrow$
	concave down			i.p.	concave up		
		2		0		-2	

5. Being a polynomial, the graph has no asymptote.
6. The x-intercept is found by solving $x^3 - 3x = 0$. The solutions are $x = 0$ and $x = \pm\sqrt{3}$. These give three points $(0, 0)$, $(-\sqrt{3}, 0)$ and $(\sqrt{3}, 0)$ on the graph.
7. $\lim_{x \to -\infty} (x^3 - 3x) = -\infty$ and $\lim_{x \to \infty} (x^3 - 3x) = \infty$.
8. Based on No. 1, 4, 5, 6 and 7, the graph is sketched as follows:

[4] $\nearrow$ means increasing; $\searrow$ means decreasing; i.p. means inflection point.



y
2
$-\sqrt{3}$
1
x
-1
O
$\sqrt{3}$
-2

Example 6.13 Sketch the curve $y = 4\sqrt{x} - x$.

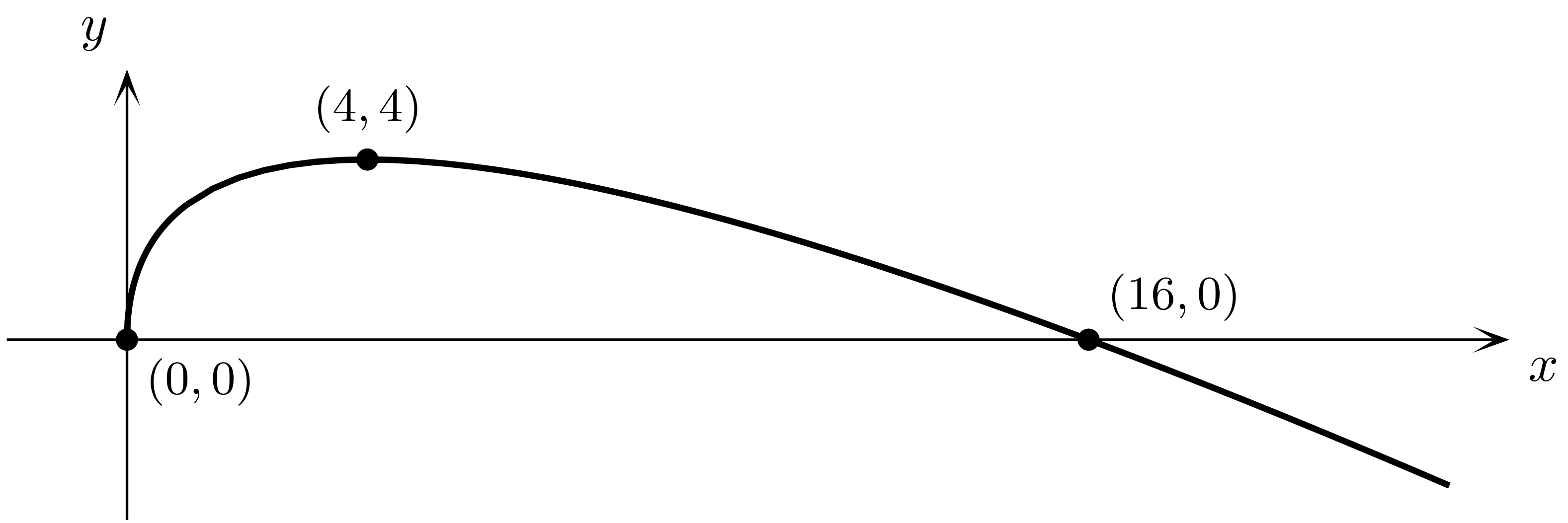
Solution

1. The domain is $x \geq 0$.
2. $y' = 2/\sqrt{x} - 1$. Solving $y' = 0$, we get $x = 4$ and $y = 4$. Also, we get $y' > 0$ if $x < 4$; and $y' < 0$ if $x > 4$. By the first derivative test, $(4, 4)$ is the point at which the local maximum is attained.
3. $y'' = -1/x^{3/2}$. Therefore the function is concave down over $x > 0$ and has no inflection point.
4.

x	$0 < x < 4$	4	$4 < x < 16$	16	$16 < x$
y'	$+$	0	$-$	$-$	$-$
y''	$-$	$-$	$-$	$-$	$-$
y	$\nearrow$	loc. max	$\searrow$	$\searrow$	$\searrow$
	concave down				
		4		0	

5. $y=0$ at $x=0$. $y\to -\infty$ as $x\to\infty$. Also $y'\to\infty$ as $x\to 0^+$.

6. The graph is sketched as follows:



Example 6.14 Sketch the curve $y = \dfrac{x}{(x-1)^2}$, $x \neq 1$.

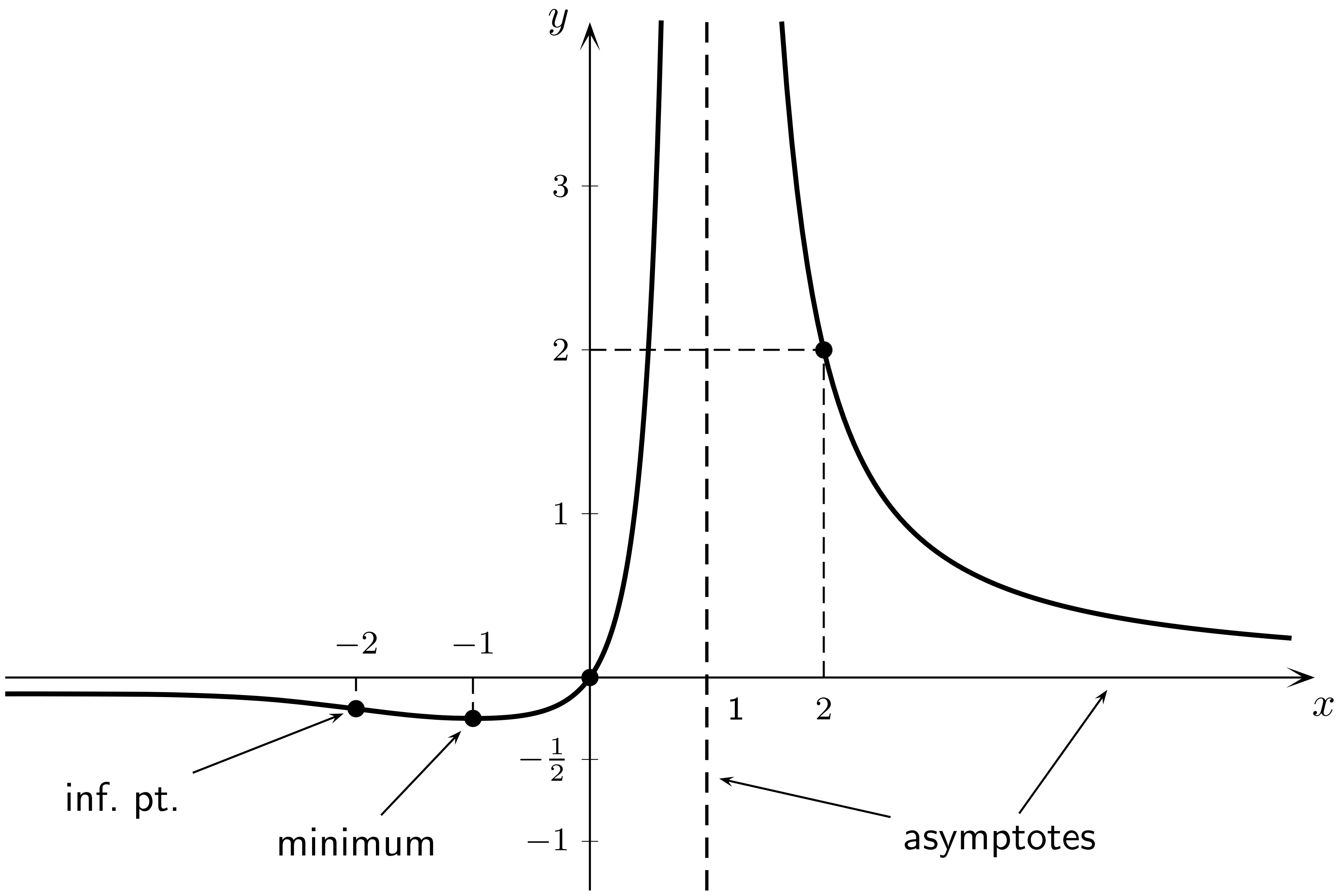
Solution

1. The domain is $(-\infty, 1) \cup (1, \infty)$.
2. $y' = \dfrac{-(x+1)}{(x-1)^3}$ for all $x \neq 1$. Thus the curve has a stationary point $x = -1$ at which $y = -1/4$. The sign of y' is the same as that of $-(x+1)(x-1)$ and is shown in the table below.
3. $y'' = -\dfrac{1}{(x-1)^3} - \dfrac{-3(x+1)}{(x-1)^4} = \dfrac{2(x+2)}{(x-1)^4}$.
   $\therefore$ at $x = -2$ we have $y'' = 0$. The sign of y'' is the same as that of $x + 2$ and is shown in the table below.

4.

x	$-\infty < x < -2$	-2	$-2 < x < 1$	-1	$-1 < x < 1$	1	$1 < x < \infty$
y'	$-$	$-$	$-$	0	$+$		$-$
y''	$-$	0	$+$	$+$	$+$		$+$
y	$\searrow$	$\searrow$	$\searrow$	loc. min	$\nearrow$		$\searrow$
	conc. down	i. p.	concave up				conc. up
		$-2/9$		$-1/4$			

5. The asymptotes are the lines $y = 0$ and $x = 1$.
6. $y = 0$ if and only if $x = 0$. Also $y = 2$ at $x = 2$.
7. $\lim_{x\to-\infty} y = 0$, $\lim_{x\to 1^-} y = \infty$, $\lim_{x\to 1^+} y = +\infty$, $\lim_{x\to\infty} y = 0$.
8. The graph is sketched as follows:


y
3
2
1
−2
−1
1
2
x
−½
−1
inf. pt.
minimum
asymptotes

Supplementary Notes

To find the closest point from an ellipse to a given point
https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/supplementary04.pdf

Two Roots (CoCalc) https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/supplementary04_CoCalcJupyter.pdf

Four Roots (CoCalc) https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/supplementary04a_CoCalcJupyter.pdf

Three Roots but still two extrema (CoCalc)

https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/supplementary04b_CoCalcJupyter.pdf

Visualize the distribution of locations giving 4 roots (green), 2 roots (red), and 3 roots (the two blue dots, right on the EVOLUTE) (CoCalc)

https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/supplementary04c_CoCalcJupyter.pdf

Increments and differentials

Consider a differentiable function $y = f(x)$ with independent variable x and dependent variable y .

Definition 6.7 Let x be fixed and let there be a small increment in x denoted by Δx. We define:

- $\mathrm{d}x = \Delta x$ (called the *differential* of x and the *increment* in x)
- $\mathrm{d}y = f'(x)\,\mathrm{d}x$ (called the *differential* of y)
- $\Delta y = f(x + \Delta x) - f(x)$ (called the *increment* in y)

Examples

Example 6.15 If $y = x^2 + 3x + 1$, find $\mathrm{d}y$.

Solution. $\mathrm{d}y = (2x + 3)\,\mathrm{d}x$. □

Example 6.16 For the function $y = x^3$, find $\mathrm{d}y$ and Δy with $x = 1$ and $\Delta x = 0.05$.

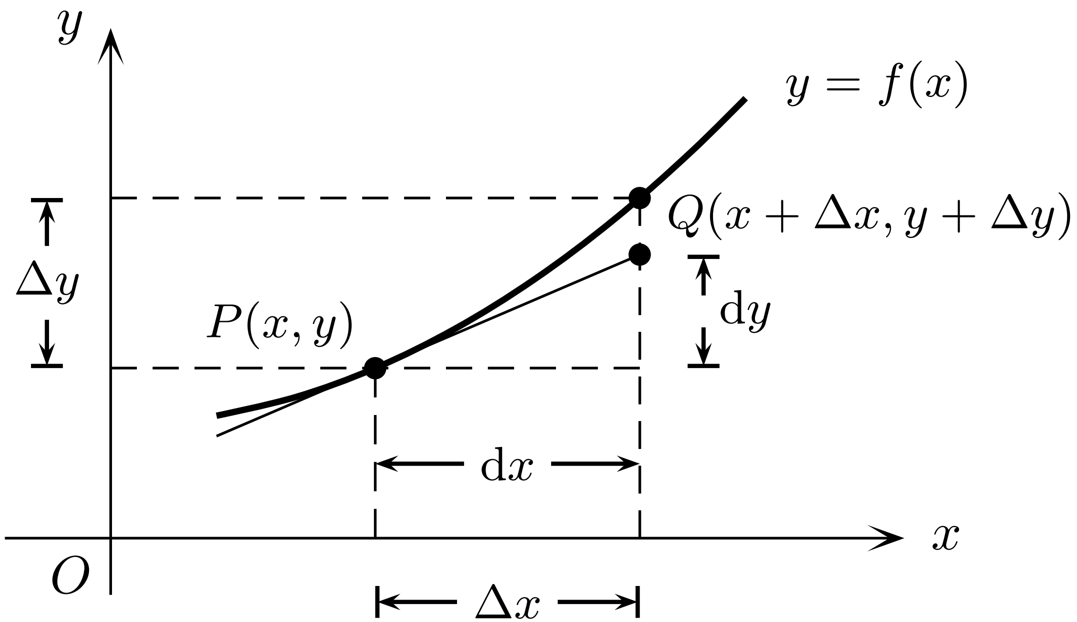
Solution. $\mathrm{d}y = 3x^2\,\mathrm{d}x$.
Therefore with $x = 1$ and $\mathrm{d}x = \Delta x = 0.05$, we get

$$\mathrm{d}y = 3 \times 1^2 \times 0.05 = 0.15 \quad \text{and} \quad \Delta y = 1.05^3 - 1^3 = 0.158.$$

□

Differentials and approximations

The meaning of $\mathrm{d}x$, $\mathrm{d}y$, Δx and Δy are illustrated in the following diagram. Note that $\mathrm{d}y$ and Δy are quantities dependent on the function $f(x)$ and the values of x and Δx.



From the above diagram, we can see the truth of the following theorem:

Theorem 6.8 $\Delta y \approx \mathrm{d}y$ *if* Δx *is small.*

For a differentiable function $y = f(x)$, this theorem allows us to get an estimate of Δy using the value $\mathrm{d}y$ when Δx is small.

Example

Example 6.17 For the function $y = f(x)$ with $f(1) = a$ and $f'(1) = m$, estimate the value of $f(1.2)$.

Solution. Taking $x = 1$ and $\mathrm{d}x = \Delta x = 0.2$, we get
$\Delta y \approx \mathrm{d}y = f'(1) \times \Delta x = (0.2)m, \quad \therefore \; f(1.2) = f(1) + \Delta y \approx a + (0.2)m.$ □

Example 6.18 For the function $f(x) = \sqrt{x}$ with $f(64) = 8$, find $f'(64)$ and hence estimate the value of $f(63)$ using differentials.

Solution. $f'(x) = \dfrac{1}{2\sqrt{x}}$. Taking $x = 64$ and $\mathrm{d}x = \Delta x = -1$, we get

$$\Delta y \approx \mathrm{d}y = f'(64) \times \mathrm{d}x = -1/16,$$

$$\therefore \ f(63) = f(64) + \Delta y \approx 8 - \tfrac{1}{16} = 7.9375.$$

□

L'Hôpital's rule for finding limits

In Chapter 2, we learned the formula

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \frac{A}{B} \tag{6.1}$$

where A and B are the limits

$$A = \lim_{x \to a} f(x) \quad \text{and} \quad B = \lim_{x \to a} g(x). \tag{6.2}$$

If $B \neq 0$, the formula (6.1) works fine and gives a finite value. If the limit B is zero and A is nonzero, A/B is ∞ or $-\infty$. What happens when both A and B are zero? In this case the following l'Hôpital's rule may help.

Theorem 6.9 (l'Hôpital's rule for type $0/0$) *If $f(x)$ and $g(x)$ are differentiable over an open interval containing a, and if A and B defined in (6.2) are zero, then*

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

if the latter limit exists or is infinite.

Example

Example 6.19 Find $\lim_{x\to 0} \dfrac{\sin x}{x}$ by l'Hôpital's rule.

Solution. Since $\lim_{x\to 0} \sin x = 0$ and $\lim_{x\to 0} x = 0$, the required limit is of type $0/0$. By l'Hôpital's rule,

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

Example 6.20 Find $\lim_{x\to 1} \frac{x^3-1}{x-1}$ by l'Hôpital's rule.

Solution. Since $\lim_{x\to 1}(x^3-1) = 0$ and $\lim_{x\to 1}(x-1) = 0$, the required limit is of type $0/0$. By l'Hôpital's rule,

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

□

Example 6.21 Find $\lim_{x \to 0} \frac{x - \sin x}{x^3 - x^2}$ by l'Hôpital's rule.

Solution. Here we apply l'Hôpital's rule twice to get the answer.

$$
\begin{aligned}
& \lim_{x \to 0} \frac{x - \sin x}{x^3 - x^2} && \left(\text{type } \frac{0}{0}\right) \\
= & \lim_{x \to 0} \frac{1 - \cos x}{3x^2 - 2x} && \left(\text{type } \frac{0}{0}\right) \\
= & \lim_{x \to 0} \frac{\sin x}{6x - 2} && \left(\text{type } \frac{0}{\text{nonzero}}\right) \\
= & \ 0.
\end{aligned}
$$

□

Another form of l'Hôpital's rule

Theorem 6.10 (l'Hôpital's rule for type ∞/∞) *If $f(x)$ and $g(x)$ are differentiable over an open interval containing a, and if A and B defined in (6.2) are infinite (i.e. ∞ or $-\infty$, then*

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

if the latter limit exists or is infinite.

Example

Example 6.22 Find $\lim\limits_{x\to\infty} \dfrac{e^x}{x^3}$ by l'Hôpital's rule.

Solution. Here we apply l'Hôpital's rule three times to get the answer.

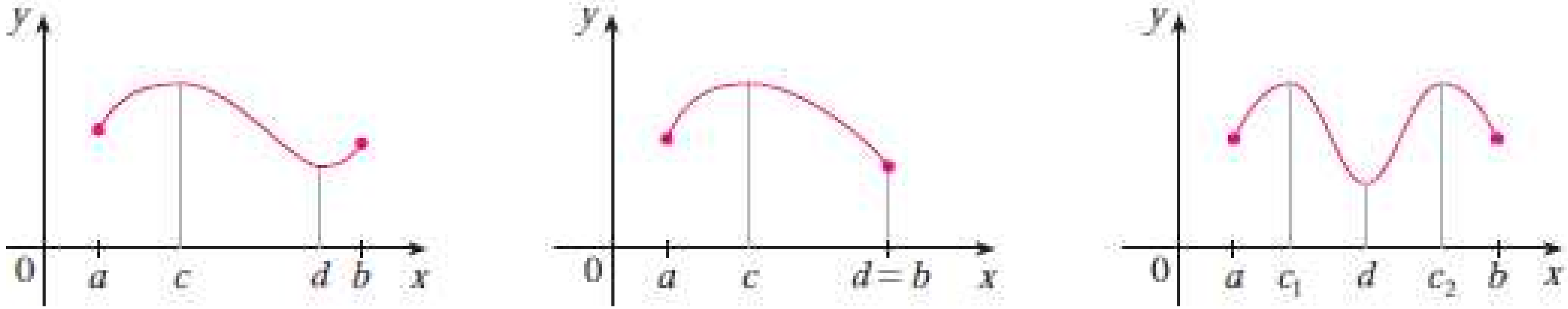
$$
\begin{aligned}
&\lim_{x\to\infty} \frac{e^x}{x^3} && \left(\text{type } \frac{\infty}{\infty}\right) \\
&= \lim_{x\to\infty} \frac{e^x}{3x^2} && \left(\text{type } \frac{\infty}{\infty}\right) \\
&= \lim_{x\to\infty} \frac{e^x}{6x} && \left(\text{type } \frac{\infty}{\infty}\right) \\
&= \lim_{x\to\infty} \frac{e^x}{6} && \left(\text{type } \frac{\infty}{\text{finite}}\right) \\
&= \infty.
\end{aligned}
$$

□

Mean Value Theorem

Extreme Value Theorem

If f is continuous on a closed interval $[a, b]$, then f attains an absolute maximum value $f(c)$ and an absolute minimum value $f(d)$ at some numbers c and d in $[a, b]$.



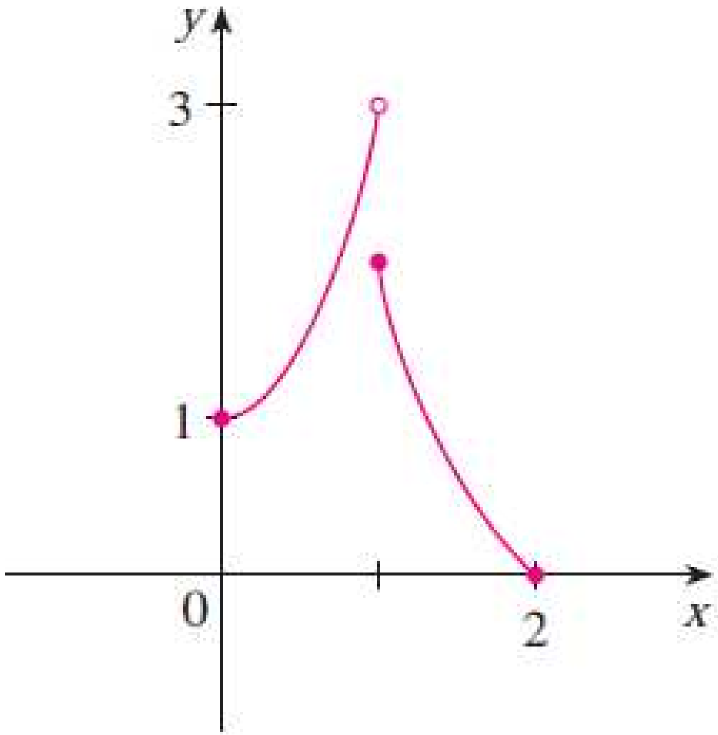


FIGURE 8

This function has minimum value $f(2) = 0$, but no maximum value.

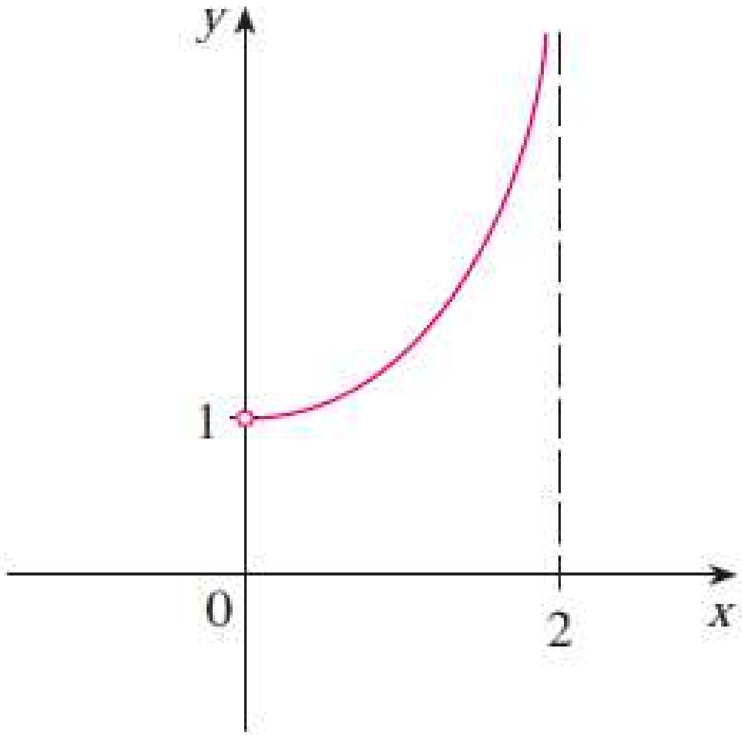


FIGURE 9

This continuous function g has no maximum or minimum.

Fermat's Theorem

If f has a local maximum or minimum at c, and if $f'(c)$ exists, then $f'(c) = 0$.

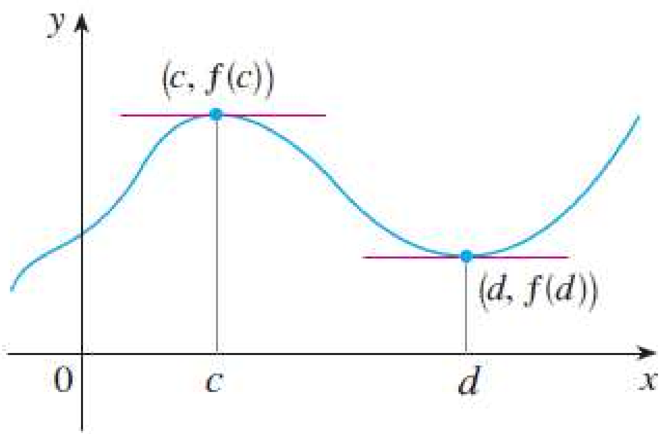


FIGURE 10

Fermat

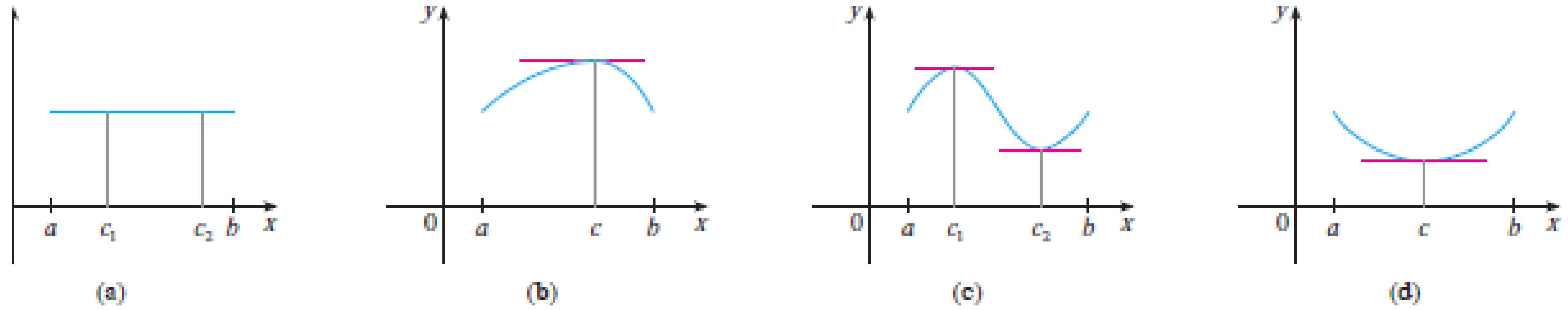
Fermat's Theorem is named after Pierre Fermat (1601–1665), a French lawyer who took up mathematics as a hobby. Despite his amateur status, Fermat was one of the two inventors of analytic geometry (Descartes was the other). His methods for finding tangents to curves and maximum and minimum values (before the invention of limits and derivatives) made him a forerunner of Newton in the creation of differential calculus.

Rolle's Theorem

Let f be a function that satisfies the following three properties

(i) f is continuous on the closed interval $[a, b]$.
(ii) f is differentiable on the open interval (a, b).
(iii) $f(a) = f(b)$.

Then, there is a number c in (a, b) such that $f'(c) = 0$.



Mean Value Theorem

Let f be a function that satisfies the following hypotheses

(i) f is continuous on the closed interval $[a, b]$.
(ii) f is differentiable on the open interval (a, b).

Then, there is a number c in (a, b) such that $f'(c) = \frac{f(b)-f(a)}{b-a}$, or equivalently, $f(b) - f(a) = f'(c)(b - a)$.

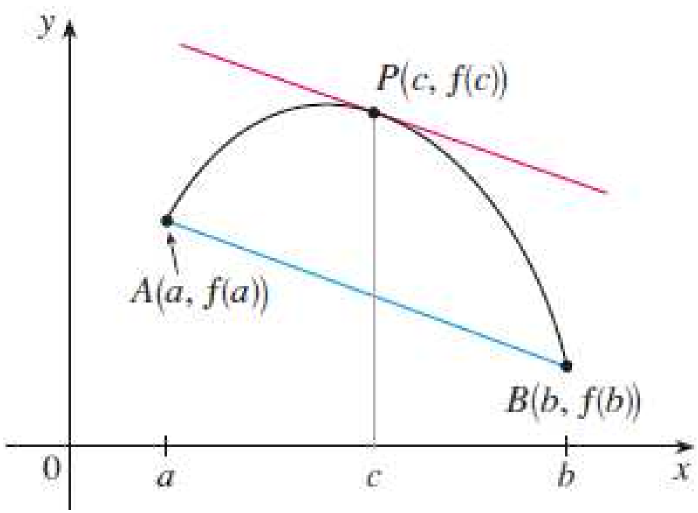


FIGURE 3

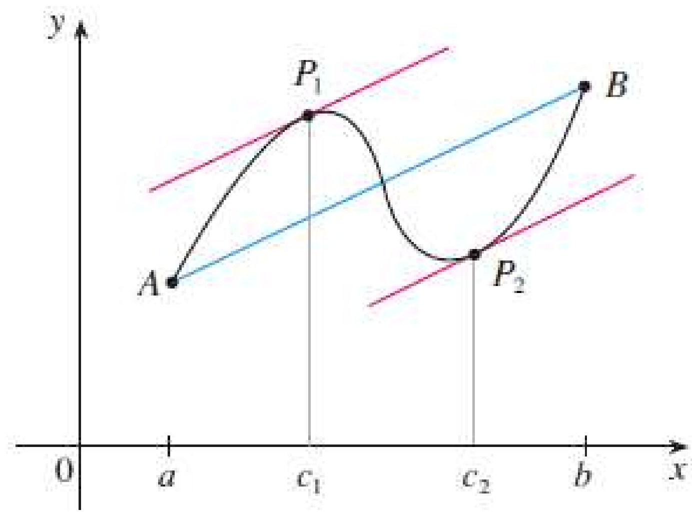


FIGURE 4

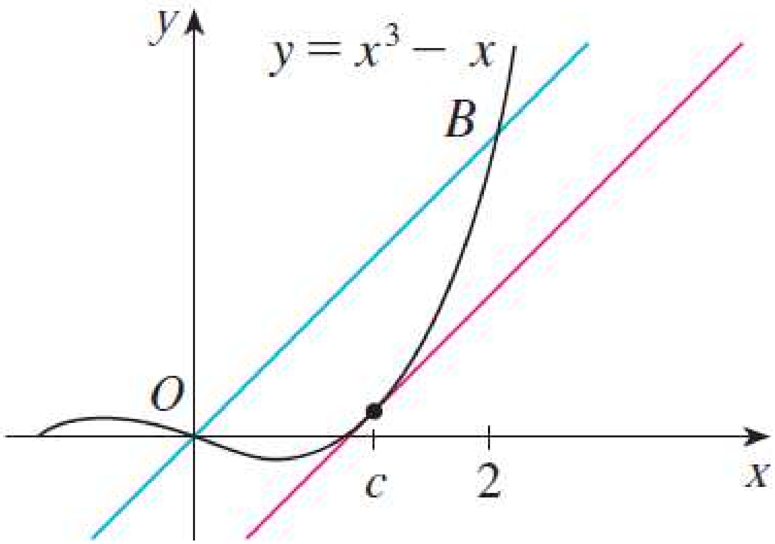
Example: Consider $f(x) = x^3 - x$ defined on the interval $[0, 2]$. Since f is a polynomial, it is continuous and differentiable. Thus, by Mean Value Theorem, there is a number c in $[0, 2]$ such that

$$f(2) - f(0) = f'(c)(2 - 0)$$

Now $f(2) = 6$, $f(0) = 0$, and $f'(x) = 3x^2 - 1$, so this equations becomes

$$6 = (3c^2 - 1)2 = 6c^2 - 2$$

which gives $c^2 = \frac{4}{3}$, that is $c = \pm\frac{2}{\sqrt{3}}$. But c must line in $(0, 2)$, so $c = \frac{2}{\sqrt{3}}$.



