Calculus (II)

Lecture Note 2021 Spring

Contents

Chapter 1

A Preview of Calculus

1.1 Preliminaries

❑ Notation:

■ Set of numbers

- $\mathbb{N}$ = set of all natural numbers = $\{1, 2, 3, \ldots\}$.
- $\mathbb{Z}$ = set of all integers = $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$.
- $\mathbb{Q}$ = set of all rational numbers = $\{x \,\big|\, x = \dfrac{p}{q}$, where $p, q \in \mathbb{Z}, q \neq 0\}$.

  $\mathbb{Q}^+ = \{x \in \mathbb{Q} \,\big|\, x > 0\}$.
- $\mathbb{R}$ = set of all real numbers,

  $\mathbb{R}^+ = \{x \in \mathbb{R} \,\big|\, x > 0\}$ and $\mathbb{R}^- = \{x \in \mathbb{R} \,\big|\, x < 0\} = \{-x \,\big|\, x \in \mathbb{R}^+\}$.

■ Intervals

- $(a, b) = \{x \,\big|\, a < x < b\}$ open interval.
- $[a, b] = \{x \,\big|\, a \leq x \leq b\}$ closed interval.
- $[a, b) = \{x \,\big|\, a \leq x < b\}$, $\quad (a, b] = \{x \,\big|\, a < x \leq b\}$.
- $[a, \infty) = \{x \,\big|\, x \geq a\}$, $\quad (a, \infty) = \{x \,\big|\, x > a\}$,

  $(-\infty, a] = \{x \,\big|\, x \leq a\}$, $\quad (-\infty, a) = \{x \,\big|\, x < a\}$.

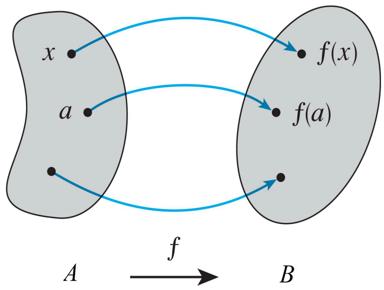
Note:

(1) ∞ and $-\infty$ do not represent real numbers.

(2) $\mathbb{R} = (-\infty, \infty)$, $\mathbb{R}^+ = (0, \infty)$ and $\mathbb{R}^- = (-\infty, 0)$.

❏ Functions

Definition 1.1.1. A function f, often called a mapping, is a rule that assigns to each element x in a set A exactly one element, called $f(x)$, in a set B.



Arrow diagram for f

The set A is called the domain of f and denoted by $Dom(f)$. The number $f(x)$ is the value of f at x.

The range (also called the image) of f, denoted by $Range(f)$ is the set of all possible values of $f(x)$ as x varies throughout the domain. That is,

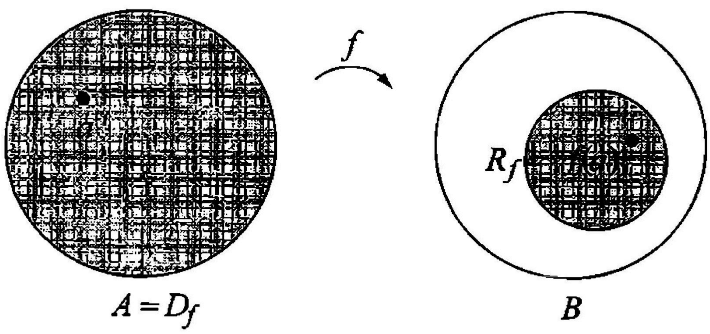
$$Range(f) = \{y \mid y = f(x) \text{ for some } x \in Dom(f)\}.$$

Note. In the class of Calculus(I), we consider those functions whose domains and ranges form subsets of $\mathbb{R}$. Thus, functions will be called real-valued functions of a real variable.

Remark.

(i) If f is a function, for each element $a \in Dom(f)$ there is exactly one element $b \in Range(f)$ such that $b = f(a)$. The value a is an independent variable and the value b is a dependent variable.

(ii) If $Dom(f) = A$ and $Range(f) \subseteq B$, then the function f from A to B is usually symbolically written as $f : A \to B$.

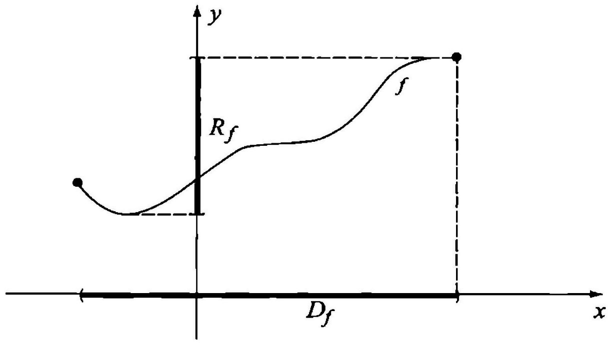


Example 1.1.2.

(1) $f(x) = \sqrt{x} : [0, 4] \to \mathbb{R}$. Then $Dom(f) = [0, 4]$ and $Range(f) = [0, 2]$.

(2) $f(x) = x^2 : (0, \infty) \to \mathbb{R}$. Then $Dom(f) = (0, \infty)$ and $Range(f) = (0, \infty)$.

Remark. If the domain of a function f is not exactly given, then $Dom(f)$ consists of all values that can have an image under f. That is, we take $Dom(f)$ as the maximal set of real number x for which $f(x)$ is a real number. In such a case $Dom(f)$ is called the natural domain.



Example 1.1.3.

(1) $f(x) = \sqrt{x}$. Then $Dom(f) = [0, \infty]$ and $Range(f) = [0, \infty]$.

(2) $f(x) = \dfrac{x^2+1}{x}$. Then $Dom(f) = (-\infty, 0) \cup (0, \infty)$ $\big($or $\{x \mid x \neq 0\}\big)$ and $Range(f) = (-\infty, -2] \cup [2, \infty)$.

■ Graph of a function

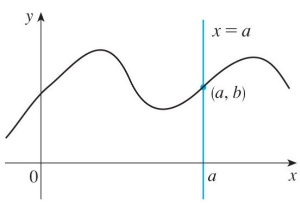
Definition 1.1.4. Let $f(x)$ be a function with domain $Dom(f)$. The graph of f consists of all points (x, y) in the coordinate plane such that $y = f(x)$ and x is in the domain of f. That is,

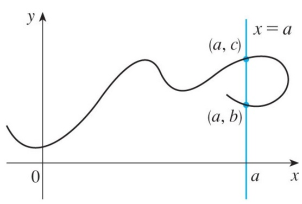
$$Graph(f) = \{(x, y) \mid y = f(x),\ x \in Dom(f)\} = \{(x, f(x)) \mid x \in Dom(f)\}.$$

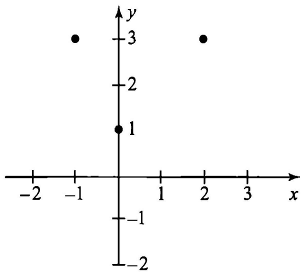
Question: Is a curve the graph a function?

■ Vertical line test

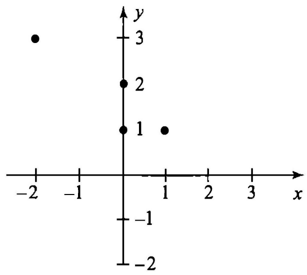
A curve C in the plane is the graph of a function if and only if no vertical line intersects C at more than one point.



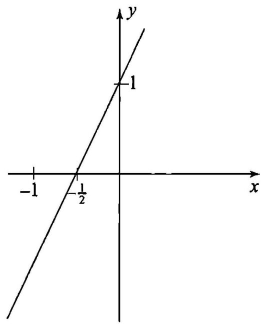




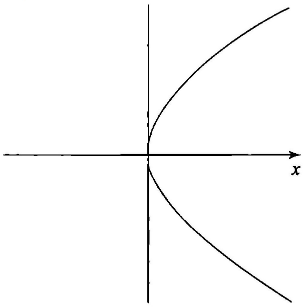
S_1



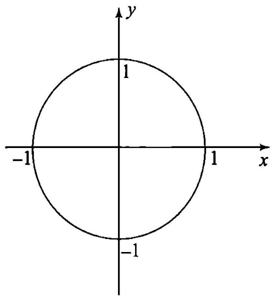
S_2



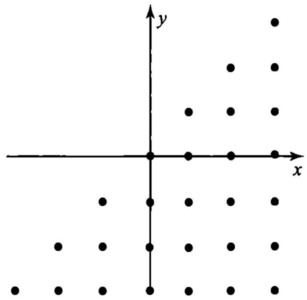
S_3



S_4



S_5



S_6

In the above diagrams, S_1 and S_3 are graphs of some functions, but S_2, S_4, S_5 and S_6 are not graphs of any function.

1.2 Bounded Sets and Functions

❑ Bounded Sets

Definition 1.2.1. (**Bounded Sets**) Let A be a set of real numbers. We say that

(a) A is bounded above if there is a number $M \in \mathbb{R}$ such that

$$a \le M \quad \text{for all } a \in A.$$

We call such a number M an "*upper bound for* A".

(b) A is bounded below if there is a number $N \in \mathbb{R}$ such that

$$a \ge N \quad \text{for all } a \in A.$$

We call such a number N a "*lower bound for* A".

(c) A is bounded if A is both bounded above and below. That is, there are $M, N \in \mathbb{R}$ such that

$$N \le a \le M \quad \text{for all } a \in A.$$

(d) If A is not bounded, we say that A is "*unbounded*".

Remark. If $A \subset \mathbb{R}$ is bounded, then there exists $M > 0$ such that

$$0 \leq |a| \leq M \quad (\text{or } -M \leq a \leq M) \quad \text{for all } a \in A$$

Definition 1.2.2. Let $A \subseteq \mathbb{R}$. We call a number a_0 a "*maximum of A*" ("*minimum of A*") if

(i) $a_0 \in A$ and

(ii) $a_0 \geq a$ ($a_0 \leq a$) for all $a \in A$.

Denoted by "max A" ("min A").

Remark.

(i) A bounded set may not have a maximum or a minimum. (ex: (0.1)).

(ii) A finite set must have a maximum and a minimum.

(iii) If a_0 is a maximum of A, then a_0 is an upper bound for A.

■ Least Upper Bound and Greatest Lower Bound

Definition 1.2.3. Let $A \subseteq \mathbb{R}$. We say that M is a "*least upper bound*" of A if

(i) M is an upper bound for A (i.e. $a \leq M$ for all $a \in A$) and

(ii) if M_1 is an upper bound for A, then $M \leq M_1$.

We denote the least upper bound for A by

$$\text{"}\sup A\text{"} \quad (\text{supremum of } A)$$

and the greatest lower bound for A by

$$\text{"}\inf A\text{"} \quad (\text{infimum of } A).$$

Lemma 1.2.4. *A least upper bound is unique. That is, if M_1 and M_2 are least upper bounds for a set A, then $M_1 = M_2$.*

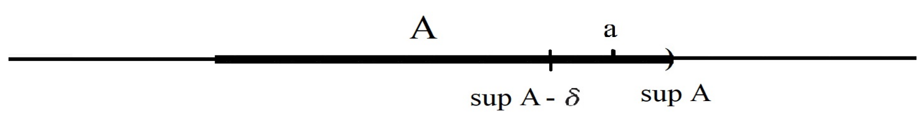
Proof. Since M_1 is a least upper bound for A and M_2 is an upper bound for A, $M_1 \leq M_2$. The converse argument is similar. We have $M_2 \leq M_1$. Therefore, $M_1 = M_2$.

□

Exercise. Prove that the supremum of A satisfies

(i) $\sup A \geq a$ for all $a \in A$

(ii) for any $\delta > 0$, there exists $a \in A$ such that $a > \sup A - \delta$.



■ Least Upper Bound Property

Theorem 1.2.5. *(**Least upper bound property**) Any nonempty set of real numbers that has an upper bound necessarily has a least upper bound.*

Proof. The proof will be postponed until Advanced Calculus.

□

Remark.

(i) A bounded set of real number may not have a maximum (or a minimum), but must have a least upper bound (and a greatest lower bound). For example, $(0, 1)$.

(ii) A maximum of a bounded set must be its supremum, but the couverse is possibly false.

(iii) In order to extend the defintion of sup and inf to more general sets, we define $\sup A = \infty$ ($\inf A = -\infty$) if A is not bounded above (A is not bounded below).

(iv) Any number is an upper bound (or a lower bound) of $\emptyset$. We define $\sup \emptyset = -\infty$ and $\inf \emptyset = \infty$.

Remark.

(i) $\mathbb{N}, \mathbb{Z}, \mathbb{Q}$ and $\mathbb{R}$ are unbounded.

(ii) For every $\varepsilon > 0$, there exists $n \in \mathbb{N}$ such that $0 < \dfrac{1}{n} < \varepsilon$.

❑ Bounded Functions

Definition 1.2.6. (**Bounded Functions**) Let $f : D \mapsto \mathbb{R}$ be a function. We say that

(a) f is bounded above (on D) if the set of the values of $f(x)$ on D is bounded above. That is,

"*the set* $\{f(x) \mid x \in D\}$ *is bounded above*"

or

"*there exists a number* $M \in \mathbb{R}$ *such that* $f(x) \leq M$ *for all* $x \in D$".

We call such a number M an "*upper bound for* $f(x)$" and "$f(x)$ *is bounded above by* M".

(b) f is bounde below on D if the set of the values of $f(x)$ on D is bounded below. That is,

"*the set* $\{f(x) \mid x \in D\}$ *is bounded below*"

or

"*there exists a number* $N \in \mathbb{R}$ *such that* $f(x) \geq N$ *for all* $x \in D$".

We call such a number N an "*lower bound for* $f(x)$" and "$f(x)$ *is bounded below by* N".

(c) f is bounded on D if f is both bounded above and below on D. That is

$$\text{“}\textit{the set } \{f(x) \mid x \in D\} \textit{ is bounded}\text{”}$$

or

$$\text{“}\textit{there exists } M, N \in \mathbb{R} \textit{ such that } N \leq f(x) \leq M \textit{ for all } x \in D\text{”}.$$

(d) If f is not bounded, we say that “*f is unbounded.*”.

Remark. We usually say that “*f is bounded (above, below)*” if f is bounded (above, below) on its domain.

Definition 1.2.7. Let f be a function and $D \subseteq Dom(f)$. We say that a number $L \in \mathbb{R}$ is

(a) “*the maximum (value) of $f(x)$ on D*” if L is the maximum of the set of the values of $f(s)$ on D. That is,

$$L = \max \left\{ f(x) \mid x \in D \right\} = \max_{x \in D} f(x)$$

or

(i) there is $a_0 \in D$ such that $f(a_0) = L$; and (ii) $f(a) \leq L$ for all $a \in D$.

(b) “*the minimum (value) of $f(x)$ on D*” if L is the minimum of the set of the values of $f(s)$ on D. That is,

$$L = \min \left\{ f(x) \mid x \in D \right\} = \min_{x \in D} f(x)$$

or

(i) there is $a_0 \in D$ such that $f(a_0) = L$; and (ii) $f(a) \geq L$ for all $a \in D$.

Similarly, we can also define the supremum and infimum of a function on a set.

Definition 1.2.8. Let f be a function and $D \subseteq Dom(f)$. Define

(a) the supremum of f on D by

$$\sup_{x \in D} f(x) = \sup \left\{ f(x) \mid x \in D \right\}.$$

and

(b) the infimum of f on D by

$$\inf_{x \in D} f(x) = \inf \left\{ f(x) \mid x \in D \right\}.$$

Remark.

(i) A bounded function may not have a maximum or a minimum.

(ii) If the range of $f(x)$ contains at most finitely many numbers, then f is bounded and contains a maximum and a minimum.

1.3 Inequalities

❏ Inequalities

■ **Triangle Inequality** Let $a, b \in \mathbb{R}$ be two numbers. Then

(i) $|a + b| \leq |a| + |b|$.

(ii) $|a| - |b| \leq |a - b|$.

(iii) $\big||a| - |b|\big| \leq |a - b|$.

Chapter 2

Sequences

In high school algebra, students already learned some basic concepts of sequence (of numbers) and knew some examples of spcific sequences. For instance, arithmetic sequence ($\{-3,-1,1,3,5,7,9,\ldots\}$ geometric sequence($\{1,3,9,27,81,\ldots\}$), sequence defined by recursion ($a_{n+2} = a_n + a_{n+1}$) etc. We realize a sequence as a set of numbers with order. In this chapter, we introduce a new viewpoint to look at a sequence. A sequence can be regarded as resulting functional values. This idea will be not only inherited the concept of functions in Chapter 1, but also generalized to the limit of general functions in Chapter 3.

We can see a sequence everywhere. For example, the irrational number π is corresponding an infinite sequence(series).

$$\begin{aligned}
\pi &= 3.1415926\ldots \\
&= 3 + 1\cdot(\frac{1}{10}) + 4\cdot(\frac{1}{10^2}) + 1\cdot(\frac{1}{10^3}) + 5\cdot(\frac{1}{10^4}) + 9\cdot(\frac{1}{10^5}) + 2\cdot(\frac{1}{10^6}) + 6\cdot(\frac{1}{10^7}) + \cdots.
\end{aligned}$$

The most basic idea in analysis is the concept of a limit. The simplest version of a limit appears in the study of sequences. In this chapter, we will study the rigorous definition and proof about sequences.

In this chapter, we will only discuss Sec2.1-Sec2.3 in the textbook and the remaining sections will be studied after Chpater 6.

2.1 Convergence

A sequence (of numbers) can be thought of as a list of numbers written in a definite order

$$a_1,\ a_2,\ a_3,\ a_4,\ \ldots,\ a_n,\ \ldots$$

It can be regarded as a list of values of a function defined on $\mathbb{N}$.

$$\begin{array}{cccccccc} & \mathbb{N} & & 1 & 2 & 3 & 4 & \dots \\ f & \downarrow & & \downarrow & \downarrow & \downarrow & \downarrow & \\ & \mathbb{R} & & f(1) & f(2) & f(3) & f(4) & \dots \end{array}$$

Note. From now on, we say "a sequence" instead of "a sequence of numbers" for the convenience.

Definition 2.1.1. An (infinite) sequence is a function whose domain is a set of the form $\{n \in \mathbb{Z} \mid n \geq m\}$, when m is a fixed integer.

Remark.

(i) The common choices for m are $0,\ 1$ or 2.

(ii) By convention, we usually write the functional value $f(n)$ as a_n and denote the sequence $\{a_n\}_{n=1}^{\infty}$ (or simply $\{a_n\}$ if n begins with 1).

(iii) The values $a_1,\ a_2,\ a_3,\ \dots,\ a_n,\ \dots$ are called the "term" (first term, second term,..., nth term, ...) of a sequence.

(iv) To distinguish the notation of a set with the one of a sequence, we use $\{a_n \mid n \in \mathbb{N}\}$ to represent a set and $\{a_n\}$ for a sequence.

Example 2.1.2.