Example: Suppose that $f(0) = -3$ and $f'(x) \le 5$ for all values of x. How large can $f(2)$ possibly be?

Solution: We are given that f is differentiable (and therefore continuous) everywhere. In particular, we can apply Mean Value Theorem on the interval $[0, 2]$. There exists a number c such that

$$f(2) - f(0) = f'(c)(2 - 0)$$

so

$$f(2) = f(0) + 2f'(c) = -3 + 2f'(c).$$

We are given that $f'(x) \le 5$ for all x, so in particular we know $f'(c) \le 5$. Multiplying both sides of this inequality by 2, we have $2f'(c) \le 10$, so

$$f(2) = = -3 + 2f'(c) \le -3 + 10 = 7.$$

The largest possible value for $f(2)$ is 7.

Indefinite Integral

Definition of indefinite integrals

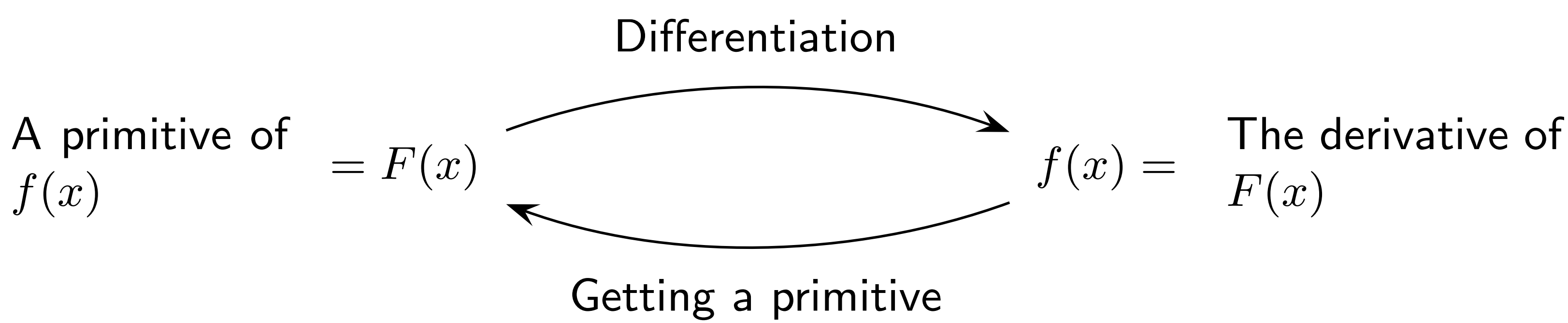
We know that if we differentiate a function $F(x)$ we get a function $f(x)$:

$$\frac{\mathrm{d}}{\mathrm{d}x}F(x) = f(x) \tag{8.1}$$

called the derivative of $F(x)$. In this chapter we start from a function $f(x)$ and proceed reversely to find an $F(x)$ satisfying (8.1).

Definition 8.1 A function $F(x)$ is called a *primitive* or *antiderivative* of a function $f(x)$ if (8.1) is true.

The definition is illustrated in the following diagram and we see that finding a primitive is just the reverse process of differentiation.



If C is any constant, then $\frac{\mathrm{d}}{\mathrm{d}x}C = 0$. It follows that if $F(x)$ is a primitive of $f(x)$ then $F(x) + C$ is also a primitive. Therefore, if $F(x)$ is a primitive of $f(x)$, we have

$$\int f(x)\,\mathrm{d}x = F(x) + C$$

where C is an arbitrary constant. Note that if, instead of x, we use another symbol, say u, for the independent variable, the above equation can be written

$$\int f(u)\,\mathrm{d}u = F(u) + C.$$

Examples

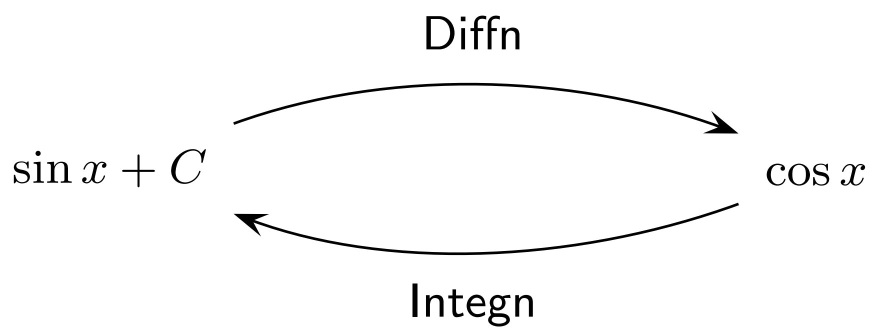
Example 8.1 Find the indefinite integral $\int \cos x \, \mathrm{d}x$.

Solution.

Since $\frac{\mathrm{d}}{\mathrm{d}x} \sin x = \cos x$, we can write at once

$$\int \cos x \, \mathrm{d}x = \sin x + C$$

where C is an arbitrary constant.



□

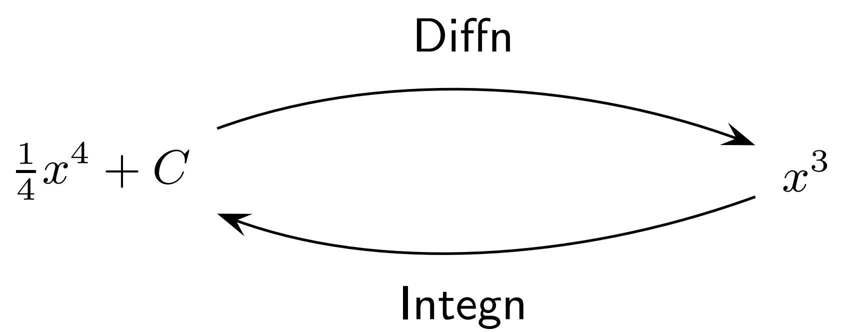
Example 8.2 Find the indefinite integral $\int x^3\,\mathrm{d}x$.

Solution.

Since $\dfrac{\mathrm{d}}{\mathrm{d}x}x^4 = 4x^3$ and hence $\dfrac{\mathrm{d}}{\mathrm{d}x}\left(\dfrac{x^4}{4}\right) = x^3$, we can write at once

$$\int x^3\,\mathrm{d}x = \frac{x^4}{4} + C$$

where C is an arbitrary constant.



A Table of Integrals

	$f(x)$	$\int f(x)\,\mathrm{d}x$
1	$x^n \quad (n \neq -1)$	$\dfrac{x^{n+1}}{n+1} + C$
2	$\dfrac{1}{x} \quad (x \neq 0)$	$\ln\lvert x\rvert + C$
3	e^x	$\mathrm{e}^x + C$
4	$\sin x$	$-\cos x + C$
5	$\cos x$	$\sin x + C$
6	$\tan x$	$-\ln\lvert\cos x\rvert = \ln\lvert\sec x\rvert + C$
7	$\cot x$	$\ln\lvert\sin x\rvert + C$
8	$\sec x$	$\ln\lvert\sec x + \tan x\rvert + C$
9	$\csc x$	$-\ln\lvert\csc x + \cot x\rvert + C$
10	$\sec x\ \tan x$	$\sec x + C$
11	$\csc x\ \cot x$	$-\csc x + C$

	$f(x)$	$\int f(x)\,\mathrm{d}x$
12	$\dfrac{1}{a^2+x^2} \quad (a \neq 0)$	$\dfrac{1}{a}\tan^{-1}\left(\dfrac{x}{a}\right)+C$
13	$\dfrac{1}{a^2-x^2} \quad (a \neq 0)$	$\dfrac{1}{2a}\ln\left\|\dfrac{a+x}{a-x}\right\|+C$
14	$\dfrac{1}{\sqrt{a^2-x^2}} \quad (a > 0)$	$\sin^{-1}\left(\dfrac{x}{a}\right)+C$
15	$\dfrac{1}{\sqrt{x^2+a^2}} \quad (a \neq 0)$	$\ln\left\|x+\sqrt{x^2+a^2}\right\|+C$
16	$\dfrac{1}{\sqrt{x^2-a^2}} \quad (a \neq 0)$	$\ln\left\|x+\sqrt{x^2-a^2}\right\|+C$
17	$\sqrt{x^2+a^2}$	$\dfrac{x}{2}\sqrt{x^2+a^2}+\dfrac{a^2}{2}\ln\left\|x+\sqrt{x^2+a^2}\right\|+C$
18	$\sqrt{x^2-a^2}$	$\dfrac{x}{2}\sqrt{x^2-a^2}-\dfrac{a^2}{2}\ln\left\|x+\sqrt{x^2-a^2}\right\|+C$
19	$\sqrt{a^2-x^2} \quad (a > 0)$	$\dfrac{x}{2}\sqrt{a^2-x^2}+\dfrac{a^2}{2}\sin^{-1}\left(\dfrac{x}{a}\right)+C$

Basic rules of integration

Each of the following rules can be easily deduced by differentiating the RHS and verifying that it is equal to the integrand of the LHS.

1. $\displaystyle\int kf(x)\,\mathrm{d}x = k\int f(x)\,\mathrm{d}x \qquad (k = \text{constant})$

2. $\displaystyle\int [f(x)+g(x)]\,\mathrm{d}x = \int f(x)\,\mathrm{d}x + \int g(x)\,\mathrm{d}x$

3. If $a,\ b$ are constants with $a \neq 0$ and if $\displaystyle\int f(x)\,\mathrm{d}x = F(x)+C$, then

$$\int f(ax+b)\,\mathrm{d}x = \frac{1}{a}F(ax+b)+C.$$

4. $\displaystyle\int \frac{f'(x)}{f(x)}\,\mathrm{d}x = \ln|f(x)|+C.$

5. $\displaystyle\int \frac{f'(x)}{\sqrt{f(x)}}\,\mathrm{d}x = 2\sqrt{f(x)}+C.$

Examples: Rule 1, 2

Example 8.3

$$\int (2x^2 + 3x - 4)\,\mathrm{d}x = \frac{2x^3}{3} + \frac{3x^2}{2} - 4x + C.$$

Example 8.4

$$\int \left(2x^3 + \frac{1}{x^2}\right)\mathrm{d}x = \frac{x^4}{2} - \frac{1}{x} + C.$$

Example 8.5

$$\int \left(3x + \frac{2}{x}\right)\mathrm{d}x = \frac{3x^2}{2} + 2\ln|x| + C.$$

See CoCalc
https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page266-CoCalcJupyter.pdf

Examples: Rule 3

Example 8.6

$$\int (ax+b)^4 \, \mathrm{d}x = \frac{(ax+b)^5}{5a} + C \quad \text{(if } a \neq 0\text{)}.$$

Example 8.7

$$\int \sin(2x-3) \, \mathrm{d}x = -\frac{1}{2}\cos(2x-3) + C.$$

Examples: Rule 4

Example 8.8

$$\int \tan x \, \mathrm{d}x = -\int \frac{-\sin x}{\cos x} \, \mathrm{d}x = -\ln|\cos x| + C.$$

Example 8.9

$$\int \frac{x}{x^2 - 3} \, \mathrm{d}x = \frac{1}{2} \int \frac{2x}{x^2 - 3} \, \mathrm{d}x = \frac{1}{2} \ln |x^2 - 3| + C.$$

Examples: Rule 5

Example 8.10

$$\int \frac{x}{\sqrt{x^2-1}}\,\mathrm{d}x = \frac{1}{2}\int \frac{2x}{\sqrt{x^2-1}}\,\mathrm{d}x = \sqrt{x^2-1} + C.$$

Techniques of integration: Substitution

Theorem 8.1 *If* $u = \phi(x)$ *with* $\phi(x)$, $\phi'(x)$ *being continuous, then*

$$\int f(\phi(x))\,\phi'(x)\,\mathrm{d}x = \int f(u)\,\mathrm{d}u \tag{8.2}$$

Approach A

Suppose we are required to find the LHS of (8.2). The theorem says that we can find the RHS of (8.2) instead. We use this approach if the RHS can be found readily. However, we seldom use (8.2) directly. Instead, we often proceed in the following way. Suppose that the RHS of (8.2) can be found readily as

$$\int f(u)\,\mathrm{d}u = F(u) + C.$$

Then the substitution $u = \phi(x)$ gives $\mathrm{d}u = \phi'(x)\,\mathrm{d}x$ and we can proceed formally as follows:

$$\int f(\phi(x))\,\phi'(x)\,\mathrm{d}x = \int f(u)\,\mathrm{d}u = F(u) + C = F(\phi(x)) + C.$$

To save the trouble of introducing new variable (the u in the above working), we may simply write

$$\int f(\phi(x))\,\phi'(x)\,\mathrm{d}x = \int f(\phi(x))\,\mathrm{d}(\phi(x)) = F(\phi(x)) + C.$$

Examples

Example 8.11 Find $\displaystyle\int \frac{\tan^{-1} x}{1+x^2}\,\mathrm{d}x$.

Solution. Putting $u = \tan^{-1} x$ and $\mathrm{d}u = \mathrm{d}x/(1+x^2)$, we have

$$\int \frac{\tan^{-1} x}{1+x^2}\mathrm{d}x = \int u\,\mathrm{d}u = \frac{1}{2}u^2 + C = \frac{1}{2}(\tan^{-1} x)^2 + C$$

or simply,

$$\int \frac{\tan^{-1} x}{1+x^2}\mathrm{d}x = \int (\tan^{-1} x)\,\mathrm{d}(\tan^{-1} x) = \frac{1}{2}(\tan^{-1} x)^2 + C.$$

□

Example 8.12 Find $I = \int x(x^2+2)^3 \, \mathrm{d}x$.

Solution.

$I = \int u^3 \cdot \frac{1}{2} \, \mathrm{d}u = \frac{1}{8}u^4 + C = \frac{1}{8}(x^2+2)^4 + C.$

$$u = x^2 + 2$$
$$\mathrm{d}u = 2x \, \mathrm{d}x$$
$$\therefore x \, \mathrm{d}x = \frac{1}{2} \, \mathrm{d}u$$

□

Example 8.13 Find $J = \displaystyle\int \sin x \cos^3 x \, \mathrm{d}x.$

Solution.

$$J = \int \sin x \cos^3 x \, \mathrm{d}x = \int u^3 \, (-\mathrm{d}u) = -\frac{1}{4} u^4 + C = -\frac{1}{4} \cos^4 x + C.$$

$$\begin{aligned} u &= \cos x \\ \mathrm{d}u &= -\sin x \, \mathrm{d}x \\ \therefore \sin x \, \mathrm{d}x &= -\mathrm{d}u. \end{aligned}$$

□

Example 8.14 Find $K = \displaystyle\int \frac{\ln x}{x}\,\mathrm{d}x.$

Solution.

$$K = \int \frac{\ln x}{x}\,\mathrm{d}x = \int u\,\mathrm{d}u = \frac{1}{2}u^2 + C = \frac{1}{2}(\ln x)^2 + C.$$

$u = \ln x$ $\mathrm{d}u = (1/x)\,\mathrm{d}x.$

□

Approach B

In this approach, the formula (8.2) is applied from the RHS to the LHS. Suppose that $f(u), \phi(x), \phi'(x)$ are continuous, and that $\phi^{-1}(u)$ exists and differentiable. Suppose further that the LHS of (8.2) can be found readily as

$$\int f(\phi(x))\, \phi'(x)\, \mathrm{d}x = G(x) + C.$$

Then the substitution $u = \phi(x)$ gives $\mathrm{d}u = \phi'(x)\, \mathrm{d}x$ and we can proceed formally as follows:

$$\int f(u)\, \mathrm{d}u = \int f(\phi(x))\, \phi'(x)\, \mathrm{d}x = G(x) + C = G(\phi^{-1}(u)) + C.$$

Examples

Example 8.15 Find $I = \int \frac{1}{\sqrt{1-u^2}}\,\mathrm{d}u$ with $|u| < 1$.

Solution.

$$I = \int \frac{1}{\sqrt{1-u^2}}\,\mathrm{d}u = \int \frac{\cos x}{|\cos x|}\,\mathrm{d}x = \int \mathrm{d}x = x + C = \sin^{-1} u + C.$$

$u = \sin x$
$\mathrm{d}u = \cos x\,\mathrm{d}x$
Take $-\pi/2 < x < \pi/2$
so that
$-1 < u < 1$ and $\cos x > 0$.

□

Example 8.16 Find $J = \displaystyle\int \frac{\mathrm{d}x}{\sqrt{x^2-1}}, \quad |x| > 1.$

Solution First assume $x > 1$.

$$J = \int \frac{\mathrm{d}x}{\sqrt{x^2-1}} = \int \frac{\sec\theta\tan\theta\,\mathrm{d}\theta}{|\tan\theta|} = \int \sec\theta\,\mathrm{d}\theta = \ln|\sec\theta + \tan\theta| + C$$
$$= \ln\left|x + \sqrt{x^2-1}\right| + C.$$

$x = \sec\theta$
$\mathrm{d}x = \tan\theta\sec\theta\,\mathrm{d}\theta$
Take $0 < \theta < \pi/2$
so that
$x > 1$ and $\tan\theta > 0$.

Next, we assume $x < -1$. Using the substitution $v = -x$ we get the same result:

$$J = \int \frac{-\mathrm{d}v}{\sqrt{v^2 - 1}} = -\ln\left|v + \sqrt{v^2 - 1}\,\right| + C \quad \text{(since } v > 0\text{)}$$

$$= -\ln\left| -x + \sqrt{x^2 - 1}\,\right| + C = \ln\left|x + \sqrt{x^2 - 1}\,\right| + C$$

since

$$\ln\left| -x + \sqrt{x^2 - 1}\,\right| + \ln\left|x + \sqrt{x^2 - 1}\,\right| = \ln\left|(x^2 - 1) - x^2\right| = \ln 1 = 0.$$

Integration by parts

Theorem 8.2 *If* u *and* v *are differentiable functions of* x, *then*

$$\int u(x)\, v'(x)\, \mathrm{d}x = u(x)\, v(x) - \int u'(x)\, v(x)\, \mathrm{d}x \tag{8.3}$$

which can be written, in a simpler form, as $\displaystyle\int u\, \mathrm{d}v = uv - \int v\, \mathrm{d}u.$

The above formula can be deduced by integrating both sides of the product rule $(uv)' = uv' + vu'$.

Integrate product of two functions

Integration by parts is often useful when the integrand is a product of two functions say, $u(x)\phi(x)$.

$$u'(x) = w(x), \qquad \int \phi(x)\,\mathrm{d}x = v(x) + C$$

Since $v'(x) = \phi(x)$, the formula (8.3) becomes

$$\int u(x)\,\phi(x)\,\mathrm{d}x = u(x)\,v(x) - \int w(x)\,v(x)\,\mathrm{d}x. \tag{8.4}$$

Thus the method works if the new product $w(x)v(x)$ on the RHS is more easily integrable than the original product $u(x)\phi(x)$ on the LHS. There are two ways to present the working.

Method A: Traditional

To get (8.4) from (8.3), we proceed formally as:

$$\begin{aligned}
\int u\phi \,\mathrm{d}x &= \int u \,\mathrm{d}v && (v' = \phi \;\therefore\; \mathrm{d}v = \phi \,\mathrm{d}x) \\
&= uv - \int v \,\mathrm{d}u && \text{by } (8.3) \\
&= uv - \int vw \,\mathrm{d}x. && (u' = w \;\therefore\; \mathrm{d}u = w \,\mathrm{d}x.)
\end{aligned}$$

The last line is (8.4).

Examples

Example 8.17 Find the integral $\int (x+2)\cos x \, \mathrm{d}x$

Solution.

$$
\begin{aligned}
&\int (x+2)\cos x \, \mathrm{d}x \\
&\quad = \int u \, \mathrm{d}v && (u = x+2, \quad \mathrm{d}v = \cos x \, \mathrm{d}x) \\
&\quad = uv - \int v \, \mathrm{d}u && (\text{using (8.3) and } v = \sin x) \\
&\quad = (x+2)\sin x - \int \sin x \, \mathrm{d}(x+2) \\
&\quad = (x+2)\sin x - \int \sin x \, \mathrm{d}x && (\mathrm{d}(x+2) = \mathrm{d}x) \\
&\quad = (x+2)\sin x + \cos x + C.
\end{aligned}
$$

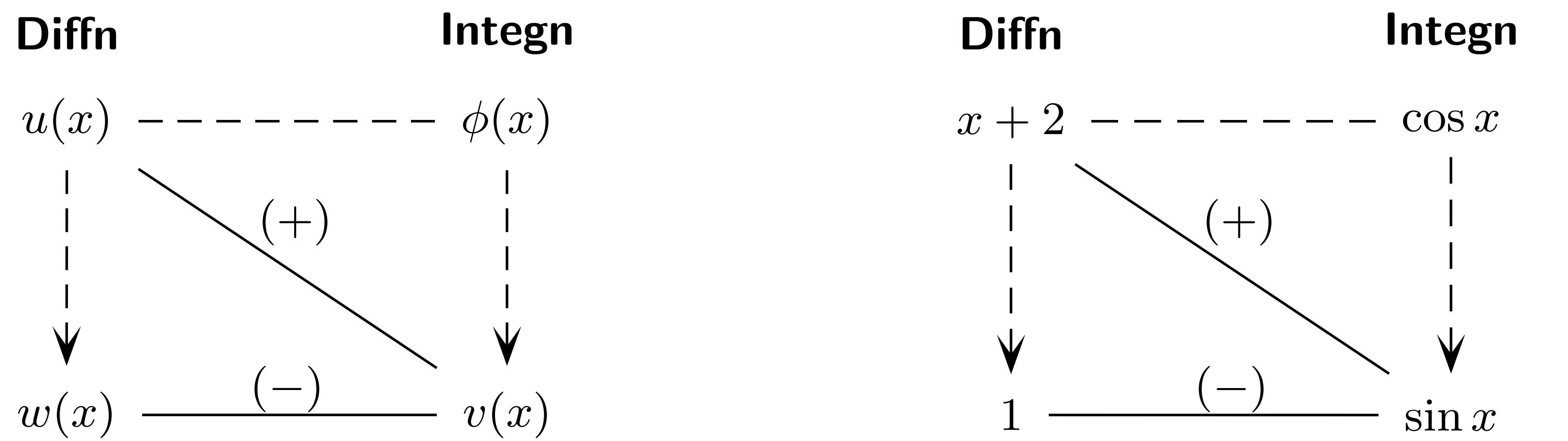
□

To save the trouble of introducing symbols like u, v, we can present the working in the following way:

$$\begin{aligned}\int (x+2)\cos x\,\mathrm{d}x &= \int (x+2)\,\mathrm{d}(\sin x)\\ &= (x+2)\sin x - \int \sin x\,\mathrm{d}(x+2)\\ &= (x+2)\sin x - \int \sin x\,\mathrm{d}x\\ &= (x+2)\sin x + \cos x + C.\end{aligned}$$

Method B: Presenting the working in a chart

We can obtain (8.4) directly while presenting the working in a chart as shown in Fig. 8.1(a). An example where $u = x + 2$ and $\phi = \cos x$ is shown in Fig. 8.1(b).



(a) For the integrand $u(x)\phi(x)$ (b) For the integrand $(x+2)\cos x$

Figure 8.1: Working chart for integration by parts.

In the working chart,

- the dashed vertical arrows indicate differentiation and integration.
- the dashed horizontal line represents the product of the original factors. This forms the LHS of (8.4), namely, $\int u\phi \, \mathrm{d}x$.
- the solid lines (one slanted with a plus sign and the other horizontal with a minus sign) joining two functions represent the product of the functions. These products form the RHS of (8.4), namely, $+uv - \int wv \, \mathrm{d}x$.

The steps of integration by parts is:

1. Write the given integrand as a product of two factors u and ϕ. Differentiate one factor u and integrate the other factor ϕ. Show these results by drawing two vertical arrows as in the above working chart. Multiply the results (and draw the solid horizontal line) and see if the product can be integrated or if it is simpler than the original product. If so, the method works and we can continue. If not, try changing the assignments of u and ϕ and start from scratch. Note that there is no guarantee that integration by parts will work.

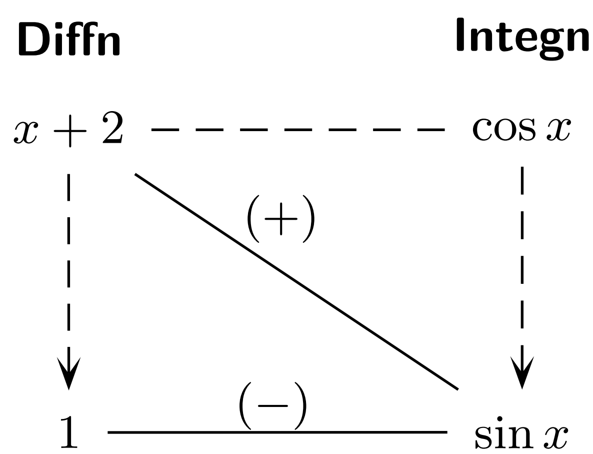
2. Complete the chart by drawing the slanted solid line. The slanted line is drawn from the "differentiation" side downward to the "integration" side. Put down (8.4) based on the chart.

Examples

Example 8.18 Find the integral $\int (x+2)\cos x \, \mathrm{d}x$.

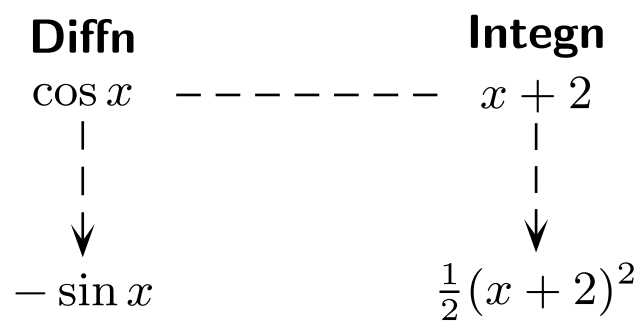
Solution. Using the chart (same as Fig. 8.1(b)), we have

$$\int (x+2)\cos x \, \mathrm{d}x = (x+2)\cdot \sin x - \int 1 \cdot \sin x \, \mathrm{d}x = (x+2)\sin x + \cos x + C.$$



□

In the above example, if we differentiate $\cos x$ and integrate $x+2$, we get the results shown on the right. As the product $\frac{1}{2}(x+2)^2 \sin x$ is more involved than the original product $(x+2)\cos x$. This way of doing integration by parts fails.

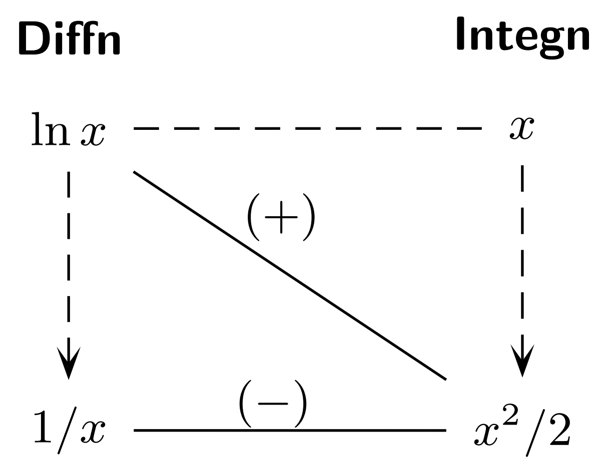


If one factor is a polynomial (like $x+1$ in the above example) and the other is not, we normally differentiate the polynomial and integrate the other factor, hoping to reduce the integral to a simpler form.

Example 8.19 Find the integral $\int x \ln x \, \mathrm{d}x \quad (x > 0)$.

Solution. If we differentiate x, then we must integrate $\ln x$. Since we do not know the integral of $\ln x$, we do the other way: differentiating $\ln x$ and integrating x. Now the method works.

$$\begin{aligned}\int x \ln x \, \mathrm{d}x &= (\ln x) \cdot \frac{x^2}{2} - \int \frac{1}{x} \cdot \frac{x^2}{2} \, \mathrm{d}x = \frac{1}{2} x^2 \ln x - \int \frac{x}{2} \, \mathrm{d}x \\ &= \frac{1}{2} x^2 \ln x - \frac{x^2}{4} + C.\end{aligned}$$



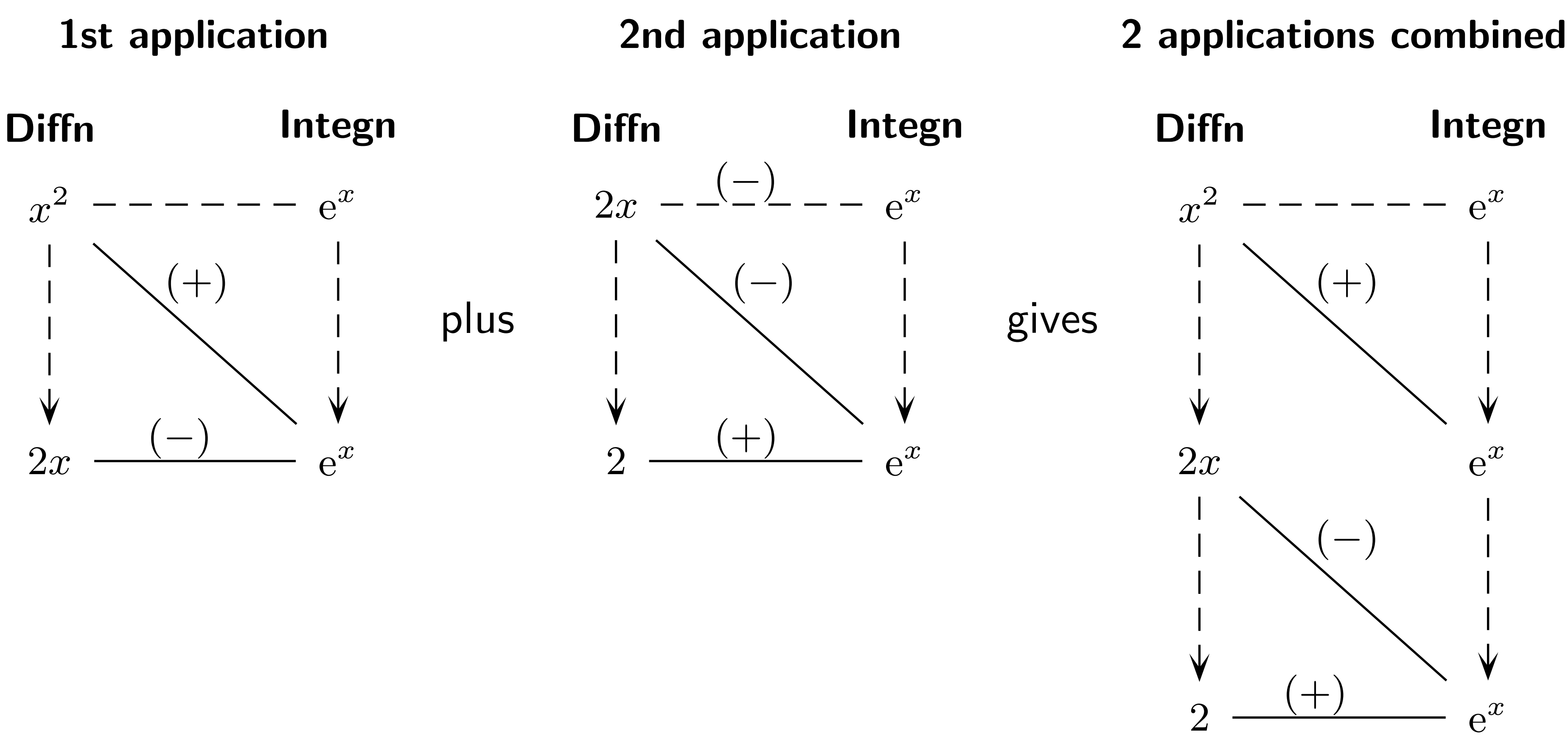
□

See CoCalc
https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page290-CoCalcJupyter.pdf

Repeated applications of integration by parts

After one application of integration by parts the result may also involve an integral of a product of functions. Sometimes a second application (or even more applications) is necessary.