(1) $\left\{\dfrac{n}{n+1}\right\}_{n=1}^{\infty}, \rightsquigarrow a_n = \dfrac{n}{n+1} \quad \Rightarrow \quad \left\{\dfrac{1}{2}, \dfrac{2}{3}, \dfrac{3}{4}, \dfrac{4}{5}, \dots\right\}.$

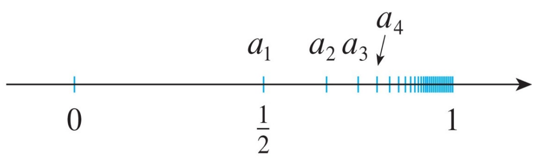
(2) $\left\{\cos\dfrac{n\pi}{6}\right\}_{n=0}^{\infty}, \rightsquigarrow a_n = \cos\dfrac{n\pi}{6}, n \geq 0 \ \Rightarrow \quad \left\{1, \dfrac{\sqrt{3}}{2}, \dfrac{1}{2}, 0, \dots\right\}.$

(3) (Fibonacci sequence)
$a_1 = 1, a_2 = 1, a_n = a_{n-1} + a_{n-2}$ for $n \geq 3 \quad \Rightarrow \quad \{1, 1, 2, 3, 5, 8, 13, 21, \dots\}.$

■ Visualization of sequence

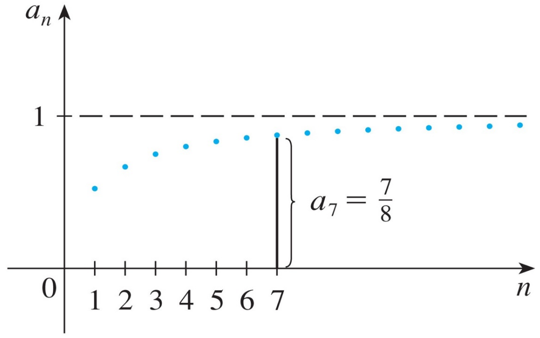
(i) Plot all terms of a sequence on number line.

Example: $a_n = \dfrac{n}{n+1}$.



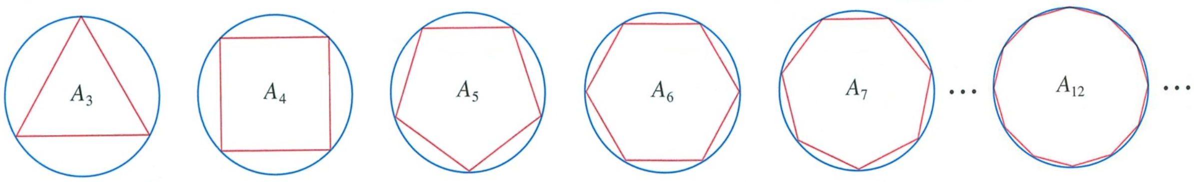
(ii) Regard a sequence a a function. $f : \mathbb{N} \mapsto \mathbb{R}$ by $a_n = f(n)$. Plot the graph of f. $(1, a_1)$, $(2, a_2), \ldots,$ (n, a_n).

Example: $f(n) = \dfrac{n}{n+1}$



Observation: From the above figures, the functional values a_n approaches as close to 1 as possible when n becomes large.

Note. People studied the limit of sequences over thousands of years. For example, to compute the are a a circle.



Question: Does A_n approach a number as n becomes large?

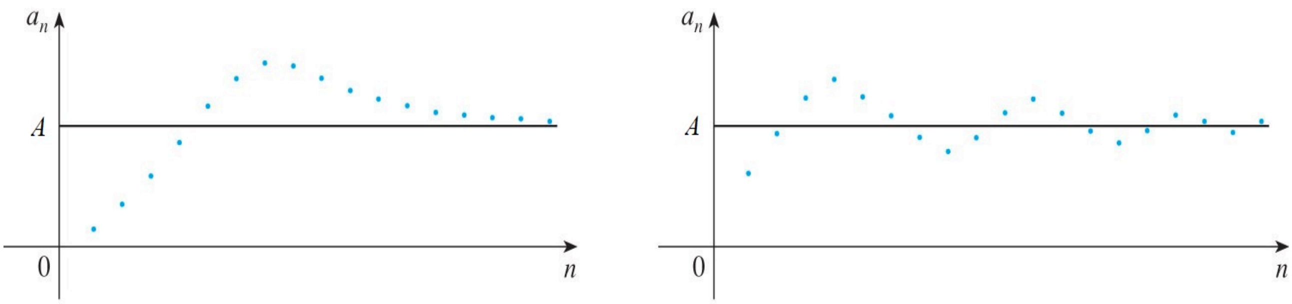
❑ Limit and Convergence

■ **Intuitive Definition:** Let $\{a_n\}$ be a sequence. We say that "the limit of $\{a_n\}$ exists" if there exists a real number $A \in \mathbb{R}$ such that we can make the term a_n as clos to A as we like by taking n sufficiently large. Denote

$$\lim_{n \to \infty} a_n = A$$

or

$$a_n \to A \quad \text{as} \quad n \to \infty$$



Graphs of two sequences with $\lim_{n \to \infty} a_n = A$

Example 2.1.3.

(1) $\left\{1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n}, \ldots\right\}$, $a_n = \frac{1}{n}$. Then $\lim_{n\to\infty} a_n = 0$.

(2) $\left\{1, -1, 1, -1, \ldots\right\}$, $a_n = (-1)^{n-1}$. Then $\lim_{n\to\infty} a_n$ does not exist (DNE).

Definition 2.1.4. (Precise) Let $\{a_n\}$ be a sequence.

(a) We say that "the limit of $\{a_n\}$ exists" if there exists a real number $A \in \mathbb{R}$ such that for every $\varepsilon > 0$ there is a corresponding integer N such that

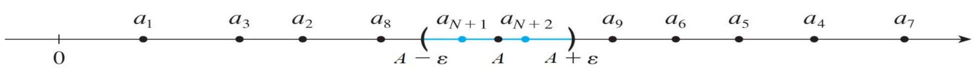
$$|a_n - A| < \varepsilon \quad \text{for all } n \geq N.$$

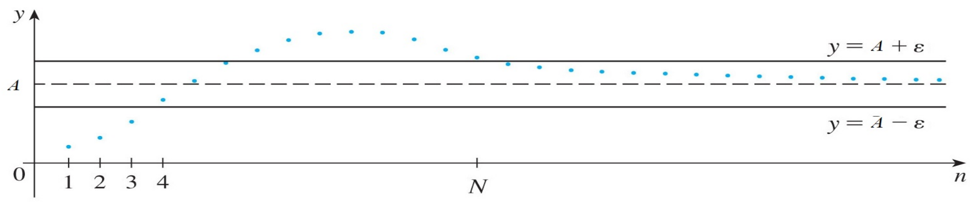
Here A is called "the limit of $\{a_n\}$" and we write

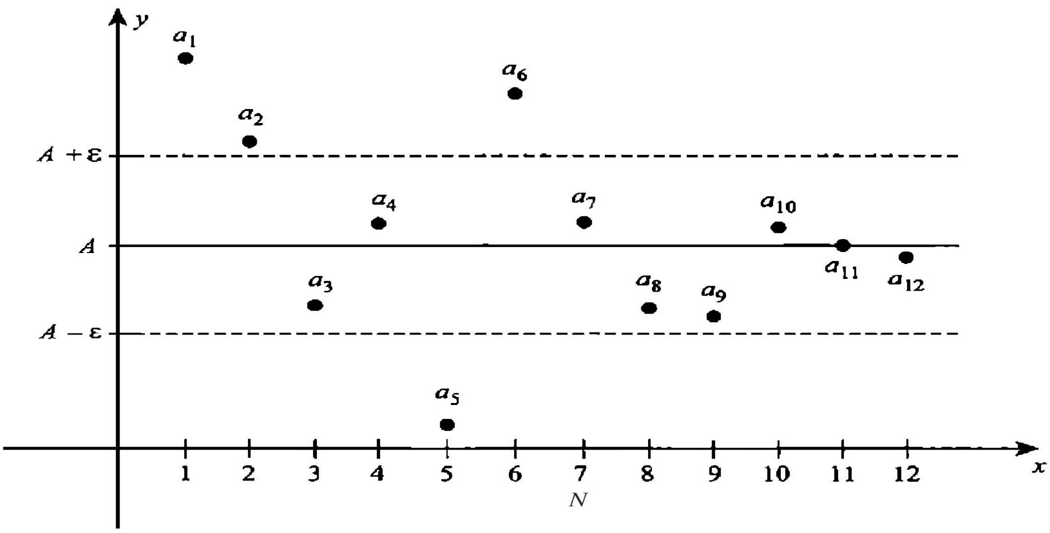
$$\lim_{n\to\infty} a_n = A$$

or

$$a_n \to A \quad \text{as} \quad n \to \infty$$







(b) If $\{a_n\}$ has a limit A. (i.e. $\lim_{n\to\infty} a_n = A$), we say that the sequence "convergs to A". Otherwise, we say that the sequence "diverges".

Example 2.1.5. Consider the sequence $\left\{\dfrac{n}{n+2}\right\}_{n=1}^{\infty}$. For $\varepsilon = 0.01$, find a positive integer N such that

$$\left|\frac{n}{n+2} - 1\right| < 0.01 \quad \text{for all } n \geq N$$

Proof. From the experience, we know (guess) that the limit, A, of the sequence is 1. Consider

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

It suffices to pick a positive integer N such that $\left|\dfrac{2}{n+2}\right| < 0.01$ for all $n \geq N$.Compute

$$\frac{2}{n+2} < 0.01 \quad \Longleftrightarrow \quad 200 < n+2.$$

Then $\left|\dfrac{n}{n+2} - 1\right| < 0.01$ when $n > 198$. Hence, we choose $N = 199$ and

$$\left|\frac{n}{n+2} - 1\right| < 0.01 \text{ for all } n \geq N.$$

□

Note. (1) In the above example, we can choose any positive integer N which is greater than 199. For instance, we choose $N = 500$. Then we stil have

$$\left|\frac{n}{n+2} - 1\right| < 0.01 \text{ for all } n \geq N.$$

(2) The above example does not prove the sequence converges to 1 since the ε is a specific number but not an arbitrary.

Example 2.1.6. Prove that the sequence $\left\{\dfrac{n}{n+2}\right\}_{n=1}^{\infty}$ converges to 1.

Proof.

By the definition, we need to show that "for any given $\varepsilon > 0$, there exists a corresponding positive integer $N = N(\varepsilon)$ such that $\left\|\dfrac{n}{n+2} - 1\right\| < \varepsilon$ for all $n \geq N$".

Given $\varepsilon > 0$, W.L.O.G. say $0 < \varepsilon < 1$, consider

$$\left|\frac{n}{n+2} - 1\right| = \frac{2}{n+2} < \varepsilon \quad \Longleftrightarrow \quad 2 - 2\varepsilon < n\varepsilon \quad \Longleftrightarrow \quad \frac{2-2\varepsilon}{\varepsilon} < n.$$

Choose $N \in \mathbb{N}$ such that $N \geq \dfrac{2-2\varepsilon}{\varepsilon}$. Then, for every $n \geq N$,

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

Hence, the sequence $\left\{\dfrac{n}{n+2}\right\}_{n=1}^{\infty}$ converges to 1.

□

Note. In the book, it provides a new estimate by using $\dfrac{2}{n+2} < \dfrac{2}{n}$. It may easily find the positive integer N.

Example 2.1.7. Prove that the sequence $\left\{\dfrac{n^2-2}{2n^3-n-1}\right\}_{n=1}^{\infty}$ converges to 0.

> **Difficulty:** In this example, it is difficult to find an exact positive integer N such that the definition holds since we should solve the inequality $n^2 - 2 < (2n^3 - n - 1)\varepsilon$. But it is not necessary. We only need to find a suitable integer.
> **Strategy:** To find a simpler middle term $(\star)$ such that for every large n,
>
> $$\left|\frac{n^2-2}{2n^3-n-1} - 0\right| < (\star) < \varepsilon.$$

Proof. For $n \geq 2$, the numerator $(n^2 - 2)$ and the denominator $(2n^3 - n - 1)$ are positive. $\big[$We observe that $2n^3 - n - 1 > n^3$ when n is large.$\big]$ Also,

$$2n^3 - n - 1 > n^3 \quad \Longleftrightarrow \quad n(n^2 - 1) > 1.$$

Then $2n^3-n-1 > n^3$ when $n \geq 2$. $\boxed{\text{Hence, for } n \geq 2,\ \dfrac{n^2}{2n^3-n-1} < \dfrac{n^2}{n^3} = \dfrac{1}{n} < \varepsilon \text{ where } (\star) = \dfrac{1}{n}.}$

Given $\varepsilon > 0$, choose a postive integer $N \geq \max(2, \frac{1}{\varepsilon})$. For every $n \geq N$,

$$\left|\frac{n^2-2}{2n^3-n-1} - 0\right| = \frac{n^2-2}{2n^3-n-1} < \frac{n^2}{n^3} = \frac{1}{n} < \frac{1}{1/\varepsilon} = \varepsilon.$$

Hence, the sequence $\left\{\dfrac{n^2-2}{2n^3-n-1}\right\}_{n=1}^{\infty}$ converges to 0.

□

Exercise.

(i) Prove that the sequence $\left\{\dfrac{1}{n}\right\}$ converges to 0

(ii) Prove that the sequence $\left\{\dfrac{1}{n^p}\right\}$ converges to 0 for every $p \in \mathbb{N}$. (Hint: you may use the fact that $\dfrac{1}{n^p} \leq \dfrac{1}{n}$.)

Recall. The following three statements are equivalent.

- *A sequence $\{a_n\}$ converges.*
- *there is a real number $A \in \mathbb{R}$ such that $\lim\limits_{n\to\infty} a_n = A$.*

- *there is a real number A such that for every $\varepsilon > 0$ there exists a corresponding integer $N \in \mathbb{N}$ such that $|a_n - A| < \varepsilon$ whenever $n \geq N$.*

Negation of Definition of Convergence:

- *A sequence $\{a_n\}$ diverges.*
- *for every real number $A \in \mathbb{R}$, $\lim\limits_{n\to\infty} a_n \neq A$.* (Careful for the notation $\lim\limits_{n\to\infty}$)
- *For every real number A, there exists corresponding $\varepsilon = \varepsilon(A) > 0$ such that for every $N \in \mathbb{N}$ there is $n = n(N) > N$ such that $|a_n - A| \geq \varepsilon$.*

Example 2.1.8. Prove that the sequence $\{\sqrt{n}\}$ diverges.

Proof. (By a contradiction)
Assume that the sequence $\{\sqrt{n}\}$ converges. Then there exist a number $A \in \mathbb{R}$ such that $\lim\limits_{n\to\infty} \sqrt{n} = A$. Let $\varepsilon = 1$. It suffices to show that, for every $N \in \mathbb{N}$, we can choose $n \geq N$ such that $|\sqrt{n} - A| > 1$.

Consider

$$\left|\sqrt{n} - A\right| \geq \sqrt{n} - |A|$$

For $N \in \mathbb{N}$, we can choose a positive integer $n > \max\left(N, (|A| + 1)^2\right)$. Then

$$\left|\sqrt{n} - A\right| \geq \sqrt{n} - |A| > \sqrt{(|A| + 1)^2} - |A| = 1.$$

Hence, the sequence $\{\sqrt{n}\}$ does not converge to A. By the contradiction, the sequence $\{\sqrt{n}\}$ diverges. □

Exercise.

(i) Prove that the sequence $\{(-1)^n\}$ diverges.

(ii) Let r be a real number with $|r| > 1$. Prove that the sequence $\{r^n\}$ diverges.

(iii) For $0 < r < 1$, prove that the sequence $\{r^n\}$ converges to 0. (Hint: write $r = \dfrac{1}{1+b}$ for some $b > 0$. Show that $0 < r^n < \dfrac{1}{nb}$ for all $n \in \mathbb{N}$ and complete the proof.)

Theorem 2.1.9. *(Uniqueness of Limit) If $\lim\limits_{n\to\infty} a_n = A$ and $\lim\limits_{n\to\infty} a_n = B$ then $A = B$.*

Proof. For a given $\varepsilon > 0$, since $\lim\limits_{n\to\infty} a_n = A$, there exists $N_1 \in \mathbb{N}$ such that

$$|a_n - A| < \frac{\varepsilon}{2}$$

whenever $n \geq N_1$. Similarly, since $\lim\limits_{n\to\infty} a_n = B$, there exists $N_2 \in \mathbb{N}$ such that

$$|a_n - B| < \frac{\varepsilon}{2}$$

f $n \geq N_2$.

Let $N = \max(N_1, N_2)$. Then

$$\begin{aligned} |A - B| &= |A - a_N + a_N - B| \\ &\leq |A - a_N| + |a_N - B| \quad \text{(triangle inequality)} \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\ &= \varepsilon \end{aligned}$$

Since ε is an arbitrary positive number, we have $A = B$. □

Note. (1) This theorem says that if the limit of a sequence exists, then the limit must be unique.

(2) It is easiler to prove this theorem by using the method of contradiction.

❑ Boundedness of Convergent Sequence

Definition 2.1.10. We say that

(a) a sequence $\{a_n\}$ is bounded above if there exists a number M such that $a_n < M$ for all $n \in \mathbb{N}$.

(b) a sequence $\{a_n\}$ is bounded below if there exists a number N such that $a_n > N$ for all $n \in \mathbb{N}$.

(c) a sequence $\{a_n\}$ is bounded if it is both bounded above and below. That is, there is a number $M > 0$ such that $|a_n| < M$ for all $n \in \mathbb{N}$.

(d) a sequence is unbounded if it is not bounded.

Theorem 2.1.11. *Every convergent sequence is bounded.*

Proof.

Idea: Every finite numbers are bounded. We only need to consider the "tail" of a sequence. The convergence of a sequence will control all terms of the tail near its limit.

Let $\{a_n\}$ be a convergent sequence with limit A. For $\varepsilon = 1$, there exists an integer $N \in \mathbb{N}$ such that $|a_n - A| < 1$ for all $n \geq N$. We obtain

$$A - 1 < a_n < A + 1$$

for all $n \geq N$. Then

$$-(|A| + 1) \leq A - 1 < a_n < A + 1 \leq |A| + 1$$

for all $n \geq N$.

Consider the bound of the first N terms. Define $M_1 := \max(|a_1|, |a_2|, \ldots, |a_{N_1}|)$, then

$$-M \leq a_1, a_2, \ldots, a_{N-1} \leq M.$$

Let $M = \max(M_1, |A| + 1) > 0$. We have $-M \leq a_n \leq M$ for all $n \in \mathbb{N}$. Hence, the sequence $\{a_n\}$ is bounded. □

Note.

(1) We could simply choose a specific bound $M = 1 + |A| + |a_1| + |a_2| + \ldots + |a_{N-1}|$.