The following is an example on integrating $x^2 e^x$. Note that in the middle chart, we started with an integral with a negative sign and hence the subsequent signs are $-$ and $+$.



Based on the combined chart, we obtain

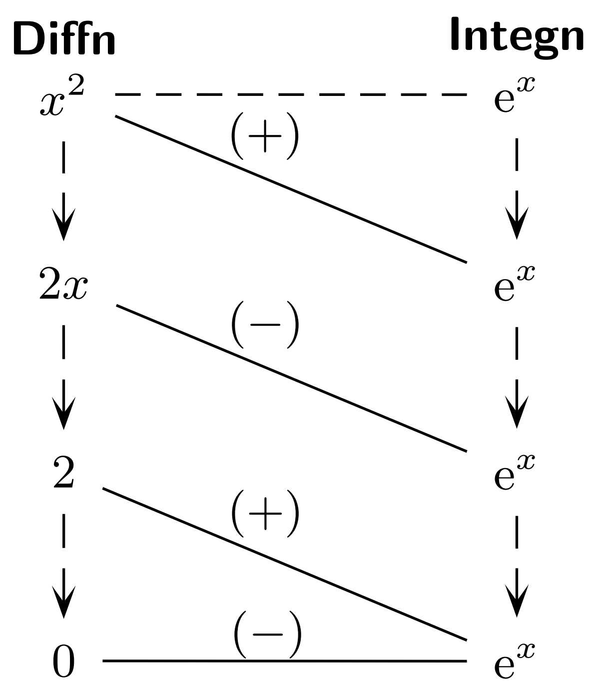
$$\int x^2 \mathrm{e}^x \,\mathrm{d}x = +x^2\mathrm{e}^x - 2x\mathrm{e}^x + \int 2\mathrm{e}^x \,\mathrm{d}x.$$

For this particular integral we can actually do one more integration by parts to finish. The complete working is shown in the next example.

Examples

Example 8.20 Find $\int x^2 e^x \, dx$

Solution. From the chart, we get $\int x^2 e^x \, dx = x^2 e^x - 2x e^x + 2e^x + C$ since the last row of the chart gives $\int 0 \, dx = C$.



Example 8.21 Find $I = \int e^{ax} \cos bx \, dx$ assuming $b \neq 0$.

Solution. From the chart, we get, with $b \neq 0$,

$$I = \int e^{ax} \cos bx \, dx = e^{ax} \frac{\sin bx}{b} + a e^{ax} \frac{\cos bx}{b^2} - \frac{a^2}{b^2} \int e^{ax} \cos bx \, dx.$$

The two integrals appearing in the above equation are of the same form but they may differ by a constant K. Therefore we have

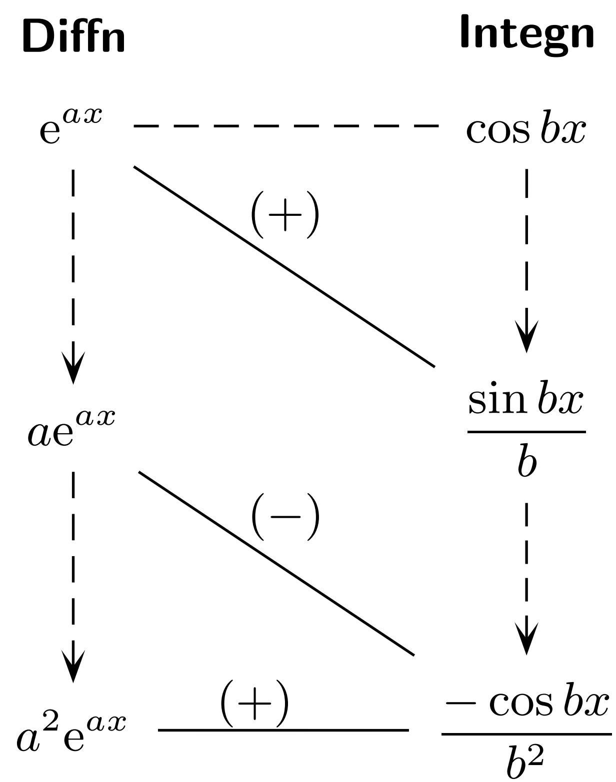
$$I = e^{ax} \frac{\sin bx}{b} + a e^{ax} \frac{\cos bx}{b^2} - \frac{a^2}{b^2}(I + K).$$

Solving for I, we get

$$I = \frac{b \sin bx + a \cos bx}{a^2 + b^2} e^{ax} + C$$

where C is the constant of integration. □

See CoCalc
https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page294-CoCalcJupyter.pdf



Note that we can get the same result by assuming $a \neq 0$.

Example 8.22 Find $J = \displaystyle\int \sec^3 x \,\mathrm{d}x$.

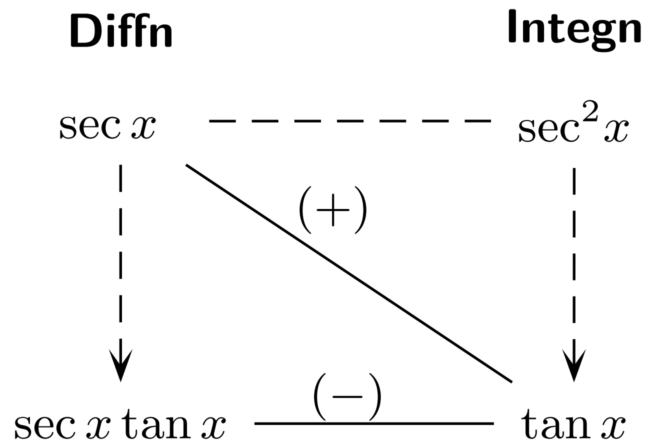
Solution.

$$\begin{aligned}
J &= \int \sec^3 x \,\mathrm{d}x = \int \sec x \sec^2 x \,\mathrm{d}x \\
&= \sec x \tan x - \int \sec x \ \tan^2 x \,\mathrm{d}x = \sec x \tan x - \int (\sec^2 x - 1) \sec x \,\mathrm{d}x \\
&= \sec x \tan x - J + \int \sec x \,\mathrm{d}x. \\
&= \sec x \tan x - J + \ln|\sec x + \tan x| + C_1
\end{aligned}$$

Solving for J, we have

$$J = \frac{1}{2} \sec x \tan x + \frac{1}{2} \ln|\sec x + \tan x| + C.$$

□


Diffn
Integn
sec x
sec² x
(+)
(−)
sec x tan x
tan x

Products of sine and cosine

If the integrand is a product of sine and cosine, we can express it in the form of a sum using the formulas:

$$\begin{aligned} 2\sin A\cos B &= \sin(A-B)+\sin(A+B) \\ 2\cos A\cos B &= \cos(A-B)+\cos(A+B) \\ 2\sin A\sin B &= \cos(A-B)-\cos(A+B) \end{aligned}$$

For an integrand that is a product of powers of sine and cosine, we may have to use formulas like:

$$\cos^2 A+\sin^2 A=1, \quad 2\sin^2 A=1-\cos 2A \quad \text{and} \quad 2\cos^2 A=1+\cos 2A.$$

Examples

Example 8.23 Find $I = \int \sin 3x \sin 2x \, \mathrm{d}x$.

Solution.
$$I = \int \frac{1}{2}[\cos(3x - 2x) - \cos(3x + 2x)] \, \mathrm{d}x$$
$$= \frac{1}{2} \int (\cos x - \cos 5x) \, \mathrm{d}x$$
$$= \frac{1}{2} \sin x - \frac{1}{10} \sin 5x + C.$$
□

Example 8.24 Find $I = \int \sin^2 x \cos^2 x \, \mathrm{d}x$.

Solution.

$$I = \frac{1}{4}\int (\sin 2x)^2 \, \mathrm{d}x = \frac{1}{8}\int (1 - \cos 4x) \, \mathrm{d}x = \frac{1}{8}\left(x - \frac{\sin 4x}{4}\right) + C.$$

□

Example 8.25 Find $I = \int \sin^2 x \cos^3 x \, \mathrm{d}x$.

Solution.

$$\begin{aligned} I &= \int \sin^2 x(1 - \sin^2 x) \, \mathrm{d}(\sin x) \\ &= \int (\sin^2 x - \sin^4 x) \, \mathrm{d}(\sin x) \\ &= \frac{1}{3} \sin^3 x - \frac{1}{5} \sin^5 x + C. \end{aligned}$$

□

Example 8.26 Find $I = \int \sin^n x \cos x \, \mathrm{d}x$.

Solution. This type of integrals can be found by substitution (cf. Example 8.13.)

$$I = \int \sin^n x \, \mathrm{d}(\sin x) = \frac{\sin^{n+1} x}{n+1} + C.$$

□

See CoCalc
https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page302-CoCalcJupyter.pdf

Rational functions

A *rational function* in x is a function of the form

$$R(x) \equiv \frac{P(x)}{Q(x)}$$

where $P(x)$ and $Q(x)$ are polynomials. A proper rational function is one with $\deg\, P < \deg\, Q$. Recall that

- Every rational function = a polynomial + a proper rational function.
- Every proper rational function = a sum of partial fractions.

Integrating a polynomial is straightforward. So we can integrate a rational function if we know how to integrate its partial fractions. Examples of a few simple cases are demonstrated as follows.

Examples

Example 8.27 Find $\displaystyle\int \frac{3x+2}{x^2+1}\,\mathrm{d}x$.

Solution. The integrand is a proper rational function.

$$\begin{aligned}\int \frac{3x+2}{x^2+1}\,\mathrm{d}x &= \frac{3}{2}\int \frac{2x}{x^2+1}\,\mathrm{d}x + 2\int \frac{1}{x^2+1}\,\mathrm{d}x \\ &= \frac{3}{2}\ln(x^2+1) + 2\tan^{-1}x + C.\end{aligned}$$

□

Example 8.28 Find $\displaystyle\int \frac{x^2(x-3)}{(x-1)(x-2)}\,\mathrm{d}x$.

Solution The integrand is not a proper rational function. By long division, we get

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

In partial fractions,

$$\frac{-2x}{(x-1)(x-2)} = \frac{A}{x-1} + \frac{B}{x-2}$$

where A and B are constants. Removing denominators,

$$-2x \equiv A(x-2) + B(x-1)$$

Comparing coefficients of x and constant terms, we get

$$-2 = A + B, \quad 0 = -2A - B \quad \text{and hence } \; A = 2, \; B = -4.$$

Therefore

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

and hence

$$\begin{aligned}\int \frac{x^2(x-3)}{(x-1)(x-2)}\,\mathrm{d}x &= \int \left(x + \frac{2}{x-1} + \frac{-4}{x-2}\right)\mathrm{d}x \\ &= \frac{x^2}{2} + 2\ln|x-1| - 4\ln|x-2| + C.\end{aligned}$$

Example 8.29 Find $\displaystyle\int \frac{x^2-2x-1}{(x-1)(x^2+1)}\,\mathrm{d}x$.

Solution We first put the integrand in partial fractions:

$$\frac{x^2-2x-1}{(x-1)(x^2+1)} = \frac{A}{x-1} + \frac{Bx+C}{x^2+1}.$$

$$\therefore \quad x^2-2x-1 \equiv A(x^2+1) + (Bx+C)(x-1).$$

Comparing coefficients of powers of x, we obtain:

$$1 = A + B, \quad -2 = -B + C, \quad -1 = A - C.$$

Solving these equations, we get $A = -1,\ B = 2,\ C = 0.$

$$\therefore \quad \int \frac{x^2-2x-1}{(x-1)(x^2+1)}\,\mathrm{d}x = \int \left(\frac{-1}{x-1} + \frac{2x}{x^2+1}\right)\mathrm{d}x$$

$$= -\ln|x-1| + \ln(x^2+1) + C.$$

Example 8.30 Find the integral $\int \frac{x^2+1}{(x-1)(x-2)(x+3)}\,\mathrm{d}x$.

Solution. We first put the integrand in partial fractions:

$$\frac{x^2+1}{(x-1)(x-2)(x+3)} = \frac{A}{x-1} + \frac{B}{x-2} + \frac{C}{x+3}$$

$$\therefore \quad x^2+1 \equiv A(x-2)(x+3) + B(x-1)(x+3) + C(x-1)(x-2).$$

Comparing coefficients of powers of x, we obtain:

$$1 = A+B+C, \quad 0 = A+2B-3C, \quad 1 = -6A-3B+2C.$$

Solving these equations, we get $A = -1/2, \quad B = 1, \quad C = 1/2.$ Therefore

$$\begin{aligned}\int \frac{x^2+1}{(x-1)(x-2)(x+3)}\,\mathrm{d}x &= \int \left(\frac{-1/2}{x-1} + \frac{1}{x-2} + \frac{1/2}{x+3}\right)\mathrm{d}x \\ &= -\frac{1}{2}\ln|x-1| + \ln|x-2| + \frac{1}{2}\ln|x+3| + C.\end{aligned}$$

□

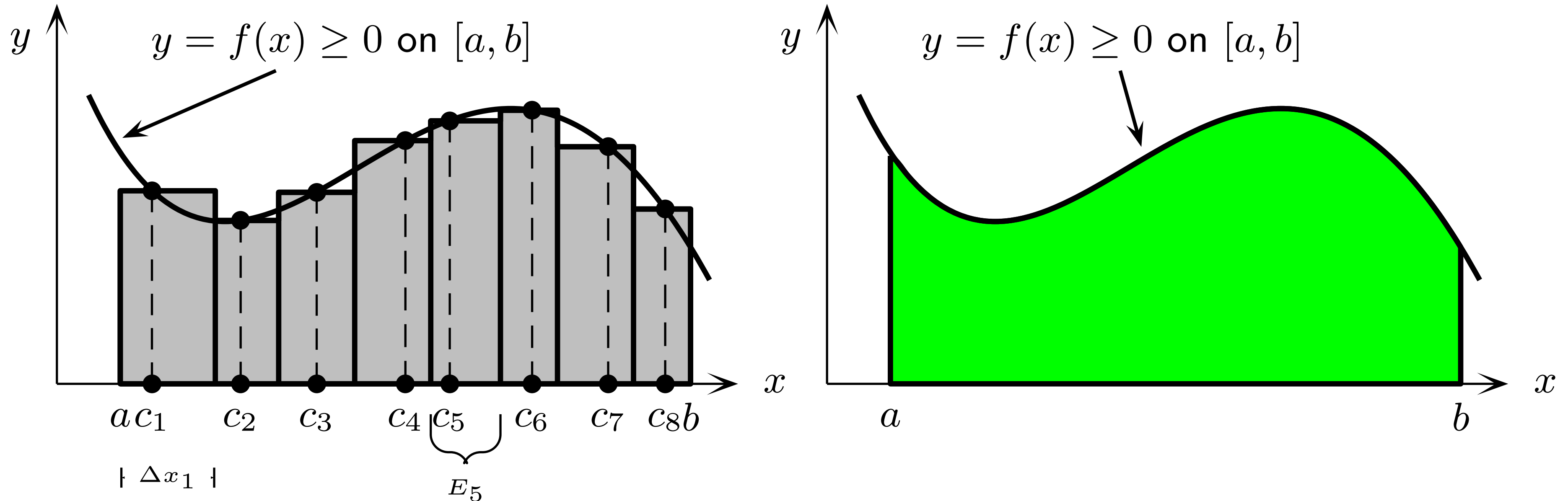
See CoCalc

https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page308-CoCalcJupyter.pdf

Definite Integral

The Riemann sum

A Riemann sum is the sum of products of the form $f(c_i) \cdot \Delta x_i$. If $f(x) \geq 0$, each product represents the area of the rectangle of length $f(c_i)$ and width Δx_i. (See Fig. 9.1(a).) In this case, the definite integral gives the area under the curve between $x = a$ and $x = b$ (see Fig. 9.1(b).)



(a) The Riemann sum $\sum_{i=1}^{8} f(c_i)\Delta x_i$ (b) $\int_a^b f(x)\,\mathrm{d}x =$ the area under the curv.

Figure 9.1: The Riemann sum and the definite integral.

Definition of definite integrals

Definition 9.1 Let $[a, b]$ $(a < b)$ be a closed and finite interval and let $f(x)$ be a continuous function defined on $[a, b]$.

- If the interval $[a, b]$ is subdivided into N small subintervals E_i $(i = 1, 2, \ldots, N)$ whose length is Δx_i and if c_i is any point inside E_i, then the sum

$$S_N = \sum_{i=1}^{N} f(c_i)\, \Delta x_i$$

  is called a *Riemann sum* of the function $f(x)$ on $[a, b]$. (See Fig. 9.1(a).)

- The *definite integral* of $f(x)$ over $[a, b]$, denoted by $\int_a^b f(x)\, \mathrm{d}x$, is defined as the limit (if it exists)

$$\lim_{\text{all } \Delta x_i \to 0} S_N = \lim_{\text{all } \Delta x_i \to 0} \sum_{i=1}^{N} f(c_i)\, \Delta x_i.$$

  (See Fig. 9.1(b).) In the limiting process we allow $N \to \infty$ with all $\Delta x_i \to 0$.

- Also, we define (note that $a < b$)

$$\int_b^a f(x)\,\mathrm{d}x = -\int_a^b f(x)\,\mathrm{d}x \quad \text{and} \quad \int_a^a f(x)\,\mathrm{d}x = 0.$$

- For the definite integral $\int_a^b f(x)\,\mathrm{d}x$, the numbers a and b are called the *lower* and *upper limits* of integration respectively. The function $f(x)$ is the *integrand*.

Basic properties of definite integrals

The following basic properties follows directly from the definition of definite integrals.

1. The definite integral $\int_a^b f(x)\,\mathrm{d}x$ is a number which depends on the function $f(x)$ and the interval $[a,b]$ only. That is to say, we may use any convenient symbol (say x or s or other) for the variable and we always have

$$\int_a^b f(x)\,\mathrm{d}x = \int_a^b f(s)\,\mathrm{d}s = \int_a^b f(t)\,\mathrm{d}t.$$

2. **Linearity**. For constants α and β,

$$\int_a^b [\alpha f(x) + \beta g(x)]\,\mathrm{d}x = \alpha \int_a^b f(x)\,\mathrm{d}x + \beta \int_a^b g(x)\,\mathrm{d}x.$$

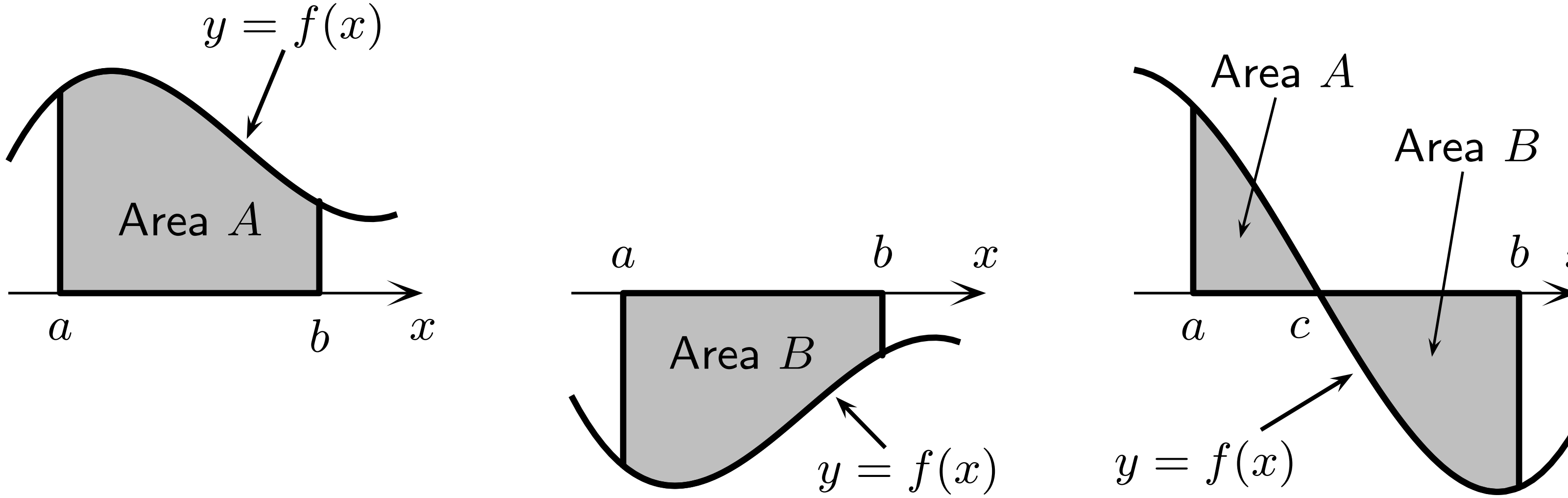
3. **Additivity over subintervals.** If $a < c < b$, then

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

Geometric interpretation

Let $f(x)$ be a continuous function on $[a, b]$ where $a < b$. The definite integral $I = \int_a^b f(x)\,\mathrm{d}x$ is a number which may be positive, zero or negative, depending on the behaviour of $f(x)$.

- $f(x) \geq 0$. Then $\int_a^b f(x)\,\mathrm{d}x = A$ where A is the area bounded between $\mathcal{C}$ and the x-axis over the interval $[a, b]$ (see Fig. 9.2(a) or Fig. 9.1(b)).
- $f(x) \leq 0$. Then the integral is negative. In fact we have $\int_a^b f(x)\,\mathrm{d}x = -B$ where B is the area bounded between $\mathcal{C}$ and the x-axis over the interval $[a, b]$ (see Fig. 9.2(b)).
- $f(x) \geq 0$ on $[a, c]$ and $f(x) \leq 0$ on $[c, b]$ where $a < c < b$. Then $\int_a^b f(x)\,\mathrm{d}x = A - B$ where A and B are the areas bounded between $\mathcal{C}$ and the x-axis over the intervals $[a, c]$ and $[c, b]$ respectively (see Fig. 9.2(c)).



(a) $\int_a^b f(x)\,\mathrm{d}x = A$ (b) $\int_a^b f(x)\,\mathrm{d}x = -B$ (c) $\int_a^b f(x)\,\mathrm{d}x = A - B$

Figure 9.2: The definite integral and the area between the curve and the x-axis.

Fundamental theorem of calculus

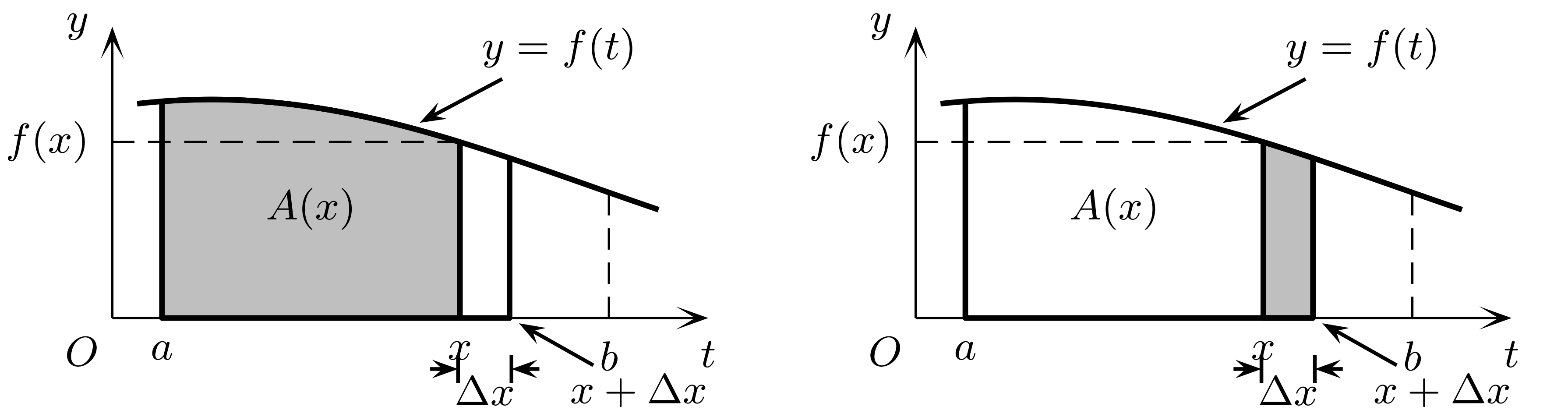
Let $f(x)$ be given. Recall that a primitive of $f(x)$ is a function $F(x)$ such that $F'(x) = f(x)$.

Theorem 9.1 (Fundamental Theorem of Calculus) *Let $F(x)$ denote any one primitive of $f(x)$. Then*

$$\int_a^b f(x)\,\mathrm{d}x = F(b) - F(a) = \Big[F(x)\Big]_a^b$$

This theorem is an important tool for evaluating a definite integral. It tells us that the value of the definite integral of $f(x)$ over $[a, b]$ is just the difference of the values of any primitive $F(x)$ at the limits b and a of integration.

Proof of the Fundamental Theorem of Calculus



(a) $A(x)$ is the shaded area. (b) $A(x+\Delta x) - A(x)$ is the shaded area.

Figure 9.3: If $A(x) = \int_a^x f(t)\,\mathrm{d}t$ then $A'(x) = f(x)$.

For simplicity, we only consider the case when $f(x) > 0$ on $[a, b]$. Then the integral $\int_a^b f(x)\,\mathrm{d}x$ is the area under the curve $y = f(x)$ over the interval $[a, b]$. For each x in $[a, b]$, we define (see Fig. 9.3(a))

$$A(x) = \int_a^x f(t)\,\mathrm{d}t = \text{ area under the curve on } [a, x]. \tag{9.1}$$

Then

$$\frac{\mathrm{d}}{\mathrm{d}x}A(x) = \lim_{\Delta x \to 0} \frac{A(x+\Delta x) - A(x)}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{\text{Area of the strip shaded in Fig 9.3(b)}}{\text{Width of the strip}}$$

$$= f(x) \qquad \text{(See Fig. 9.3(b).)}$$

Therefore, in addition to the given $F(x)$, $A(x)$ is also a primitive of $f(x)$.
It follows that

$$\frac{\mathrm{d}}{\mathrm{d}x}(A(x) - F(x)) = f(x) - f(x) = 0 \quad \text{and} \quad \therefore A(x) = F(x) + C$$

where C is a constant. By (9.1), $A(x) = 0$ when $x = a$.
Therefore $0 = A(a) = F(a) + C$ giving $C = -F(a)$ and hence $A(x) = F(x) - F(a)$.

Putting $x = b$, we have $A(b) = F(b) - F(a)$, i.e. $\displaystyle\int_a^b f(x)\,\mathrm{d}x = F(b) - F(a)$.

□

Examples

Example 9.1 Find $I = \int_0^2 (x^3 + 3x^2 - 4)\,\mathrm{d}x$.

Solution. Since $\int (x^3 + 3x^2 - 4)\mathrm{d}x = x^4/4 + x^3 - 4x + C$, we have

$$I = \left[\frac{x^4}{4} + x^3 - 4x\right]_0^2 = \left(\frac{2^4}{4} + 2^3 - 4 \times 2\right) - 0 = 4.$$

□

Example 9.2 Find $I = \int_0^1 (x^2 - 3\sin x)\,\mathrm{d}x$.

Solution. $I = \left[\frac{x^3}{3} + 3\cos x\right]_0^1 = \frac{1}{3} + 3(\cos 1 - 1) = 3\cos 1 - 8/3 \approx -1.046.$ □

Example 9.3 Find $I = \int_0^2 \frac{1}{2x-5}\,dx$

Solution.

$$I = \frac{1}{2}\Big[\ln|2x-5|\Big]_0^2 = \frac{1}{2}\Big(\ln 1 - \ln 5\Big) = -\frac{\ln 5}{2}.$$

□

Finding definite integrals: Substitution

In the following examples, we change the variable of integration from x to u using a substitution $u = \phi(x)$. Also, the limits of integration are changed accordingly.

Example 9.4 Find $I = \int_0^2 x(x^2+2)^3 \, \mathrm{d}x$.

Solution.

$$I = \int_2^6 u^3 \cdot \frac{1}{2} \, \mathrm{d}u = \frac{1}{8}\left[u^4\right]_2^6 = 160.$$

$$\begin{aligned}
&u = x^2 + 2 \\
&\mathrm{d}u = 2x \, \mathrm{d}x \\
&\therefore x \, \mathrm{d}x = \frac{1}{2} \, \mathrm{d}u \\
&x = 0 \implies u = 2, \\
&x = 2 \implies u = 6.
\end{aligned}$$

□

Example 9.5 Find $I = \int_0^{\sqrt{3}} x^5 \sqrt{x^2+1}\,\mathrm{d}x.$

Solution.

$$I = \int_1^4 (u-1)^2 \cdot u^{1/2} \cdot \frac{1}{2}\,\mathrm{d}u = \frac{1}{4}\int_1^2 (u^{5/2} - 2u^{3/2} + u^{1/2})\,\mathrm{d}u$$
$$= \left[\frac{u^{7/2}}{7} - 2\frac{u^{5/2}}{5} + \frac{u^{3/2}}{3}\right]_1^4 = 8\frac{8}{105}.$$

$$u = x^2 + 1$$
$$\mathrm{d}u = 2x\,\mathrm{d}x$$
$$x\,\mathrm{d}x = \frac{1}{2}\,\mathrm{d}u$$
$$x^2 = u - 1.$$
$$x = 0,\ \ u = 1$$
$$x = \sqrt{3},\ \ u = 4.$$

□

See CoCalc

https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page323-CoCalcJupyter.pdf

Example 9.6 Find $I = \displaystyle\int_0^{\pi/3} \sec^2 x \, \tan x \, dx.$

Solution.

$$I = \int_1^2 u \, du = \left[\frac{u^2}{2}\right]_1^2 = \frac{3}{2}.$$

$u = \sec x$ $du = \sec x \, \tan x \, dx$ $x = 0, \; u = 1$ $x = \pi/3, \; u = 2.$

□

Two More Examples:

https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/TwoIntegrationExamples.pdf

Integration by parts

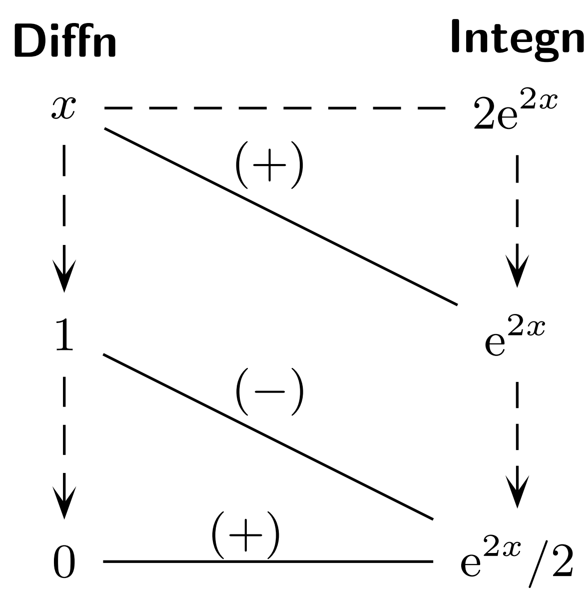
Meaning of Integration by parts

https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/Integration_by_Parts.pdf

Example 9.7 Find $I = \int_0^2 2x\mathrm{e}^{2x}\,\mathrm{d}x$.