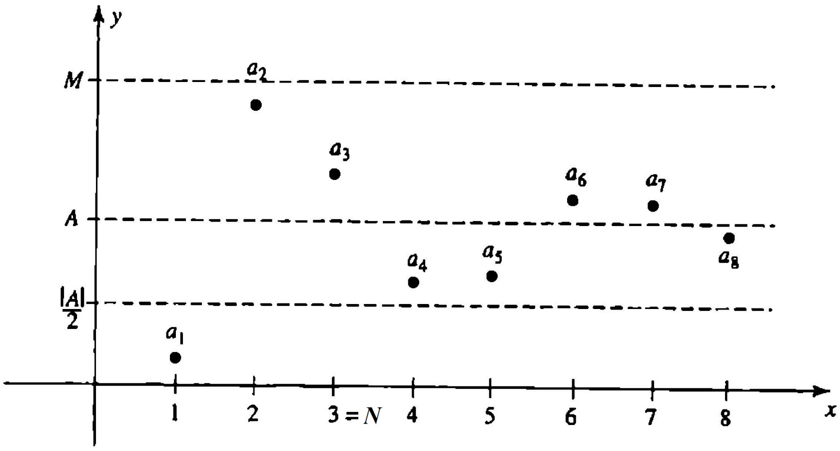
(2) A divergent sequence may be bounded $\big($e.g. $\{(-1)^n\}\big)$ or unbounded $\big($e.g. $\{n\}\big)$.

(3) A bounded sequence may not be convergnet.

Corollary 2.1.12. *Every unbounded sequence is divergent.*

Theorem 2.1.13. *Suppose that $\{a_n\}$ be a sequence which converges to A where $A \neq 0$. Then there exists $N \in \mathbb{N}$ such that $a_n \neq 0$ for all $n \geq N$. In fact, $|a_n| \geq \frac{1}{2}|A|$ for all $n \geq N$.*

Proof.



Since $\{a_n\}$ converges to A and $A \neq 0$, for $\varepsilon = \dfrac{|A|}{2} > 0$, there exists $N \in \mathbb{N}$ such that $|a_n - A| < \dfrac{|A|}{2}$ for all $n \geq N$. Then

$$|a_n| = |a_n - A + A| \geq |A| - |a_n - A| > |A| - \frac{|A|}{2} = \frac{|A|}{2}$$

for all $n \geq N$. □

Note. Heuristically, if a sequence converges to a nonzero number, then the term a_n will be "bounded away from 0 for sufficiently large n.

Theorem 2.1.14. *The sequence $\{a_n\}$ converges to* 0 *if and only if the sequence $\{|a_n|\}$ converges to* 0.

Proof. Observe that

$$\big|\,|a_n| - 0\,\big| = |a_n| = |a_n - 0| \tag{2.1}$$

($\Rightarrow$) Given $\varepsilon > 0$, since $\lim_{n\to\infty} a_n = 0$, there exists $N \in \mathbb{N}$ such that $|a_n - 0| < \varepsilon$ for all $n \geq N$. Then, by (2.1),

$$\big|\,|a_n| - 0\,\big| < \varepsilon$$

for all $n \geq N$. Hence, the sequence $\{|a_n|\}$ converges to 0.

($\Leftarrow$) Similarly, if $\big|\,|a_n| - 0\,\big| < \varepsilon$ for all $n \geq N$, we can also use (2.1) to prove

$$|a_n - 0| < \varepsilon$$

for all $n \geq N$. Hence, the sequence $\{a_n\}$ converges to 0. $\square$

Remark. We can observe that some qualitative problems, such as convergence, divergence, boundedness etc, of a sequence only depends on the "*tails* " of the sequence. Any finitely many terms do not change those properties

2.2 Limit Theorems

In this section, we will discuss some properties of limits.

❑ Limit Laws

Theorem 2.2.1. *If sequences $\{a_n\}$ and $\{b_n\}$ converge to A and B, respectively and C is a constant number, then*

(a) $\lim\limits_{n\to\infty}(a_n + b_n) = A + B$. $\left[\lim\limits_{n\to\infty}(a_n + b_n) = \lim\limits_{n\to\infty} a_n + \lim\limits_{n\to\infty} b_n\right]$.

(b) $\lim\limits_{n\to\infty} Ca_n = CA$. $\left[\lim\limits_{n\to\infty}(Ca_n) = C\lim\limits_{n\to\infty} a_n\right]$.

(c) $\lim\limits_{n\to\infty}(a_n - b_n) = A - B$. $\left[\lim\limits_{n\to\infty}(a_n + b_n) = \lim\limits_{n\to\infty} a_n - \lim\limits_{n\to\infty} b_n\right]$

(d) $\lim\limits_{n\to\infty}(a_nb_n) = AB$. $\left[\lim\limits_{n\to\infty}(a_nb_n) = (\lim\limits_{n\to\infty} a_n)(\lim\limits_{n\to\infty} b_n)\right]$

(e) $\lim\limits_{n\to\infty}\left(\dfrac{a_n}{b_n}\right) = \dfrac{A}{B}$ *provided* $B \neq 0$. $\left[\lim\limits_{n\to\infty}\dfrac{a_n}{b_n} = \dfrac{\lim_{n\to\infty} a_n}{\lim_{n\to\infty} b_n}\right]$

(f) $\lim\limits_{n\to\infty}(a_n)^p = A^p$ *for all* $p \in \mathbb{N}$.

(g) $\lim\limits_{n\to\infty}\sqrt[k]{a_n} = \sqrt[k]{A}$ *if A and a_n are nonnegative for all n with $k \in \mathbb{N}$.*

(h) if $a_n \leq b_n$ for all $n \geq N \in \mathbb{N}$, then $A \leq B$.

Proof. We will prove part(a), (b), (d), (f) and (h) here, and skip (c), (e), and (g).

(a) Since $\lim\limits_{n\to\infty} a_n = A$ and $\lim\limits_{n\to\infty} b_n = B$, for given $\varepsilon > 0$ there are integers $N_1, N_2 \in \mathbb{N}$ such that

$$|a_n - A| < \frac{\varepsilon}{2} \quad \text{for all } n \geq N_1$$

and

$$|b_n - B| < \frac{\varepsilon}{2} \quad \text{for all } n \geq N_2.$$

Let $N = \max(N_1, N_2)$. Then, for $n \geq N$,

$$\big|(a_n + b_n) - (A + B)\big| = \big|(a_n - A) + (b_n - B)\big| \leq \big|a_n - A\big| + \big|b_n - B\big| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Hence, $\lim\limits_{n\to\infty}(a_n + b_n) = A + B$.

(b) For $C = 0$, $Ca_n = 0$ for all $n \in \mathbb{N}$. Then $\lim_{n\to\infty} Ca_n = \lim_{n\to\infty} 0 = 0 = CA$.
Suppose that $C \neq 0$. Since $lim_{n\to\infty} a_n = A$, for given $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$|a_n - A| < \frac{\varepsilon}{|C|} \quad \text{for all } n \geq N.$$

Then

$$\left|Ca_n - CA\right| = \left|C(a_n - A)\right| = |C|\left|a_n - A\right| < |C|\frac{\varepsilon}{|C|} = \varepsilon \quad \text{for all } n \geq N.$$

Hence, $\lim_{n\to\infty} Ca_n = CA$.

(d)

> **A priori estimate:**
>
> $$|a_n b_n - AB| = |a_n b_n - Ab_n + Ab_n - AB| = |(a_n - A)b_n| + |A(b_n - B)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2}.$$

Since $\{b_n\}$ is a convergent sequence, it is bounded. There exists a number $M > 0$ such that $|b_n| < M$ for all $n \in \mathbb{N}$. Also, since $\lim_{n\to\infty} a_n = A$ and $\lim_{n\to\infty} b_n = B$, for given $\varepsilon > 0$ there are integers $N_1, N_2 \in \mathbb{N}$ such that

$$|a_n - A| < \frac{\varepsilon}{2M} \quad \text{for all } n \geq N_1$$

and

$$|b_n - B| < \frac{\varepsilon}{2(|A| + 1)} \quad \text{for all } n \geq N_2.$$

Let $N = \max(N_1, N_2)$. Then, for $n \geq N$,

$$\begin{aligned}
\left|a_n b_n - AB\right| &= \left|a_n b_n - Ab_n + Ab_n - AB\right| \\
&\leq \left|(a_n - A)b_n\right| + \left|A(b_n - B)\right| \\
&\leq \left|a_n - A\right||b_n| + |A|\left|b_n - B\right| \\
&\leq \frac{\varepsilon}{2M} \cdot M + \frac{\varepsilon}{2(|A| + 1)} \cdot |A| \\
&\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.
\end{aligned}$$

Hence, $\lim_{n\to\infty}(a_n b_n) = AB$.

(e) (Exercise!) Skip the proof here

> **A priori estimate:** (Hint!)
>
> $$\begin{aligned}
> \left|\frac{a_n}{b_n} - \frac{A}{B}\right| &= \left|\frac{a_n B - Ab_n}{Bb_n}\right| = \left|\frac{a_n B - AB + AB - Ab_n}{Bb_n}\right| \\
> &\leq \left|\frac{(a_n - A)B}{Bb_n}\right| + \left|\frac{A(B - b_n)}{Bb_n}\right| = \left|\frac{(a_n - A)}{b_n}\right| + \left|\frac{A(B - b_n)}{Bb_n}\right|
> \end{aligned}$$

(f)

A priori estimate:

$$a^p - b^p = (a-b)(a^{p-1} + a^{p-2}b + a^{p-3}b^2 + \ldots + ab^{p-2} + b^{p-1}) \quad \text{for all } p \in \mathbb{N}.$$

Since $\lim_{n\to\infty} a_n = A$, for given $\varepsilon > 0$, there exists a number $N \in \mathbb{N}$ such that

$$|a_n - A| < \frac{\varepsilon}{p(|A|+1)^{p-1}}$$

for all $n \geq N$. W.L.O.G, we may assume $\varepsilon < 1$. Then $|a_n - A| < \dfrac{\varepsilon}{p(|A|+1)^{p-1}} < 1$ and hence $|a_n| < |A| + 1$. For $n \geq N$,

$$\begin{aligned}
|a^p - A^p| &= \left|(a_n - A)(a_n^{p-1} + a_n^{p-2}A + a_n^{p-3}A^2 + \ldots + a_nA^{p-2} + A^{p-1})\right| \\
&= |a_n - A|\left|a_n^{p-1} + a_n^{p-2}A + a_n^{p-3}A^2 + \ldots + a_nA^{p-2} + A^{p-1}\right| \\
&\leq |a_n - A|\left[|a_n|^{p-1} + |a_n|^{p-2}|A| + |a_n|^{p-3}|A|^2 + \ldots + |a_n||A|^{p-2} + |A|^{p-1}\right] \\
&\leq |a_n - A|\left[(|A|+1)^{p-1} + (|A|+1)^{p-1} + \ldots + (|A|+1)^{p-1}\right] \\
&= |a_n - A| \cdot p(|A|+1)^{p-1} < \varepsilon.
\end{aligned}$$

Thus, $\lim_{n\to\infty}(a_n)^p = A^p$.

(h)

<u>Exercise:</u> (by using a contradiction)

$$A \leq B \quad \Longleftrightarrow \quad A < B + \varepsilon \quad \text{for every } \varepsilon > 0.$$

Since $\lim_{n\to\infty} a_n = A$ and $\lim_{n\to\infty} b_n = B$, for given $\varepsilon > 0$ there are integers $N_1, N_2 \in \mathbb{N}$ such that

$$|a_n - A| < \frac{\varepsilon}{2} \quad \text{for all } n \geq N_1$$

and

$$|b_n - B| < \frac{\varepsilon}{2} \quad \text{for all } n \geq N_2.$$

Then, for $N = \max(N_1, N_2)$,

$$A < a_N + \frac{\varepsilon}{2} \quad \text{and} \quad B > b_N - \frac{\varepsilon}{2}.$$

We have

$$A < a_N + \frac{\varepsilon}{2} \leq b_N + \frac{\varepsilon}{2} < B + \varepsilon.$$

Since ε is an arbitrary positive number, we have $A \leq B$.
(<u>Note:</u> We can also directly use the method of contradiction to prove this part.)

□

Remark.

(i) The convergence of the two sequences is necessary. (Exercise: Give examples that the laws are false if losing convergence.)

(ii) In part (h), the strict inequality is not preserved by limits.

Example 2.2.2. (1) Does the limit $\lim_{n\to\infty} \frac{2n-1}{\sqrt{10+n^2}}$ converge? If yes, find the limit.

Proof.

$$\begin{aligned}
\lim_{n\to\infty} \frac{2n-1}{\sqrt{10+n^2}} &= \lim_{n\to\infty} \Big(\frac{2n-1}{\sqrt{10+n^2}} \cdot \frac{\frac{1}{n}}{\frac{1}{n}} \Big) = \lim_{n\to\infty} \frac{2-\frac{1}{n}}{\sqrt{\frac{10}{n^2}+1}} \\
&= \frac{\lim_{n\to\infty}(2-\frac{1}{n})}{\lim_{n\to\infty} \sqrt{\frac{10}{n^2}+1}} = \frac{\lim_{n\to\infty} 2 - \lim_{n\to\infty} \frac{1}{n}}{\sqrt{\lim_{n\to\infty} \frac{10}{n^2} + \lim_{n\to\infty} 1}} \\
&= \frac{2+0}{\sqrt{0+1}} = 2
\end{aligned}$$

□

(2) Find the limit $\lim_{n\to\infty}(\sqrt{n^2+1}-n)$ if it exists.

Proof.

$$\begin{aligned}
\lim_{n\to\infty}(\sqrt{n^2+1}-n) &= \lim_{n\to\infty} \Big[(\sqrt{n^2+1}-n) \cdot \frac{\sqrt{n^2+1}+n}{\sqrt{n^2+1}+n} \Big] = \lim_{n\to\infty} \frac{1}{\sqrt{n^2+1}+n} \\
&= \lim_{n\to\infty} \Big(\frac{1}{\sqrt{n^2+1}+n} \cdot \frac{\frac{1}{n}}{\frac{1}{n}} \Big) = \lim_{n\to\infty} \frac{\frac{1}{n}}{\sqrt{1+\frac{1}{n^2}}} \\
&= \frac{\lim_{n\to\infty} \frac{1}{n}}{\lim_{n\to\infty} \sqrt{1+\frac{1}{n^2}}} = \frac{\lim_{n\to\infty} \frac{1}{n}}{\sqrt{\lim_{n\to\infty} 1 + \lim_{n\to\infty} \frac{1}{n^2}}} \\
&= \frac{0}{1} = 0.
\end{aligned}$$

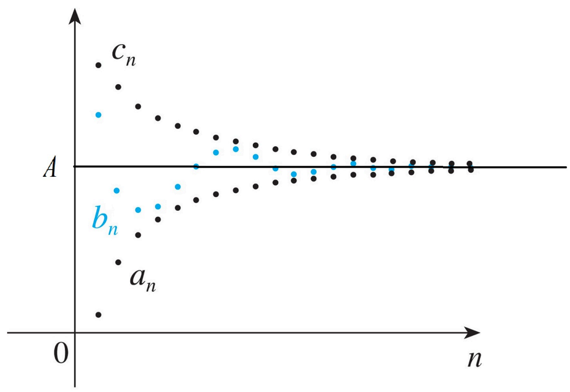
□

❑ Squeeze (Sandwich, Pinching) Theorem for Sequences

Theorem 2.2.3. *(Squeeze Theorem) Suppose that $\{a_n\}$, $\{b_n\}$ and $\{c_n\}$ are three sequences, and suppose that there exists $N \in \mathbb{N}$ such that*

$$a_n \le b_n \le c_n$$

for all $n \ge N$. If both $\{a_n\}$ and $\{c_n\}$ converge to A, then $\{b_n\}$ must also converge to A.



The sequence $\{b_n\}$ is squeezed between the sequences $\{a_n\}$ and $\{c_n\}$.

Proof. Since both $\{a_n\}$ and $\{c_n\}$ converge to A, for given $\varepsilon > 0$, there exist $N_1, N_2 \in \mathbb{N}$ such that

$$|a_n - A| < \varepsilon \text{ for all } n \geq N_1 \quad \text{and} \quad |c_n - A| < \varepsilon \text{ for all } n \geq N_2.$$

Let $N = \max(N_1, N_2)$. Then, for $n \geq N$,

$$-\varepsilon < a_n - A \leq b_n - A \leq c_n - A < \varepsilon.$$

Hence, the sequence $\{b_n\}$ converges to A. □

Corollary 2.2.4. *Suppose that $\{a_n\}$ and $\{b_n\}$ are two sequences and $|a_n| \leq b_n$ for all sufficiently large n. If $\lim_{n\to\infty} b_n = 0$ then $\lim_{n\to\infty} a_n = 0$.*

Exercise.

(i) Discuss the convergence of the sequence $a_n = \dfrac{n!}{n^n}$, where $n! = 1 \cdot 2 \cdot 3 \cdot \cdots \cdot n$.

(ii) Evaluate $\lim_{n\to\infty} \dfrac{(-1)^n}{n}$ if it exists.

Theorem 2.2.5. *If a sequence $\{a_n\}$ converges to 0, and a sequence $\{b_n\}$ is bounded, then the sequence $\{a_n b_n\}$ converges to 0.*

Proof. (Exercise) □

2.3 Infinite Limits

The properties and diversities of divergent sequences are much more complicated than convergent sequences. Divergent sequneces can be subdivided into categories.

❑ Infinite Limits

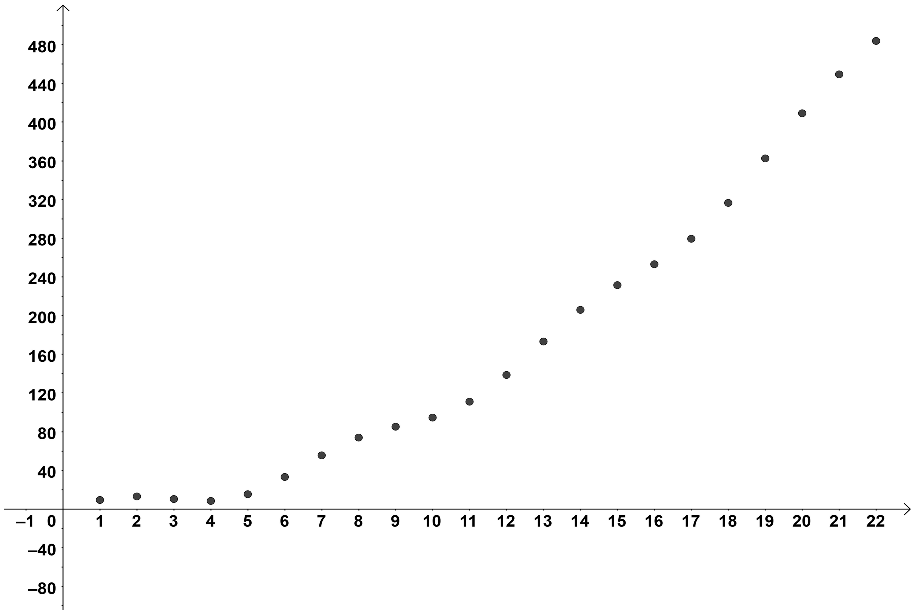
Some types of divergent sequences have nice properties. For example, $\{n\}$, $\{2^n\}$.

■ **Intuitive Definition:** We say that a sequence $\{a_n\}$ diverges to $+\infty$ (approaches to $+\infty$) (as n tends to ∞) if we can make the term $\{a_n\}$ as large as we like by taking n sufficiently large. Denote

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

or

$$a_n \to \infty \quad \text{as} \quad n \to \infty$$



■ **Precise Definition:**

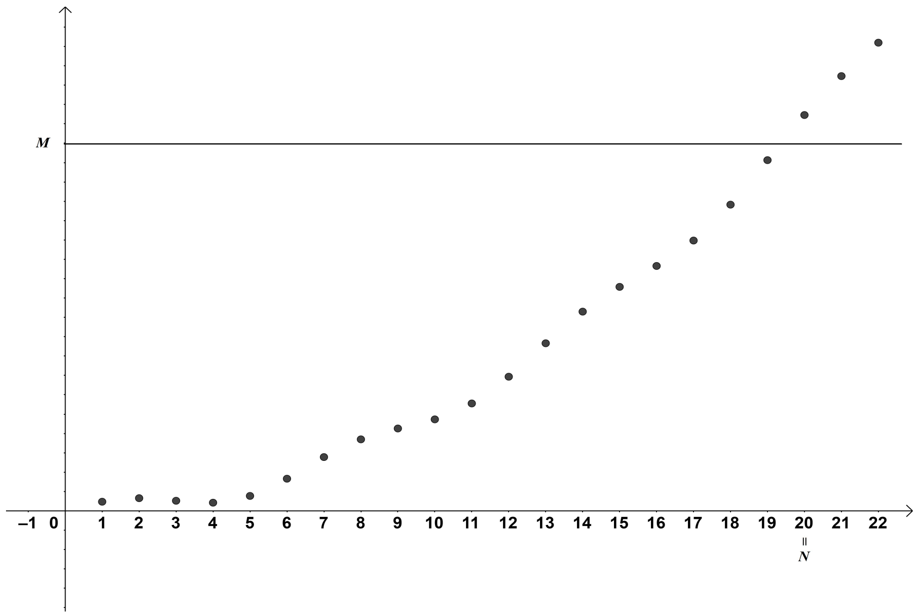
Definition 2.3.1. We say that

(a) a sequence $\{a_n\}$ diverges to $+\infty$ (approaches to $+\infty$) as n tends to ∞ if for any $M > 0$, there exists $N \in \mathbb{N}$ such that $a_n > M$ for all $n \geq N$. Denote

$$\lim_{n\to\infty} a_n = +\infty$$

or

$$a_n \to +\infty \quad \text{as} \quad n \to \infty$$



(b) $\{a_n\}$ diverges to $-\infty$ (approaches to $-\infty$) as n tends to ∞ if for any $M < 0$, there exists $N \in \mathbb{N}$ such that $a_n < M$ for all $n \geq N$. Denote

$$\lim_{n\to\infty} a_n = -\infty$$

or

$$a_n \to -\infty \quad \text{as} \quad n \to \infty$$

Note.

(1) If $\lim_{n\to\infty} a_n = \infty$ then $\lim_{n\to\infty}(-a_n) = -\infty$.

(2) Since $+\infty$ and $-\infty$ are not real numbers, if $\lim_{n\to\infty} a_n = \pm\infty$, we will say that the limit of $\{a_n\}$ does not exist(DNE). (It is different from the note in the book, p81.)

Example 2.3.2. Prove that the sequence $\left\{\dfrac{5n^2 - 2n - 10}{3n + 100}\right\}$ diverges to ∞.

Proof. Consider $5n^2 - 2n - 10 = (5n^2 - 5n) + (3n - 10) > 5n(n - 1)$ for all $n \geq 4$. Also, $3n + 100 < 5n$ for all $n \geq 50$. Hence,

$$\frac{5n^2 - 2n - 10}{3n + 100} > \frac{5n(n-1)}{5n} = n - 1$$

for all $n \geq 50$.
Given $M \in \mathbb{R}$, we choose $N \in \mathbb{N}$ such that $N > \max(50, M + 1)$. Then, for all $n \geq N$,

$$\frac{5n^2 - 2n - 10}{3n + 100} > n - 1 \geq N - 1 > M$$

Since M is an arbitrary number, $\lim_{n\to\infty} \dfrac{5n^2 - 2n - 10}{3n + 100} = \infty$. □

Example 2.3.3. For $r > 1$, prove that the sequence $\{r^n\}$ diverges to ∞.

Proof. Since $r > 1$, we can write $r = 1 + h$ for some $h > 0$. Then

$$r^n = (1+h)^n = 1 + nh + \frac{n(n-1)}{2}h^2 + \ldots + h^n > 1 + nh$$

Given $M \in \mathbb{R}$, choose $N \in \mathbb{N}$ such that $N > \dfrac{|M|}{h}$. Then, for $n \geq N$,

$$r^n > 1 + nh > 1 + \frac{|M|}{h} \cdot h = 1 + |M| > M.$$

Since M is an arbitrary number, $\lim_{n\to\infty} r^n = \infty$.

□

Theorem 2.3.4. *(Comparison Theorem) If a sequence $\{a_n\}$ diverges to ∞ and $a_n \leq b_n$ for all $n \leq N$, then the sequence $\{b_n\}$ must also diverge to ∞.*

Proof. Exercise! □

Example 2.3.5. Use the Comparison Theorem to prove that above two examples.

Insight: Observe the rational function $\dfrac{5n^2 - 2n - 10}{3n + 100} \sim \dfrac{5n^2}{3n}$. Hence, we have possibility to adust it as

$$\frac{5n^2 - 2n - 10}{3n + 100} > \frac{4.8n^2}{3.2n} = \frac{3}{2}n \quad \text{when } n \text{ is sufficiently large.}$$

Theorem 2.3.6. *Suppose that $\{a_n\}$ is a sequence satisfying $a_n > 0$ for all $n \in \mathbb{N}$. Then $\{a_n\}$ diverges to ∞ if and only if the sequence $\left\{\dfrac{1}{a_n}\right\}$ converges to* 0.

Proof. (Exercise!) □

Question: Suppose that the hypothesis $a_n > 0$ is replaced by $a_n \neq 0$. Is the theorem still true? If not, are both sides false or just one side?

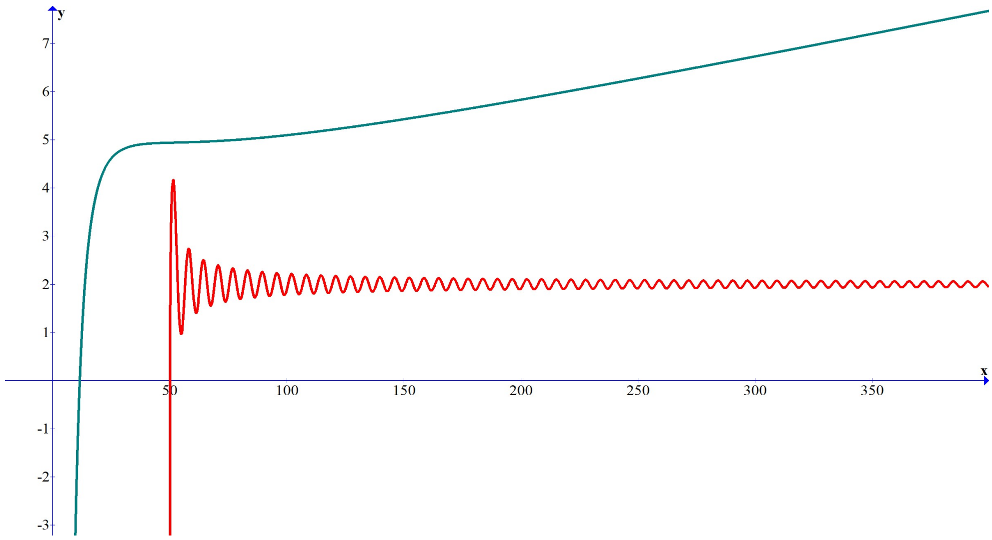
Chapter 3

Limits of Functions

In Chapter 2, we regard a sequence as values of a function whose domain is a set of $\mathbb{N}$ and consider its limit as n tends to infinity. In this chapter, we generalize the concept of a limit to functions with a domain that can contain values other than integers.

3.1 Limit at Infinity

Consider a function f with domain which contains arbitrarily large values. We want to study the behavior of the function when x becomes larger and larger.



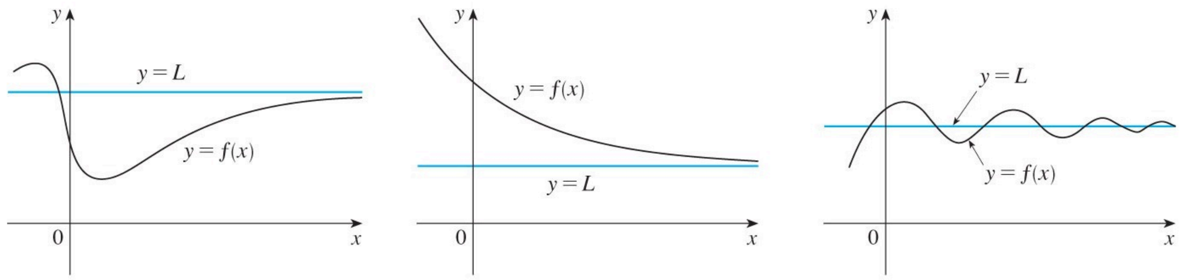
■ **Intuitive Definition:** Let $f(x)$ be a function defined on some interval (a, ∞). We say that "*the limit of $f(x)$, as x approaches ∞, exists*" if there exists a number $L \in \mathbb{R}$ such that the values of

$f(x)$ can be made arbitrarily close to L by requiring x to be sufficiently large. Denote

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

or

$$f(x) \to L \quad \text{as} \quad x \to \infty$$



Examples illustrating $\lim_{x\to\infty} f(x) = L$

Definition 3.1.1. (Precise) Let $f(x)$ be a function defined on some interval (a, ∞).

(a) We say that "*the limit of $f(x)$, as x approaches ∞, exists*" if there exists a number $L \in \mathbb{R}$ such that for every $\varepsilon > 0$, there exists a real number $M > 0$ such that

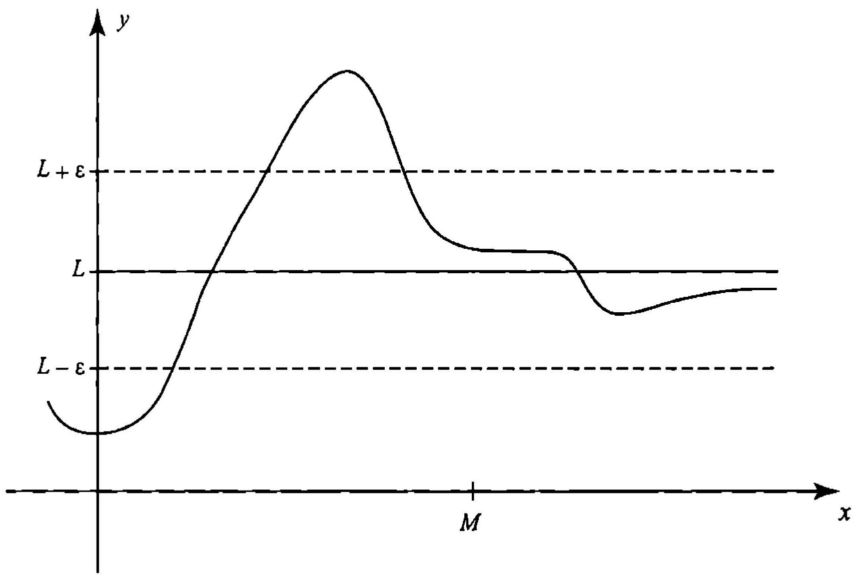
$$|f(x) - L| < \varepsilon \quad \text{if } x \geq M \text{ and } x \in (a, \infty)$$

Here, L is called "*the limit of $f(x)$, as x tends to ∞*", and we write

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

or

$$f(x) \to L \quad \text{as} \quad x \to \infty$$



(b) If $f(x)$ has a limit L as x tends to ∞, we say that f *"converges" to L (as x tends to* ∞*)*. Otherwise, we say that the function *"diverges" (as x tends to* ∞*)*.

(c) Similarly, let f be a function defined on $(-\infty, b)$. We write $\lim\limits_{x\to-\infty} f(x) = L$ if for every $\varepsilon > 0$ there exists a real number $M > 0$ such that

$$|f(x) - L| < \varepsilon \quad \text{if } x \le -M \text{ and } x \in (-\infty, b).$$

Note. In the book, the function f only need to be defined on a set $D \subseteq \mathbb{R}$ which contains arbitrarily large values. For example, $D = \mathbb{R} - \mathbb{Q}$. In our definition, we only consider the simpler situation that D contains the interval (a, ∞) for some number a.

Definition 3.1.2. If $\lim\limits_{x\to\infty} f(x) = L$, then f has a "horizontal asymptote at ∞" and the line $y = L$ is called a "horizontal asymptote" for the function f.

Example 3.1.3. (1) Let $f(x) = \dfrac{1}{x}$. Prove that $\lim\limits_{x\to\infty} f(x) = 0$.

Proof. Given $\varepsilon > 0$, choose a number $M > \dfrac{1}{\varepsilon}$ (e.g. $M = \frac{1}{2\varepsilon}$). Then for $x \ge M$,

$$\left|f(x) - 0\right| = \left|\frac{1}{x} - 0\right| = \frac{1}{x} \le \frac{1}{M} < \varepsilon.$$

Hence, $\lim\limits_{x\to\infty} f(x) = 0$. □

(2) Prove that $\lim\limits_{x\to\infty} \dfrac{1}{x^p} = 0$ for all $p \in \mathbb{N}$.

Proof. Exercise □

(3) Let $f(x) = \dfrac{2x^2 - 3}{x^2 + 3x - 4}$. Does f converge as x tends to ∞? If yes, find the limit.

Proof. According to our experience, we expect that the limit is 2. Let's try to prove our guess is true.
Consider

$$\left|\frac{2x^2 - 3}{x^2 + 3x - 4} - 2\right| = \left|\frac{-6x + 5}{x^2 + 3x - 4}\right| = \left|\frac{-\frac{6}{x} + \frac{5}{x^2}}{1 + \frac{3}{x} - \frac{4}{x^2}}\right| \quad \text{for } x \ne 0.$$

For $x \ge 2$,

$$\frac{3}{x} - \frac{4}{x^2} = \frac{1}{x}(3 - \frac{4}{x}) > 0$$

and

$$\left| -\frac{6}{x} + \frac{5}{x^2}\right| = \left|\frac{1}{x}(-6 + \frac{5}{x})\right| \le \left|\frac{1}{x}(6 + 5)\right| = \frac{11}{x}.$$

Given $\varepsilon > 0$, choose $M > \max(2, \dfrac{11}{\varepsilon})$. Then for every $x \geq M$,

$$\left|\frac{2x^2-3}{x^2+3x-4} - 2\right| = \left|\frac{-\frac{6}{x}+\frac{5}{x^2}}{1+\frac{3}{x}-\frac{4}{x^2}}\right| \leq \left|\frac{\frac{11}{x}}{1}\right| \leq \left|\frac{11}{M}\right| < \varepsilon.$$

Hence, $\lim_{x\to\infty} f(x) = 2$. □

Example 3.1.4. A horizontal asymptote of the function $f(x) = \dfrac{2x^2-3}{x^2+3x-4}$ is $y = 2$.

Theorem 3.1.5. *Suppose that $\lim_{x\to\infty} f(x) = L$. Then*

(a) the limit is unique,

(b) $\lim_{x\to\infty}[f(x) - L] = 0$,

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

Proof. (Exercise) □

Negation of Definition of Convergence:

Let f be defined on some interval (a, ∞). "*The limit of f, as x approaches ∞ does not exist*" if for every number L, there exists $\varepsilon > 0$ such that for every $M > 0$ there exists a number $x \geq M$ such that $|f(x) - L| > \varepsilon$.

Note. We usually prove the divergence of a function by using the method of contradiction.

Example 3.1.6. Verify that $f(x) = \sin x$ has no limit at infinity.

Proof. Assume that $f(x)$ has a limit L at infinity. For $\varepsilon = \dfrac{1}{2}$, there exists $M > 0$ such that for every $x > M$ $|f(x) - L| < \dfrac{1}{2}$. Choose a sufficiently large integer n such that $\dfrac{\pi}{2} + 2n\pi > M$. Then

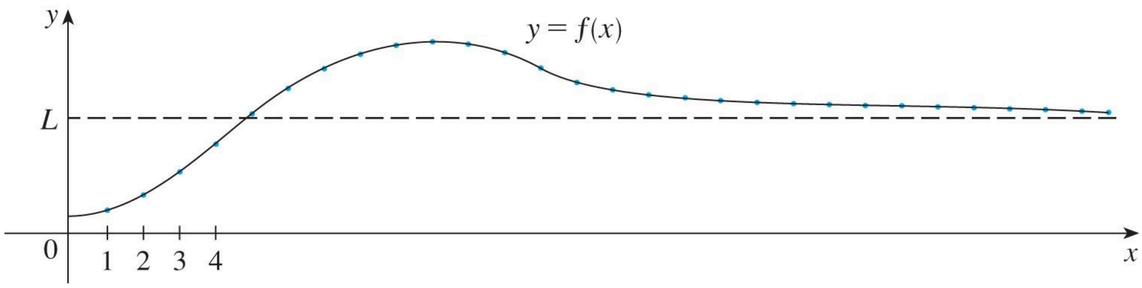
$$|L-1| = \left|f(\frac{\pi}{2}+2n\pi) - L\right| < \frac{1}{2} \quad \text{and} \quad |L-(-1)| = \left|f(\frac{3\pi}{2}+2n\pi) - L\right| < \frac{1}{2}$$

We have $L < -\dfrac{1}{2}$ and $L > \dfrac{1}{2}$. It implies a contradiction and hence f has no limit at infinity. □

■ Limits of sequences and functions at infinity

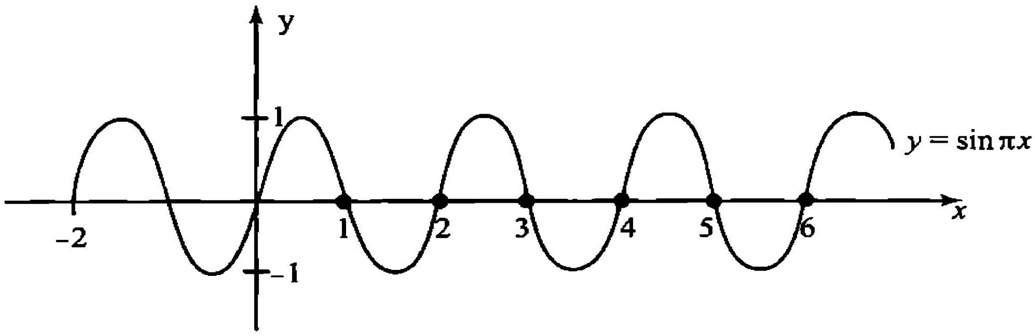
Let f be a function defined on $[1, \infty)$ and $a_n = f(n)$ for $n = 1, 2, \ldots$. Consider the limit $\lim_{x\to\infty} f(x)$ and $\lim_{n\to\infty} a_n [= \lim_{n\to\infty} f(n)]$.

Theorem 3.1.7. *Suppose that f is defined on $[1, \infty)$ and $a_n = f(n)$ for $n = 1, 2, \ldots$. If $\lim_{x\to\infty} f(x) = L$, then $\lim_{n\to\infty} a_n = L$.*



Proof. (Exercise) □

Remark. The converse of the theorem is false. For example, $f(x) = \sin(\pi x)$.