Solution. Using the working chart, we get

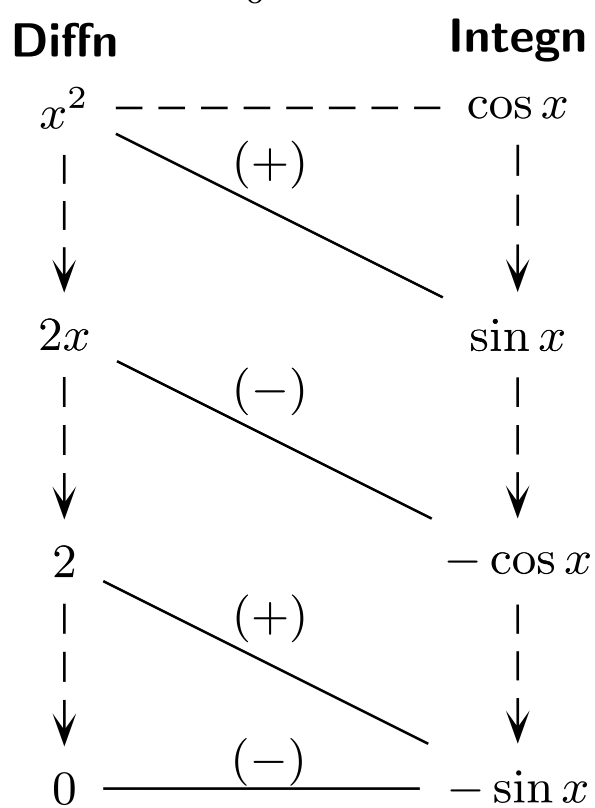
$$I = \left[x\mathrm{e}^{2x} - \frac{1}{2}\mathrm{e}^{2x}\right]_0^2 = 2\mathrm{e}^4 - 0 - \frac{1}{2}(\mathrm{e}^4 - 1) = \frac{3}{2}\mathrm{e}^4 + \frac{1}{2}.$$



Example 9.8 Evaluate $I = \int_0^1 x^2 \cos x \, \mathrm{d}x$.

Solution. Using the working chart, we get

$$I = \left[x^2 \sin x + 2x \cos x - 2 \sin x\right]_0^1 = 2\cos 1 - \sin 1 \approx 0.239.$$



□

More examples

Example 9.9 Find $I = \int_0^1 \cos^2 x \, \mathrm{d}x$.

Solution. By the trigonometric formula: $\cos^2 x = \frac{1}{2}(1 + \cos 2x)$, we have

$$I = \frac{1}{2} \int_0^1 (1 + \cos 2x) \, \mathrm{d}x = \frac{1}{2} \left[x + \frac{1}{2} \sin 2x \right]_0^1 = \frac{1}{2} \left(1 + \frac{\sin 2}{2} \right) \approx 0.7273.$$

□

Example 9.10 Find $J = \int_0^{\pi/2} \sin 2x \, \sin 3x \, \mathrm{d}x.$

Solution. By the trigonometric formula: $2 \sin 2x \, \sin 3x = \cos x - \cos 5x$, we have

$$J = \frac{1}{2} \int_0^{\pi/2} (\cos x - \cos 5x) \, \mathrm{d}x = \frac{1}{2} \Big[\sin x - \frac{1}{5} \sin 5x \Big]_0^{\pi/2}$$
$$= \frac{1}{2} \Big(1 - \frac{1}{5} \Big) = \frac{2}{5}.$$

□

Example 9.11 Find $K = \int_{-2}^{-1} \frac{1}{x(x-1)} \, \mathrm{d}x.$

Solution. Resolving into partial fractions: $\frac{1}{x(x-1)} = \frac{1}{x-1} - \frac{1}{x}$, we have

$$K = \int_{-2}^{-1} \left(\frac{1}{x-1} - \frac{1}{x} \right) \mathrm{d}x = \Big[\ln|x-1| - \ln|x| \Big]_{-2}^{-1}$$

$$= \ln 2 - \ln 3 - \ln 1 + \ln 2 = 2\ln 2 - \ln 3 = \ln(4/3).$$

□

Reduction formulas for definite integrals

Consider the definite integrals $\int_0^1 x^n \mathrm{e}^x \, \mathrm{d}x$ where n is a non-negative integer. As this integral involves an integer n, we can denote it by J_n. With this notation, we can show that (see Example 9.12)

$$J_n = \mathrm{e} - nJ_{n-1}, \quad n = 1, 2, \ldots.$$

Such a formula is called a *reduction formula*.
If J_0 is known, we can find inductively (using the above reduction formula) the values of J_1 and then J_2, J_3, ...
Usually we can establish reduction formulas using integration by parts.

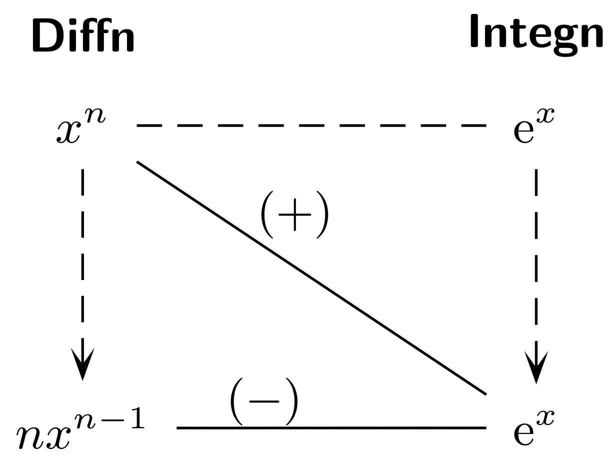
Examples

Example 9.12 Establish a reduction formula for the integral $J_n = \int_0^1 x^n e^x \, dx$. Hence find the value of J_3.

Soltion For $n = 1, 2, \ldots$

$$J_n = \int_0^1 x^n e^x \, dx = \Big[x^n e^x\Big]_0^1 - \int_0^1 n x^{n-1} e^x \, dx = e - nJ_{n-1}.$$

which is the reduction formula for J_n.



Now J_0 can be found directly as

$$J_0 = \int_0^1 \mathrm{e}^x \,\mathrm{d}x = \mathrm{e} - 1.$$

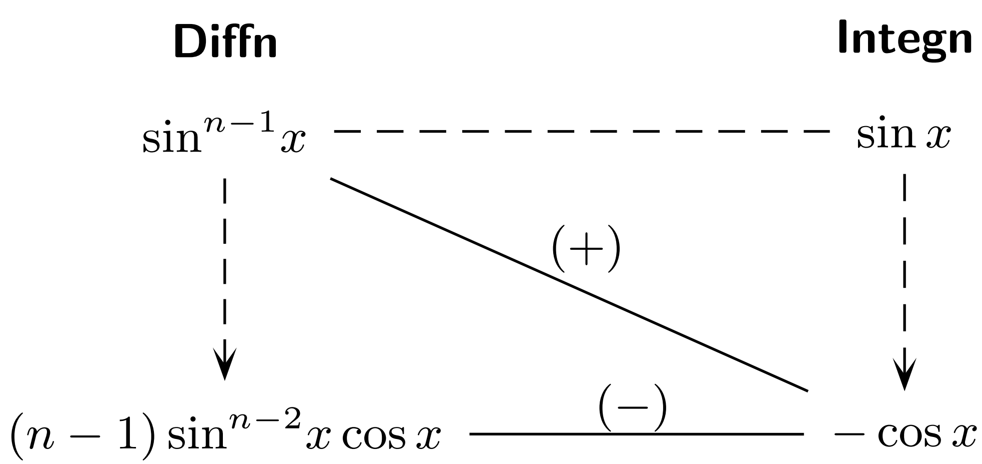
Therefore, inductively we get

$$\begin{aligned} J_1 &= \mathrm{e} - J_0 = \mathrm{e} - (\mathrm{e} - 1) = 1, \\ J_2 &= \mathrm{e} - 2J_1 = \mathrm{e} - 2, \\ J_3 &= \mathrm{e} - 3J_2 = \mathrm{e} - 3(\mathrm{e} - 2) = -2\mathrm{e} + 6. \end{aligned}$$

Example 9.13 Establish a reduction formula for the integral $J_n = \int_0^{\pi/2} \sin^n x \, \mathrm{d}x$. Hence find the values of J_4 and J_5.

Solution For $n = 2, 3, 4, \ldots$,

$$
\begin{aligned}
J_n &= \int_0^{\pi/2} \sin^n x \, \mathrm{d}x = \int_0^{\pi/2} \sin^{n-1} x \, \sin x \, \mathrm{d}x \\
&= \Big[-\sin^{n-1} x \, \cos x \Big]_0^{\pi/2} + \int_0^{\pi/2} (n-1) \sin^{n-2} x \, \cos^2 x \, \mathrm{d}x \\
&= 0 + (n-1) \int_0^{\pi/2} \sin^{n-2} x \, (1 - \sin^2 x) \, \mathrm{d}x = (n-1)(J_{n-2} - J_n).
\end{aligned}
$$



Now J_0 and J_1 can be found directly as

$$J_0 = \int_0^{\pi/2} \sin^0 x \, \mathrm{d}x = \int_0^{\pi/2} \mathrm{d}x = \frac{\pi}{2}, \quad J_1 = \int_0^{\pi/2} \sin x \, \mathrm{d}x = \Big[-\cos x \Big]_0^{\pi/2} = 1.$$

Therefore, inductively we get

$$J_2 = \frac{1}{2} J_0 = \frac{\pi}{4}, \qquad J_3 = \frac{2}{3} J_1 = \frac{2}{3},$$

$$J_4 = \frac{3}{4} J_2 = \frac{3}{4} \cdot \frac{\pi}{4} = \frac{3\pi}{16}, \qquad J_5 = \frac{4}{5} J_3 = \frac{4}{5} \cdot \frac{2}{3} = \frac{8}{15}.$$

Even and odd functions

- a function $f(x)$ is an even function iff $f(-x) = f(x)$ for all x.
- a function $f(x)$ is an odd function iff $f(-x) = -f(x)$ for all x.

For an even or odd function $f(x)$, the integral over an interval of the form $[-a, a]$ (where $a > 0$) can be simplified. By considering the area of the graph of $f(x)$, the following results are obvious.

Theorem 9.2

- *If $f(x)$ is even then $\int_{-a}^{a} f(x)\,\mathrm{d}x = 2\int_{0}^{a} f(x)\,\mathrm{d}x$.*
- *If $f(x)$ is odd then $\int_{-a}^{a} f(x)\,\mathrm{d}x = 0$.*

Examples

Example 9.14 Find the integral $I = \int_{-1}^{1} (x^4 - 3x^2 + 1)\,\mathrm{d}x$.

Solution. Let $f(x) = x^4 - 3x^2 + 1$. Then $f(-x) = (-x)^4 - 3(-x)^2 + 1 = f(x)$. Therefore $f(x)$ is even. By Theorem 9.2, we have

$$\begin{aligned} I &= \int_{-1}^{1} (x^4 - 3x^2 + 1)\,\mathrm{d}x = 2\int_{0}^{1} (x^4 - 3x^2 + 1)\,\mathrm{d}x \\ &= 2\left[\frac{x^5}{5} - x^3 + x\right]_0^1 = 2\left(\frac{1}{5} - 1 + 1\right) = \frac{2}{5}. \end{aligned}$$

□

Example 9.15 Find the integral $I = \int_{-1}^{1} (x^3 - 6x)\,\mathrm{d}x$.

Solution. Let $f(x) = x^3 - 6x$. Then $f(-x) = (-x)^3 - 6(-x) = -f(x)$. Therefore $f(x)$ is odd. By Theorem 9.2, we have $I = 0$. $\square$

Example 9.16 Find the integral $I = \int_{-2}^{2} x^3 \cos 5x \, \mathrm{d}x$.

Solution. Let $f(x) = x^3 \cos 5x$. Then

$$f(-x) = (-x)^3 \cos(-5x) = -x^3 \cos 5x = -f(x).$$

Therefore $f(x)$ is odd and hence $I = 0$. $\square$

Example 9.17 Find the integral $I = \int_{-4}^{4} \tan(x^3 - 6x)\,\mathrm{d}x$.

Solution. Let $f(x) = \tan(x^3 - 6x)$. Then

$$f(-x) = \tan[(-x)^3 - 6(-x)] = \tan(-x^3 + 6x) = -\tan(x^3 - 6x) = -f(x).$$

Therefore $f(x)$ is odd and hence $I = 0$. □

Example 9.18 Find the integral $I = \int_{-2}^{2} (x^2 + \sin x)\,\mathrm{d}x$.

Solution. Since x^2 is even and $\sin x$ is odd, we have

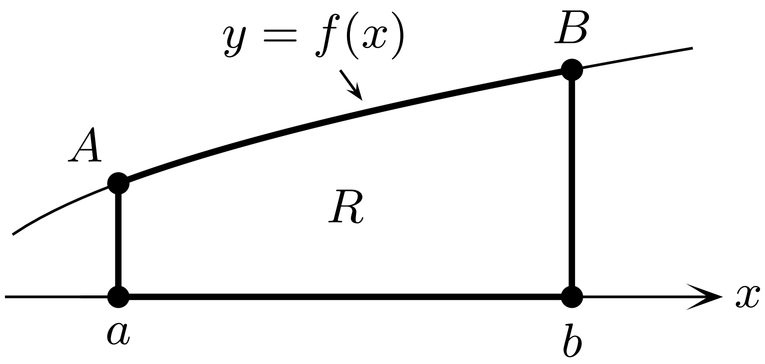
$$I = \int_{-2}^{2} (x^2 + \sin x)\,\mathrm{d}x = 2\int_{0}^{2} x^2\,\mathrm{d}x + 0 = 2\left[\frac{x^3}{3}\right]_0^2 = \frac{16}{3}.$$

□

Applications of Definite Integral

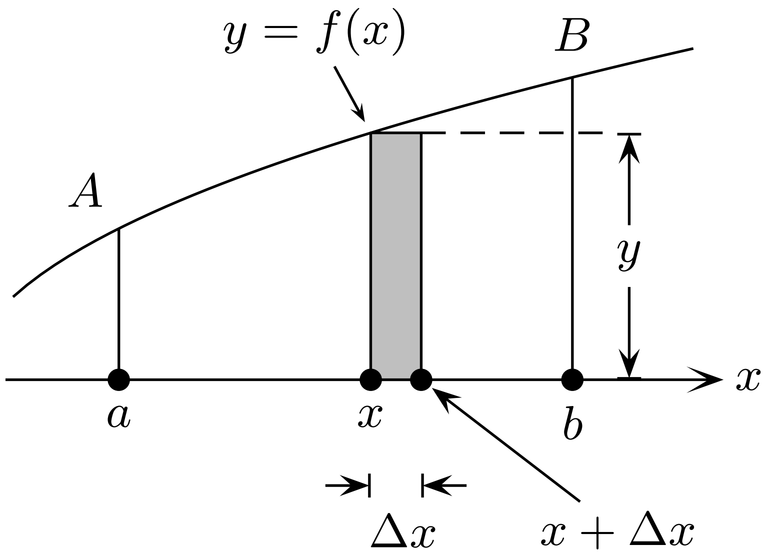
Area bounded by curves

Suppose that $f(x) \geq 0$ on $[a, b]$ and that we want to find the area A of a region R bounded by the graph of $y = f(x)$ and the x-axis over the interval $[a, b]$.



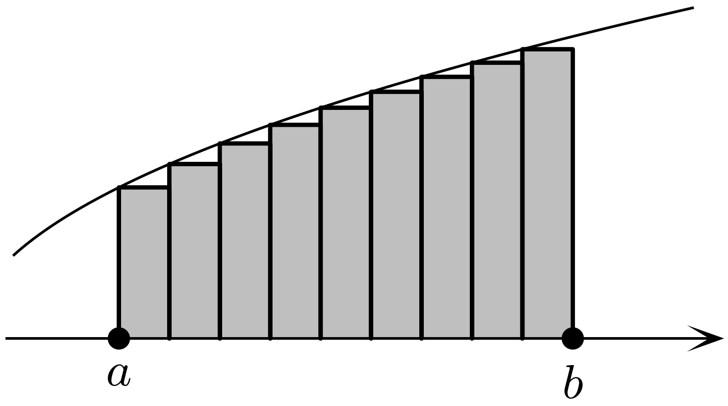
Area of rectangle

First we consider two nearby points x and $x + \Delta x$ in the interval $[a, b]$ with $\Delta x > 0$. With height equal to $y(= f(x))$, a rectangle of small width Δx is drawn standing on the interval $[x, x + \Delta x]$. The area of this rectangle is $f(x)\,\Delta x$.



Riemann sum

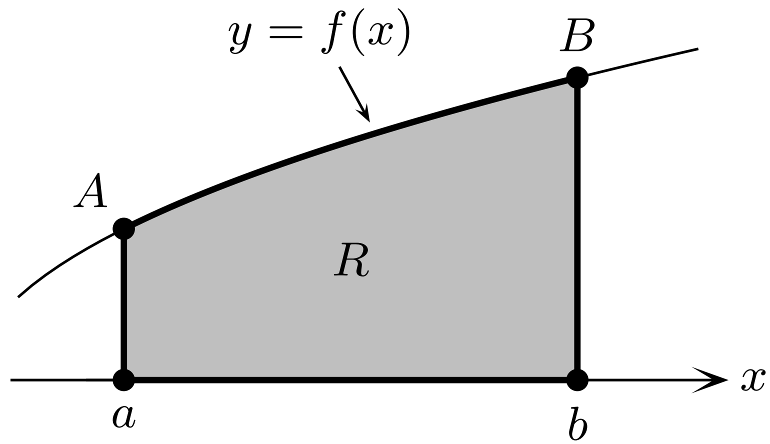
If we partition the interval $[a, b]$ into N subintervals of equal length Δx, we can form N rectangles as described above. The sum of all the areas of such rectangles therefore forms a Riemann sum.



Taking limit

To get the area of R, we let $\Delta x \to 0$ and hence obtain the formula:

$$\boxed{\text{Area of the region } R \text{ is } \int_a^b f(x)\,\mathrm{d}x.} \tag{10.1}$$

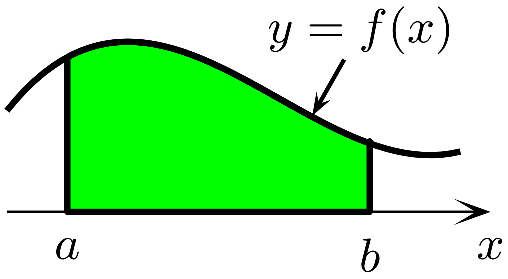


Area of R is

$\lim \sum y\,\Delta x = \int_a^b y\,\mathrm{d}x.$

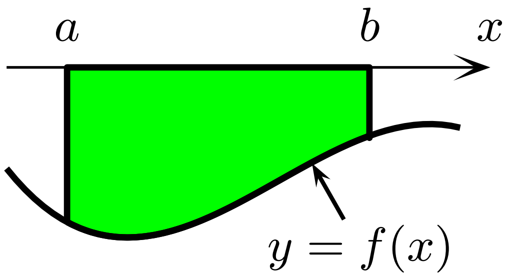
Modifications

R is bounded by the curve $y = f(x)$ ($f(x) \geq 0$) and the x-axis over $[a, b]$.



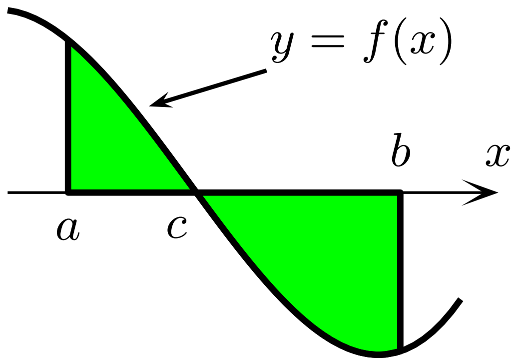
$$A = \int_a^b f(x)\,\mathrm{d}x$$

R is bounded by the curve $y = f(x)$ $(f(x) \leq 0)$ and the x-axis over $[a, b]$.



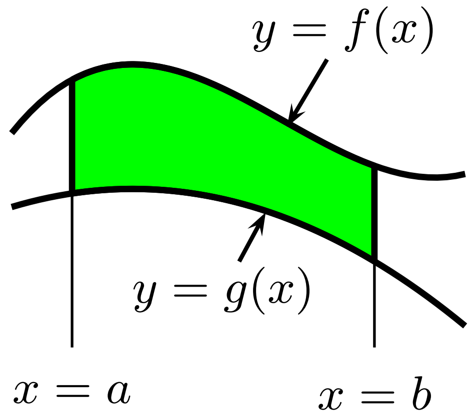
$$A = -\int_a^b f(x)\,\mathrm{d}x$$

R is bounded by the curve $y = f(x)$ and the x-axis over $[a, b]$. $a < c < b$. $f(x) \geq 0$ on $[a, c]$ and $f(x) \leq 0$ on $[c, b]$.



$$A = \int_a^c f(x)\,\mathrm{d}x - \int_c^b f(x)\,\mathrm{d}x$$

R is bounded by the curves $y = f(x)$, $y = g(x)$ over $[a, b]$ with $f(x) \geq g(x)$ on $[a, b]$.



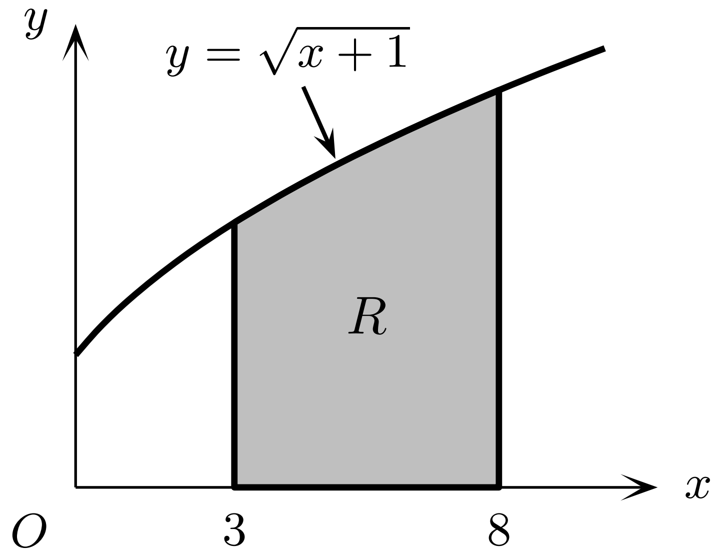
$$A = \int_a^b [f(x) - g(x)]\,\mathrm{d}x$$

Examples

Example 10.1 Find the area A of the region R between the curve $y = \sqrt{x+1}$ and the x-axis over the interval $3 \le x \le 8$.

Solution. The function $\sqrt{x+1}$ is positive on the interval $[3, 8]$. Therefore

$$A = \int_3^8 \sqrt{x+1}\,\mathrm{d}x = \frac{2}{3}\Big[(x+1)^{3/2}\Big]_3^8 = \frac{2}{3}\Big[9^{3/2} - 4^{3/2}\Big] = \frac{38}{3}.$$

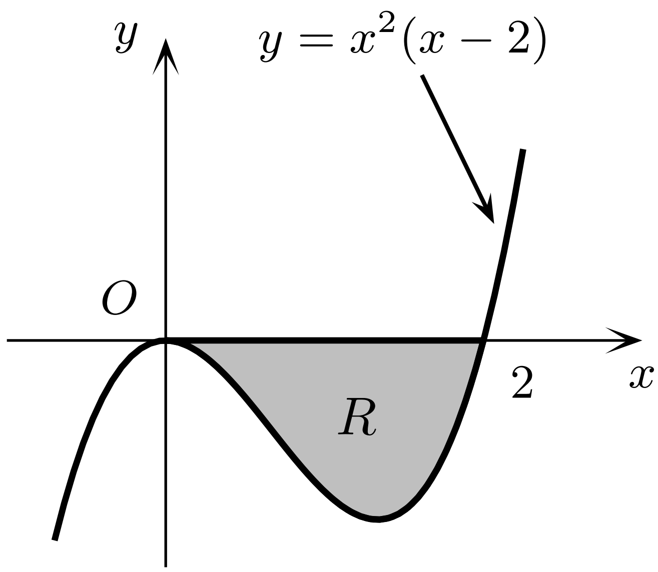


□

Example 10.2 Find the area A of the region R bounded by the curve $y = x^2(x-2)$ and the x-axis.

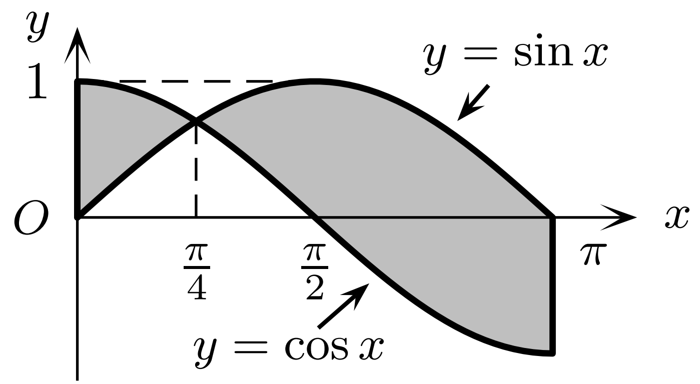
Solution. The curve $y = x^2(x-2)$ intersects the x-axis at only two points where $x = 0$ and $x = 2$. The given region therefore lies between $x = 0$ and $x = 2$ and is sketched in the diagram. Over the interval $[0, 2]$ the function $y = x^2(x-2)$ is non-positive. Therefore

$$A = -\int_0^2 x^2(x-2)\,\mathrm{d}x = -\int_0^2 (x^3 - 2x^2)\,\mathrm{d}x = -\left[\frac{x^4}{4} - \frac{2x^3}{3}\right]_0^2 = -\frac{2^4}{4} + \frac{2^4}{3} = \frac{4}{3}.$$



Example 10.3 Find the area between the curves $y = \sin x$ and $y = \cos x$ over the interval $0 \le x \le \pi$.

Solution.



The curves intersect when $\sin x = \cos x$. The solution of this equation is $x = \pi/4$. Since $\cos x \ge \sin x$ on $0 \le x \le \pi/4$ and $\cos x \le \sin x$ on $\pi/4 \le x \le \pi$,

$$\text{the area} = \int_0^{\pi/4} (\cos x - \sin x)\,\mathrm{d}x + \int_{\pi/4}^{\pi} (\sin x - \cos x)\,\mathrm{d}x$$

$$= \Big[\sin x + \cos x\Big]_0^{\pi/4} + \Big[-\cos x - \sin x\Big]_{\pi/4}^{\pi} = 2\sqrt{2}.$$

□

Volume of revolution

A solid can be generated in space when a plane area is rotated about an axis. This solid is called a *solid of revolution*. For example, if the area of a semi-circle is rotated about its diameter it will generate a solid sphere.

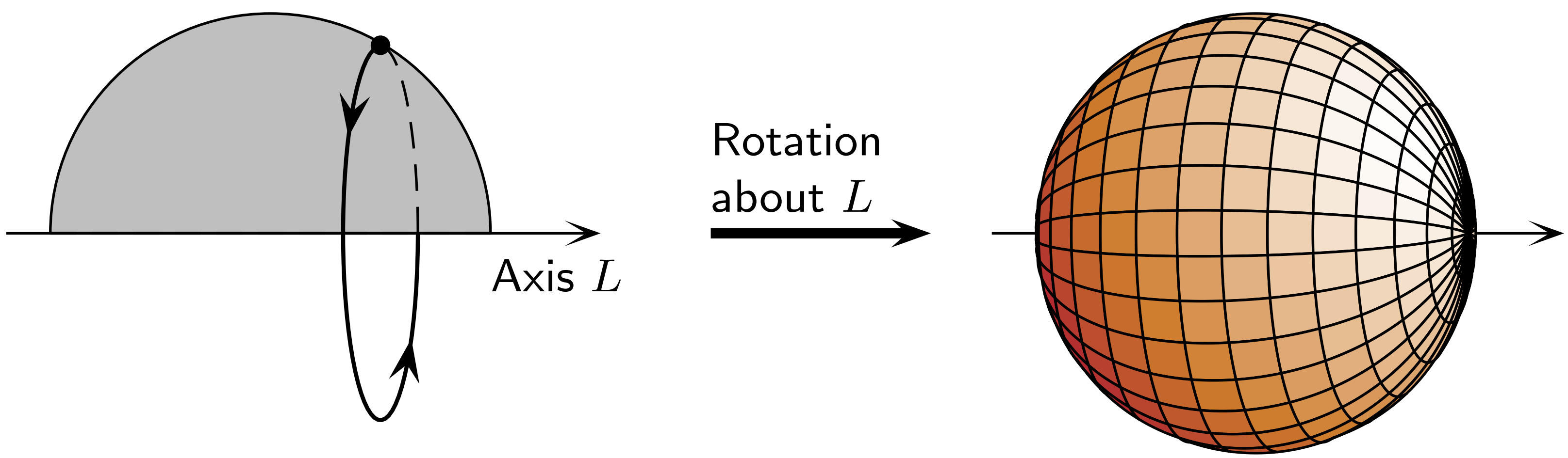
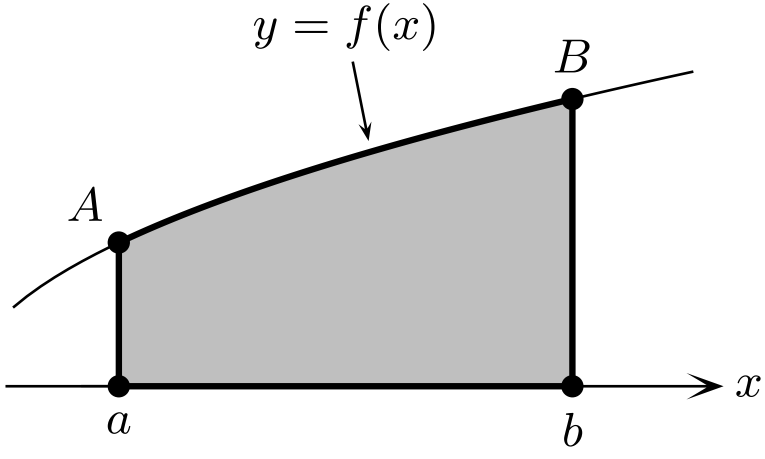


Figure 10.1: Rotation of a semi-circular region about its diameter to get a solid sphere.

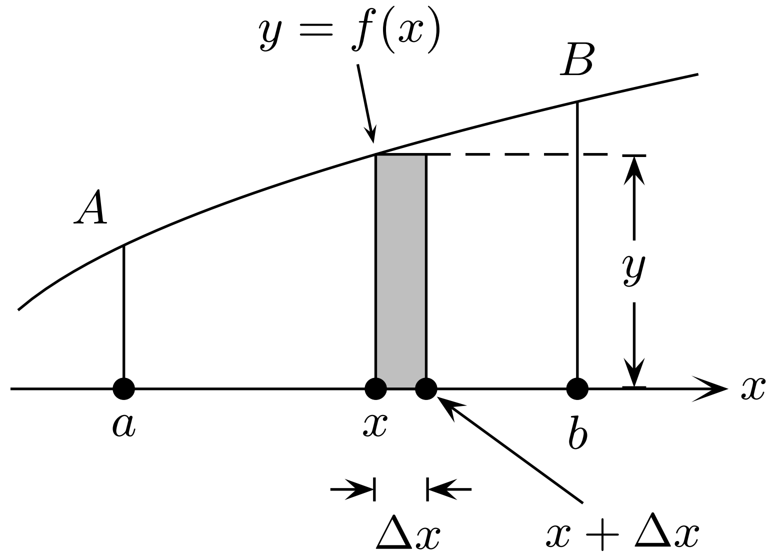
Step 1: Area under an arc

Suppose that $f(x) \geq 0$ on $[a, b]$. Consider the part $\widehat{AB}$ of the curve $y = f(x)$ between $x = a$ and $x = b$.



Step 2: A rectangle of width Δx

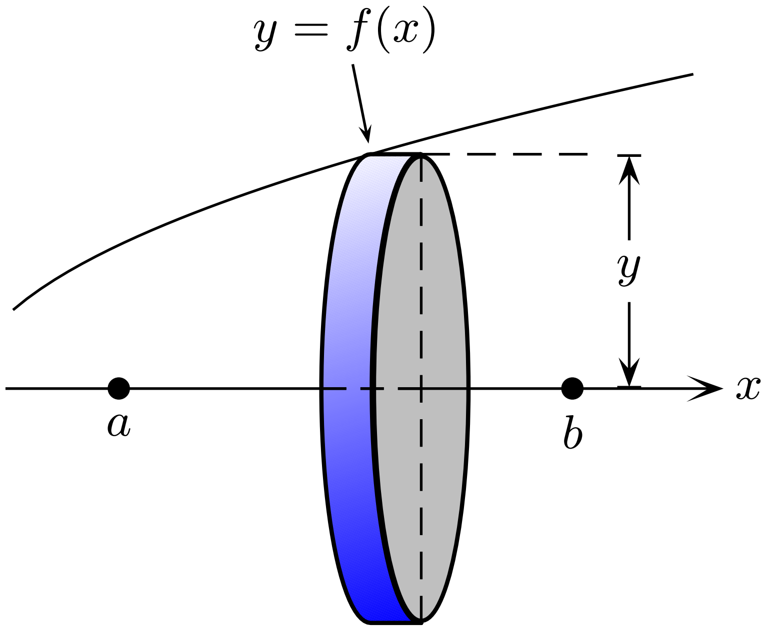
Consider two nearby points x and $x + \Delta x$ in the interval $[a, b]$ with $\Delta x > 0$. With height equal to $y = f(x)$, a rectangle of width Δx is standing on the interval $[x, x + \Delta x]$.