Question: How about other sequence defined by a function?

Suppose that $\{x_n\} \subset Dom(f)$ is a sequence with $\lim_{n\to\infty} x_n = \infty$. Define the functional values $b_n = f(x_n)$ for $n = 1, 2, \ldots$. [Compare with the sequence $a_n = f(n)$ for $n = 1, 2, \cdots$, b_n could be defined at any number $x_n \in Dom(f)$ with $x_n \to \infty$ rather than positive integer n.]

Question: Is Theorem 3.1.7 still true for $\{b_n\}$?

Answer: Yes. (Exercise)

Theorem 3.1.8. *Suppose that f is defined on $(1, \infty]$. Then $\lim_{x\to\infty} f(x) = L$ if and only if for every sequence $\{x_n\} \subset Dom(f)$ with $\lim_{n\to\infty} x_n = \infty$, the sequence $\{f(x_n)\}$ converges to L.*

Proof.
(⇒) Exercise!
(⇐) **Idea:** If false, we can construct a counterexample.

Assume that $\lim_{x\to\infty} f(x) \neq L$. Then there exists a number $\varepsilon > 0$ such that, for every $M > 0$, there exists $x_M > M$ such that $|f(x_M) - L| \geq \varepsilon$.

Fix the above number $\varepsilon > 0$. Let $M_1 = 1$ and there exists $x_1 > M_1$ such that $|f(x_1) - L| \geq \varepsilon$. Define $M_2 = \max(2, x_1)$ and there exists $x_2 > M_2$ such that $|f(x_2) - L| \geq \varepsilon$. Continue this process, we can define $M_n = \max(n, x_{n-1})$ and we can find a sequence $\{x_n\}$ such that $x_n > M_n$ and

$|f(x_n) - L| \geq \varepsilon$.

Since $x_n \geq M_n \geq n$, $\lim\limits_{n\to\infty} x_n = \infty$. Also, $|f(x_n) - L| \geq \varepsilon$ for all $n = 1, 2, \ldots$. Then $\lim\limits_{n\to\infty} f(x_n) \neq L$. It contradicts the hypothesis. Therefore, $\lim\limits_{x\to\infty} f(x) = L$. □

Example 3.1.9. Evaluate $\lim\limits_{x\to\infty}(\sqrt{x^2+2x} - x)$ and $\lim\limits_{x\to-\infty}(\sqrt{x^2+2x} + x)$.

❏ Limit Laws

Theorem 3.1.10. *Suppose that the functions f, g are defined on (a, ∞), and $\lim\limits_{x\to\infty} f(x) = L$ and $\lim\limits_{x\to\infty} g(x) = M$, and C is a constant number. Then*

(a) $\lim\limits_{x\to\infty}(f \pm g)(x) = \lim\limits_{x\to\infty} f(x) \pm \lim\limits_{x\to\infty} g(x) = L \pm M$.

(b) $\lim\limits_{x\to\infty} \big(Cf\big)(x) = C \lim\limits_{x\to\infty} f(x) = CL$.

(c) $\lim\limits_{x\to\infty}(fg)(x) = [\lim\limits_{x\to\infty} f(x)][\lim\limits_{x\to\infty} g(x)] = LM$.

(d) $\lim\limits_{x\to\infty}[f(x)]^n = [\lim\limits_{x\to\infty} f(x)]^n = L^n$ *for all* $n \in \mathbb{N}$.

(e) $\lim\limits_{x\to\infty} \left(\dfrac{f}{g}\right)(x) = \dfrac{L}{M}$ *provided* $M \neq 0$.

(f) $\lim\limits_{x\to\infty} \sqrt[n]{f(x)} = \sqrt[n]{\lim\limits_{x\to\infty} f(x)} = \sqrt[n]{L}$ *if* $L \geq 0$ *and* $f(x) \geq 0$ *with* $n \in \mathbb{N}$.

(g) $\lim\limits_{x\to\infty} C = C$ *where C is a constant.*

(h) If $f(x) \leq g(x)$ for all sufficiently large x, then $L \leq M$.

Proof. (Exercise) □

Remark. If $\lim\limits_{x\to\infty} f_1(x) = L_1, \ldots, \lim\limits_{x\to\infty} f_n(x) = L_n$, then

(i) $\lim\limits_{x\to\infty}(f_1 + \cdots + f_n)(x) = L_1 + \cdots + L_n$ and

(ii) $\lim\limits_{x\to\infty}(f_1 \cdots f_n)(x) = L_1 \cdots L_n$.

Remark. In the hypothesis, the convergence of f and g are important. The limit law (a), (b), (d) are false if without the condition of convergence.

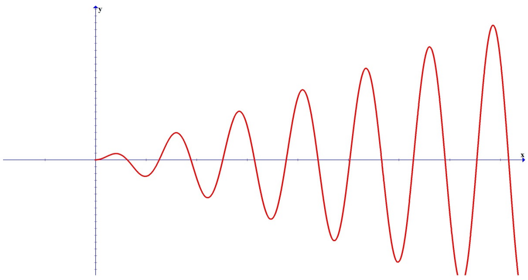
❏ Squeeze (Sandwich, Pinching) Theorem for Functions at Infinity

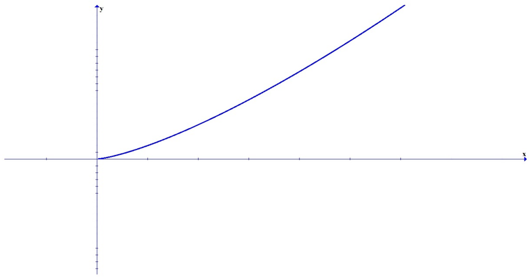
Theorem 3.1.11. *(Squeeze Theorem) Suppose that f, g and h are three functions defined on (a, ∞), and $f(x) \leq g(x) \leq h(x)$ for all sufficiently large x. If $\lim\limits_{x\to\infty} f(x) = L = \lim\limits_{x\to\infty} h(x)$, then the limit of g, as x tends to ∞, exists and moreover $\lim\limits_{x\to\infty} g(x) = L$.*

Proof. (Exercise)(Postpone until the squeeze theorem for function at a point) □

❑ Infinite Limit

Let f be a function defined on (a, ∞). Observe that if the limit of f, as x tends to ∞, exists then $f(x)$ is bounded above and below when x is sufficiently large. Hence, if f is not bounded for all large x, it must diverge at infinity. (e.g. $f(x) = x$.) Some situation may happen. For example, $f(x) = x \sin x$, $f(x) = x^2$.





Note. That f is bounded above and below for all large x does not imply f converges at infinity (e.g. $f(x) = \sin x$).

Definition 3.1.12.

(a) Let f be a function defined on (a, ∞). We say that f *tends to* ∞*, as* x *tends to* ∞, if for any $K > 0$, there exists a number $M > 0$ such that $f(x) > K$ for every $x \geq M$. Denote

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

or

$$f(x) \to \infty \quad \text{as} \quad x \to \infty$$

(b) Let f be a function defined on $(-\infty, b)$. We say that f *tends to* ∞*, as* x *tends to* $-\infty$, if for any $K > 0$, there exists a number $M > 0$ such that $f(x) > K$ for every $x \leq -M$. Denote

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

or

$$f(x) \to \infty \quad \text{as} \quad x \to -\infty$$

Remark. If f tends to $\pm\infty$, as x tends to ∞, we say that the f diverges to $\pm\infty$ and the limit does not exist(DNE). It is different from the textbook.

Example 3.1.13. For $n \in \mathbb{N}$, the n degree polynomial

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

with $a_n > 0$ has infinite limit at ∞. That is, $\lim_{x \to \infty} f(x) = \infty$.

Proof. (Exercise) □

Question: How about the limit of a rational functions at ∞?

Theorem 3.1.14. *Suppose that a function, f, is defined by $f(x) = \dfrac{p(x)}{q(x)}$, where*

$$\begin{aligned} p(x) &= a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \quad \textit{and} \\ q(s) &= b_m x^m + b_{m-1}x^{m-1} + \cdots + b_1 x + b_0 \end{aligned}$$

where $a_n, b_m \neq 0$, are n and m degree polynomials, respectively. Then

(a) if $n < m$, then $\lim\limits_{x\to\pm\infty} f(x) = 0$.

(b) if $n = m$, then $\lim\limits_{x\to\pm\infty} f(x) = \dfrac{a_n}{b_n}$.

(c) if $n > m$, then $\lim\limits_{x\to\pm\infty} f(x)$ is infinite.

Proof. (a) For $x \neq 0$,

$$f(x) = \frac{\frac{a_n}{x^{m-n}} + \frac{a_{n-1}}{x^{m-n+1}} + \cdots + \frac{a_1}{x^{m-1}} + \frac{a_0}{x^m}}{b_m + \frac{b_{m-1}}{x} + \cdots + \frac{b_1}{x^{m-1}} + \frac{b_0}{x^m}}$$

Then

$$\begin{aligned} \lim_{x\to\infty} f(x) &= \lim_{x\to\infty} \frac{\frac{a_n}{x^{m-n}} + \frac{a_{n-1}}{x^{m-n+1}} + \cdots + \frac{a_1}{x^{m-1}} + \frac{a_0}{x^m}}{b_m + \frac{b_{m-1}}{x} + \cdots + \frac{b_1}{x^{m-1}} + \frac{b_0}{x^m}} \\ &= \frac{\lim_{x\to\infty} \frac{a_n}{x^{m-n}} + \lim_{x\to\infty} \frac{a_{n-1}}{x^{m-n+1}} + \cdots + \lim_{x\to\infty} \frac{a_1}{x^{m-1}} + \lim_{x\to\infty} \frac{a_0}{x^m}}{b_m + \lim_{x\to\infty} \frac{b_{m-1}}{x} + \cdots + \lim_{x\to\infty} \frac{b_1}{x^{m-1}} + \lim_{x\to\infty} \frac{b_0}{x^m}} \\ &= \frac{0}{b_m} = 0 \end{aligned}$$

The proof of (b)and (c) are left to the readers. □

■ Oblique Asymptotes

Observe that if a function $f(x)$ has an oblique asymptote $L: \; y = ax + b$ where $a \neq 0$. Then the graph $y = f(x)$ are as close to L as we like by taking x sufficiently large. This means that

$$\lim_{x\to\infty} [f(x) - (ax + b)] = 0.$$

Example 3.1.15. Find oblique asymptotes for the function $f(x) = \dfrac{x^2 - 1}{2x + 4}$, if any exists.

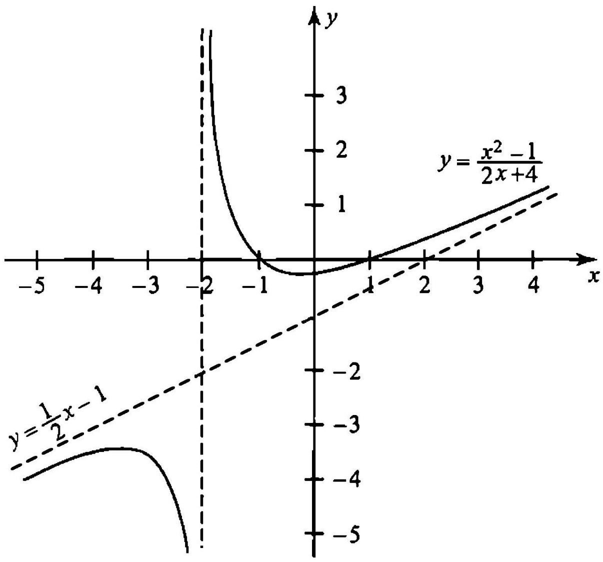
Proof. Observe that the degree of the numerator is greater than the degree of the denominator by 1. Thus, the function may have an oblique asymptote. Consider their leading coefficients of the denominator and numerator. We expect that the equation of the oblique asymptote is supposed to be $y = \dfrac{1}{2}x + b$. Then

$$\left|f(x) - (\frac{1}{2}x + b)\right| = \left|\frac{x^2 - 1}{2x + 4} - (\frac{1}{2}x + b)\right| = \left|\frac{-2(1 + b)x - (1 + 4b)}{2x + 4}\right|$$

In order to obtain the above term tends to 0 as x tends to ∞, b should be -1. Hence, we have

$$\lim_{x\to\infty} \left|\frac{x^2-1}{2x+4} - (\frac{1}{2}x - 1)\right| = \lim_{x\to\infty} \left|\frac{-5}{2x+4}\right| = 0$$

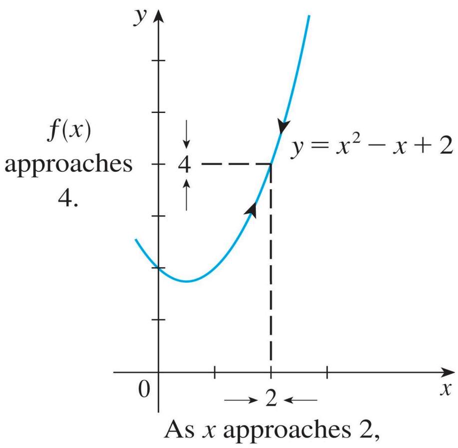
Then, the oblique asymptote of f is $y = \frac{1}{2}x - 1$.



□

3.2 Limit at a Real Number

Consider the function $f(x) = x^2 - x + 2$. What is the behavior of $f(x)$ for values of x near 2?



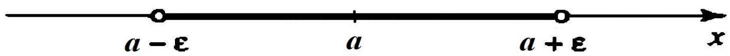
When x tends to 2, $f(x)$ approaches 4. We can make the value of $f(x)$ as close to 4 as we like by taking x sufficiently close to 2.

In the world, lots of problems involve the "tendency" of a function near a number. For example, tangent line problem, instantaneous velocity etc.

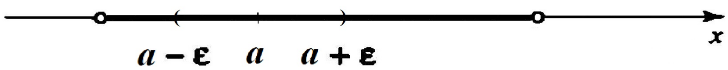
To study the behaviors of a function near a point, the function should be defined near this point (possibly except the point itself).

Definition 3.2.1. Let $a \in \mathbb{R}$ be a real number.

(a) For $\varepsilon > 0$, we call the set $B(a, \varepsilon) := \left\{x \in \mathbb{R} \,\middle|\, |x - a| < \varepsilon\right\}$ a "*ball of a*"



(b) We call a set $N \subseteq \mathbb{R}$ a "*neighborhood of a*" if it contains a ball, $B(a, \varepsilon)$, of a for some $\varepsilon > 0$.



Note. Any open interval containing a is a neighborhood. Specifically, a ball of a is also a neighborhood of a.

■ **Intuitive Definition** (limits of functions at a number):

Suppose $f(x)$ is defined in a neighborhood of a (except possibly at a itself). We say that "*the limit of $f(x)$, as x approaches a, exists*" if there is a number $L \in \mathbb{R}$ such that the values of $f(x)$ can be arbitrarily close to L by taking x to be sufficiently close to a (on either side of a) but not equal to a. We write

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

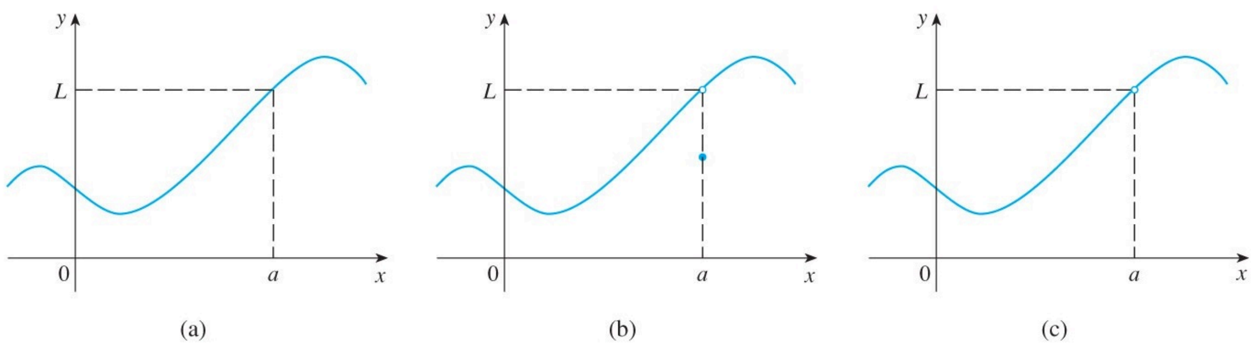
or

$$x \to L \quad \text{as } x \to a.$$

Note.

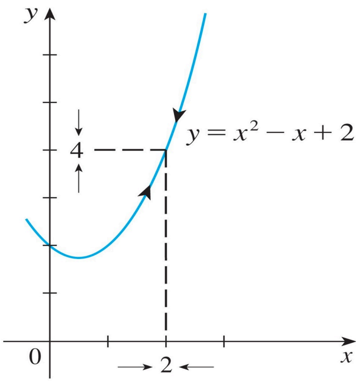
(1) In the definition, if there is no such number, we say that "*the limit of $f(x)$, as x approaches a, does not exist* (DNE)."

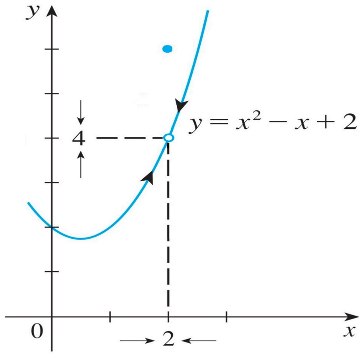
(2) The words "but not equal to a" means that we never consider the value of f at $x = a$.



$\lim_{x \to a} f(x) = L$ in all three cases.