Area of rectangle $= y\Delta x$.

Step 3: Revolving the rectangle

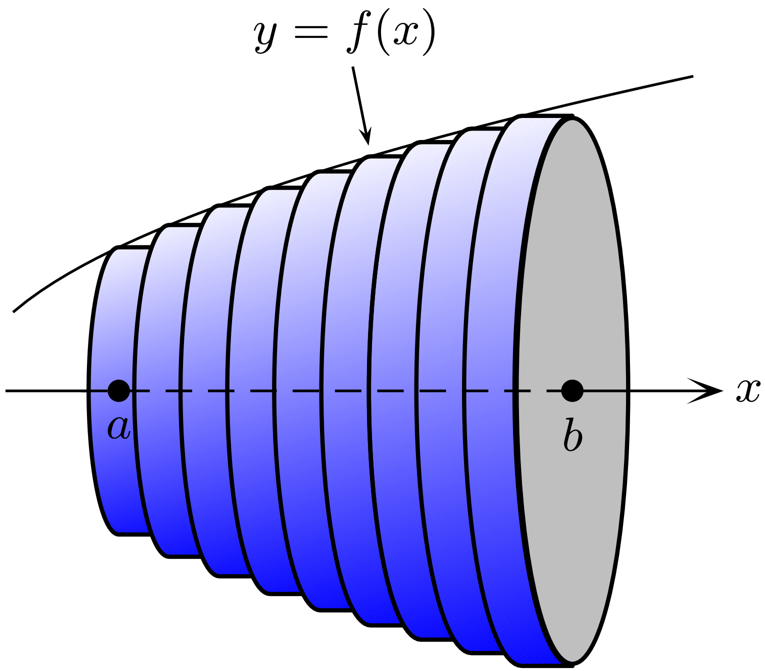
On rotation (through 360° about the x-axis), the rectangle generates a thin solid disk of thickness Δx (see Box 3 of Fig. 7.4.) The volume of the disc is $\pi y^2 \Delta x$ where $y = f(x) \geq 0$.



Volume of disc $= \pi y^2 \Delta x$.

Step 4: Add the volumes of elements

If we partition the interval $[a, b]$ into N subintervals of equal length Δx, we can form N discs as described above. The sum of the volumes of all these discs is $\sum \pi y^2 \, \Delta x$.



Total volume of disc $= \sum \pi y^2 \Delta x$.

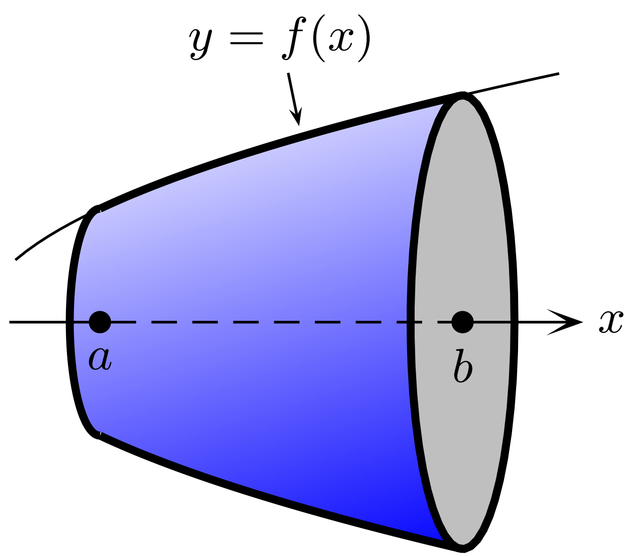
Step 5: Taking limit as $\Delta x \to 0$

The volume V of revolution is the limit as $\Delta x \to 0$ of the sum of all the volumes of the discs, i.e.

$$V = \lim_{\Delta x \to 0} \sum \pi y^2 \, \Delta x.$$

This gives the formula for the volume of revolution about the x-axis:

$$\boxed{V = \pi \int_a^b [f(x)]^2 \, \mathrm{d}x.} \qquad (10.2)$$



Volume of revolution $= \lim \sum \pi y^2 \Delta x = \pi \int_a^b y^2 \, \mathrm{d}x.$

Examples

Example 10.4 Find the volume of revolution about the x-axis of the region R which is bounded by the curve $y = \sin x$, the x-axis, the vertical lines $x = 0$ and $x = 2$.

Solution. Here $f(x) = \sin x$ and $f(x) \geq 0$ on $[0, 2]$. Therefore by (10.2) the volume is

$$\begin{aligned} V &= \pi \int_0^2 \sin^2 x \, \mathrm{d}x = \frac{\pi}{2} \int_0^2 (1 - \cos 2x) \, \mathrm{d}x \\ &= \frac{\pi}{2} \left[x - \frac{\sin 2x}{2} \right]_0^2 = \frac{\pi}{2} \left(2 - \frac{\sin 4}{2} \right) \\ &\approx 3.736. \end{aligned}$$

□

Volume of revolution bounded between curves

If the region R is bounded between two curves $y = f(x)$ and $y = g(x)$ with $f(x) \geq g(x) \geq 0$ on $[a, b]$, then by the same principle, we can get the following formula for the volume of revolution of R about the x-axis:

$$\boxed{V = \pi \int_a^b (f^2 - g^2)\,\mathrm{d}x.} \tag{10.3}$$

Example 10.5 Find the volume of revolution about the x-axis of the region R which is bounded by the curves $y = \sin x$ and $y = \cos x$ over the interval $[0, \pi/4]$.

Solution. $\cos x \geq \sin x$ on $[0, \pi/4]$. Therefore by (10.3) the volume is

$$V = \pi \int_0^{\pi/4} (\cos^2 x - \sin^2 x)\,\mathrm{d}x = \pi \int_0^{\pi/4} \cos 2x\,\mathrm{d}x$$
$$= \frac{\pi}{2}\Big[\sin 2x\Big]_0^{\pi/4} = \frac{\pi}{2}.$$

See CoCalc □
https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page361-CoCalcJupyter.pdf

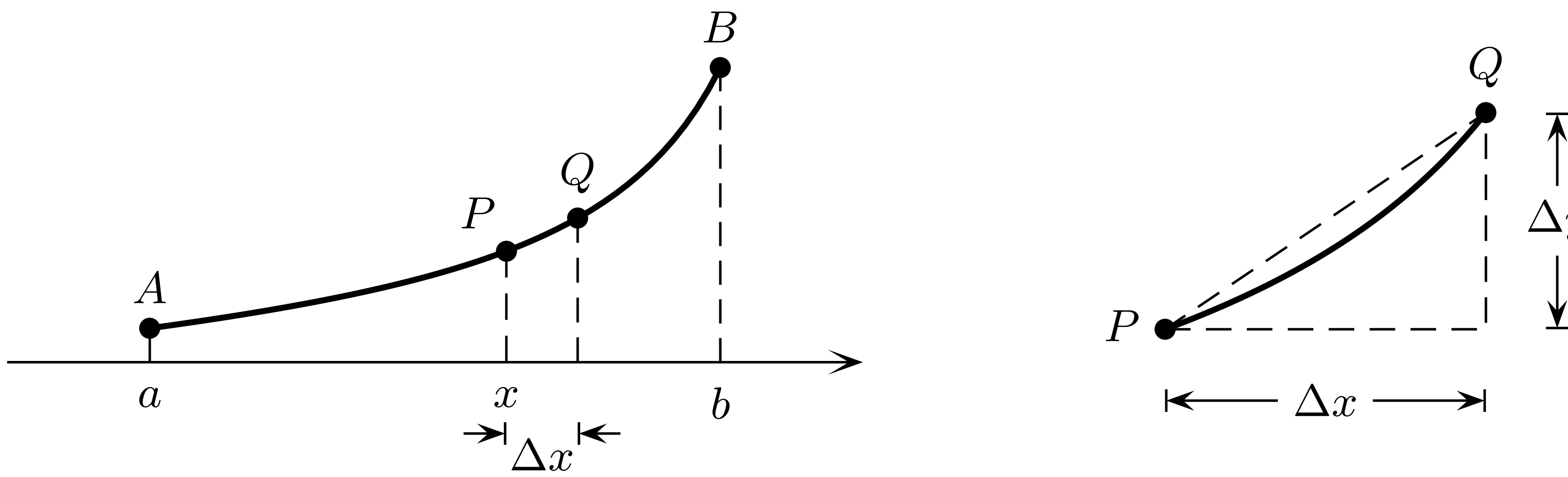
Cylindrical Shell Method to find the volume of rotation

https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/supplementary06.pdf

See CoCalc https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/supplementary06_CoCalcJupyter.pdf

Length of curves

Let $y = f(x)$ be the equation of the curve and consider the part $\widehat{AB}$ of the curve between $x = a$ and $x = b$.
Let $P(x, y)$ and $Q(x + \Delta x, y + \Delta y)$ be two points on $\widehat{AB}$ with $\Delta x > 0$ and small.
Let s be the length of the arc from A to P and $s + \Delta s$ be the length from A to Q.
Then the length of the arc $\widehat{PQ}$ is Δs.



(a) $\Delta s =$ length of $\widehat{PQ}$ (b) $\Delta s \approx PQ$

Figure 10.2: Getting the formula for the arc length.

By Pythagoras' theorem,

$$\Delta s \approx PQ = \sqrt{(\Delta x)^2 + (\Delta y)^2}.$$

Therefore, the length of the arc $\widehat{AB}$ is

$$L = \lim_{\Delta x \to 0} \sum \sqrt{(\Delta x)^2 + (\Delta y)^2} = \lim_{\Delta x \to 0} \sum \sqrt{1 + \left(\frac{\Delta y}{\Delta x}\right)^2}\, \Delta x.$$

This gives the formula

$$\boxed{\text{Arc length of } \widehat{AB} = \int_a^b \sqrt{1 + \left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2}\, \mathrm{d}x} \tag{10.4}$$

Examples

Example 10.6 Find the length of the arc of the parabola $y = \frac{1}{2}x^2$ between $x = 0$ and $x = 1$.

Solution. $y = \frac{1}{2}x^2$. Therefore $\dfrac{\mathrm{d}y}{\mathrm{d}x} = x$. Hence by (10.4) the arc length is

$$\begin{aligned} L &= \int_0^1 \sqrt{1 + \left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2}\,\mathrm{d}x = \int_0^1 \sqrt{1 + x^2}\,\mathrm{d}x \\ &= \frac{1}{2}\left[x\sqrt{1 + x^2} + \ln\left|\frac{x + \sqrt{x^2 + 1}}{2}\right|\right]_0^1 \\ &= \frac{1}{2}[\sqrt{2} + \ln(1 + \sqrt{2})] \approx 1.1478. \end{aligned}$$

□

Example 10.7 Use (10.4) to find the length of the quarter-circle $y = \sqrt{a^2 - x^2}$, $(0 \le x \le a)$ of radius a.

Solution. $y = \sqrt{a^2 - x^2}$. Therefore $\dfrac{\mathrm{d}y}{\mathrm{d}x} = \dfrac{-x}{\sqrt{a^2 - x^2}}$. Hence by (10.4) the arc length of the quarter-circle is

$$\begin{aligned} L &= \int_0^a \sqrt{1 + \left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2}\,\mathrm{d}x = \int_0^a \sqrt{1 + \frac{x^2}{a^2 - x^2}}\,\mathrm{d}x \\ &= \int_0^a \frac{a}{\sqrt{a^2 - x^2}}\,\mathrm{d}x = a\left[\sin^{-1}\frac{x}{a}\right]_0^a \\ &= \frac{\pi a}{2}. \end{aligned}$$

See CoCalc
https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page365-CoCalcJupyter.pdf □

To find the area of the surface by rotation

https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/supplementary05.pdf

See CoCalc https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/supplementary05_CoCalcJupyter.pdf

Improper Integrals

Improper Integrals of Type 1

Definition

(1) If $\int_a^t f(x)dx$ exists for every number $t \geq a$, then,

$\int_a^\infty f(x)dx = \lim_{t\to\infty} \int_a^t f(x)dx$ provided this limit exists (as a finite number).

(2) If $\int_t^b f(x)dx$ exists for every number $t \leq b$, then,

$\int_{-\infty}^b f(x)dx = \lim_{t\to-\infty} \int_t^b f(x)dx$ provided this limit exists (as a finite number).

The improper integrals $\int_a^t f(x)dx$ and $\int_t^b f(x)dx$ are called **convergent** if the corresponding limit exists, and **divergent** if the limit does not exist.

(3) If both $\int_a^\infty f(x)dx$ and $\int_{-\infty}^a f(x)dx$ are convergent, then we define

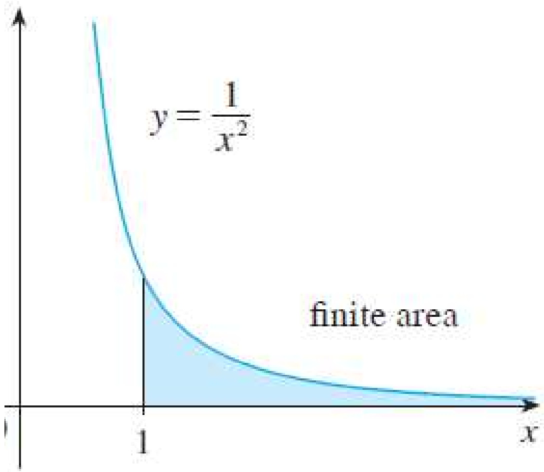
$$\int_{-\infty}^\infty f(x)dx = \int_a^\infty f(x)dx + \int_{-\infty}^a f(x)dx$$

for any real number a.

Example Determine the convergence the integrals $\int_1^\infty \frac{1}{x^2}dx$ and $\int_1^\infty \frac{1}{x}dx$.
See CoCalc
https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page368-CoCalcJupyter.pdf
Solution

$$\int_1^\infty \frac{1}{x^2}dx = \lim_{t\to\infty}\int_1^t \frac{1}{x^2}dx = \lim_{t\to\infty} -\frac{1}{x}\Big]_1^t = \lim_{t\to\infty}\left(1-\frac{1}{t}\right) = 1.$$

$$\int_1^\infty \frac{1}{x}dx = \lim_{t\to\infty}\int_1^t \frac{1}{x}dx = \lim_{t\to\infty} \ln|x|\Big]_1^t = \lim_{t\to\infty}(\ln t - \ln 1) = \lim_{t\to\infty}\ln t = \infty$$



URE 4 $\int_1^\infty (1/x^2)\,dx$ converges

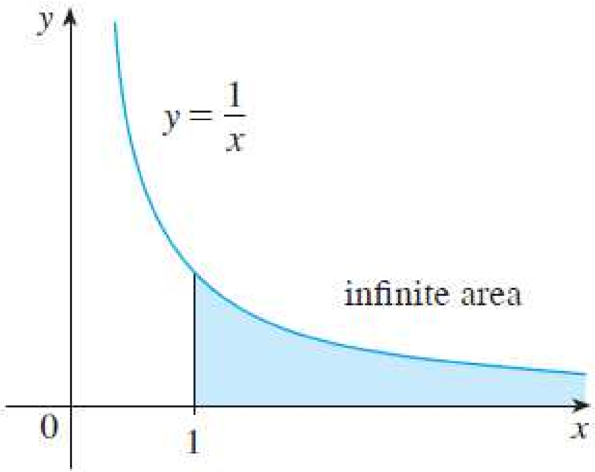


FIGURE 5 $\int_1^\infty (1/x)\,dx$ diverges

Theorem $\displaystyle\int_1^\infty \frac{1}{x^p}dx$ is convergent if $p > 1$, and divergent if $p \leq 1$.

Proof

$$\int_1^\infty \frac{1}{x^p}dx = \lim_{t\to\infty}\int_1^t x^{-p}dx = \lim_{t\to\infty} \left.\frac{x^{-p+1}}{-p+1}\right]_{x=1}^{x=t} = \lim_{t\to\infty} \frac{1}{1-p}\left[\frac{1}{t^{p-1}} - 1\right]$$

If $p > 1$, then, $p - 1 > 0$, so as $t \to \infty$, $t^{p-1} \to \infty$ and $\frac{1}{t^{p-1}} \to 0$.

Therefore $\displaystyle\int_1^\infty \frac{1}{x^p}dx = \frac{1}{p-1}$, and so the integral converges for $p > 1$.

But if $p < 1$, then, $p - 1 < 0$, and so $\displaystyle\frac{1}{t^{p-1}} = t^{1-p} \to \infty$ as $t \to \infty$, and the integral diverges.

Improper Integrals of Type 2

Definition

(1) If f is continuous on $[a, b)$ and is discontinuous at b, then
$$\int_a^b f(x)dx = \lim_{t \to b^-} \int_a^t f(x)dx \quad \text{if this limit exists (as a finite number).}$$

(2) If f is continuous on $(a, b]$ and is discontinuous at a, then
$$\int_a^b f(x)dx = \lim_{t \to a^+} \int_t^b f(x)dx \quad \text{if this limit exists (as a finite number).}$$

The improper integral $\displaystyle\int_a^b f(x)dx$ is called **convergent** if the corresponding limit exists, and **divergent** if the limit does not exist.

(3) If f has a discontinuity at c, where $a < c < b$, and both $\int_a^c f(x)dx$ and $\int_c^b f(x)dx$ are convergent, then we define
$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx.$$

Example Evaluate $\int_0^1 \ln x dx$.

Solution Function $f(x) = \ln x$ has a vertical asymptote at 0, $\lim_{x\to 0^+} \ln x = -\infty$. Thus, it is an improper integral

$$\int_0^1 \ln x dx = \lim_{t\to 0^+} \int_t^1 \ln x dx$$

Using Integration by Parts with $u = \ln x$, $v = x$, and thus, $du = \frac{dx}{x}$, $dv = dx$

$$\int_t^1 \ln x dx = [x \ln x]_t^1 - \int_t^1 dx = 1 \ln 1 - t \ln t - (1 - t) = -t \ln t - 1 + t.$$

To find the limit of the first term we use L'Hospital's Rule

$$\lim_{t\to 0^+} t \ln t = \lim_{t\to 0^+} \frac{\ln t}{1/t} = \lim_{t\to 0^+} \frac{1/t}{-1/t^2} = \lim_{t\to 0^+} (-t) = 0.$$

Therefore, $\int_0^1 \ln x dx = \lim_{t\to 0^+} (-t \ln t - 1 + t) = -0 - 1 + 0 = -1.$

See CoCalc

https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page371-CoCalcJupyter.pdf

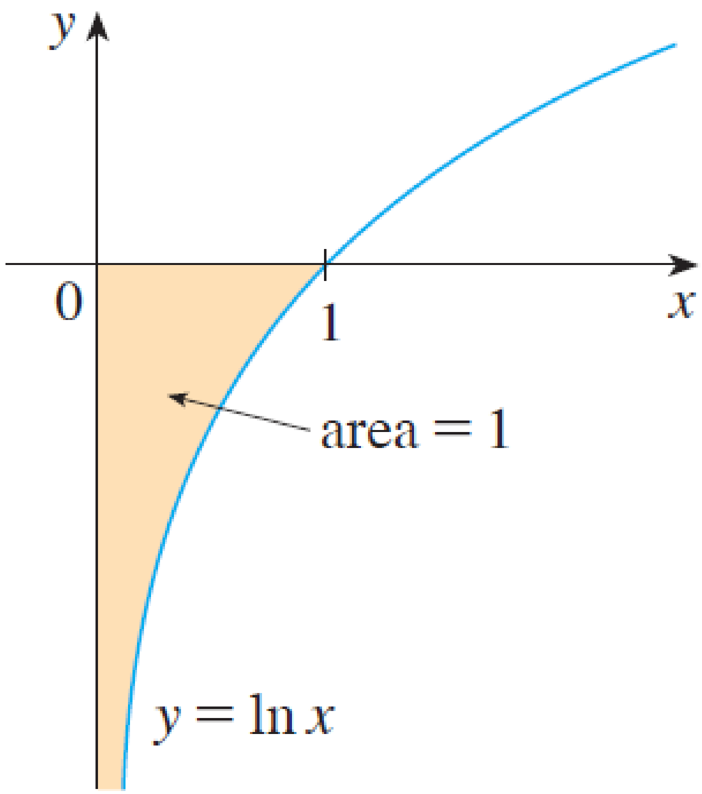


FIGURE 11

Infinite Series

Infinite series

An infinite series (or simply series) is an expression of the form

$$a_1 + a_2 + a_3 + \cdots$$

where a_1, a_2, $a_3, \ldots$ are real numbers. It is convenient to use the notation $\sum_{n=1}^{\infty} a_n$ to represent the series. For every positive integer n, we define the n-th partial sum of the series by $S_n = a_1 + \cdots + a_n$.

Definition: The infinite series $\sum_{n=1}^{\infty} a_n$ is said to be convergent if $\lim_{n\to\infty} S_n$ exists (and is finite). This limit is called the sum of the series. A series which is not convergent is called a divergent series.

Example: Consider the series $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n(n+1)}$. Observe that

$$\begin{aligned} S_n &= \frac{1}{1\times 2} + \frac{1}{2\times 3} + \cdots + \frac{1}{n\times(n+1)} \\ &= \left(1 - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \cdots + \left(\frac{1}{n} - \frac{1}{n+1}\right) = 1 - \frac{1}{n+1}, \end{aligned}$$

which tends to 1 as $n \to \infty$. Therefore, the series is convergent with its sum equal to 1.

See CoCalc

https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page374-CoCalcJupyter.pdf

Example: The series $\sum_{n=0}^{\infty} \alpha^n = 1 + \alpha + \alpha^2 + \alpha^3 + \cdots$, where α is a given real number, is called a geometric series with common ratio α. Since

$$S_n = 1 + \alpha + \alpha^2 + \cdots + \alpha^{n-1} = \frac{1 - \alpha^n}{1 - \alpha} \text{ (if } \alpha \neq 1),$$

we conclude that the series converges to $\dfrac{1}{1-\alpha}$ if $-1 < \alpha < 1$. On the other hand, if $|\alpha| \geq 1$, then $\sum_{n=0}^{\infty} \alpha^n$ is divergent.

Example: $\displaystyle\sum_{n=0}^{\infty} \frac{1}{3^n}$ is a geometric series with common ratio $\alpha = \dfrac{1}{3}$. Thus, the series is convergent with its sum equal to $\dfrac{3}{2}$.

The following facts on convergence of series follow from the corresponding results on limit of sequences.

Proposition: If $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ are convergent series and t is any scalar, then both $\sum_{n=1}^{\infty}(a_n + b_n)$ and $\sum_{n=1}^{\infty} t\, a_n$ are convergent. Moreover, one has

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

and

$$\sum_{n=1}^{\infty} t\, a_n = t \sum_{n=1}^{\infty} a_n.$$

We have the following necessary condition for the convergence of a series.

Theorem: If $\sum_{n=1}^{\infty} a_n$ is convergent, then $a_n \to 0$ as $n \to \infty$.

Proof: Convergence of the series $\sum_{n=1}^{\infty} a_n \Rightarrow \lim_{n\to\infty} S_n = \lim_{n\to\infty} S_{n-1}$. Since $a_n = S_n - S_{n-1}$, one concludes that

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} (S_n - S_{n-1}) = \lim_{n\to\infty} S_n - \lim_{n\to\infty} S_{n-1} = 0.$$

This Theorem tells us that if the general term of a series does not tend to zero, then the series must be divergent. For instance, the series $\displaystyle\sum_{n=1}^{\infty} \frac{n}{n+2}$ is divergent since $\displaystyle\lim_{n\to\infty} \frac{n}{n+2} \neq 0$. However, it should be emphasized that this necessary condition is not *sufficient* for the convergence of a series. For example, even though $\displaystyle\frac{1}{n} \to 0$, the series $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n}$ is divergent.

Test of convergence

Positive series
In general, it is very difficult to tell whether a given series is convergent or otherwise. In this connection, "positive series" are a lot easier to handle.

Definition: A series $\sum_{n=1}^{\infty} a_n$ is said to be positive if $a_n \geq 0$ for every n.

It is evident that the partial sums of a positive series form a monotonically increasing sequence of non-negative real numbers, i.e., they satisfy the inequalities $0 \leq S_1 \leq S_2 \leq S_3 \leq \cdots$. This simple observation proves the following

Theorem: A positive series $\sum_{n=1}^{\infty} a_n$ is convergent if and only if $\{S_n\}$ is bounded.

Using this Theorem, we obtain the following tests of convergence for positive series.

The Comparison Test

Suppose $\sum_{n=1}^{\infty} a_n$ is a positive series and that $a_n \leq b_n$ for all except finitely many positive integer n. If $\sum_{n=1}^{\infty} b_n$ is known to be convergent, then $\sum_{n=1}^{\infty} a_n$ is also convergent.

Example Since $n^2 \geq n^2 - n = n(n-1)$ for any $n \geq 2$, we have $0 < \dfrac{1}{n^2} \leq \dfrac{1}{(n-1)n}$.

Since $\sum\limits_{n=2}^{\infty} \dfrac{1}{(n-1)n}$ is convergent, the infinite series $\displaystyle\sum_{n=2}^{\infty} \frac{1}{n^2}$ is convergent by the Comparison Test. Thus, $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^2} = 1 + \sum_{n=2}^{\infty} \frac{1}{n^2}$ is convergent.

See CoCalc

https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page379-CoCalcJupyter.pdf

The Integral Test

Suppose $f(x)$ is a decreasing and nonnegative function for $x \geq 1$. If $f(n) = a_n$ for every positive integer n, then the series $\sum_{n=1}^{\infty} a_n$ is convergent if and only if

$$\lim_{n\to\infty} \int_1^n f(x)\, dx < \infty.$$

Example Test the series $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^2+1}$ for convergence or divergence.

Solution The function $f(x) = \frac{1}{x^2+1}$ is continuous, positive, and decreasing on $[1, \infty)$ so we use Integral Test:

$$\int_1^{\infty} \frac{1}{x^2+1}\, dx = \lim_{t\to\infty} \int_1^t \frac{1}{x^2+1}\, dx = \lim_{t\to\infty} \tan^{-1}\Big|_1^t = \lim_{t\to\infty} \left(\tan^{-1} t - \frac{\pi}{4}\right)$$
$$= \frac{\pi}{2} - \frac{\pi}{4} = \frac{\pi}{4}.$$

Thus $\displaystyle\int_1^{\infty} \frac{1}{x^2+1} dx$ is a convergent integral and so, by the Integral Test, the series $\displaystyle\sum^{\infty} \frac{1}{n^2+1}$ is convergent.

The Ratio Test

Suppose $a_n > 0$ and let $\rho = \lim_{n\to\infty} \dfrac{a_{n+1}}{a_n}$. Then

1. If $\rho < 1$, $\sum_{n=1}^{\infty} a_n$ is convergent;
2. If $\rho > 1$, $\sum_{n=1}^{\infty} a_n$ is divergent;
3. If $\rho = 1$, the ratio test is inconclusive.

Example Test the convergence of the series $\displaystyle\sum_{n=1}^{\infty} \frac{n^n}{n!}$.

Solution

$$\frac{a_{n+1}}{a_n} = \frac{(n+1)^{n+1}}{(n+1)!}\,\frac{n!}{n^n} = \frac{(n+1)(n+1)^n}{(n+1)n!}\,\frac{n!}{n^n} = \left(\frac{n+1}{n}\right)^n = \left(1+\frac{1}{n}\right)^n.$$

Since $\displaystyle\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = e > 1$, the given series is divergent by the Ratio Test.

Absolute convergence

Convergence of non-positive series is much more difficult to handle. Fortunately, in most applications, the more useful series are "*absolutely convergent*".

Definition $\sum_{n=1}^{\infty} a_n$ is said to be absolutely convergent if $\sum_{n=1}^{\infty} |a_n|$ is convergent.

Theorem If $\sum_{n=1}^{\infty} |a_n|$ is convergent, then $\sum_{n=1}^{\infty} a_n$ is also convergent. In other words, an absolutely convergent series is convergent.

Proof We define two sequences $\{b_n\}$ and $\{c_n\}$ as follows:

$$b_n = \begin{cases} a_n, & \text{if } a_n \geq 0; \\ 0, & \text{if } a_n < 0. \end{cases} \quad \text{and } c_n = \begin{cases} 0, & \text{if } a_n \geq 0; \\ -a_n, & \text{if } a_n < 0. \end{cases}$$

It is clear that $\{b_n\}$ and $\{c_n\}$ are positive sequences such that $a_n = b_n - c_n$. Furthermore, the inequalities $0 \leq b_n \leq |a_n|$ and $0 \leq c_n \leq |a_n|$, together with the convergence of $\sum_{n=1}^{\infty} |a_n| \Rightarrow$ the convergence of the series $\sum_{n=1}^{\infty} b_n$ and $\sum_{n=1}^{\infty} c_n$. Therefore, $\sum_{n=1}^{\infty} a_n$ is convergent.

Example Since $\sum_{n=1}^{\infty} \frac{1}{n^2}$ is convergent, the Theorem implies the convergence of the non-positive series

$$1 + \frac{1}{2^2} - \frac{1}{3^2} + \frac{1}{4^2} + \frac{1}{5^2} - \frac{1}{6^2} + \cdots .$$

Alternating series – Leibniz's test
The converse of the Theorem is false, i.e., a convergent series may not converge absolutely. The following theorem, due to Leibniz, gives useful information regarding convergence of alternating series, $\sum_{n=1}^{\infty}(-1)^{n+1}a_n$ where $a_n \geq 0$.

Theorem (Leibniz's test) Let $\{a_n\}$ be a sequence of real numbers such that

1. $a_n > 0$ for every positive integer n;
2. $a_1 \geq a_2 \geq a_3 \geq a_4 \geq \ldots$ and
3. $a_n \to 0$ as $n \to \infty$.

Then the alternating series $a_1 - a_2 + a_3 - a_4 + \ldots$ is convergent.

Example Leibniz's test implies the convergence of the following series

1. $$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^2} = 1 - \frac{1}{2^2} + \frac{1}{3^2} - \frac{1}{4^2} + \cdots;$$

2. $$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots;$$

3. $$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{\sqrt{n}} = 1 - \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} - \frac{1}{\sqrt{4}} + \cdots;$$

4. $$\sum_{n=2}^{\infty} \frac{(-1)^{n}}{\ln n} = \frac{1}{\ln 2} - \frac{1}{\ln 3} + \frac{1}{\ln 4} - \frac{1}{\ln 5} + \cdots .$$

Note that (1) is absolutely convergent, while the remaining are not.

Power series

Radius of convergence A power series in x is a series of the form

$$\sum_{n=0}^{\infty} a_n\, x^n = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots\ ,$$

where a_0, a_1, a_2, a_3, ... are real constants and x is a real variable. More generally, we may consider power series in $(x - x_0)$, where x_0 is a fixed number. In other words, we may consider

$$\sum_{n=0}^{\infty} a_n\, (x - x_0)^n = a_0 + a_1\, (x - x_0) + a_2\, (x - x_0)^2 + a_3 (x - x_0)^3 + \cdots\ .$$

Note: a power series in $(x - x_0)$ can be transformed into a power series in x by a simple change of variable.

The power series in x may converge for some values of x and diverge for others. For example, every power series in x is convergent when $x = 0$. Using the *Ratio Test*, we have the following useful result governing the range of convergence of a given power series.