Example 3.2.2. (Find the limit by graphing)



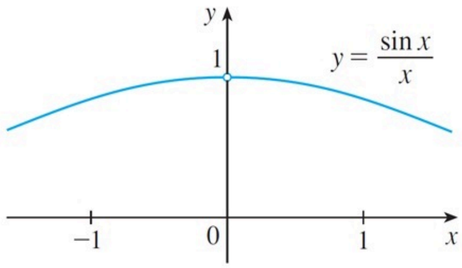


$f(x) = x^2 - x + 2$ $\qquad$ $f(x) = \begin{cases} x^2 - x + 2 & x \neq 2 \\ 6 & x = 2 \end{cases}$

$\lim_{x\to 2} f(x) = 4.$ $\qquad$ $\lim_{x\to 2} f(x) = 4.$

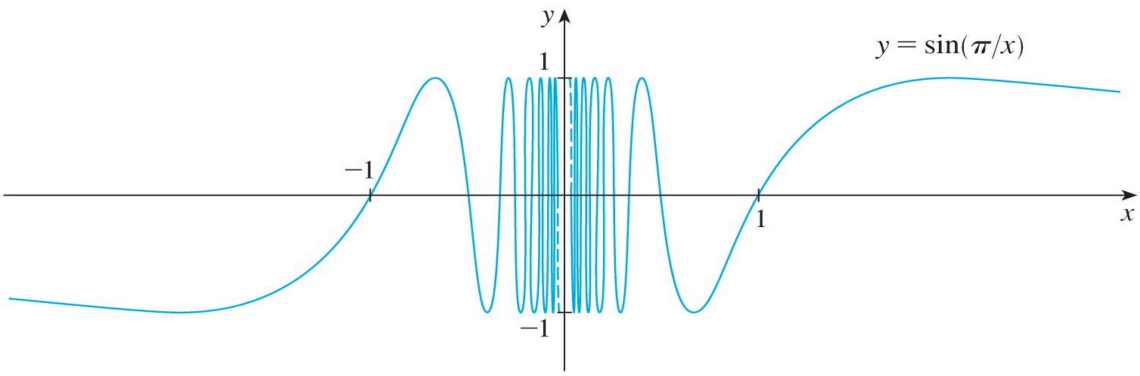
Example 3.2.3. (Guess by taking values) Evaluate $\lim_{x\to 0} \frac{\sin x}{x}$.

x	$\frac{\sin x}{x}$
± 1.0	0.84147098
± 0.5	0.95885108
± 0.4	0.97354586
± 0.3	0.98506736
± 0.2	0.99334665
± 0.1	0.99833417
± 0.05	0.99958339
± 0.01	0.99998333
± 0.005	0.99999583
± 0.001	0.99999983

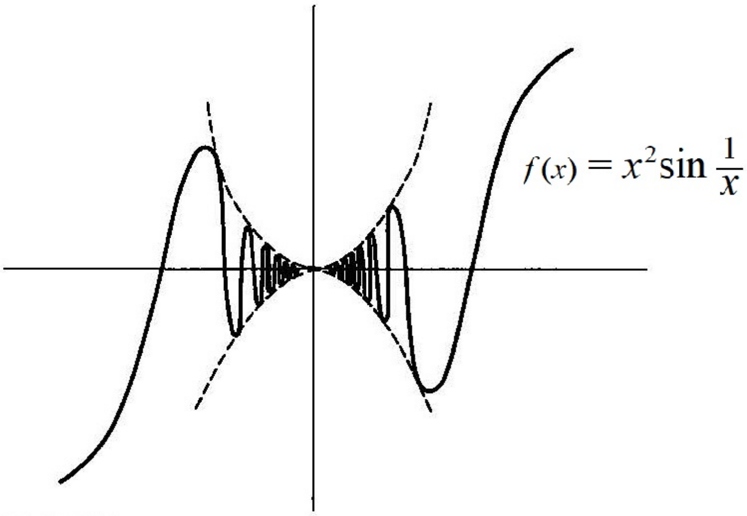


Guess: $\lim_{x\to 0} \frac{\sin x}{x} = 1$.

Note. The wrong evaluation may happen by graphing or taking testing values. For example $f(x) = \sin \frac{\pi}{x}$ for $x \neq 0$. Then $f(1) = f(\frac{1}{2}) = f(\frac{1}{3}) = f\frac{1}{10} = f(\frac{1}{1000000}) = \cdots = 0$. But $\lim_{x\to 0} f(x)$ does not exist.



Example 3.2.4. For the function $f(x) = x^2 \sin \dfrac{1}{x}$, it seems that f approaches 0 as x tends to 0.



We need to check whether we can make f arbitrarily close to 0 by taking x to be sufficiently close to 0.

- For the error $= \frac{1}{10}$, how much close to 0 should we choose x such that $|x^2 \sin \frac{1}{x} - 0| < \frac{1}{10}$? Choose $|x - 0| < \frac{1}{5}$, then $|x^2| < \frac{1}{25}$. Hence, $|f(x) - 0| = |x^2 \sin \frac{1}{x}| \le x^2 < \frac{1}{25} < \frac{1}{10}$.
- For the error $= \frac{1}{10000}$, how much close to 0 should we choose x such that $|x^2 \sin \frac{1}{x} - 0| < \frac{1}{10000}$? Choose $|x - 0| < \frac{1}{100}$, then $|x^2| < \frac{1}{10000}$. Hence, $|f(x) - 0| = |x^2 \sin \frac{1}{x}| \le x^2 < \frac{1}{10000}$.
- For the error $= \varepsilon > 0$ to be an arbitrarily small number, choose $|x - 0| < \sqrt{\varepsilon}$. Then $|f(x) - 0| = |x^2 \sin \frac{1}{x}| < x^2 < \varepsilon$.

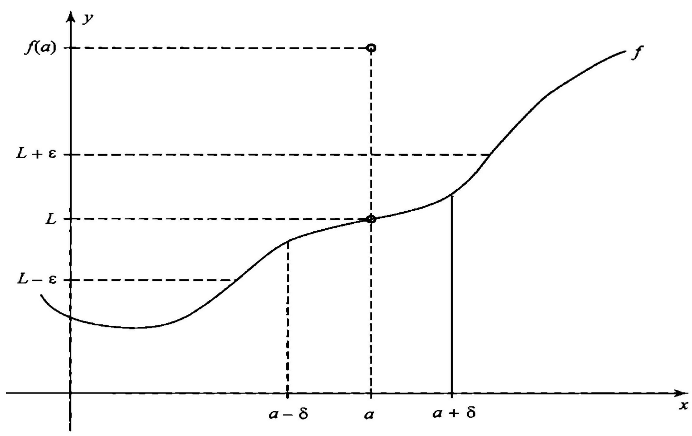
To give a suitable definition of limit at a number a, it is supposed to check that, for every "error", the values of f is close to L within this error whenever x is close to a within a certain range.

Definition 3.2.5. (Precise)(δ-ε definition)

Let f be a function defined on some neighborhood of a (except possibly at a itself).

(a) We say that *"the limit of $f(x)$, as x approaches a, exists"* if there is a number $L \in \mathbb{R}$ satisfying for every $\varepsilon > 0$, there is a number $\delta > 0$ such that

$$|f(x) - L| < \varepsilon \quad \text{whenever} \quad 0 < |x - a| < \delta$$



(b) If f has a limit L as x tends to a, we say that f *"converges" to L (as x tends to a)*. Otherwise, we say that f *"diverges" (as x tends to a)*.

Remark.

(i) That the words "$0 < |x - a|$" rather than $0 \le |x - a|$ reflects when we consider the limit of f as x approaches a, we only concern the values of f "near" a. It is not essential that the function f be defined at $x = a$.

(ii) The number $\delta = \delta(\varepsilon)$ usually depends on the chosen number ε. For a given $\varepsilon > 0$, it suffices to show that the corresponding number δ exists but not necessary to find the exact number.

(iii) For a given $\varepsilon > 0$, if δ satisfies the statement of definition, any smaller number $0 < \delta_1 < \delta$ must also satisfy the statement for the same ε. In other words, δ can be replaced by δ_1 (if necessary).

Example 3.2.6. Prove that $\lim_{x\to 3}(4x - 5) = 7$.

Proof.

In this problem, the limit 7 is given. Otherwise, we should guess a possible limit and then prove it.

Like the proof of limit of a sequence, we usually need some priori estimates before proving.

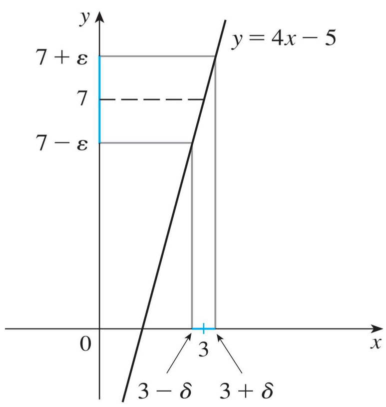
Consider

$$|(4x - 5) - 7| = |4x - 12| = 4|x - 3|.$$

Hence, $4|x - 3| < \varepsilon$ if $|x - 3| < \dfrac{\varepsilon}{4}$. For given $\varepsilon > 0$, choose $\delta = \dfrac{\varepsilon}{4}$. Then, for every $|x - 3| < \delta$,

$$|(4x - 5) - 7| = |4x - 12| = 4|x - 3| < 4\delta = 4 \cdot \frac{\varepsilon}{4} = \varepsilon.$$

Thus, $\lim_{x\to 3}(4x - 5) = 7$.



$\square$

Example 3.2.7. Prove that $\lim_{x\to 2} x^2 = 4$.

Proof.

Consider $|x^2 - 4| = |(x+2)(x-2)| = |x+2||x-2| < \varepsilon$. If $|x+2| < M$ for some constant number M. Then we can choose $\delta = \dfrac{\varepsilon}{M}$ and $|x^2 - 4| \le M \cdot |x-2| < M \cdot \dfrac{\varepsilon}{M} = \varepsilon$.

To obtain an upper bound of $|x+2|$, we consider that if $|x-2| < 1$, then $1 < x < 3$ and hence $|x+2| < 5$.

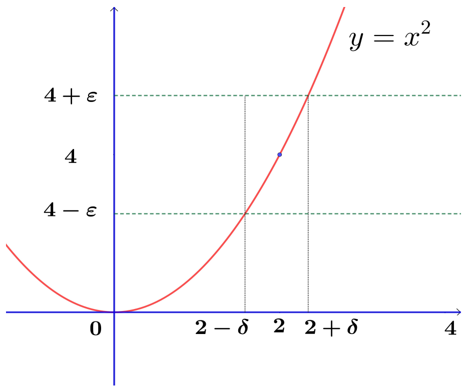
Given $\varepsilon > 0$, let $\delta = \min(1, \dfrac{\varepsilon}{5})$. For all x with $|x-2| < \delta$,

$$|x+2| = |x-2+4| \le |x-2| + 4 < \delta + 4 \le 5.$$

Then, for $|x-2| < \delta$,

$$|x^2 - 4| = |(x+2)(x-2)| = |x+2||x-2| \le 5|x-2| < 5\delta \le 5 \cdot \frac{\varepsilon}{5} = \varepsilon.$$

Hence, $\lim\limits_{x\to 2} x^2 = 4$.



$\square$

Exercise. Prove that $\lim\limits_{x\to a} x^2 = a^2$ for every $a \in \mathbb{R}$.

Negation of Definition of Convergence:

Let f be defined on some neighborhood of a (except possibly at a itself). "*The limit of f, as x approaches a, does not exist*" if for every number L, there exists $\varepsilon > 0$ satisfying for every $\delta > 0$ there exists a number x with $|x-a| < \delta$ such that $|f(x) - L| > \varepsilon$.

Note. We prove the convergence of a function by using $\delta - \varepsilon$ definition. However, in order to prove the divergence of a function, we usually use the method of contradiction.

Example 3.2.8. Prove that $\lim\limits_{x\to 0} \sin \dfrac{1}{x}$ does not exist.

Proof. (The function $f(x) = \sin \frac{1}{x}$ is defined near 0.)

Assume that there is a number L such that $\lim\limits_{x\to 0} \sin \dfrac{1}{x} = L$. For $\varepsilon = \frac{1}{2}$, there is a number $\delta > 0$ such that for $0 < |x-0| < \delta$, $|\sin \frac{1}{x} - L| < \frac{1}{2}$.

Let $x_1 = \dfrac{1}{(2N+\frac{1}{2})\pi}$ and $x_2 = \dfrac{1}{(2N+\frac{3}{2})\pi}$ for some sufficiently large $N \in \mathbb{N}$ such that $|x_1| < \delta$ and $|x_2| < \delta$. Then

$$\left|1 - L\right| = \left|\sin\frac{1}{x_1} - L\right| < \frac{1}{2} \quad \text{and} \quad \left|(-1) - L\right| = \left|\sin\frac{1}{x_2} - L\right| < \frac{1}{2}.$$

We have

$$2 = |1-(-1)| = \left|1 - L + L - (-1)\right| < \left|1 - L\right| + \left|L - (-1)\right| < \frac{1}{2} + \frac{1}{2} = 1 \quad (\textit{Contradiction}).$$

Therefore, $\lim\limits_{x\to 0} \sin\dfrac{1}{x}$ does not exist.

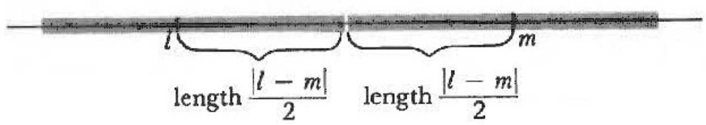
□

Exercise. Let $f(x) = \begin{cases} 0, & x \text{ is irrational} \\ 1, & x \text{ is rational} \end{cases}$. Prove that the limit of f does not exist at every point.

Theorem 3.2.9. *(Uniqueness of a limit)*
If the limit of a function exists, as x approaches a, then it is unique.

Proof.

In Sec3.1, we proved that the uniqueness of the limit of a function, as x approaches ∞ by using the fact that "*if two numbers, L and M, satisfy* $|L - M| < \varepsilon$ *for every* $\varepsilon > 0$*, then* $L = M$. Here, we keep the same spirit and use the method of contradiction to prove it.



Let $\lim\limits_{x\to a} f(x) = L$ and $\lim\limits_{x\to a} f(x) = M$. Assume that $L \neq M$.

For $\varepsilon = \frac{1}{2}|L - M| > 0$, there exist $\delta_1, \delta_2 > 0$ such that

$$|f(x) - L| < \varepsilon \quad \text{for all} \quad |x - a| < \delta_1$$

and

$$|f(x) - M| < \varepsilon \quad \text{for all} \quad |x - a| < \delta_2.$$

Let $\delta = \min(\delta_1, \delta_2)$. For $0 < |x - a| < \delta$,

$$|L - M| = |L - f(x) + f(x)_M| \leq |L - f(x)| + |f(x)_M| < \varepsilon + \varepsilon = 2\varepsilon = |L - M|.$$

It implies a contradiction and hence $L = M$. □

Theorem 3.2.10. $\lim\limits_{x\to a} f(x) = 0$ *if and only if* $\lim\limits_{x\to a} |f(x)| = 0$.

Proof. (Exercise) □

Note. In the above thorem, the direction ($\Longleftarrow$) is false if the limit is nonzero. (Why?)

Exercise. $\lim_{x\to a} f(x) = L$ if and only if $\lim_{x\to a} |f(x) - L| = 0$.

❑ Limit Laws

Theorem 3.2.11. *Suppose that the functions f, g are defined on a neighborhood of a, and $\lim_{x\to a} f(x) = L$ and $\lim_{x\to a} g(x) = M$ and C is a constant number. Then*

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

(b) $\lim_{x\to a} \left(Cf\right)(x) = C \lim_{x\to a} f(x) = CL$.

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

(d) $\lim_{x\to a}[f(x)]^n = [\lim_{x\to a} f(x)]^n = L^n$ *for all* $n \in \mathbb{N}$.

(e) $\lim_{x\to a} \left(\frac{f}{g}\right)(x) = \frac{L}{M}$ *provided* $M \neq 0$.

(f) $\lim_{x\to a} \sqrt[n]{f(x)} = \sqrt[n]{\lim_{x\to a} f(x)} = \sqrt[n]{L}$ *if* $L \geq 0$ *and* $f(x) \geq 0$ *with* $n \in \mathbb{N}$.

(g) $\lim_{x\to a} C = C$ *where* C *is a constant.*

(h) *If* $f(x) \leq g(x)$ *for all* x *near* a*, then* $L \leq M$.

Proof. (Exercise) □

Remark. If $\lim_{x\to a} f_1(x) = L_1, \ldots, \lim_{x\to a} f_n(x) = L_n$, then

(i) $\lim_{x\to a}(f_1 + \cdots + f_n)(x) = L_1 + \cdots + L_n$ and

(ii) $\lim_{x\to a}(f_1 \cdots f_n)(x) = L_1 \cdots L_n$.

Example 3.2.12. (Polynomial functions)

(1) For $n = 0, 1, 2, \ldots$,

$$\lim_{x\to a} x^n = \lim_{x\to a} \left(\overbrace{x \cdot x \cdot x \cdots x}^{n}\right) = \overbrace{\left(\lim_{x\to a} x\right) \cdots \left(\lim_{x\to a} x\right)}^{n} = \overbrace{a \cdot a \cdots a}^{n} = a^n.$$

(2) Let c be a constant.

$$\lim_{x\to a}(cx^n) = \lim_{x\to a} c \cdot \lim_{x\to a} x^n = ca^n.$$

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

$$\begin{aligned}\lim_{x\to a} P_n(x) &= \lim_{x\to a}\left(c_n x^n + c_{n-1}x^{n-1} + \cdots + c_1 x + c_0\right)\\ &= \lim_{x\to a} c_n x^n + \lim_{x\to a} c_{n-1}x^{n-1} + \cdots + \lim_{x\to a} c_1 x + \lim_{x\to a} c_0\\ &= c_n a^n + c_{n-1}a^{n-1} + \cdots + c_1 a + c_0.\end{aligned}$$

The limits of polynomial functions exist everywhere and the limits are equal to the values of the polynomial at those points. That is, $\lim_{x\to a} P_n(x) = P_n(a)$.

Example 3.2.13. (Rational functions)

A rational function has thr form $R(x) = \dfrac{P(x)}{Q(x)}$, where $P(x)$ and $Q(x)$ are polynomials. Consider $\lim_{x\to a} R(x)$.

(1) If $Q(a) \neq 0$, then $\lim_{x\to a} R(x) = \dfrac{\lim_{x\to a} P(x)}{\lim_{x\to a} Q(x)} = \dfrac{P(a)}{Q(a)} = R(a)$. For every number a in the domain of $R(x)$, $\lim_{x\to a} R(x) = R(a)$.

(2) If $Q(a) = 0$,

- when $P(a) \neq 0$, we will show later that the limit $\lim_{x\to a} R(x) = \pm\infty$.
- when $P(a) = 0$, we will factorize $P(x)$ and $Q(x)$. After dividing their common factors, the problem will reduce to the above two cases.

Example 3.2.14.

(1) Compute $\lim_{x\to 3} x^3 + 2x - 6$.

Proof. Since $x^3 + 2x - 6$ is a polynomial function,

$$\lim_{x\to 3} x^3 + 2x - 6 = 3^3 + 2\cdot 3 - 6 = 27.$$

□

(2) Compute $\lim_{x\to 3} \dfrac{x^2+5}{3x^2-1}$.

Proof. Since $\dfrac{x^2+5}{3x^2-1}$ is a rational function defined at 3,

$$\lim_{x\to 3} \frac{x^2+5}{3x^2-1} = \frac{3^2+5}{3\cdot 3^2-1} = \frac{7}{13}.$$

□

(3) Compute $\lim_{x\to 3} x^3 + 2x - 6$.

Proof. □

(4) If $\lim\limits_{x\to a} f(x) = 3$, $\lim\limits_{x\to a} g(x) = 2$ and $\lim\limits_{x\to a} h(x) = 4$, compute $\lim\limits_{x\to a} \dfrac{f(x)+g(x)}{h^2(x)-f(x)}$.

Proof.

$$\lim_{x\to a}\frac{f(x)+g(x)}{h^2(x)-f(x)} = \frac{\lim_{x\to a}\big(f(x)+g(x)\big)}{\lim_{x\to a}\big(h^2(x)-f(x)\big)} = \frac{\lim_{x\to a} f(x) + \lim_{x\to a} g(x)}{\lim_{x\to a} h^2(x) - \lim_{x\to a} f(x)} = \frac{5}{13}.$$

□

(5) Compute $\lim\limits_{x\to 3}\dfrac{x^3-27}{x-3}$.