Theorem For the power series in x, we define $R = \lim_{n\to\infty} \left|\frac{a_n}{a_{n+1}}\right|$.

1. If $-R < x < R$, then the series converges absolutely;
2. if $|x| > R$, then the series diverges.

Note that $0 \le R \le +\infty$. The number R is called the *radius of convergence* of the power series, and the interval $(-R, R)$ is known as the *interval of convergence*. Note that if $R = +\infty$, then the power series converges for all x.

Example Radii of convergence for the following series are

1. $\sum_{n=0}^{\infty} x^n,\ R = 1;$

2. $\sum_{n=0}^{\infty} nx^n,\ R = 1;$

3. $\sum_{n=0}^{\infty} \frac{x^n}{n+1},\ R = 1;$

4. $\sum_{n=0}^{\infty} \frac{x^n}{n!},\ R = \infty.$

The following theorem says that term-by-term differentiation and integration of a power series is legitimate within its interval of convergence.

Theorem Let R be the radius of convergence of the power series in x, and let f be defined by $f(x) = \sum_{n=0}^{\infty} a_n x^n$ for every x in the interval $(-R, R)$. Then

1. $f'(x) = \sum_{n=1}^{\infty} n a_n x^{n-1} = \sum_{n=0}^{\infty} (n+1) a_{n+1} x^n$ for every x in the interval $(-R, R)$;

2. if $-R < \alpha \leq \beta < R$, then

$$\int_{\alpha}^{\beta} f(x)\, dx = \sum_{n=0}^{\infty} a_n \left[\int_{\alpha}^{\beta} x^n dx \right] = \sum_{n=0}^{\infty} \frac{a_n}{n+1} \left[\beta^{n+1} - \alpha^{n+1} \right].$$

Example

It follows from previous discussions that the power series $\sum_{n=0}^{\infty} x^n$ converges to $\dfrac{1}{1-x}$ whenever $-1 < x < 1$. Using (i) of the above Theorem, we may differentiate term-by-term to obtain $\sum_{n=0}^{\infty} (n+1)\, x^n = \dfrac{1}{(1-x)^2}$ for $-1 < x < 1$. On the other hand, we may use the 2nd result of the Theorem to obtain

$$\ln \frac{1}{1-x} = x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \cdots ,$$

for $-1 < x < 1$.

See CoCalc
https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page388-CoCalcJupyter.pdf

Taylor Series

Theorem
If f has a power series representation (expansion) at a, that is, if

$$f(x) = \sum_{n=0}^{\infty} c_n (x-a)^n, \quad |x-a| < R$$

then its coefficients are given by the formula

$$c_n = \frac{f^{(n)}(a)}{n!}.$$

That is to say,

$$\begin{aligned} f(x) &= \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n \\ &= f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f^{(3)}(a)}{3!}(x-a)^3 + \dots \end{aligned}$$

This series is called the Taylor Series of the function f at a. In the special case of $a = 0$, it is called the Maclaurin Series.

Example Maclaurin expansions of the following functions are useful.

1. $e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!}$;

2. $\cos x = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \frac{1}{6!}x^6 + \cdots = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}$;

3. $\sin x = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \frac{1}{7!}x^7 + \cdots = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}$;

4. $\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots = \sum_{n=0}^{\infty} x^n$;

5. $\frac{1}{1+x} = 1 - x + x^2 - x^3 + \cdots \sum_{n=0}^{\infty} (-1)^n x^n$;

6. $\ln(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + \cdots = \sum_{n=0}^{\infty} (-1)^n \frac{x^{n+1}}{n+1}$;

7. $\tan^{-1} x = x - \frac{1}{3}x^3 + \frac{1}{5}x^5 - \frac{1}{7}x^7 + \cdots = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{2n+1}$.

See CoCalc
https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page391-CoCalcJupyter.pdf

Note Expansions of e^x, $\cos x$ and $\sin x$ are valid for all real x, while those of the remaining functions are for $-1 < x < 1$. Note that the Maclaurin expansion of $\sin x$ may be obtained from differentiating the expansion of $\cos x$, and the expansion of $\ln(1+x)$ follows from term-by-term integration of the expansion of $\dfrac{1}{1+x}$.

Supplementary Notes on Power Series of Rational Functions

https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/supplementary10.pdf

See CoCalc https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/supplementary10-CoCalcJupyter.pdf

Theorem

If $f(x) = T_n(x) + R_n(x)$, where T_n is the n-th degree Taylor polynomial of f at a, and

$$\lim_{n\to\infty} R_n(x) = 0$$

for $|x-a| < R$, then f is equal to the sum of its Taylor series on the interval $|x-a| < R$.

Taylor's Inequality

If $|f^{(n+1)}(x)| \leq M$ for $|x-a| \leq d$, then the remainder $R_n(x)$ of the Taylor series satisfies the inequality

$$|R_n(x)| \leq \frac{M}{(n+1)!}|x-a|^{n+1} \quad , \quad |x-a| \leq d.$$

Determinants

Two linear equations

Consider a system of two linear equations in two unknowns x and y:

$$\begin{cases} a_1x + b_1y = k_1 \\ a_2x + b_2y = k_2 \end{cases} \tag{14.1}$$

An ordered pair of numbers (x, y) is said to be a *solution* of the system if x and y satisfy the equations in the system.
To find x we eliminate y from the equations to get

$$(a_1b_2 - a_2b_1)x = k_1b_2 - k_2b_1. \tag{14.2}$$

To find y we eliminate x from the equations to get

$$(a_1b_2 - a_2b_1)y = a_1k_2 - a_2k_1. \tag{14.3}$$

Therefore if $a_1b_2 - a_2b_1 \neq 0$, the solution is given by

$$x = \frac{k_1b_2 - k_2b_1}{a_1b_2 - a_2b_1}, \qquad y = \frac{a_1k_2 - a_2k_1}{a_1b_2 - a_2b_1}. \tag{14.4}$$

The solution is unique. Indeed, if (x, y) is a solution, the above elimination method shows that x and y cannot be numbers other than those given by (14.4). Hence the solution is uniquely determined.

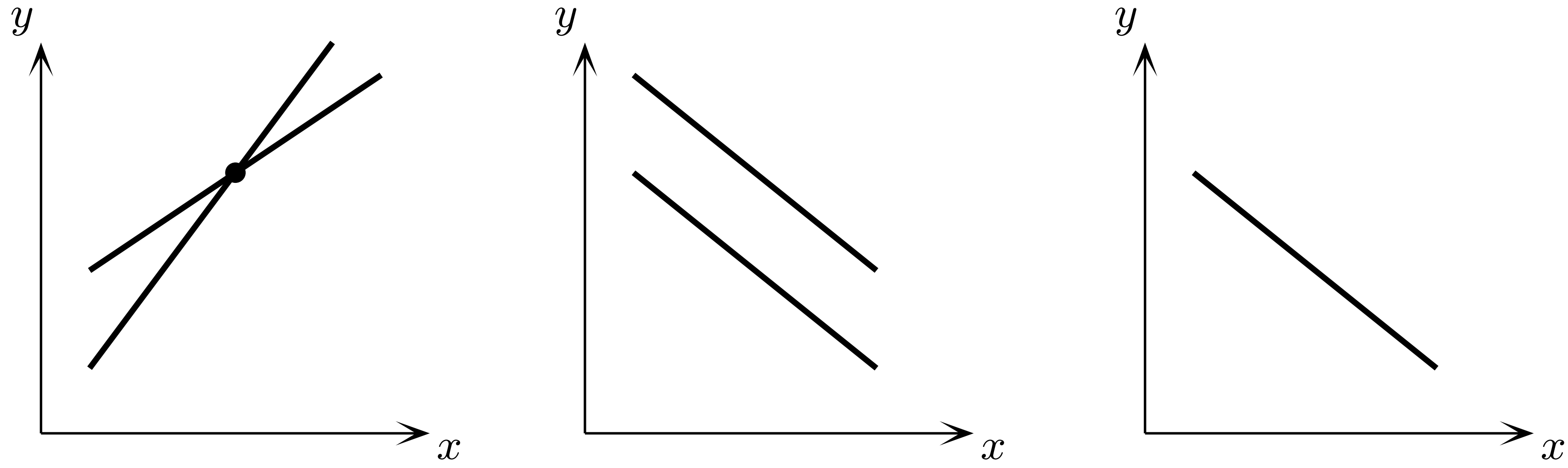
What happens if $a_1b_2 - a_2b_1 = 0$? In this case, (14.2) reduces to

$$0 \cdot x = k_1b_2 - k_2b_1.$$

Therefore if the RHS$= k_1b_2 - k_2b_1 \neq 0$, the equation has no solution for x; while if $k_1b_2 - k_2b_1 = 0$, any x satisfies the equation. We can get similar conclusion for y based on (14.3). We therefore have the following theorem:

Theorem 14.1

- *If* $a_1b_2 - a_2b_1 \neq 0$, *the system* (14.1) *has a unique solution given by* (14.4).
- *If* $a_1b_2 - a_2b_1 = 0$, *the system* (14.1) *has no solution or infinitely many solutions.*



(a) Lines intersect at a point. Exactly one solution.

(b) Two parallel lines. No solution.

(c) Two identical lines. Infinitely many solution.

Figure 14.1: Three possibilities for the solutions of two linear equations.

Each equation in the system (14.1) is a straight line in the xy-plane. Theorem 14.1 says that if $a_1b_2 - a_2b_1 \neq 0$, the two straight lines intersect at exactly one point in the xy-plane (see Fig. 14.1(a)). The following are examples of linear systems which have no solution or have infinitely many solutions.

Examples

Example 14.1 Let $a_1 = 2$, $b_1 = 3$, $k_1 = 9$, $a_2 = 4$, $b_2 = 6$, $k_2 = 12$. Then $a_1 b_2 - a_2 b_1 = 2 \times 6 - 4 \times 3 = 0$ and the system (14.1) becomes

$$\begin{cases} 2x + 3y = 9 \\ 4x + 6y = 12. \end{cases}$$

This system has no solution, otherwise we would have $2 \times 9 = 12$. On the xy-plane, the two linear equations represent two parallel lines that do not intersect at any point. (See Fig. 14.1(b))

Example 14.2 As in the previous example, we let $a_1 = 2$, $b_1 = 3$, $a_2 = 4$, $b_2 = 6$, $k_2 = 12$. In this example, we let $k_1 = 6$. Then $a_1 b_2 - a_2 b_1 = 0$ and the system (14.1) becomes

$$\begin{cases} 2x + 3y &= 6 \\ 4x + 6y &= 12. \end{cases}$$

This system represent two identical lines in the xy-plane. Any point (x, y) on the line satisfies both equations and therefore the system has infinitely many solutions. (See Fig. 14.1(c))

Second-order determinants

By a determinant of the second order we mean the symbol

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$$

which represents the number $a_1 b_2 - a_2 b_1$ evaluated in the following way:

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$$

The second order determinant has two (horizontal) *rows* and two (vertical) *columns*[5].

[5] We write R_i for row i and C_j for column j in this book.

Cramer's rule

With the notation of determinants, we may now write (14.4) as

$$x = \begin{vmatrix} k_1 & b_1 \\ k_2 & b_2 \end{vmatrix} \div \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}, \qquad y = \begin{vmatrix} a_1 & k_1 \\ a_2 & k_2 \end{vmatrix} \div \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}. \tag{14.5}$$

The formulas (14.5) are called *Cramer's rule* in which we assume that the denominator

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} \neq 0.$$

Examples

Example 14.3 Solve the following linear system by Cramer's rule.

$$\begin{cases} 2\,x + 3\,y = 4 \\ 4\,x + \quad y = -2 \end{cases}$$

Solution. Using Cramer's rule (14.5), we get

$$x = \begin{vmatrix} 4 & 3 \\ -2 & 1 \end{vmatrix} \div \begin{vmatrix} 2 & 3 \\ 4 & 1 \end{vmatrix} = \frac{10}{-10} = -1, \quad y = \begin{vmatrix} 2 & 4 \\ 4 & -2 \end{vmatrix} \div (-10) = \frac{-20}{-10} = 2.$$

□

Third-order determinants

A third-order determinant consists of three rows and three columns. It is denoted by

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$$

and is a number which can be found as follows:

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = a_1 \begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix} - b_1 \begin{vmatrix} a_2 & c_2 \\ a_3 & c_3 \end{vmatrix} + c_1 \begin{vmatrix} a_2 & b_2 \\ a_3 & b_3 \end{vmatrix} \tag{14.6}$$

$$= a_1(b_2c_3 - b_3c_2) - b_1(a_2c_3 - a_3c_2) + c_1(a_2b_3 - a_3b_2) \tag{14.7}$$

$$= a_1b_2c_3 + a_3b_1c_2 + a_2b_3c_1 - a_3b_2c_1 - a_1b_3c_2 - a_2b_1c_3. \tag{14.8}$$

First row expansion

Formula (14.6) or (14.7) is the result obtained by *expanding the determinant along the first row*. In this method, in order to find the coefficient of b_1, say, we imagine the row and column containing b_1 erased and the determinant of the remaining entries put down as they stand. Also we put the $+$ sign to a_1 (the upper-left entry), $-$ sign to b_1 (the one adjacent to a_1), etc. so that the signs associated with the entries of the determinant appear alternately as shown below. Thus we have $+a_1$, $-b_1$, $+c_1$, $-c_2$, $+b_2$, etc. with

$$\begin{vmatrix} + & - & + \\ - & + & - \\ + & - & + \end{vmatrix} \quad \text{overlaying on} \quad \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}.$$

Row and Column expansion

We can also get the same value of the determinant by expanding along any row or column. For example, if we expand along the second column, we get

$$\begin{aligned}\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} &= -b_1 \begin{vmatrix} a_2 & c_2 \\ a_3 & c_3 \end{vmatrix} + b_2 \begin{vmatrix} a_1 & c_1 \\ a_3 & c_3 \end{vmatrix} - b_3 \begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix} \\ &= -b_1(a_2c_3 - a_3c_2) + b_2(a_1c_3 - a_3c_1) - b_3(a_1c_2 - a_2c_1) \\ &= a_1b_2c_3 + a_3b_1c_2 + a_2b_3c_1 - a_3b_2c_1 - a_1b_3c_2 - a_2b_1c_3.\end{aligned}$$

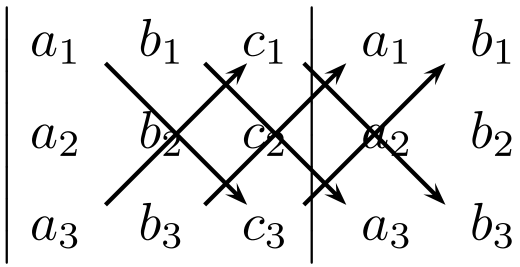
which is the same as (14.8). It can be proved that all values obtained by expansion along a row or column are equal.

Sarrus' rule for 3rd order determinant:

This rule says that we evaluated the determinant directly as

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = a_1b_2c_3 + b_1c_2a_3 + c_1a_2b_3 - a_3b_2c_1 - b_3c_2a_1 - c_3a_2b_1.$$

The rule can be easily memorized using the following diagram. The six terms on the RHS are obtained by multiplying the entries following the arrows. A downward right arrow is associated with a positive sign while an upward right arrow a negative sign.



Examples

Example 14.4 Evaluate the determinant

$$D = \begin{vmatrix} 3 & 1 & 2 \\ 2 & 3 & 4 \\ 5 & 3 & 1 \end{vmatrix}$$

by

(a) expanding along the first row;
(b) expanding along the second column;
(c) expanding along the third row;
(d) Sarrus' rule.

See CoCalc
https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page407-CoCalcJupyter.pdf

Solution.

(a) $$D = 3\begin{vmatrix} 3 & 4 \\ 3 & 1 \end{vmatrix} - 1\begin{vmatrix} 2 & 4 \\ 5 & 1 \end{vmatrix} + 2\begin{vmatrix} 2 & 3 \\ 5 & 3 \end{vmatrix} \quad \text{(expansion along R}_1\text{)}$$

$$= 3(-9) - (-18) + 2(-9) = -27.$$

(b) $$D = -1\begin{vmatrix} 2 & 4 \\ 5 & 1 \end{vmatrix} + 3\begin{vmatrix} 3 & 2 \\ 5 & 1 \end{vmatrix} - 3\begin{vmatrix} 3 & 2 \\ 2 & 4 \end{vmatrix} \quad \text{(expansion along C}_2\text{)}$$

$$= -(-18) + 3(-7) - 3(8) = -27.$$

(c) $$D = 5\begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix} - 3\begin{vmatrix} 3 & 2 \\ 2 & 4 \end{vmatrix} + 1\begin{vmatrix} 3 & 1 \\ 2 & 3 \end{vmatrix} \quad \text{(expansion along R}_3\text{)}$$

$$= 5(-2) - 3(8) + (7) = -27.$$

(d) By Sarrus' rule

$$D = (3)(3)(1) + (1)(4)(5) + (2)(2)(3) - (5)(3)(2) - (3)(4)(3) - (1)(2)(1)$$
$$= -27.$$

□

Basic properties

1. The value of a determinant is not changed if we put its rows as columns, in the same consecutive order, e.g.

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} \qquad \text{(Transposition)}$$

2. Interchanging any two columns (or any two rows) changes the sign of the determinant, e.g.

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = - \begin{vmatrix} c_1 & b_1 & a_1 \\ c_2 & b_2 & a_2 \\ c_3 & b_3 & a_3 \end{vmatrix} \qquad (\mathsf{C}_1 \sim \mathsf{C}_3)$$

where the first and third columns ($\mathsf{C}_1,\ \mathsf{C}_3$) of the first determinant have been interchanged to form the second determinant.

3. If a determinant has two rows (or two columns) identical, its value is zero, e.g.

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_2 & b_2 & c_2 \end{vmatrix} = 0 \qquad\qquad (\mathsf{R}_2 = \mathsf{R}_3)$$

since the second and third rows $(\mathsf{R}_2,\ \mathsf{R}_3)$ are identical.

4. A common factor of any row or column can be taken out to multiply the remaining determinant value, e.g.

$$\begin{vmatrix} kma_1 & mb_1 & mc_1 \\ ka_2 & b_2 & c_2 \\ ka_3 & b_3 & c_3 \end{vmatrix} = km \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}.$$

5. If each element of a row (or a column) is the sum of two or more terms, the determinant can be expressed as the sum of two or more determinants, e.g.

$$\begin{vmatrix} a_1 + m_1 & b_1 & c_1 \\ a_2 + m_2 & b_2 & c_2 \\ a_3 + m_3 & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} + \begin{vmatrix} m_1 & b_1 & c_1 \\ m_2 & b_2 & c_2 \\ m_3 & b_3 & c_3 \end{vmatrix}.$$

A more general formula is

$$\begin{vmatrix} ka_1 + lm_1 & b_1 & c_1 \\ ka_2 + lm_2 & b_2 & c_2 \\ ka_3 + lm_3 & b_3 & c_3 \end{vmatrix} = k \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} + l \begin{vmatrix} m_1 & b_1 & c_1 \\ m_2 & b_2 & c_2 \\ m_3 & b_3 & c_3 \end{vmatrix}. \qquad (14.9)$$

6. By choosing $l = -a_3/b_3$ (possible if $b_3 \neq 0$) so that $a_3 + lb_3 = 0$, we can simplify the determinant as

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 + lb_1 & b_1 & c_1 \\ a_2 + lb_2 & b_2 & c_2 \\ 0 & b_3 & c_3 \end{vmatrix}. \qquad (\mathsf{C}_1 + l\mathsf{C}_2 \to \mathsf{C}_1)$$

Note that in the result, a_3 is replaced by 0 since $a_3 + lb_3 = 0$. This result follows by putting $k = 1$ and $m_i = b_i$ in (14.9). This helps to change one or more entries of the determinant to 0 without changing the value of the determinant.

Remark

Remark 14.1

- In No. 6 above, we used the abbreviation $\mathsf{C}_1 + l\mathsf{C}_2 \to \mathsf{C}_1$ to represent the *column operation*:

  New column 1 is formed by adding old column 1 to l times column 2

  For convenience, we will simply write $\mathsf{C}_1 + l\mathsf{C}_2$ to represent this operation in the future. The convention being used is: when we write $k\mathsf{C}_i + m\mathsf{C}_j$ $(i \neq j)$ to represent a column operation, the result is used to replace C_i, the column being first written.
- *Row operations* $k\mathsf{R}_i + m\mathsf{R}_j$ $(i \neq j)$ are defined similarly.
- If a determinant D is transformed to another determinant E by the operation: $k\mathsf{C}_i + m\mathsf{C}_j$ or $k\mathsf{R}_i + m\mathsf{R}_j$ $(i \neq j)$, then $E = kD$. This fact follows from (14.9) above.

Example

Example 14.5 Evaluate the determinant $\Delta = \begin{vmatrix} 2 & 3 & 23 \\ 1 & 6 & 16 \\ 3 & 8 & 38 \end{vmatrix}$.

Solution 1. By expansion along a row or a column, or using Sarrus' rule, we can get $\Delta = 0$. For details, see Example 14.4 on page 407.

Solution 2.

$$\Delta = \begin{vmatrix} 2 & 3 & 23 \\ 1 & 6 & 16 \\ 3 & 8 & 38 \end{vmatrix} = \begin{vmatrix} 0 & -9 & -9 \\ 1 & 6 & 16 \\ 0 & -10 & -10 \end{vmatrix} = 90 \begin{vmatrix} 0 & 1 & 1 \\ 1 & 6 & 16 \\ 0 & 1 & 1 \end{vmatrix} = 0.$$

Solution 3. Obviously the third column is 10 times the first plus the second. Therefore

$$\Delta = \begin{vmatrix} 2 & 3 & 23 \\ 1 & 6 & 16 \\ 3 & 8 & 38 \end{vmatrix} = 10 \begin{vmatrix} 2 & 3 & 2 \\ 1 & 6 & 1 \\ 3 & 8 & 3 \end{vmatrix} + \begin{vmatrix} 2 & 3 & 3 \\ 1 & 6 & 6 \\ 3 & 8 & 8 \end{vmatrix} = 10 \times 0 + 0 = 0.$$

Example 14.6 Factorize the determinant $\Delta = \begin{vmatrix} 1 & a & a^3 \\ 1 & b & b^3 \\ 1 & c & c^3 \end{vmatrix}$.

Solution.

$$\Delta = \begin{vmatrix} 1 & a & a^3 \\ 1 & b & b^3 \\ 1 & c & c^3 \end{vmatrix} = \begin{vmatrix} 1 & a & a^3 \\ 0 & b-a & b^3-a^3 \\ 0 & c-a & c^3-a^3 \end{vmatrix} = (b-a)(c-a)\begin{vmatrix} 1 & a & a^3 \\ 0 & 1 & b^2+a^2+ab \\ 0 & 1 & c^2+a^2+ac \end{vmatrix}$$

$$= (b-a)(c-a)\begin{vmatrix} 1 & a & a^3 \\ 0 & 1 & b^2+a^2+ab \\ 0 & 0 & c^2-b^2+a(c-b) \end{vmatrix}$$

$$= (b-a)(c-a)(c-b)\begin{vmatrix} 1 & a & a^3 \\ 0 & 1 & b^2+a^2+ab \\ 0 & 0 & a+b+c \end{vmatrix}$$

$$= (b-a)(c-a)(c-b)(a+b+c).$$

□

See CoCalc

https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page415-CoCalcJupyter.pdf

Cramer's rule for linear equations

Consider the linear system of three linear equations in three unknowns x, y, z:

$$\begin{cases} a_1\,x + b_1\,y + c_1\,z &= k_1 \\ a_2\,x + b_2\,y + c_2\,z &= k_2 \\ a_3\,x + b_3\,y + c_3\,z &= k_3 \end{cases} \tag{14.10}$$

Theorem 14.2 (Cramer's Rule) *Let*

$$D = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}. \tag{14.11}$$

If $D \neq 0$, the system (14.10) has a unique solution (x, y, z) given by

$$x = \begin{vmatrix} k_1 & b_1 & c_1 \\ k_2 & b_2 & c_2 \\ k_3 & b_3 & c_3 \end{vmatrix} \div D, \quad y = \begin{vmatrix} a_1 & k_1 & c_1 \\ a_2 & k_2 & c_2 \\ a_3 & k_3 & c_3 \end{vmatrix} \div D, \quad z = \begin{vmatrix} a_1 & b_1 & k_1 \\ a_2 & b_2 & k_2 \\ a_3 & b_3 & k_3 \end{vmatrix} \div D.$$

These formulas are known as *Cramer's Rule*.

Similar to Theorem 14.1 on page 396, we have the following theorem for three linear equations.

Theorem 14.3 *Let D denote the determinant in (14.11).*

- *If $D \neq 0$, the system (14.10) has a unique solution given by Cramer's rule.*
- *If $D = 0$, the system (14.10) has no solution or infinitely many solutions.*

Example

Example 14.7 Solve the following linear system by Cramer's Rule.

$$\begin{cases} -2\,x + 3\,y - \ z = \ \ 1 \\ \ \ \ \ \ x + 2\,y - \ z = \ \ 4 \\ -2\,x - \ \ y + \ z = -3 \end{cases}$$

Solution. We have $D = \begin{vmatrix} -2 & 3 & -1 \\ 1 & 2 & -1 \\ -2 & -1 & 1 \end{vmatrix} = -2 \neq 0$. Therefore

$$x = \begin{vmatrix} 1 & 3 & -1 \\ 4 & 2 & -1 \\ -3 & -1 & 1 \end{vmatrix} \div D = \frac{-4}{-2} = 2. \qquad z = \begin{vmatrix} -2 & 3 & 1 \\ 1 & 2 & 4 \\ -2 & -1 & -3 \end{vmatrix} \div D = \frac{-8}{-2} = 4.$$

$$y = \begin{vmatrix} -2 & 1 & -1 \\ 1 & 4 & -1 \\ -2 & -3 & 1 \end{vmatrix} \div D = \frac{-6}{-2} = 3.$$

See CoCalc □

https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page418-CoCalcJupyter.pdf

Cross Product as a determinant

https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/CrossProduct.pdf

Volume of a parallelepiped, 3$\times$3 determinant

See CoCalc

https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page425b-CoCalcJupyter.pdf

To determine if three given points on the xy-plane are collinear

https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/supplementary12.pdf

See CoCalc https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/supplementary12-CoCalcJupyter.pdf

Homogeneous linear system

Definition 14.1 The linear system (14.10) is said to be *homogeneous* if the numbers k_1, k_2, k_3 on the RHS are zero. That is

$$\left\{\begin{array}{rcl} a_1\,x + b_1\,y + c_1\,z &=& 0 \\ a_2\,x + b_2\,y + c_2\,z &=& 0 \\ a_3\,x + b_3\,y + c_3\,z &=& 0 \end{array}\right.$$

Theorem 14.4 *If the*

$$\begin{cases} a_1\,x + b_1\,y + c_1\,z \;=\; 0 \\ a_2\,x + b_2\,y + c_2\,z \;=\; 0 \\ a_3\,x + b_3\,y + c_3\,z \;=\; 0 \end{cases} \tag{14.12}$$

has a solution $(x, y, z) \neq (0, 0, 0)$ *then*

$$D = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = 0. \tag{14.13}$$

Proof. Since $(0, 0, 0)$ is obviously a solution, the existence of a solution $(x, y, z) \neq (0, 0, 0)$ implies that there are at least two solutions. Therefore by Theorem 14.2, we must have $D = 0$. □

Examples

Example 14.8 Find the values of the constant λ for which the system of equations

$$\begin{cases} x \qquad\quad + 6z = 0 \\ \quad - \quad y + \lambda z = 0 \\ x + \lambda y + 2z = 0 \end{cases}$$

has solutions $(x, y, z) \neq (0, 0, 0)$.

Solution. By Theorem 14.4, we have

$$D = \begin{vmatrix} 1 & 0 & 6 \\ 0 & -1 & \lambda \\ 1 & \lambda & 2 \end{vmatrix} = 0.$$

which can be reduced to

$$\lambda^2 = 4 \quad \text{and hence} \quad \lambda = \pm 2.$$

$\square$

Example 14.9 Find the values of the constant k for which the system of equations

$$\begin{cases} x + ky + 2 = 0 \\ 2x + y + k = 0 \\ 2x + ky + 1 = 0 \end{cases}$$

is consistent, i.e. it has at least one solution (x, y).

Solution. Let $z = 1$. Then $(x, y, z) = (x, y, 1) \neq (0, 0, 0)$ satisfies the homogeneous system. By Theorem 14.4, we must have $D = \begin{vmatrix} 1 & k & 2 \\ 2 & 1 & k \\ 2 & k & 1 \end{vmatrix} = 0$ which can be reduced to

$$(k + 3)(k - 1) = 0 \quad \text{and hence} \quad k = -3 \text{ or } 1.$$

□

See CoCalc
https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page422-CoCalcJupyter.pdf

Determinants of higher order

The determinants of order n ($n = 1, 2, 3, \ldots$) can be defined using permutations.[6] For simplicity, we just point out that:

- A determinant of order n can be evaluated by expansion along any row or column.
- All Properties **1** to **6** apply to determinants of order n.

[6] See for example, B. Kolman and D. Hill, *Introductory Linear Algebra*, 8th edition, 2005

Examples

Example 14.10 Evaluate the determinant

$$D = \begin{vmatrix} 2 & 3 & 3 & 1 \\ 1 & 4 & 1 & 3 \\ 3 & 2 & 3 & 1 \\ 2 & 1 & 4 & 2 \end{vmatrix}.$$

Solution 1. By expanding along the first column,

$$\begin{aligned} D &= 2\begin{vmatrix} 4 & 1 & 3 \\ 2 & 3 & 1 \\ 1 & 4 & 2 \end{vmatrix} - \begin{vmatrix} 3 & 3 & 1 \\ 2 & 3 & 1 \\ 1 & 4 & 2 \end{vmatrix} + 3\begin{vmatrix} 3 & 3 & 1 \\ 4 & 1 & 3 \\ 1 & 4 & 2 \end{vmatrix} - 2\begin{vmatrix} 3 & 3 & 1 \\ 4 & 1 & 3 \\ 2 & 3 & 1 \end{vmatrix} \\ &= 2(20) - 2 + 3(-30) - 2(-8) = -36. \end{aligned}$$

Solution 2. Here we simplify the determinant using the properties **1** to **6**.

$$D = \begin{vmatrix} 2 & 3 & 3 & 1 \\ 1 & 4 & 1 & 3 \\ 3 & 2 & 3 & 1 \\ 2 & 1 & 4 & 2 \end{vmatrix}$$

$$= \begin{vmatrix} 0 & -5 & 1 & -5 \\ 1 & 4 & 1 & 3 \\ 0 & -10 & 0 & -8 \\ 0 & -7 & 2 & -4 \end{vmatrix} \qquad (\mathsf{R}_1 - 2\mathsf{R}_2,\ \mathsf{R}_3 - 3\mathsf{R}_2,\ \mathsf{R}_4 - 2\mathsf{R}_2)$$

$$= -\begin{vmatrix} -5 & 1 & -5 \\ -10 & 0 & -8 \\ -7 & 2 & -4 \end{vmatrix} \qquad (\text{Expansion along } \mathsf{C}_1)$$

$$= 2\begin{vmatrix} 5 & -1 & 5 \\ 5 & 0 & 4 \\ 7 & -2 & 4 \end{vmatrix} = 2(-28 - 50 + 40 + 20) = -36.$$

Matrices

Matrices

A *matrix* is a rectangular array of scalars of the form

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}. \tag{15.1}$$

This matrix is called an $m \times n$ *matrix*. It consists of m *rows* and n *columns*, the i-th row (or simply row i) being the $1 \times n$ matrix $[a_{i1} \quad a_{i2} \quad \cdots \quad a_{in}]$ and the j-th column (or column j) the $m \times 1$ matrix

$$\begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{bmatrix}.$$

The entry a_{ij} which belongs to row i and column j is called the (i, j)-*entry* of the matrix. The matrix (15.1) can be written in abbreviated forms such as $[a_{ij}]_{m \times n}$ or $[a_{ij}]$ or $[A]$ or simply A. The 1×1 matrix $[\alpha]$ is just the scalar α.

Examples

Example 15.1 The following are examples of matrices.

$$A = \begin{bmatrix} 4.5 & 1.4 \\ 2.3 & 5.9 \end{bmatrix}, \quad B = \begin{bmatrix} 5 & 1 \\ 2 & -9 \\ 3 & 2 \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} 2 \\ 6 \\ 5 \end{bmatrix}.$$

The first, denoted by A, is a 2×2 matrix. The second, denoted by B, is a 3×2 matrix. The third, denoted by $\mathbf{c}$ is a 3×1 matrix. If the entries of the matrix A above are denoted by a_{ij}, then $a_{11} = 4.5,\ a_{12} = 1.4,\ a_{21} = 2.3,\ a_{22} = 5.9$.

Transposition

If $A = [a_{ij}]$ is $m \times n$, then its *transpose* is the $n \times m$ matrix, denoted by A^T, obtained by interchanging the rows and columns of A, i.e.

$$A^T = [b_{ij}] \quad \text{where} \quad b_{ij} = a_{ji}$$

for all $i = 1, \ldots, n;\ j = 1, \ldots, m$.

Example 15.2 If A, B and $\mathbf{c}$ are the matrices given in the previous example, then

$$A^T = \begin{bmatrix} 4.5 & 2.3 \\ 1.4 & 5.9 \end{bmatrix}, \quad B^T = \begin{bmatrix} 5 & 2 & 3 \\ 1 & -9 & 2 \end{bmatrix}, \quad \mathbf{c}^T = [\,2\ 6\ 5\,].$$

Vectors

An $m \times 1$ matrix is called a *column* m*-vector* or an m*-vector* or a *column-vector* or simply a *vector*. Column-vectors are denoted by bold-faced lower-case letters like $\mathbf{a}$, $\mathbf{x}$ in printed form and is hand-written as $\underset{\sim}{a}$, $\underset{\sim}{x}$, etc.

A $1 \times n$ matrix is called a *row-vector*. As row-vectors are transposes of column-vectors, they are denoted by $\mathbf{a}^T$, $\mathbf{x}^T$, etc. and are written as $\underset{\sim}{a}^T$, $\underset{\sim}{x}^T$, etc.

Equality of matrices

Two $m \times n$ matrices A and B are said to be *equal*, written $A = B$, if $a_{ij} = b_{ij}$ for all $i = 1, \ldots, m$; $j = 1, \ldots, n$.

Example 15.3 If $E = \begin{bmatrix} 4 & 2 \\ 1 & 5 \end{bmatrix}$ and $F = \begin{bmatrix} 4 & 1 \\ 2 & 5 \end{bmatrix}$, then $E \neq F$ but $E = F^T$.

Addition and subtraction

We can add two $m \times n$ matrices. The *sum*, denoted by $A + B$ is an $m \times n$ matrix $C = [c_{ij}]$ such that $c_{ij} = a_{ij} + b_{ij}$ for all i, j. Similarly, the *difference*, denoted by $A - B$ is an $m \times n$ matrix $D = [d_{ij}]$ such that $d_{ij} = a_{ij} - b_{ij}$ for all i, j.

Example 15.4

$$\begin{bmatrix} 4 & 2 \\ 3 & 5 \end{bmatrix} + \begin{bmatrix} -2 & 1 \\ 2 & -2 \end{bmatrix} = \begin{bmatrix} 4-2 & 2+1 \\ 3+2 & 5-2 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 5 & 3 \end{bmatrix}.$$

$$\begin{bmatrix} 4 & 2 \\ 3 & 5 \end{bmatrix} - \begin{bmatrix} -2 & 1 \\ 2 & -2 \end{bmatrix} = \begin{bmatrix} 4+2 & 2-1 \\ 3-2 & 5+2 \end{bmatrix} = \begin{bmatrix} 6 & 1 \\ 1 & 7 \end{bmatrix}.$$

Scalar multiplication

If k is a scalar (real or complex), then kA is the matrix $C = [c_{ij}]$ such that $c_{ij} = ka_{ij}$ for all i, j.

Example 15.5

$$3\begin{bmatrix} 4 & 2 \\ 3 & -5 \end{bmatrix} = \begin{bmatrix} 12 & 6 \\ 9 & -15 \end{bmatrix}, \qquad k\begin{bmatrix} -2 & 1 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} -2k & k \\ 3k & 4k \end{bmatrix}.$$

Zero matrix

This is a matrix all whose entries are 0. Zero matrices are denoted by O. However, zero vectors are usually denoted by $\mathbf{0}$.

Example 15.6

$$O = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \qquad O = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad \mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Multiplication of vectors

If a row n-vector is multiplied by a column n-vector *on the right*, then the product is a scalar which is equal to the sum of the products of the corresponding entries in the two vectors. Using symbols, if

$$\mathbf{a}^T = [a_1 \quad a_2 \quad \cdots \quad a_n] \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

then

$$\mathbf{a}^T\mathbf{b} = a_1b_1 + a_2b_2 + \cdots + a_nb_n.$$

Example 15.7

$$\begin{bmatrix} 3 & -2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \\ -3 \end{bmatrix} = 3 \times 2 + (-2) \times 4 + 1 \times (-3) = -5.$$

Multiplication of matrices

If $A = [a_{ij}]$ is $m \times n$, $B = [b_{ij}]$ is $n \times p$ then the product AB is the $m \times p$ matrix $C = [c_{ij}]$ such that

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}$$

for all i, j, i.e. the entry c_{ij} is equal to row i of A right-multiplied by column j of B. The product AB is well-defined only when the number of columns of A is equal to the number of rows of B.

Example 15.8 If $A = \begin{bmatrix} 5 & 1 & 2 & 3 \\ 2 & 2 & -1 & 3 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 2 & 3 \\ -2 & 3 & -1 \\ 3 & -3 & -1 \\ -4 & 2 & 2 \end{bmatrix}$ then

$$AB = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \end{bmatrix} = \begin{bmatrix} -3 & 13 & 18 \\ -17 & 19 & 11 \end{bmatrix}$$

where the entries c_{ij} are computed as follows:

$$\begin{aligned}
c_{11} &= 5 \times 1 + 1 \times (-2) + 2 \times 3 + 3 \times (-4) = -3. \\
c_{12} &= 5 \times 2 + 1 \times 3 + 2 \times (-3) + 3 \times 2 = 13. \\
c_{13} &= 5 \times 3 + 1 \times (-1) + 2 \times (-1) + 3 \times 2 = 18. \\
c_{21} &= 2 \times 1 + 2 \times (-2) + (-1) \times 3 + 3 \times (-4) = -17. \\
c_{22} &= 2 \times 2 + 2 \times 3 + (-1) \times (-3) + 3 \times 2 = 19. \\
c_{23} &= 2 \times 3 + 2 \times (-1) + (-1) \times (-1) + 3 \times 2 = 11.
\end{aligned}$$

See CoCalc
https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page437-CoCalcJupyter.pdf

Properties

If the sums are well-defined, we have:

1. $A + B = B + A$
2. $(A + B) + C = A + (B + C)$
3. $A + O = A$
4. $A + (-A) = O$ where $-A = (-1)A$
5. $(A^T)^T = A$
6. $(A + B)^T = A^T + B^T$

If the products are well-defined, then

7. $A(BC) = (AB)C$
8. $A(B + C) = AB + AC, \quad (A + B)C = AC + BC$
9. $(AB)^T = B^T A^T$
10. There are examples of A and B for which $AB \neq BA$.

Square matrices

An $n \times n$ matrix is a *square matrix*. Its *order* is n. Note that a square matrix of order 1 is just a number.

Example 15.9 The following matrices A, B, C are square matrices.

$$A = \begin{bmatrix} 5 & 2 \\ 1 & 3 \end{bmatrix}, \qquad B = \begin{bmatrix} 2 & 0 & 3 \\ 4 & 1 & 0 \\ 7 & 2 & 4 \end{bmatrix}, \qquad C = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

Diagonal elements

If $A = [a_{ij}]$ is a square matrix of order n, then the entries a_{11}, $a_{22}, \ldots, a_{nn}$ are the *diagonal elements* of A. For example, the diagonal elements of the matrix

$$A = \begin{bmatrix} 2 & 5 & -3 \\ 3 & -1 & 7 \\ 2 & 0 & 4 \end{bmatrix}$$

are $2, -1, 4$.

Diagonal matrix

A *diagonal matrix* is a square matrix whose entries are all zero except possibly the diagonal elements.

Example 15.10 The matrices

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 4 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

are diagonal matrices.

We use diag$(a_{11}, \ldots, a_{nn})$ to denote the diagonal matrix of order n whose diagonal elements are $a_{11}, \ldots, a_{nn}$. For example, the matrix B above can be represented by diag$(2, 1, 0)$.

Identity matrix

A diagonal matrix whose diagonal elements are all unity is called an *identity matrix*. It is denoted by I_n if the specification of its order n is necessary. Usually we denote it simply by I.

Example 15.11 The matrices

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

are identity matrices.

Theorem 15.1 *If A is a square matrix of order n and I the identity matrix of the same order then $AI = IA = A$.*

Determinant

We have learned that a determinant is defined for numbers given in a square array. Using matrix terminology, a determinant is defined for every square matrix. If A is a square matrix, then the determinant formed this way is called the *determinant of* A and is denoted by $\det A$ or $\det(A)$.
For example, let

$$A = \begin{bmatrix} 3 & 2 & -1 \\ 2 & -1 & 0 \\ 2 & 3 & 1 \end{bmatrix}. \tag{15.2}$$

Then

$$\det A = \begin{vmatrix} 3 & 2 & -1 \\ 2 & -1 & 0 \\ 2 & 3 & 1 \end{vmatrix} = -15. \tag{15.3}$$

Theorem 15.2 *For square matrices A, B of order n, we have*

- $\det A^T = \det A$.
- $\det(AB) = \det(A)\det(B)$.

The first equality is just Property 1 of determinants stated on page 409.
The second equality is a useful property of determinants. Its proof is quite involved for matrices of general order n though we can verify the formula for small values of n.

Example 15.12 Verify the formula $\det(AB) = \det(A)\det(B)$ using the matrices

$$A = \begin{bmatrix} 2 & 3 \\ 4 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 3 & 2 \\ 2 & 5 \end{bmatrix}.$$

Solution.

$$\det A = \begin{vmatrix} 2 & 3 \\ 4 & 1 \end{vmatrix} = -10, \qquad \det B = \begin{vmatrix} 3 & 2 \\ 2 & 5 \end{vmatrix} = 11.$$

$$AB = \begin{bmatrix} 2 & 3 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 2 & 5 \end{bmatrix} = \begin{bmatrix} 12 & 19 \\ 14 & 13 \end{bmatrix}$$

$$\therefore \ \det(AB) = \begin{vmatrix} 12 & 19 \\ 14 & 13 \end{vmatrix} = -110 = (-10) \times 11 = \det(A)\det(B).$$

$\square$

Minors

Given a square matrix A of order n, there are n^2 scalars which are the determinants of order $n-1$ obtained by deleting one row and one column of A. These scalars are called the *minors* and will be denoted by M_{ij} if row i and column j are deleted in the computation. For examples, for the matrix

$$A = \begin{bmatrix} 3 & 2 & -1 \\ 2 & -1 & 0 \\ 2 & 3 & 1 \end{bmatrix} \tag{15.4}$$

(the same A as in (15.2)), there are 9$(= 3 \times 3)$ minors:

$$M_{11} = \begin{vmatrix} -1 & 0 \\ 3 & 1 \end{vmatrix} = -1, \qquad M_{12} = \begin{vmatrix} 2 & 0 \\ 2 & 1 \end{vmatrix} = 2, \qquad M_{13} = \begin{vmatrix} 2 & -1 \\ 2 & 3 \end{vmatrix} = 8,$$

$$M_{21} = \begin{vmatrix} 2 & -1 \\ 3 & 1 \end{vmatrix} = 5, \qquad M_{22} = \begin{vmatrix} 3 & -1 \\ 2 & 1 \end{vmatrix} = 5, \quad M_{23} = \begin{vmatrix} 3 & 2 \\ 2 & 3 \end{vmatrix} = 5,$$

$$M_{31} = \begin{vmatrix} 2 & -1 \\ -1 & 0 \end{vmatrix} = -1, \quad M_{32} = \begin{vmatrix} 3 & -1 \\ 2 & 0 \end{vmatrix} = 2, \quad M_{33} = \begin{vmatrix} 3 & 2 \\ 2 & -1 \end{vmatrix} = -7.$$

Cofactors

The *cofactors* of a square matrix A are the numbers defined by

$$A_{ij} = (-1)^{i+j} M_{ij}, \quad i, j = 1, 2, \ldots, n.$$

For the matrix A in (15.4), by matching the above minors with the corresponding signs in the following sign pattern

$$\begin{bmatrix} + & - & + \\ - & + & - \\ + & - & + \end{bmatrix}$$

we obtain the following cofactors for the matrix A.

$$\begin{aligned} A_{11} &= -1, & A_{12} &= -2, & A_{13} &= 8, \\ A_{21} &= -5, & A_{22} &= 5, & A_{23} &= -5, \\ A_{31} &= -1, & A_{32} &= -2, & A_{33} &= -7. \end{aligned}$$

Adjoint

The *adjoint* of an $n \times n$ matrix A is the transpose of the $n \times n$ matrix whose (i, j)-entry is the cofactor A_{ij} of A. The adjoint of A is denoted by $\operatorname{adj} A$ or $\operatorname{adj}(A)$. For the matrix A in (15.4), we have found its cofactors and therefore we can put down the adjoint at once as follows:

$$\operatorname{adj} A = \begin{bmatrix} -1 & -2 & 8 \\ -5 & 5 & -5 \\ -1 & -2 & -7 \end{bmatrix}^T = \begin{bmatrix} -1 & -5 & -1 \\ -2 & 5 & -2 \\ 8 & -5 & -7 \end{bmatrix}. \tag{15.5}$$

See CoCalc
https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page448-CoCalcJupyter.pdf

Example

The following example demonstrates how to find the same adjoint from scratch.

Example 15.13 Given the matrix $A = \begin{bmatrix} 3 & 2 & -1 \\ 2 & -1 & 0 \\ 2 & 3 & 1 \end{bmatrix}$, find $\operatorname{adj} A$, showing all intermediate steps.

Solution.

$$\operatorname{adj} A = \begin{bmatrix} +\begin{vmatrix} -1 & 0 \\ 3 & 1 \end{vmatrix} & -\begin{vmatrix} 2 & 0 \\ 2 & 1 \end{vmatrix} & +\begin{vmatrix} 2 & -1 \\ 2 & 3 \end{vmatrix} \\ -\begin{vmatrix} 2 & -1 \\ 3 & 1 \end{vmatrix} & +\begin{vmatrix} 3 & -1 \\ 2 & 1 \end{vmatrix} & -\begin{vmatrix} 3 & 2 \\ 2 & 3 \end{vmatrix} \\ +\begin{vmatrix} 2 & -1 \\ -1 & 0 \end{vmatrix} & -\begin{vmatrix} 3 & -1 \\ 2 & 0 \end{vmatrix} & +\begin{vmatrix} 3 & 2 \\ 2 & -1 \end{vmatrix} \end{bmatrix}^T = \begin{bmatrix} -1 & -2 & 8 \\ -5 & 5 & -5 \\ -1 & -2 & -7 \end{bmatrix}^T = \begin{bmatrix} -1 & -5 & -1 \\ -2 & 5 & -2 \\ 8 & -5 & -7 \end{bmatrix}.$$

□

Inverse

The *inverse* of a square matrix A, if exists, is the unique matrix, denoted by A^{-1}, such that

$$AA^{-1} = A^{-1}A = I.$$

If A^{-1} exists, then A is said to be *invertible*.

Theorem 15.3 *If A and B are invertible matrices of the same order, then*

- $(A^{-1})^{-1} = A$.
- $(AB)^{-1} = B^{-1}A^{-1}$.

The next theorem gives a formula for the inverse. This allows us to find the inverse A^{-1} (if exists) of A when its order is small.

Theorem 15.4 *If $\det A \neq 0$,*

$$\boxed{A^{-1} = \frac{\operatorname{adj} A}{\det A}.}$$

If $\det A = 0$, the inverse of A does not exist.

Examples

Example 15.14 Use the above theorem to find the inverse A^{-1} for the matrix A given in (15.4).

Solution.

$$A^{-1} = \frac{\operatorname{adj} A}{\det A} \qquad \text{(Theorem 15.4)}$$

$$= \frac{1}{-15}\begin{bmatrix} -1 & -5 & -1 \\ -2 & 5 & -2 \\ 8 & -5 & -7 \end{bmatrix} \qquad \text{(By (15.3) and (15.5))}$$

$$= \begin{bmatrix} 1/15 & 1/3 & 1/15 \\ 2/15 & -1/3 & 2/15 \\ -8/15 & 1/3 & 7/15 \end{bmatrix}.$$

□

See CoCalc
https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page452-CoCalcJupyter.pdf

Planes and Lines in 3D

https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/supplementary09.pdf

Systems of n linear equations in n unknowns. We consider the case when the number of unknowns is equal to the number of equations assuming that the system has a unique solution. We learned Cramer's rule for solving this kind of equations. In this section, we introduce the method based on matrix inversion.

Matrix form

Consider, for example, the system of linear equations

$$\begin{cases} 2x_1 + 3x_2 + \ \ x_3 = \ \ 5, \\ \ \ x_1 + 7x_2 + 2x_3 = \ \ 5, \\ 2x_1 + 5x_2 - 3x_3 = 15. \end{cases} \tag{15.6}$$

The system can be put in matrix form: $\begin{bmatrix} 2 & 3 & 1 \\ 1 & 7 & 2 \\ 2 & 5 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \\ 15 \end{bmatrix}$ or in short

$$A\mathbf{x} = \mathbf{b} \tag{15.7}$$

where $A = \begin{bmatrix} 2 & 3 & 1 \\ 1 & 7 & 2 \\ 2 & 5 & -3 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$,$\mathbf{b} = \begin{bmatrix} 5 \\ 5 \\ 15 \end{bmatrix}$. The matrix A is called the *coefficient matrix*, $\mathbf{x}$ the *unknown vector* and $\mathbf{b}$ the *constant vector* of the system. If $\mathbf{x}$ satisfies the matrix equation, it is a *solution* of the matrix equation or the linear system.

Augmented matrix

Another convenient way to display the system is the use of the *augmented matrix*:

$$\left[\begin{array}{ccc|c} 2 & 3 & 1 & 5 \\ 1 & 7 & 2 & 5 \\ 2 & 5 & -3 & 15 \end{array}\right]$$

Solving by matrix inversion

In this method, we first find the inverse of the coefficient matrix A. One way to do so is to apply the formula

$$A^{-1} = \frac{\operatorname{adj} A}{\det A}.$$

Order 2 case: $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. The adjoint is $\operatorname{adj} A = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$. Therefore

$$A^{-1} = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \div \begin{vmatrix} a & b \\ c & d \end{vmatrix}.$$

Then the solution of $A\mathbf{x} = \mathbf{b}$, i.e.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

is $\mathbf{x} = A^{-1}\mathbf{b}$ or

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

Examples

Example 15.15 Solve by matrix inversion the linear system

$$\begin{bmatrix} 2 & 3 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 12 \\ 13 \end{bmatrix}$$

Solution.

$$\begin{bmatrix} 2 & 3 \\ 1 & 5 \end{bmatrix}^{-1} = \begin{bmatrix} 5 & -3 \\ -1 & 2 \end{bmatrix} \div \begin{vmatrix} 2 & 3 \\ 1 & 5 \end{vmatrix} = \frac{1}{7} \begin{bmatrix} 5 & -3 \\ -1 & 2 \end{bmatrix}$$

Therefore

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{7} \begin{bmatrix} 5 & -3 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 12 \\ 13 \end{bmatrix} = \frac{1}{7} \begin{bmatrix} 21 \\ 14 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}.$$

□

Example 15.16 Solve the linear system by matrix inversion.

$$\begin{bmatrix} 2 & 3 & 1 \\ 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 11 \\ 11 \\ 14 \end{bmatrix}$$

Solution Let A denote the coefficient matrix.

$$\det A = \begin{vmatrix} 2 & 3 & 1 \\ 1 & 2 & 3 \\ 3 & 1 & 2 \end{vmatrix} = 8 + 27 + 1 - 6 - 6 - 6 = 18.$$

$$\operatorname{adj} A = \begin{bmatrix} +\begin{vmatrix} 2 & 3 \\ 1 & 2 \end{vmatrix} & -\begin{vmatrix} 1 & 3 \\ 3 & 2 \end{vmatrix} & +\begin{vmatrix} 1 & 2 \\ 3 & 1 \end{vmatrix} \\ -\begin{vmatrix} 3 & 1 \\ 1 & 2 \end{vmatrix} & +\begin{vmatrix} 2 & 1 \\ 3 & 2 \end{vmatrix} & -\begin{vmatrix} 2 & 3 \\ 3 & 1 \end{vmatrix} \\ +\begin{vmatrix} 3 & 1 \\ 2 & 3 \end{vmatrix} & -\begin{vmatrix} 2 & 1 \\ 1 & 3 \end{vmatrix} & +\begin{vmatrix} 2 & 3 \\ 1 & 2 \end{vmatrix} \end{bmatrix}^T = \begin{bmatrix} 1 & 7 & -5 \\ -5 & 1 & 7 \\ 7 & -5 & 1 \end{bmatrix}^T = \begin{bmatrix} 1 & -5 & 7 \\ 7 & 1 & -5 \\ -5 & 7 & 1 \end{bmatrix}.$$

Therefore $A^{-1} = \frac{1}{18}\begin{bmatrix} 1 & -5 & 7 \\ 7 & 1 & -5 \\ -5 & 7 & 1 \end{bmatrix}$ and

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \frac{1}{18}\begin{bmatrix} 1 & -5 & 7 \\ 7 & 1 & -5 \\ -5 & 7 & 1 \end{bmatrix}\begin{bmatrix} 11 \\ 11 \\ 14 \end{bmatrix} = \frac{1}{18}\begin{bmatrix} 54 \\ 18 \\ 36 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}.$$

See CoCalc

https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page458-CoCalcJupyter.pdf

General systems of linear Equations

The general system, with m equations and n unknowns and in the form of an augmented matrix, can be solved systematically in two steps.

- The first is to reduce (by elimination) the original system to a new equivalent system of the so-called *echelon form* from which we can see right away whether the system has a solution.
- In the second step, we solve by *back-substitution* this equivalent system by assuming known parameter values to some of the unknowns if the system has a solution.

Systems of Linear Equations

Consider a system of m linear equations in n unknowns $x_1, x_2, \ldots, x_n$:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2 \\ \quad\vdots \qquad\quad \vdots \qquad\qquad\qquad \vdots \qquad\quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m \end{cases}$$

This linear system can be conveniently written in matrix form as $\mathbf{Ax} = \mathbf{b}$ by putting

$$\mathbf{A} = [a_{ij}] = \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \ldots & a_{mn} \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}.$$

- $\mathbf{v}$ is a solution if $\mathbf{Av} = \mathbf{b}$.
- System is consistent if such a $\mathbf{v}$ exist.
- Inconsistent if no such $\mathbf{v}$ exist.

Example

The system

$$\begin{cases} x_1 - 2x_2 = -1 \\ -x_1 + 3x_2 = 3 \end{cases}$$ is consistent

since $\mathbf{x} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$ is a solution.

In fact, this is the unique solution.

Geometrically, it means two linear straight lines having a point of intersection.

On the other hand, the system

$$\begin{cases} x_1 - 2x_2 = -1 \\ -3x_1 + 6x_2 = 13 \end{cases}$$ has no solution,

thus, is inconsistent.

Geometrically, it means two linear straight lines are not intersecting.

Problem

Determine whether a given system of linear equations $\mathbf{Ax} = \mathbf{b}$ is consistent, and find all the solution in case it is consistent.

Definition Two systems of linear equations are said to be *equivalent* if their solution sets are identical.

Operations giving equivalent systems

(i) interchange any 2 equations of a system of linear equations;
(ii) multiply both sides of any equation in a system by a *nonzero* scalar;
(iii) add a multiple of one equation to another equation within the system.

The *augmented matrix*

Notation : $$[\mathbf{A}|\mathbf{b}] = \left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right].$$

(i) interchange any two rows of the augmented matrix $[\mathbf{A}\,|\mathbf{b}]$;
(ii) multiply any row of the augmented matrix $[\mathbf{A}\,|\mathbf{b}]$ by a nonzero scalar;
(iii) add a scalar multiple of one row of the augmented matrix $[\mathbf{A}\,|\mathbf{b}]$ to another row.

These operations are called *elementary row operations* on matrices.

Row operations

The elimination steps are implemented using the following *row operations* which are equivalent to operations on the equations themselves:

Row operations	Meaning	Purpose
$\mathsf{R}_i \sim \mathsf{R}_j$	Interchanging Row i and Row j	Move the nonzero entry to an upper position if necessary.
$k\mathsf{R}_i$	k times Row i	Reduce the coefficients of the equation to a simpler form, e.g. multiply the common denominator or divide by the common factor.
$\mathsf{R}_i - m\mathsf{R}_j$	Row i minus m times Row j to replace Row i	Change a nonzero entry in Row i to zero so that an unknown in equation i is eliminated.

Reduced row-echelon form

Definition
A matrix is said to be in *reduced row-echelon* form if it has the following properties :

(1) If a row does not consist entirely of zeros, then the 1^{st} non-zero entry of this row is equal to 1 (known as the leading 1's);
(2) All the rows that consist entirely of zeros are grouped together at the bottom of the matrix;
(3) If the leading 1 of *i*-th row occurs at the *p*-th column and if the leading 1 of row $(i+1)$ occurs at the *q*-th column, then $p < q$;
(4) Each column that contains a leading 1 has zeros elsewhere.

A matrix that satisfies (1), (2) and (3) but not necessarily (4) is said to be in *row-echelon form*.

Example

$\begin{bmatrix} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & -9 \end{bmatrix}$ and $\begin{bmatrix} 0 & 1 & -2 & 0 & 15 \\ 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ are in reduced row-echelon form, while

$\begin{bmatrix} 1 & 4 & 3 & 7 \\ 0 & 1 & 6 & 2 \\ 0 & 0 & 1 & 5 \end{bmatrix}$ and $\begin{bmatrix} 0 & 1 & 2 & 6 & 0 \\ 0 & 0 & 1 & -1 & 18 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$ are in row-echelon form.

Theorem Every matrix $\mathbf{A}$ can be reduced to a matrix in *reduced row-echelon form* by applying to $\mathbf{A}$ a sequence of elementary row operations.

Example

Consider the system of linear equations

$$\begin{cases} x_1 - 2x_2 + x_3 = 0 \\ 2x_2 - 8x_3 = 6 \\ -4x_1 + 5x_2 + 9x_3 = -9 \end{cases}$$

The augmented matrix is given by

$$\left[\begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 6 \\ -4 & 5 & 9 & -9 \end{array}\right]$$

The elimination process goes as follows:

$$4 \times (1) + (3) \rightarrow \quad \left[\begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 6 \\ 0 & -3 & 13 & -9 \end{array}\right]$$

$$\frac{1}{2} \times (2) \rightarrow \quad \left[\begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 3 \\ 0 & -3 & 13 & -9 \end{array}\right]$$

$$3 \times (2) + (3) \rightarrow \quad \left[\begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 3 \\ 0 & 0 & 1 & 0 \end{array}\right]$$

$$\begin{array}{r} -1 \times (3) + (1) \rightarrow \\ 4 \times (3) + (2) \rightarrow \\ \end{array} \quad \left[\begin{array}{ccc|c} 1 & -2 & 0 & 0 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 0 \end{array}\right]$$

$$2 \times (2) + (1) \rightarrow \quad \left[\begin{array}{ccc|c} 1 & 0 & 0 & 6 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 0 \end{array}\right].$$

Thus the given linear system is consistent and has $[6 \quad 3 \quad 0]^T$ as its (unique) solution.

Recall that a system of linear equations is either *inconsistent* (no solution) or *consistent* (exactly one solution or infinitely many solutions).
See CoCalc
https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page468-CoCalcJupyter.pdf

Row Equivalent

Definition Two matrices are *row equivalent* if there is a sequence of row operations that can transform one matrix into the other.

For example, the matrices $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$ and $\begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$ are row equivalent.

Example

Consider the system

$$\begin{cases} \quad\quad x_2 - 4x_3 = 8 \\ 2x_1 - 3x_2 + 2x_3 = 1 \\ 5x_1 - 8x_2 + 7x_3 = 1 \end{cases}.$$

Solution

$$\left[\begin{array}{ccc|c} 0 & 1 & -4 & 8 \\ 2 & -3 & 2 & 1 \\ 5 & -8 & 7 & 1 \end{array}\right] \to \left[\begin{array}{ccc|c} 2 & -3 & 2 & 1 \\ 0 & 1 & -4 & 8 \\ 5 & -8 & 7 & 1 \end{array}\right]$$

$$\to \left[\begin{array}{ccc|c} 2 & -3 & 2 & 1 \\ 0 & 1 & -4 & 8 \\ 0 & -1/2 & 2 & -3/2 \end{array}\right] \to \left[\begin{array}{ccc|c} 2 & -3 & 2 & 1 \\ 0 & 1 & -4 & 8 \\ 0 & 0 & 0 & 5/2 \end{array}\right].$$

This corresponds to the **inconsistent** system

$$\begin{cases} 2x_1 - 3x_2 + 2x_3 = & 1 \\ 0x_1 + x_2 - 4x_3 = & 8 \\ 0x_1 + 0x_2 + 0x_3 = & \frac{5}{2} \quad \longleftarrow \end{cases}$$

Definition If certain variables in a system of linear equations can be expressed in terms of the remaining variables, then the former are called *basic variables* while the latter are known as *free variables*.

Example Consider the system

$\begin{cases} x_1 - 5x_3 = 1 \\ x_2 + x_3 = 4 \end{cases}$, we see that x_1 and x_2 are *basic variables* while x_3 is a *free variable*.

Let $x_3 = t$, we obtain $x_1 = 1 + 5t$ and $x_2 = 4 - t$. Thus

$$\mathbf{x} = \begin{bmatrix} 1+5t \\ 4-t \\ t \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 0 \end{bmatrix} + t \begin{bmatrix} 5 \\ -1 \\ 1 \end{bmatrix}.$$

Geometrically, this solution set is a straight line in space.
See CoCalc
https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page472-CoCalcJyputer.pdf

Gaussian elimination and Gauss-Jordan method

Example 2.6 Solve the system

$$\begin{cases} x_1 + x_2 + 2x_3 = 9 \\ 2x_1 + 4x_2 - 3x_3 = 1 \\ 3x_1 + 6x_2 - 5x_3 = 0 \end{cases}$$

Solution:

$$\left[\begin{array}{ccc|c} 1 & 1 & 2 & 9 \\ 2 & 4 & -3 & 1 \\ 3 & 6 & -5 & 0 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 1 & 2 & 9 \\ 0 & 2 & -7 & -17 \\ 1 & 2 & -2 & -1 \end{array}\right]$$

$$\to \left[\begin{array}{ccc|c} 1 & 1 & 2 & 9 \\ 0 & 1 & -7/2 & -17/2 \\ 0 & 1 & -4 & -10 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 1 & 2 & 9 \\ 0 & 1 & -7/2 & -17/2 \\ 0 & 0 & -1/2 & -3/2 \end{array}\right]$$

$$\to \left[\begin{array}{ccc|c} 1 & 1 & 2 & 9 \\ 0 & 1 & -7/2 & -17/2 \\ 0 & 0 & 1 & 3 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 0 & 11/2 & 35/2 \\ 0 & 1 & -7/2 & -17/2 \\ 0 & 0 & 1 & 3 \end{array}\right].$$

See CoCalc
https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page473-CoCalcJupyter.pdf
See CoCalc https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page491-stepbystepsolve-CoCalcJupyter.pdf

$$\begin{cases} x_1 \qquad + \frac{11}{2}x_3 = \frac{35}{2} \\ \qquad x_2 - \frac{7}{2}x_3 = -\frac{17}{2} \\ \qquad\qquad\quad x_3 = 3 \end{cases}$$, which implies

$x_3 = 3,$

$x_2 = \dfrac{7}{2} \times 3 - \dfrac{17}{2} = 2$ and

$x_1 = \dfrac{-11}{2} \times 3 + \dfrac{35}{2} = 1.$

Thus $\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ is the only solution of the original linear system.