Proof. For $x \neq 3$, $\dfrac{x^3-27}{x-3} = \dfrac{(x-3)(x^2+3x+9)}{x-3} = x^2+3x+9$. Also, x^2+3x+9 is a polynomial functon, then

$$\lim_{x\to 3}\frac{x^3-27}{x-3} = \lim_{x\to 3}(x^2+3x+9) = 3^2 + 3\cdot 3 + 9 = 27.$$

Hence, $\lim\limits_{x\to 3}\dfrac{x^3-27}{x-3} = 27$.

□

(6) Compute $\lim\limits_{x\to 1}\dfrac{1-\sqrt{x}}{1-x}$.

Proof. For $x \neq 1$,

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

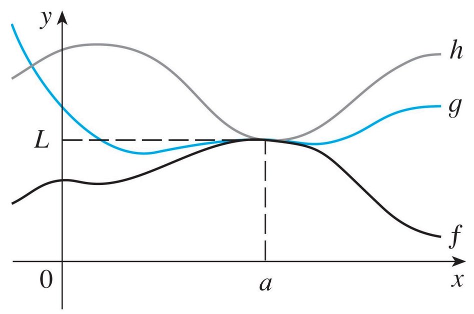
Then,

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

□

❑ Squeeze (Sandwich, Pinching) Theorem for Functions at a number

Theorem 3.2.15. *(Squeeze Theorem) Suppose that f, g and h are three functions defined on $(a-\sigma, a+\sigma)$ (except possibly at a itself), and $f(x) \le g(x) \le h(x)$ for all numbers near a. If $\lim\limits_{x\to a} f(x) = L = \lim\limits_{x\to a} h(x)$, then the limit of g, as x tends to a, exists and moreover $\lim\limits_{x\to a} g(x) = L$.*



Proof.

> To prove that for given $\varepsilon > 0$, there exists $\delta > 0$ such that for all $0 < |x - a| < \delta$, then $|g(x) - L| < \varepsilon$.

Given $\varepsilon > 0$, since $\lim\limits_{x \to a} h(x) = L$, there exists $0 < \delta_1 < \sigma$ such that for every $0 < |x - a| < \delta_1$, $|h(x) - L| < \varepsilon$. Then, for $0 < |x - a| < \delta_1$, $h(x) < L + \varepsilon$. Hence,

$$g(x) \leq h(x) < L + \varepsilon \quad \text{for all} \quad 0 < |x - a| < \delta_1$$

Similarly, since $\lim\limits_{x \to a} f(x) = L$, there exists $0 < \delta_2 < \sigma$ such that for every $0 < |x - a| < \delta_2$, $|f(x) - L| < \varepsilon$. Then, for $0 < |x - a| < \delta_2$, $f(x) > L - \varepsilon$. Hence,

$$g(x) \geq f(x) < L - \varepsilon \quad \text{for all} \quad 0 < |x - a| < \delta_2$$

Choose $\delta = \min \delta_1, \delta_2 > 0$. For $0 < |x - a| < \delta$,

$$L - \varepsilon < g(x) < L + \varepsilon.$$

Hence, $\lim\limits_{x \to a} g(x) = L$. □

Example 3.2.16.

(1) Find $\lim\limits_{x \to 0} \sin x$.

Proof.
For $x > 0$, from the figure, $\sin x = \dfrac{\overline{BC}}{\overline{OB}} = \overline{BC}$. Then

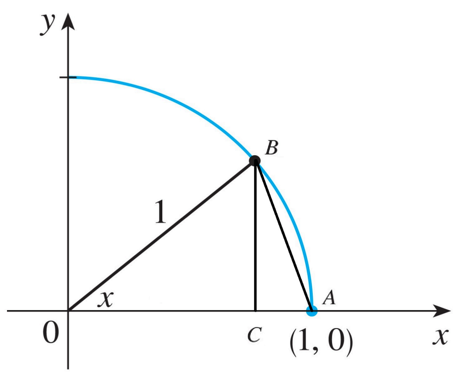
$$0 < \sin x = \overline{BC} < \overline{AB} < \overset{\frown}{AB} = x$$

Similarly, for $x < 0$, we have $x < \sin x < 0$. We have

$$-|x| < \sin x < |x| \qquad \text{for every } x.$$

Also, $\lim\limits_{x \to 0}(-|x|) = \lim\limits_{x \to 0} x = 0$. By the squeeze theorem,

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



□

(2) Prove that $\lim\limits_{x \to 0} \cos x = 1$. (Hint: $\cos x = \sqrt{1 - \sin^2 x}$ when x is near 0.)

(3) Find $\lim\limits_{x \to a} \sin x$.

Proof. Since $\lim\limits_{h \to 0} \sin h = 0$ and $\lim\limits_{h \to 0} \cos h = 1$,

$$\lim_{x \to a} \sin x = \lim_{h \to 0} \sin(a + h) = \lim_{h \to 0} \big(\sin a \cos h + \cos a \sin h \big) = \sin a \cdot 1 + \cos a \cdot 0 = \sin a.$$

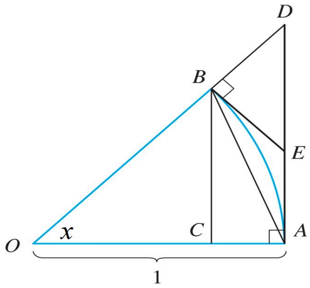
□

(4) Prove that $\lim_{x \to a} \cos x = \cos a$.

(5) Prove that the limits of all trigonometric functions are equal to the vlues of the trigonometric functions at that points. That is, assuming that a is in the domain of the below functions, $\lim_{x \to a} \tan x = \tan a$, $\lim_{x \to a} \cot x = \cot a$, $\lim_{x \to a} \sec x = \sec a$, $\lim_{x \to a} \csc x = \csc a$.

(6) Find $\lim_{x \to 0} \dfrac{\sin x}{x}$.

Proof. W.L.O.G, we may assume $x > 0$ and the case $x < 0$ is similiar. From the figure,



$$\triangle OAB = \frac{1}{2} \sin x, \quad \text{sector } OAB = \frac{1}{2} x \quad \text{and} \quad \triangle OAD = \frac{1}{2} \tan x.$$

Then,

$$\frac{1}{2} \sin x < \frac{1}{2} x < \frac{1}{2} \tan x.$$

Hence,

$$0 < \frac{\sin x}{x} < 1 < \frac{\tan x}{x} = \frac{\sin x}{x} \cdot \frac{1}{\cos x} \quad \text{for } x \in (0, \frac{\pi}{2}).$$

We have $\cos x < \dfrac{\sin x}{x} < 1$ for $x \in (0, \frac{\pi}{2})$. Since $\lim_{x \to 0} \cos x = 1$ and $\lim_{x \to 0} 1 = 1$, by the squeeze theorem,

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

□

(7) Prove that $\lim_{x \to 0} \dfrac{\cos x - 1}{x} = 0$.

(8) Find $\lim_{x \to 0} \dfrac{\sin(3x)}{5x}$

Proof.

$$\lim_{x \to 0} \frac{\sin(3x)}{5x} = \lim_{x \to 0} \left(\frac{\sin(3x)}{3x} \cdot \frac{3}{5} \right) = \frac{3}{5} \cdot \lim_{x \to 0} \frac{\sin(3x)}{3x} = \frac{3}{5}.$$

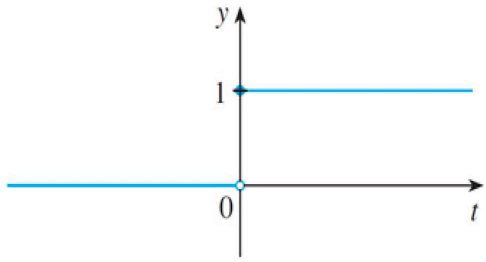
□

3.3 Sided Limits and Infinite Limits

3.3.1 Sided Limits

Consider the Heaviside function

$$H(t) = \begin{cases} 0 & \text{if } t < 0 \\ 1 & \text{if } t \geq 0 \end{cases}$$



There is no number that $H(t)$ approaches as t approaches 0. Therefore, $\lim_{t \to 0} H(t)$ does not exist. But if we only consider the number that $H(t)$ approaches as t approaches 0 from the right (left) side, such a number exists.

■ **<u>Intuitive Definition</u>** (one-sided limits):
Let f be a function whose domain contains an open interval $(a, a+\sigma)$ for some samll number $\sigma > 0$. We say that the "*right-hand limit of $f(x)$ as x approaches a from the right, exists*" if there exists a number L such that we can make the values of $f(x)$ arbitrarily close to L by taking x to be sufficiently close to a with x greater than a. We write

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

or

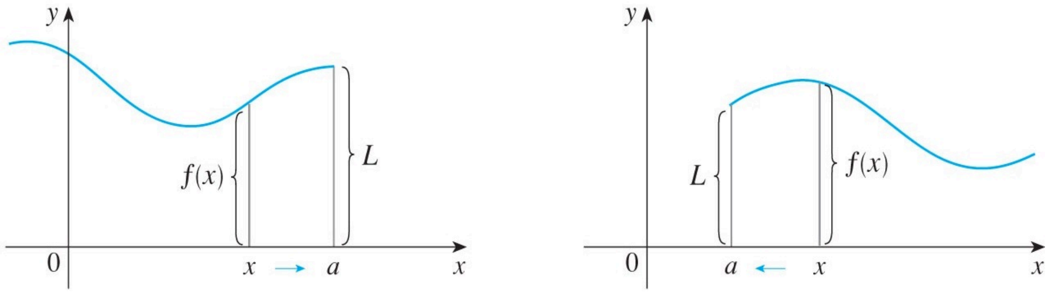
$$x \to L \quad \text{as } x \to a^+.$$

Similarly, if we require that x be less than a, we get the "*right-hand limit of $f(x)$ as x approaches* a is equal to L and we write

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

or

$$x \to L \quad \text{as } x \to a^-.$$



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

(b) $\lim_{x \to a^+} f(x) = L$

Definition 3.3.1. (Precise) Suppose $f(x)$ is defined when x is near a from the right side (except possibly at a itself). We say that "*the right-hand limit (or limit from the right) of $f(x)$, as x approaches a, exists*" if there is a umber $L \in \mathbb{R}$ satisfying for every $\varepsilon > 0$, there is a number $\delta > 0$ such that

$$|f(x) - L| < \varepsilon \quad \text{whenever } 0 < x - a < \delta.$$

Denote

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

or

$$x \to L \quad \text{as } x \to a^+.$$

Example 3.3.2. Let $H(t)$ be the Heaviside function. Then $\lim_{t \to 0^+} H(t) = 1$ and $\lim_{t \to 0^-} H(t) = 0$

Remark. If $\lim_{x \to a} f(x)$ exists, then both $\lim_{x \to a^+} f(x)$ and $\lim_{x \to a^-} f(x)$ exists . But the converse could be false. For example, the Heaviside function.

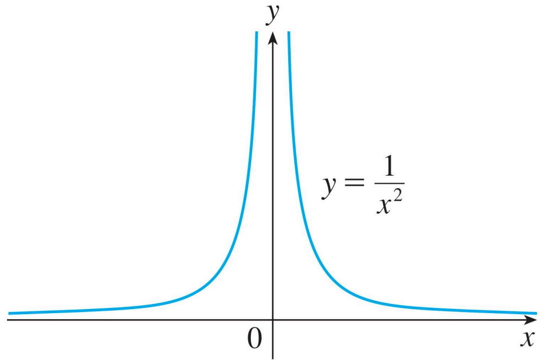
Theorem 3.3.3. *$\lim_{x \to a} f(x) = L$ if and only if $\lim_{x \to a^+} f(x) = L$ and $\lim_{x \to a^-} f(x) = L$ (where L could be $\pm\infty$).*

Proof. (Exercise) □

3.3.2 Infinite Limits

Consider $\lim_{x \to 0} \dfrac{1}{x^2}$ if it exists

x	$\dfrac{1}{x^2}$
± 1	1
± 0.5	4
± 0.2	25
± 0.1	100
± 0.05	400
± 0.01	10000
± 0.001	1000000



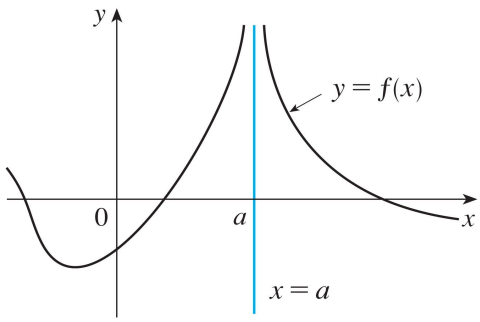
As x becomes close to 0, $1/x^2$ becomes vergy large. In fact, the values of $f(x)$ can be made arbitrarily large by taing x close enough to 0. Thus, the values of $f(x)$ do not approach a number, so $\lim_{x \to 0} \dfrac{1}{x^2}$ do not exist.

■ **Intuitive Definition:** Let $f(x)$ be a function defined on a neighborhood of a (except possibly at a itself). We say that f approaches (tends to) ∞ $(-\infty)$, as x approaches a, if the values of $f(x)$ can be made arbitrarily (negative) large by taking x sufficiently close to a, but not equal to a. Denote

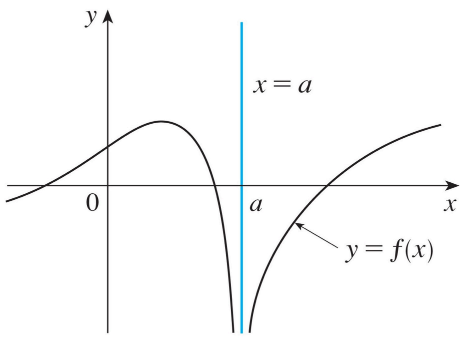
$$\lim_{x \to a} f(x) = \infty \ (-\infty)$$

or

$$f(x) \to \infty \ (-\infty) \quad \text{as} \quad x \to a$$



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



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

Definition 3.3.4. (Precise) Suppose that f is a function defined on a neighborhood of a (except possibly at a itself). We say that "*f approaches ∞ $(-\infty)$, as x approaches a*," if for every $M > 0$ there exists a number $\delta > 0$ such that

$$f(x) > M \quad \big(f(x) < -M\big) \quad \text{whenever} \quad 0 < |x - a| < \delta.$$

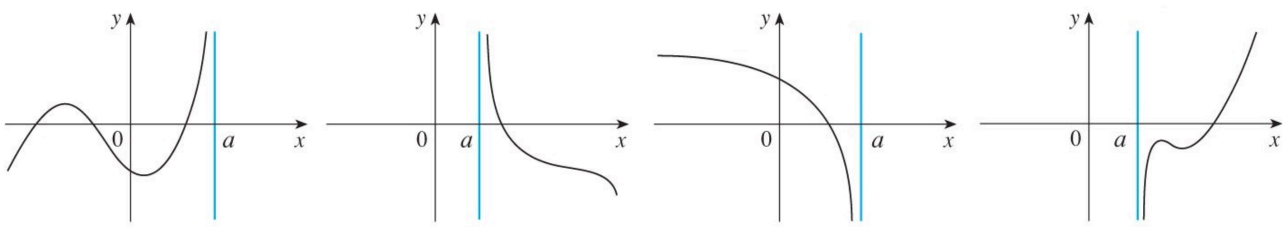
We write

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

or

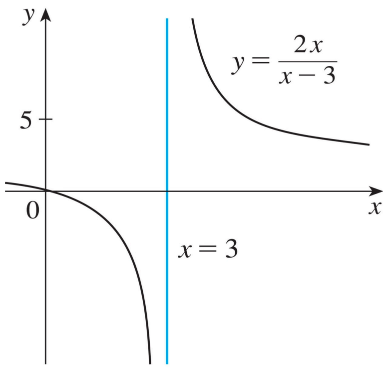
$$f(x) \to \infty \ (-\infty) \quad \text{as} \quad x \to a$$

Note. We can define the sided infinite limits, $\lim_{x \to a^+} f(x) = \infty$ by replacing "$0 < |x - a| < \delta$" by "$0 < x - a < \delta$". The other three limits, $\lim_{x \to a^-} f(x) = \infty$, $\lim_{x \to a^+} f(x) = -\infty$ and $\lim_{x \to a^-} f(x) = -\infty$ can be defined in a similar fashion.



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

Example 3.3.5. Find $\lim_{x \to 3^-} \dfrac{2x}{x - 3}$



Proof. Observe that if $0 < 3 - x < \frac{1}{2}$, then $5 < 2x < 6$ and the quotient $\frac{2x}{x-3}$ is a negative number. Moreover,

$$\frac{2x}{x-3} < \frac{5}{x-3}$$

For given $M > 0$, $\frac{5}{x-3} < -M$ if and only if $x > 3 - \frac{5}{M}$. Thus, choose $\delta = \min(\frac{1}{2}, \frac{5}{M})$. For every number x with $0 < 3 - x < \delta$,

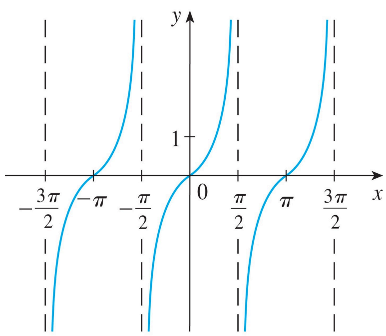
$$\frac{2x}{x-3} < \frac{5}{x-3} < -\frac{5}{5/M} = -M.$$

Since M is an arbitrarily positive number, $\lim_{x\to 3^-} \frac{2x}{x-3} = -\infty$. □

■ **Vertical Asymptote:** The vertical line $x = a$ is called a "*vertical asymptote*" of the curve $y = f(x)$ if at least one of the following statements is true:

$$\begin{array}{lll} \lim_{x\to a} f(x) = \infty & \lim_{x\to a^+} f(x) = \infty & \lim_{x\to a^-} f(x) = \infty \\ \lim_{x\to a} f(x) = -\infty & \lim_{x\to a^+} f(x) = -\infty & \lim_{x\to a^-} f(x) = -\infty \end{array}$$

Example 3.3.6. Find the vertical asymptotes of $f(x) = \tan x$.



$y = \tan x$

Proof. Check that $\lim x \to (\pi/2 + n\pi)^- \tan x = \infty$ or $\lim x \to (\pi/2 + n\pi)^+ \tan x = -\infty$. Then the lines $x = \frac{\pi}{2} + n\pi$, where $n \in \mathbb{Z}$ are all vertical asymptotes of $f(x) = \tan x$.

□

Exercise. Let a function f be defined on $(0, a)$ for some $a > 0$. Prove that either both of the limits

$$\lim_{x \to 0^+} f(x) \quad \text{and} \quad \lim_{t \to \infty} f\left(\frac{1}{t}\right)$$

exist and are equal, or both of them diverge.

Exercise. If $\lim_{x \to a} g(x) = \infty$ and $|f(x)| \leq M$ for all x, then $\lim_{x \to a} \dfrac{f(x)}{g(x)} = 0$.
(Students are supposed to have the ability of writing down the rigorous proof if a is replaced by a^+, a^-, $\pm\infty$.)