Another Example

https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/to_get_to_reduced_row_echelon_form.pdf

Another Example

https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/to_get_to_reduced_row_echelon_form_2.pdf

Example

Consider

$$\begin{cases} x_1 + 3x_2 - 2x_3 + 2x_5 = 0 \\ 2x_1 + 6x_2 - 5x_3 - 2x_4 + 4x_5 - 3x_6 = -1 \\ 5x_3 + 10x_4 + 15x_6 = 5 \\ 2x_1 + 6x_2 + 8x_4 + 4x_5 + 18x_6 = 6 \end{cases}.$$

The augmented matrix can be row reduced to $\left[\begin{array}{cccccc|c} 1 & 3 & 0 & 4 & 2 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$.

The corresponding system of equations is given by $\begin{cases} x_1 + 3x_2 + 4x_4 + 2x_5 = 0 \\ x_3 + 2x_4 = 0 \\ x_6 = \frac{1}{3} \end{cases}$.

Thus the solutions are
$x_1 = -3x_2 - 4x_4 - 2x_5$,
$x_3 = -2x_4$
$x_6 = 1/3$
and x_2, x_4 and x_5 are free variables which may take arbitrary values. The solution set thus consists of all vectors of the form

$$\mathbf{x} = \begin{bmatrix} -3\alpha - 4\beta - 2\gamma \\ \alpha \\ -2\beta \\ \beta \\ \gamma \\ 1/3 \end{bmatrix}$$

, where α, β and γ are arbitrary scalars (known as parameters).

Theorem Suppose the augmented matrix $[\mathbf{A}|\mathbf{b}]$ of the linear system $\mathbf{Ax}=\mathbf{b}$ is reduced to $[\mathbf{R}|\mathbf{c}]$ by elementary row operations, where $\mathbf{R}$ is an $m \times n$ matrix in reduced row-echelon form (or in row-echelon form) and $\mathbf{c} = [c_1 \quad c_2 \quad \ldots \quad c_m]^T$. If $\mathbf{R}$ has r non-zero rows, then the system $\mathbf{Ax}=\mathbf{b}$ is consistent if and only if $c_j = 0$ for $r < j \leq m$.

Example 2.8 Consider the linear system

$\mathbf{Ax}=\mathbf{b}$ where $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$.

The augmented matrix of the system can be row reduced as follows:

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & b_1 \\ 4 & 5 & 6 & b_2 \\ 7 & 8 & 9 & b_3 \end{array}\right] \rightarrow \left[\begin{array}{ccc|c} 1 & 2 & 3 & b_1 \\ 0 & -3 & -6 & b_2 - 4b_1 \\ 0 & -6 & -12 & b_3 - 7b_1 \end{array}\right] \rightarrow$$

$\left[\begin{array}{ccc|c} 1 & 2 & 3 & b_1 \\ 0 & 1 & 2 & \frac{4}{3}b_1 - \frac{1}{3}b_2 \\ 0 & 0 & 0 & b_3 - 2b_2 + b_1 \end{array}\right]$. The matrix $\mathbf{R} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$ is in row-echelon with has 2 nonzero rows. Therefore, the system $\mathbf{Ax}=\mathbf{b}$ is consistent if and only if $b_3 - 2b_2 + b_1 = 0$.

Examples

Example 15.17 Solve the linear system $\begin{cases} 2x + y - z = 1 \\ 4x - y - 3z = 0 \\ 6x \phantom{{}-y} - 4z = 3\ . \end{cases}$

Solution. Step 1. Applying elimination steps to the system to eliminate the unknowns x_1, x_2, successively, we obtain:

$$\left[\begin{array}{ccc|c} 2 & 1 & -1 & 1 \\ 4 & -1 & -3 & 0 \\ 6 & 0 & -4 & 3 \end{array}\right] \xrightarrow[\mathsf{R}_3-3\mathsf{R}_1]{\mathsf{R}_2-2\mathsf{R}_1} \left[\begin{array}{ccc|c} 2 & 1 & -1 & 1 \\ 0 & -3 & -1 & -2 \\ 0 & -3 & -1 & 0 \end{array}\right] \xrightarrow{\mathsf{R}_3-\mathsf{R}_2} \left[\begin{array}{ccc|c} 2 & 1 & -1 & 1 \\ 0 & -3 & -1 & -2 \\ 0 & 0 & 0 & 2 \end{array}\right].$$

The last augmented matrix is in *echelon form* (or *row-echelon form*) because the number of leading zeros (counting from left to right) in a row is increasing from one row to the next row. In this example, the numbers of leading zeros are 0,1,3 and hence the matrix is in echelon form[7].

From the last augmented matrix, we see that the last equation is $0 = 2$ which is absurd. Therefore the given system has no solution and Step 2 is not applicable. □

[7]Some authors require the first nonzero entry in each row of an echelon matrix be unity. We do not impose this requirement in this book.

Example 15.18 Solve the linear system $\begin{cases} x + 2y + z - w = 1 \\ x + 2y + 5z + w = 2 \\ x + 2y - 3z - 3w = 0 \ . \end{cases}$

Solution
Step 1.

$$\left[\begin{array}{cccc|c} 1 & 2 & 1 & -1 & 1 \\ 1 & 2 & 5 & 1 & 2 \\ 1 & 2 & -3 & -3 & 0 \end{array}\right] \xrightarrow[\mathsf{R}_3 - \mathsf{R}_1]{\mathsf{R}_2 - \mathsf{R}_1} \left[\begin{array}{cccc|c} 1 & 2 & 1 & -1 & 1 \\ 0 & 0 & 4 & 2 & 1 \\ 0 & 0 & -4 & -2 & -1 \end{array}\right] \xrightarrow{\mathsf{R}_3 + \mathsf{R}_2} \left[\begin{array}{cccc|c} 1 & 2 & 1 & -1 & 1 \\ 0 & 0 & 4 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] .$$

The last system is in echelon form and is *consistent* (which means that there exists at least one solution).
Step 2. Transposing the unknowns y and w to the right-hand side, we obtain

$$\begin{cases} x + z = -2y + w + 1 \\ 4z = -2w + 1 \ . \end{cases}$$

By assuming $y = \lambda$ and $w = \mu$ (both are parameters which are considered fixed at the moment), the above can be regarded as a triangular system with two unknowns x and z.

By *back-substitution*, we obtain $\begin{cases} z = & -\frac{1}{2}\mu + \frac{1}{4} \\ x = -2\lambda + \frac{3}{2}\mu + \frac{3}{4} \end{cases}$. Therefore the solution of the given system is

$$\begin{cases} x = -2\lambda + \frac{3}{2}\mu + \frac{3}{4} \\ y = \lambda \\ z = -\frac{1}{2}\mu + \frac{1}{4} \\ w = \mu \end{cases}$$

which can be put into vector form as

$$\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \lambda \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \mu \begin{bmatrix} 3/2 \\ 0 \\ -1/2 \\ 1 \end{bmatrix} + \begin{bmatrix} 3/4 \\ 0 \\ 1/4 \\ 0 \end{bmatrix}.$$

Since λ and μ can be arbitrarily assigned, we see that the given system has infinitely many solutions.
See CoCalc
https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page479-CoCalcJupyter.pdf

Remark 15.1

- The above solution in vector form represents *all* solutions of the linear system.
- Right at the beginning of Step 2, we can transpose the unknowns x, z or x, w or y, z to the right-hand side to get other triangular systems. Solving these we get solutions in terms of parameters representing unknowns other than y and w. Solutions in vector form obtained these ways look quite differently from one another but they all represent the same solution set of the linear system.

Find the linear map matrix in 2-D

https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/supplementary08.pdf

See CoCalc

https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/supplementary08_CoCalcJupyter.pdf

To find the closest points between 2 lines in 3D

https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/supplementary02.pdf

See CoCalc

https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/supplementary02_CoCalcJupyter.pdf

Systems of Homogeneous Equations

$$\mathbf{Ax} = \mathbf{0},$$

The solution set either has only the trivial solution, or infinitely many solutions.

Theorem If $\mathbf{A}$ is an $m \times n$ matrix where $m < n$, then the homogeneous system $\mathbf{Ax} = \mathbf{0}$ always has non-trivial solutions. In other words, if the number of equations is less than the number of unknowns, then the system has non-trivial solutions.

Example Solve the homogeneous system $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$

Solution The augmented matrix of the linear system may be simplified by elementary row operations as follows:

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 4 & 5 & 6 & 0 \\ 7 & 8 & 9 & 0 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 0 & -3 & \text{-6} & 0 \\ 0 & -6 & -12 & 0 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & -6 & -12 & 0 \end{array}\right]$$

$$\to \left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

The last matrix $\left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$ is the augmented matrix of the linear system $\begin{cases} x_1 \qquad -x_3 = 0 \\ \qquad x_2 + 2x_3 = 0 \end{cases}$.

By putting the free variable $x_3 = t$, we conclude that solutions of the linear system are given by $\mathbf{x} = \begin{bmatrix} t \\ -2t \\ t \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$ where t is any scalar.

Nonsingular Matrices

Definition A square matrix $\mathbf{A}$ is said to be *nonsingular* (or *invertible*) if there is a square matrix $\mathbf{B}$ such that $\mathbf{AB} = \mathbf{I}$ and $\mathbf{BA} = \mathbf{I}$. The matrix $\mathbf{B}$ is called an inverse of $\mathbf{A}$, denoted by the symbol $\mathbf{A}^{-1}$.

Example
Since $\begin{bmatrix} 3 & 2 \\ 7 & 5 \end{bmatrix} \begin{bmatrix} 5 & -2 \\ -7 & 3 \end{bmatrix} = \mathbf{I}$ and $\begin{bmatrix} 5 & -2 \\ -7 & 3 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 7 & 5 \end{bmatrix} = \mathbf{I}$, we conclude that the matrix $\begin{bmatrix} 3 & 2 \\ 7 & 5 \end{bmatrix}$ is nonsingular, with $\begin{bmatrix} 5 & -2 \\ -7 & 3 \end{bmatrix}$ as its inverse.

Example Let $\mathbf{A} = \begin{bmatrix} 0 & a & b \\ 0 & c & d \\ 0 & e & f \end{bmatrix}$.
If $\mathbf{B}$ is any 3×3 matrix, then the first column of the product $\mathbf{BA}$ consists entirely of zeros. Therefore $\mathbf{BA} \neq \mathbf{I}$ and thus $\mathbf{A}$ has no inverse.

Property If $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are $n \times n$ matrices such that $\mathbf{AC} = \mathbf{I}$ and $\mathbf{BA} = \mathbf{I}$, then $\mathbf{B} = \mathbf{C}$.
Proof: $\mathbf{B} = \mathbf{BI} = \mathbf{B}(\mathbf{AC}) = (\mathbf{BA})\mathbf{C} = \mathbf{IC} = \mathbf{C}$.

Property 4.2 If $\mathbf{A}$ and $\mathbf{B}$ are nonsingular matrices of the same order, then $\mathbf{AB}$ is nonsingular and $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.

Proof: By the associativity of matrix multiplication, we have

$$\begin{aligned}
\mathbf{AB}(\mathbf{B}^{-1}\mathbf{A}^{-1}) &= \mathbf{A}(\mathbf{B}(\mathbf{B}^{-1}\mathbf{A}^{-1})) = \mathbf{A}((\mathbf{BB}^{-1})\mathbf{A}^{-1}) \\
&= \mathbf{A}(\mathbf{IA}^{-1}) = \mathbf{AA}^{-1} = \mathbf{I} \ \text{ and} \\
(\mathbf{B}^{-1}\mathbf{A}^{-1})\mathbf{AB} &= \mathbf{B}^{-1}(\mathbf{A}^{-1}(\mathbf{AB})) = \mathbf{B}^{-1}(\mathbf{A}^{-1}\mathbf{A})\mathbf{B} \\
&= \mathbf{B}^{-1}(\mathbf{IB}) = \mathbf{B}^{-1}\mathbf{B} = \mathbf{I}.
\end{aligned}$$

Therefore $\mathbf{AB}$ is nonsingular and $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.

Property 4.3 If $\mathbf{A}$ is nonsingular and k is a nonzero scalar, then

(i) $\mathbf{A}^{-1}$ is nonsingular and $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$;
(ii) $k\mathbf{A}$ is nonsingular and $(k\mathbf{A})^{-1} = (\frac{1}{k})\mathbf{A}^{-1}$.

Theorem (4) Let $\mathbf{A}$ be an $n \times n$ nonsingular matrix. Then

(i) the homogeneous system $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution;
(ii) the system $\mathbf{Ax} = \mathbf{b}$ has a unique solution for any $\mathbf{b}$ in $\mathbb{R}_n$.

Proof

(i) if $\mathbf{v}$ in $\mathbb{R}_n$ is a solution of $\mathbf{Ax} = \mathbf{0}$, then $\mathbf{Av} = \mathbf{0}$ and therefore $\mathbf{v} = \mathbf{Iv} = (\mathbf{A}^{-1}\mathbf{A})\mathbf{v} = \mathbf{A}^{-1}(\mathbf{Av}) = \mathbf{A}^{-1}\mathbf{0} = \mathbf{0}$.
(ii) Since $\mathbf{A}(\mathbf{A}^{-1}\mathbf{b}) = (\mathbf{A}\mathbf{A}^{-1})\mathbf{b} = \mathbf{Ib} = \mathbf{b}$, $\mathbf{A}^{-1}\mathbf{b}$ is a solution of the linear system $\mathbf{Ax} = \mathbf{b}$. On the other hand, if $\mathbf{v}$ is a solution of $\mathbf{Ax} = \mathbf{b}$, then $\mathbf{Av} = \mathbf{b}$. Hence $\mathbf{v} = \mathbf{Iv} = (\mathbf{A}^{-1}\mathbf{A})\mathbf{v} = \mathbf{A}^{-1}(\mathbf{Av}) = \mathbf{A}^{-1}\mathbf{b}$.

The simplest nonsingular matrices are the so-called *elementary matrices*.

Definition: An $n \times n$ matrix is called an *elementary matrix* if it can be obtained from the $n \times n$ identity matrix $\mathbf{I}$ by performing a single elementary row operation.

Example 4.3

$$\mathbf{E} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \mathbf{F} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix} \text{ and } \mathbf{G} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ are elementary.}$$

(1) If $\mathbf{E}$ is the result from performing a certain elementary row operation on $\mathbf{I_m}$, then the product $\mathbf{EA}$ is the matrix that results when this same elementary row operation is performed on $\mathbf{A}$.

(2) If an elementary row operation is applied to an identity matrix $\mathbf{I}$ to produce an elementary matrix $\mathbf{E}$, then there exists another elementary row operation which, when applied to $\mathbf{E}$, produces $\mathbf{I}$.

(3) Every elementary matrix is nonsingular, and the inverse of an elementary matrix is also an elementary matrix.

Example: Consider $\mathbf{E} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $\mathbf{F} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix}$ and $\mathbf{G} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

(a) the inverse of $\mathbf{E}$ is $\begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & 1 \end{bmatrix}$, which is the elementary matrix obtained from $\mathbf{I}$ by multiplying its second row by $\frac{1}{3}$;

(b) the inverse of $\mathbf{F}$ is $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}$, which is the elementary matrix obtained from $\mathbf{I}$ by adding 2 times the first row to the third row;

(c) the inverse of $\mathbf{G}$ is given by $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, which is the elementary matrix obtained from $\mathbf{I}$ by interchanging the first and second row.

Theorem (5) The following statements are *equivalent*:

(i) $\mathbf{A}$ is nonsingular;
(ii) the homogeneous system $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution;
(iii) $\mathbf{A}$ can be reduced to $\mathbf{I}$ by a sequence of elementary row operations.

Proof: "(i) $\Rightarrow$ (ii)": Follows from Theorem (4).

"(ii) $\Rightarrow$ (iii)": Suppose $\mathbf{R}$ is the reduced row-echelon form of $\mathbf{A}$. If $\mathbf{R} \neq \mathbf{I}$, then the number of non-zero rows of $\mathbf{R}$ is less than n, thus, $\mathbf{Ax} = \mathbf{0}$ has nontrivial solutions. Therefore $\mathbf{A}$ can be reduced to $\mathbf{I}$ by elementary row operations.

"(iii) $\Rightarrow$ (i)": Suppose that there are elementary row operations $\rho_1, \rho_2, \ldots, \rho_k$ such that

$$\mathbf{A} \stackrel{\rho_1}{\to} \mathbf{A}_1 \stackrel{\rho_2}{\to} \mathbf{A}_2 \stackrel{\rho_3}{\to} \cdots \stackrel{\rho_{k-1}}{\to} \mathbf{A}_{k-1} \stackrel{\rho_k}{\to} \mathbf{A}_k = \mathbf{I}.$$

If $\mathbf{E}_j$ is the elementary matrix obtained by applying ρ_j to $\mathbf{I}$, i.e., $\mathbf{I} \stackrel{\rho_j}{\to} \mathbf{E}_j$, then by "Fact (1)", one has

$$\mathbf{A} \stackrel{\rho_1}{\to} \mathbf{E}_1\mathbf{A} \stackrel{\rho_2}{\to} \mathbf{E}_2\mathbf{E}_1\mathbf{A} \stackrel{\rho_3}{\to} \cdots \stackrel{\rho_k}{\to} \mathbf{E}_k \cdots \mathbf{E}_2\mathbf{E}_1\mathbf{A} = \mathbf{I}.$$

We conclude that $(\mathbf{E}_k \ldots \mathbf{E}_1)\mathbf{A} = \mathbf{I}$.
Since $\mathbf{E}_1, \mathbf{E}_2, \ldots, \mathbf{E}_k$ are elementary matrices, they are also nonsingular. Therefore, their product
$\mathbf{E}_k \ldots \mathbf{E}_2\mathbf{E}_1$ is nonsingular. Hence, $\mathbf{A} = (\mathbf{E}_k \ldots \mathbf{E}_1)^{-1}$ is nonsingular, and $\mathbf{A}^{-1} = \mathbf{E}_k \ldots \mathbf{E}_2\mathbf{E}_1$.

Remark

We shall describe a practical method to find $\mathbf{A}^{-1}$. In fact, it is evident from the proof of "(iii)⇒(i)" in Theorem (5) that if we find a sequence of elementary row operations that reduce $\mathbf{A}$ to $\mathbf{I}$, and then perform this same sequence of elementary operations on $\mathbf{I}$, we will be able to obtain $\mathbf{A}^{-1}$. Symbolically, we have

$$[\mathbf{A}\,|\mathbf{I}] \stackrel{\rho_1}{\to} [\mathbf{A}_1|\mathbf{E}_1\mathbf{I}] \stackrel{\rho_2}{\to} [\mathbf{A}_2\,|\mathbf{E}_2\mathbf{E}_1\mathbf{I}] \stackrel{\rho_3}{\to} \ldots\ldots \stackrel{\rho_k}{\to} [\mathbf{I}\,|\mathbf{E}_k \cdots \mathbf{E}_2\mathbf{E}_1\mathbf{I}] = [\mathbf{I}\,|\mathbf{A}^{-1}].$$

Example 4.6 Find $\begin{bmatrix} 1 & -2 & 2 \\ 2 & -3 & 6 \\ 1 & 1 & 7 \end{bmatrix}^{-1}$.

Solution

$$\left[\begin{array}{ccc|ccc} 1 & -2 & 2 & 1 & 0 & 0 \\ 2 & -3 & 6 & 0 & 1 & 0 \\ 1 & 1 & 7 & 0 & 0 & 1 \end{array}\right] \rightarrow \left[\begin{array}{ccc|ccc} 1 & -2 & 2 & 1 & 0 & 0 \\ 0 & 1 & 2 & -2 & 1 & 0 \\ 0 & 3 & 5 & -1 & 0 & 1 \end{array}\right]$$

$$\rightarrow \left[\begin{array}{ccc|ccc} 1 & -2 & 2 & 1 & 0 & 0 \\ 0 & 1 & 2 & -2 & 1 & 0 \\ 0 & 0 & 1 & -5 & 3 & -1 \end{array}\right] \rightarrow \left[\begin{array}{ccc|ccc} 1 & -2 & 0 & 11 & -6 & 2 \\ 0 & 1 & 0 & 8 & -5 & 2 \\ 0 & 0 & 1 & -5 & 3 & -1 \end{array}\right]$$

$$\rightarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 27 & -16 & 6 \\ 0 & 1 & 0 & 8 & -5 & 2 \\ 0 & 0 & 1 & -5 & 3 & -1 \end{array}\right].$$

See CoCalc https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page491-CoCalcJupyter.pdf

Another Example

https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/to_get_to_reduced_row_echelon_form_3.pdf

Another Application: Hill Cipher

https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/supplementary07.pdf
See CoCalc https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/supplementary07_CoCalcJupyter.pdf

Theorem (6) Let $\mathbf{A}$ and $\mathbf{B}$ be $n \times n$ matrices. Then $\mathbf{BA} = \mathbf{I}$ if and only if $\mathbf{AB} = \mathbf{I}$.

Proof: Suppose that $\mathbf{BA} = \mathbf{I}$. Consider the homogeneous system $\mathbf{Ax} = \mathbf{0}$. If $\mathbf{v}$ is a solution of this system, then $\mathbf{Av} = \mathbf{0}$ and

$$\mathbf{v} = \mathbf{Iv} = (\mathbf{BA})\mathbf{v} = \mathbf{B}(\mathbf{Av}) = \mathbf{B0} = \mathbf{0}.$$

Therefore, the homogeneous system $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution, and Theorem (5) implies that $\mathbf{A}$ is nonsingular, and $\mathbf{B} = \mathbf{A}^{-1}$. Thus $\mathbf{AB} = \mathbf{AA}^{-1} = \mathbf{I}$.

Theorem (7)
For any $n \times n$ matrix $\mathbf{A}$, the following statements are equivalent.

(i) $\mathbf{A}$ is nonsingular.
(ii) The system of homogeneous equations $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution.
(iii) $\mathbf{A}$ can be reduced to $\mathbf{I}$ by a sequence of elementary row operations.
(iv) The non-homogeneous system $\mathbf{Ax} = \mathbf{b}$ is consistent for every vector $\mathbf{b}$ in $\mathbb{R}_n$.

Proof: Equivalence of (i), (ii) and (iii) follows from Theorem (5).

"(i) $\Rightarrow$(iv)" is a consequence of Theorem (4). We only need to prove "(iv) $\Rightarrow$(i)".

For $1 \leq k \leq n$, let $\mathbf{e^k}$ be the k-th column of the identity matrix $\mathbf{I}$, i.e., $\mathbf{e^k}$ is the column vector whose k-th entry is equal to 1 while all other entries are zero. By the hypothesis of (iv), the system of linear equation $\mathbf{Ax} = \mathbf{e^k}$ has a solution (say $\mathbf{c^k}$) for every k.

We now construct a $n \times n$ matrix $\mathbf{B}$ whose k-th column is equal to $\mathbf{c^k}$. It is then clear from the definition of matrix multiplication that $\mathbf{AB} = \mathbf{I}$. Theorem (6) then implies that $\mathbf{A}$ is nonsingular.

Eigenvalue Problem

Eigenvalue problem

Let $\mathbf{A}$ be an $n \times n$ matrix.

Consider the system

$$\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$$

for a non-zero vector $\mathbf{v}$. Then, $\mathbf{v}$ is called the eigenvector and λ is its corresponding eigenvalue.

Example: Let $\mathbf{A} = \begin{bmatrix} 3 & 0 \\ 8 & -1 \end{bmatrix}$. The vector $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ is an eigenvector of $\mathbf{A}$ because $\mathbf{A}\mathbf{v} = \begin{bmatrix} 3 & 0 \\ 8 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \end{bmatrix} = \mathbf{3v}$.

However, $\mathbf{w} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ is not an eigenvector of $\mathbf{A}$, as $\mathbf{A}\mathbf{w} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$, which cannot be a scalar multiple of $\mathbf{w}$.

Remark

It is clear that if $\mathbf{v}$ is an eigenvector of $\mathbf{A}$ corresponding to the eigenvalue λ, then $t\mathbf{v}$ is also an eigenvector of $\mathbf{A}$ corresponding to the same eigenvalue λ, provided that $t \neq 0$.

Let us rewrite $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ as $\mathbf{A}\mathbf{v} - \lambda\mathbf{v} = \mathbf{0}$, or $(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$. Thus to find an eigenvector of A is equivalent to finding nontrivial solutions of the homogeneous system $(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$.

The homogeneous system, $(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$ has nontrivial solutions if and only if $\det(\mathbf{A} - \lambda\mathbf{I}) = \mathbf{0}$.

Theorem Let $\mathbf{A}$ be an $n \times n$ matrix with real entries. $\mathbf{A}$ real number λ is an eigenvalue of $\mathbf{A}$ if and only if $\det(\mathbf{A} - \lambda\mathbf{I}) = \mathbf{0}$.

Once we obtain an eigenvalue of $\mathbf{A}$, one can use Gaussian elimination to find the corresponding eigenvector $\mathbf{v}$ since $(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$. As the homogeneous system has infinitely many non-trivial solutions, we only need to find eigenvectors that are linearly independent. All other eigenvectors may be expressed as linear combinations of these linearly independent eigenvectors.

Remark

Let $f(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}) = \det \begin{bmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{bmatrix}$.

If follows by induction on n that $f(\lambda)$ is a polynomial of degree n with leading coefficient $(-1)^n$. $f(\lambda)$ is called the characteristic polynomial of $\mathbf{A}$. Therefore, eigenvalues of the matrix $\mathbf{A}$ are simply the real roots of the equation $f(\lambda) = 0$. $\mathbf{A}$ has *at most* n eigenvalues.

Example

Find the eigenvalues and eigenvectors of $\mathbf{A} = \begin{bmatrix} 5 & 4 \\ 1 & 2 \end{bmatrix}$.

We have characteristic polynomial

$$f(\lambda) = det\left(\begin{bmatrix} 5 - \lambda & 4 \\ 1 & 2 - \lambda \end{bmatrix}\right) = \lambda^2 - 7\lambda + 6 = 0.$$

Therefore, the eigenvalues of A are $\lambda_1 = 6$, $\lambda_2 = 1$.

Case (i) : For $\lambda_1 = 6$, the system of equations for $(\mathbf{A} - \lambda_{\mathbf{1}}\mathbf{I})\mathbf{v} = \mathbf{0}$ is

$$\left.\begin{array}{r} -v_1 + 4v_2 = 0 \\ v_1 - 4v_2 = 0 \end{array}\right\} \Rightarrow v_1 = 4v_2.$$

We therefore obtain $\mathbf{v} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$ as an eigenvector corresponding to $\lambda_2 = 6$.

Case (ii) : For $\lambda_2 = 1$, the system of equations for $(\mathbf{A} - \lambda_{\mathbf{2}}\mathbf{I})\mathbf{v} = \mathbf{0}$ is

$$\left.\begin{array}{r} 4v_1 + 4v_2 = 0 \\ v_1 + v_2 = 0 \end{array}\right\} \Rightarrow v_1 = -v_2.$$

We thus obtain $\mathbf{v} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ as an eigenvector corresponding to $\lambda_2 = 1$.

Example 7.3

If $\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 6 & -11 & 6 \end{bmatrix}$, then the characteristic equation is given by

$$f(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}) = \begin{vmatrix} -\lambda & 1 & 0 \\ 0 & -\lambda & 1 \\ 6 & -11 & 6-\lambda \end{vmatrix}$$
$$= -(\lambda-1)(\lambda-2)(\lambda-3) = 0.$$

Therefore, $\lambda_1 = 1$, $\lambda_2 = 2$, $\lambda_3 = 3$.

For $\lambda_1 = 1$, using Gaussian elimination on
$(\mathbf{A} - \lambda_1\mathbf{I})\mathbf{v} = \mathbf{0}$, i.e. $\begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 6 & -11 & 5 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \mathbf{0}$
we obtain $\mathbf{v} = [1 \quad 1 \quad 1]^T$ as an eigenvector of $\mathbf{A}$ corresponding to the eigenvalue $\lambda_1 = 1$.

For $\lambda_2 = 2$, using Gaussian elimination on

$(\mathbf{A} - \lambda_1 \mathbf{I})\mathbf{v} = \mathbf{0}$, i.e., $\begin{bmatrix} -2 & 1 & 0 \\ 0 & -2 & 1 \\ 6 & -11 & 4 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \mathbf{0}$

we obtain $\mathbf{v} = [1 \quad 2 \quad 4]^T$.

For $\lambda_3 = 3$, using Gaussian elimination on

$(\mathbf{A} - \lambda_3 \mathbf{I})\mathbf{v} = \mathbf{0}$, i.e., $\begin{bmatrix} -3 & 1 & 0 \\ 0 & -3 & 1 \\ 6 & -11 & 3 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \mathbf{0}$

we obtain $\mathbf{v} = [1 \quad 3 \quad 9]^T$.

See CoCalc

https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page499-CoCalcJupyter.pdf

Example 7.4

Let $\mathbf{A} = \begin{bmatrix} 3 & -2 & 0 \\ -2 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix}$. It then follows that

$$f(\lambda) = \begin{vmatrix} 3-\lambda & -2 & 0 \\ -2 & 3-\lambda & 0 \\ 0 & 0 & 5-\lambda \end{vmatrix} = (5-\lambda)^2(1-\lambda).$$

Therefore the eigenvalues are $\lambda_1 = 1$, $\lambda_2 = \lambda_3 = 5$.

For $\lambda_1 = 1$, solving $(\mathbf{A} - \lambda_1 \mathbf{I})\mathbf{v} = \mathbf{0}$,

i.e., $\begin{bmatrix} 2 & -2 & 0 \\ -2 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \mathbf{0}$

we obtain $\mathbf{v} = [1 \quad 1 \quad 0]^T$.

For $\lambda_2 = 5$, solving $(\mathbf{A} - \lambda_2 \mathbf{I})\mathbf{v} = \mathbf{0}$,

i.e., $\begin{bmatrix} -2 & -2 & 0 \\ -2 & -2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \mathbf{0}$ we obtain

$v_1 = t$, $v_2 = -t$ and $v_3 = s$, where t and s are arbitrary scalars. By taking $t = 1$, $s = 0$ and $t = 0$, $s = 1$ respectively, we obtain two linearly independent eigenvectors $[1 \quad -1 \quad 0]^T$ and $[0 \quad 0 \quad 1]^T$.

Example 7.5

Let $\mathbf{A} = \begin{bmatrix} 1 & 1 & -1 \\ -1 & 3 & -1 \\ -1 & 2 & 0 \end{bmatrix}$. Then $f(\lambda) = \begin{vmatrix} 1-\lambda & 1 & -1 \\ -1 & 3-\lambda & -1 \\ -1 & 2 & -\lambda \end{vmatrix} = (2-\lambda)(1-\lambda)^2$.

Therefore the eigenvalues are $\lambda_1 = 2$, $\lambda_2 = \lambda_3 = 1$.

Solving the linear systems $(\mathbf{A} - \lambda \mathbf{I})\mathbf{v} = \mathbf{0}$ for $\lambda = 2$ and $\lambda = 1$ (multiplicity 2), we obtain respectively two linearly independent eigenvectors $\mathbf{v}^1 = [0 \quad 1 \quad 1]^T$ and $\mathbf{v}^2 = [1 \quad 1 \quad 1]^T$.

See CoCalc

https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/page502-CoCalcJupyter.pdf

Geometric Progression in matrix form

https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/supplementary11.pdf

See CoCalc https://www.polyu.edu.hk/ama/profile/hwlee/AMA1007/supplementary11-CoCalcJupyter.pdf