Chapter 4

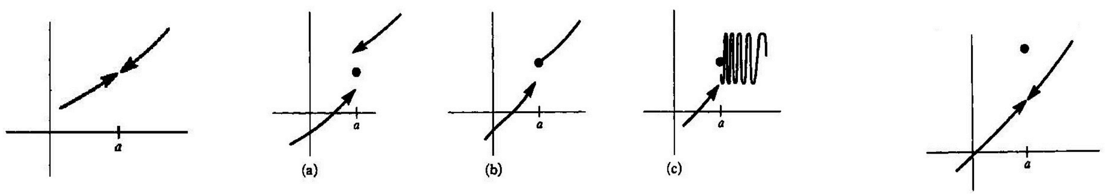
Continuity

Continuity of functions is an important concept in physics and mathematics. From the macroscopic scale, the motion of an object is smooth. By using Euclidian Algorithm to approach a root of an equation, the continuity is also necessary.

4.1 Continuity of a Function

Heuristically, the graph of a continuous function contains no breaks, jumps, or wild oscillation. There are many ways which may make a function fail to be continuous. For example,

- f is not defined at a.
- $\lim\limits_{x \to a} f(x)$ does not exist.
- The limit exists but there is a jump at a.



undefined at **a**   limit does not exist at **a**   limit exists but jump at **a**

Definition 4.1.1. Let f be a function whose domain D contains a number a. We say that

(a) f is "*continuous at a*" if

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

(b) if f is not continuous at a, then f is "*discontinuous at* a.

(c) f is continuous on a set $E \subseteq D$ if it is continuous at every point in E. If f is continuous at every point in D, we say that f is continuous.

Note. The ε-δ expression of the definition is that $f : D \to \mathbb{R}$ is continuous at $a \in D$ if for every $\varepsilon > 0$, there exists $\delta > 0$ such that

$$|f(x) - f(a)| < \varepsilon$$

for every $x \in D$ with $|x - a| < \delta$.

Remark. If $f(x)$ is continuous at a, then

(i) f is defined at a.

(ii) the limit of f exist at a. ($\lim\limits_{x \to a} f(x)$ exists).

(iii) the limit at a is equal to the value of f at a. $\left(\lim\limits_{x \to a} f(x) = f(\lim\limits_{x \to a} x) = f(a)\right)$.

Example 4.1.2.

(1) Any polynomial function is continuous on $\mathbb{R}$.

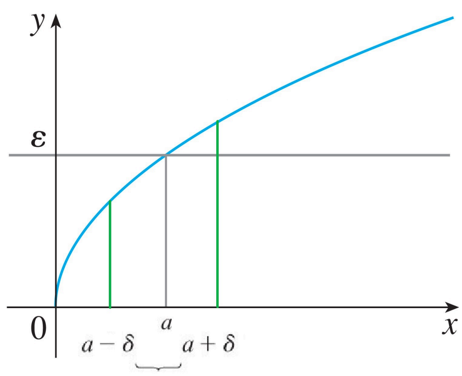
(2) Any rational function is continuous on its domian.

(3) $f(x) = |x|$ is continuous on $\mathbb{R}$.

(4) $f(x) = \sqrt{x}$ is continuous on $\mathbb{R}^+$.

(5) Any trigonometric function is continuos on its domain.

Theorem 4.1.3. *Suppose that $f(x)$ is continuous at a and $f(a) > 0$ ($f(a) < 0$). Then there is $\delta > 0$ such that $f(x) > 0$ ($f(x) > 0$) for all x with $|x - a| < \delta$.*



Proof. Choose $\varepsilon = f(a) > 0$. Since $f(x)$ is continuous at a, there exists $\delta > 0$ such that for every x with $|x - a| < \delta$, then

$$|f(x) - f(a)| < \varepsilon.$$

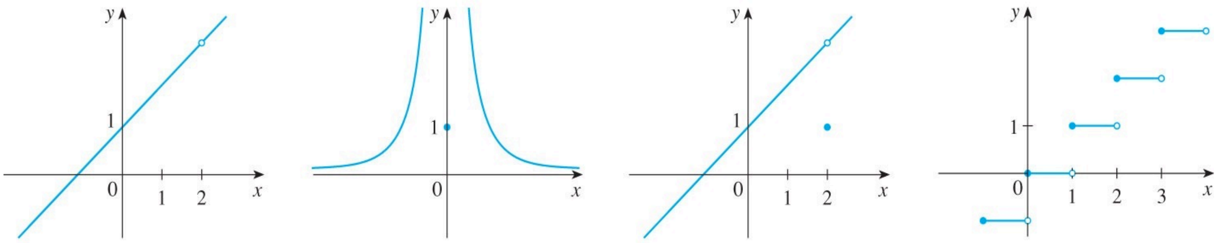
Then, $f(x) - f(a) > -\varepsilon$ and hence $f(x) > f(a) - \varepsilon = f(a) - f(a) = 0$.

□

Example 4.1.4. Suppose $f(x)$ is continuous at a and $f(a) > 0$. Then there is $\delta > 0$ such that $f(x) > \dfrac{f(a)}{2}$ for every x with $|x - a| < \delta$.

■ **Discontinuities**

There are some different types of discontinuities.



(a) $f(x) = \dfrac{x^2 - x - 2}{x - 2}$ — removable discontinuity

(b) $f(x) = \begin{cases} \dfrac{1}{x^2} & \text{if } x \neq 0 \\ 1 & \text{if } x = 0 \end{cases}$ — infinite discontinuity

(c) $f(x) = \begin{cases} \dfrac{x^2 - x - 2}{x - 2} & \text{if } x \neq 2 \\ 1 & \text{if } x = 2 \end{cases}$ — removable discontinuity

(d) $f(x) = [\![x]\!]$ — jump discontinuities

Example 4.1.5. The Heaviside function $H(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$ is (jump) discontinuous at 0 and continuous elsewhere.

Remark. There are some methods for proving discontinuity at $x = a$. See Remark 4.1.11 in the textbook.

❑ Laws of Continuous Functions

Theorem 4.1.6. *(The sums, differences, products, quotients and scalar products of continuous functions are continuous.)*

If $f(x)$ and $g(x)$ are continuous at a and c is a constant , then

(a) $(f \pm g)(x)$ *is continuous at a.*

(b) $(cf)(x)$ *is continuous at a.*

(c) $(f \cdot g)(x)$ *is continuous at a.*

(d) $\left(\dfrac{f}{g}\right)(x)$ *is continuous at a provided* $g(a) \neq 0$.

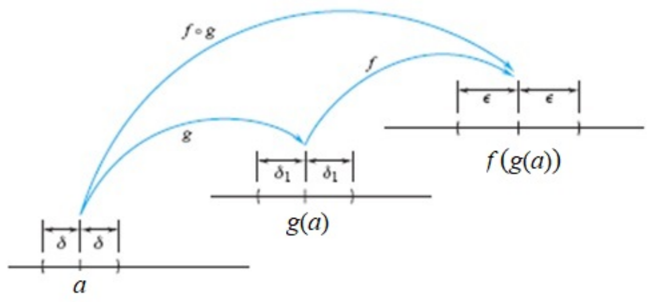
Proof. (Exercise) □

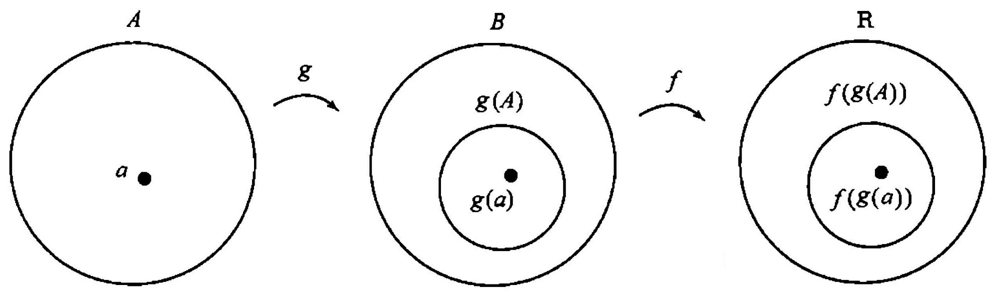
Lemma 4.1.7. *If f is continuous at b and $\lim_{x\to a} g(x) = b$, then $\lim_{x\to a} f(g(x)) = f(b)$. In other words,*

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

Proof. (Exercise) □

Theorem 4.1.8. *(Composite of continuous functions is continuous) If g is continuous at a and f is continuous at $g(a)$, then the composite function $f \circ g$ given by $(f \circ g)(x) = f(g(x))$ is continuous at a.*





Proof. (Exercise) $\square$

Note. $f(x)$ is continuous at "$g(a)$" rather than "a".

Example 4.1.9. Prove that the function $f(x) = \begin{cases} x\sin\frac{1}{x}, & x \neq 0 \\ 0, & x = 0 \end{cases}$ is continuous.

Example 4.1.10. Evaluate $\lim\limits_{x\to 1}\left(\dfrac{3\sqrt{x^2+3x-1}}{\sqrt{|x-2|\cos x}}\right)^3$.

Proof. Let $f(x) = x^2+3x-1$, $g(x) = \sqrt{x}$, $h(x) = |x-2|$, $k(x) = \cos x$ and $F(x) = x^3$. Since f, h and k are continuous at 1, g is continuous at 3 and cos 1 which is nonzero there. Also, F is continuous at $\dfrac{3\sqrt{3}}{\sqrt{\cos 1}}$. Hence,

$$\left(\frac{3\sqrt{x^2+3x-1}}{\sqrt{|x-2|\cos x}}\right)^3 = F\left(\frac{3g\big(f(x)\big)}{g\big(h(x)k(x)\big)}\right)$$

is continuous at 1 and

$$\lim_{x\to 1}\left(\frac{3\sqrt{x^2+3x-1}}{\sqrt{|x-2|\cos x}}\right)^3 = F\left(\frac{3g\big(f(1)\big)}{g\big(h(1)k(1)\big)}\right) = \frac{81\sqrt{3}}{\cos^3 1}.$$

$\square$

■ **One-sided Continuity**

We recall the defintion of continuity that the limit of f at a from both sides. But some functions are only defined on one side. For example, $f(x) = \sqrt{x}$ is defined on $[0, \infty)$. What's the continity of f at 0?

Definition 4.1.11. Let f be a function. We say that f is "*right continuous at a*" (or "*continuous from the right at a*") if $\lim_{x \to a^+} f(x) = f(a)$.
Rephase as ε-δ defintion that "*for given* $\varepsilon > 0$ *there exists* $\delta > 0$ *such that* $|f(x) - f(a)| < \varepsilon$, *provided that* $0 \leq x - a < \delta$".

Similary, we can define the "*left continuous at a*" by $\lim_{x \to a^-} f(x) = f(a)$ and replacing "$0 \leq x - a < \delta$" by "$0 \leq a - x > \delta$".

■ **Continuous on an interval with endpoint(s)**

Definition 4.1.12. Suppose that f is a function defined on $[a, b]$. We say that f is continuous on $[a, b]$ if

(i) f is continuous on (a, b), and

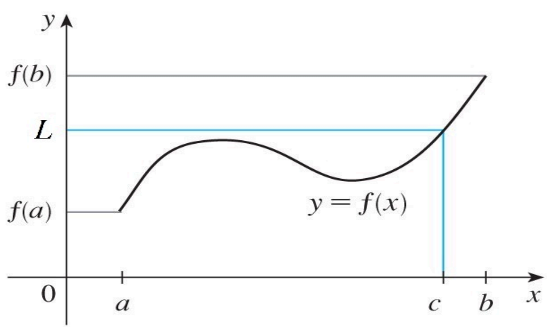
(ii) $\lim_{x \to a^+} f(x) = f(a)$ and $\lim_{x \to b^-} f(x) = f(b)$

Example 4.1.13. The Heaviside function $H(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$ is (jump) discontinuous at 0 and continuous elsewhere. $H(x)$ is right continuous at 0 but not left continuous there.

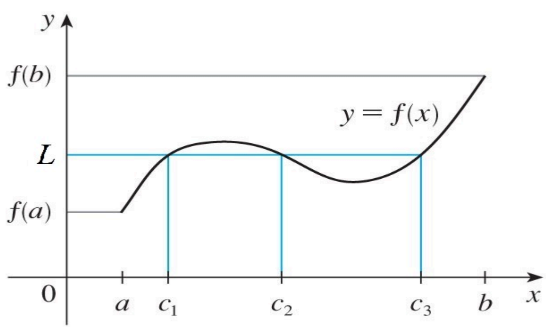
4.2 Properties of Continuous Functions

❑ Intermediate Value Theorem

Theorem 4.2.1. *(Intermediate Value Theorem) Suppose that f is continuous on the closed interval $[a, b]$ and let L be any number between $f(a)$ and $f(b)$, where $f(a) \neq f(b)$. Then there exists a number c in (a, b) such that $f(c) = L$.*



(a)

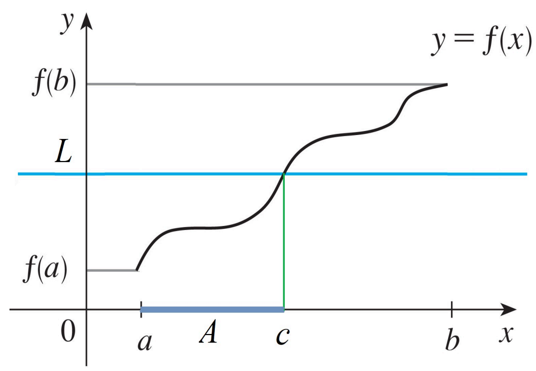


(b)

Proof.

W.L.O.G, we may assume $f(a) < L < f(b)$. Define

$$A = \left\{x \in [a,b] \;\middle|\; f(y) < L \text{ for all } y < x\right\}.$$



Since $f(a) < L$ and f is continuous at a and $f(a) < L$, there exists $\delta_1 > 0$ such that $f(x) < L$ for all $x \in [a, a+\delta_1)$. Hence, A is nonempty.

Clearly, b is an upper bound for A. By the least upper bound property, there exists a number $c \in [a,b]$ such that $c = \sup A$.

We claim that $c \neq a, b$. Since f is continuous at b and $f(b) > L$, there exists $\delta_2 > 0$ such that $f(x) > L$ for all $x \in (b - \delta_2, b]$. Then $b - \delta_2/2$ is an upper bound for A and b is not a least upper bound for A. Thus, $c \neq b$. Similarly, $c \neq a$.

Now, we want to prove $f(c) = L$. Assume that $f(c) \neq L$, then either $f(c) < L$ or $f(c) > L$.

(i) If $f(c) < L$, since f is continuous at c, there exists $\delta_3 > 0$ such that $f(x) < L$ for all $x \in (c - \delta_3, c + \delta_3)$. Then there is $x_0 > c$ such that $f(x) < L$ for all $x < x_0$. Hence, it contradicts that c is an upper bound for A.

(ii) if $f(c) > L$, since f is continuous at c, there exists $\delta_4 > 0$ such that $f(x) > L$ for all $x \in (c - \delta_4, c + \delta_4)$. Then there is $x_1 < c$ such that x_1 is an upper bound for A. It contradicts that c is a least upper bound for A.

Therefore, $f(c) = L$ and the theorem is proved.

□

Remark. The theorem is false if one of the following condition happens.

(i) f is not continuous.

(ii) f is continuous on (a,b) but not continuous at the endpoint(s).

(iii) $f(a) = f(b)$ (No number is between them.)

Example 4.2.2. Let $f(x) = 4x^3 - 6x^2 + 3x - 2$. Prove that there is a root of $f(x)$ between 1 and 2.

Proof. Since f is a polynomial, it is continuous on $[1,2]$. Also,

$$\begin{aligned} f(1) &= -1 < 0 \\ f(2) &= 12 > 0 \end{aligned}$$

By the Intermediate Value Theorem, there exists a number $c \in (1,2)$ such that $f(c) = 0$ and c is a root of f between 1 and 2.

□

Exercise.

(1) Let n be an even positive integer. Prove that every positive number has a n-th root.

(2) Let n be an odd integer. Prove that the polynomial $P(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ has a root.

(3) Let n be an odd integer. Prove that for any number $\alpha \in \mathbb{R}$ there is a number c such that $P(c) = \alpha$.

❑ Boundedness of Continuous Functions

Theorem 4.2.3. *If f is continuous at a, then there exists $\delta > 0$ such that f is bounded on $(a - \delta, a + \delta)$.*

Proof. Since f is continuous at a, for $\varepsilon = 1$, there exists $\delta > 0$ such that $|f(x) - f(a)| < 1$ for all $|x - a| < \delta$. Therefore,

$$f(a) - 1 < f(x) < f(a) + 1 \quad \text{for all } x \in (a - \delta, a + \delta).$$

Hence, f is bounded on $(a - \delta, a + \delta)$. □

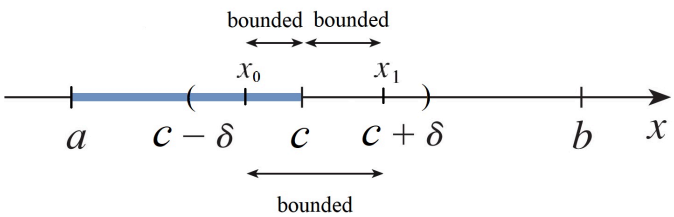
❑ Extreme Value Theorem

Theorem 4.2.4. *If f is continuous on $[a, b]$. Then f is bounded on $[a, b]$.*

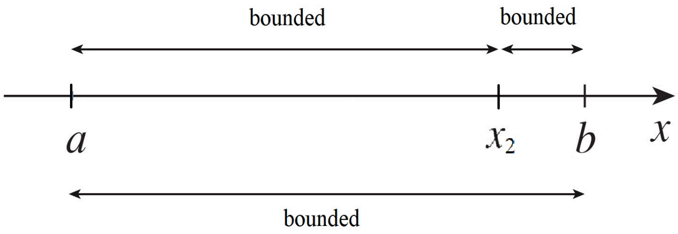
Proof. Define $A = \{x \in [a, b] \mid f \text{ is bounded on } [a, x]\}$. $\boxed{\text{We want to prove } A = [a, b]}$.

(i) **Step1:** To prove that A has a least upper bound.
Clearly, $a \in A$. Hence, A is nonempty. Since b is an upper bound for A, by the least upper bound property, there exists $c \in [a, b]$ such that $c = \sup A$.

(ii) **Step2:** To prove $b = c$ $(= \sup A)$.
Assume that $c < b$. Since f is continuous at c, there exists $\delta > 0$ such that $f(x)$ is bounded on $(c - \delta, c + \delta) \subseteq [a, b)$. Moreover, since c is a least upper bound for A, there exists $x_0 \in A$ with $c - \delta < x_0 < c$. Hence, f is bounded on $[a, x_0]$. Also, there exists $x_1 \in (c, c + \delta)$ such that $f(x)$ is bounded on $[x_0, x_1]$. Then f is bounded on $[a, x_1]$. This implies that $x_1 \in A$ and we obtain a contradiction that c is an upper bound for A. Then $c = b$.



(iii) **Step3:** To prove that f is bounded on $[a, b]$.
Since f is continuous at b, there exists $\delta_1 > 0$ such that f is bounded on $(b - \delta_1, b]$. That $b = \sup A$ implies that there exists $x_2 \in (b - \delta_1, b]$ and $x_2 \in A$. Then f is bounded on $[a, x_2]$. Also, f is bounded on $[x_2, b]$. We have f is bounded on $[a, b]$.



∎

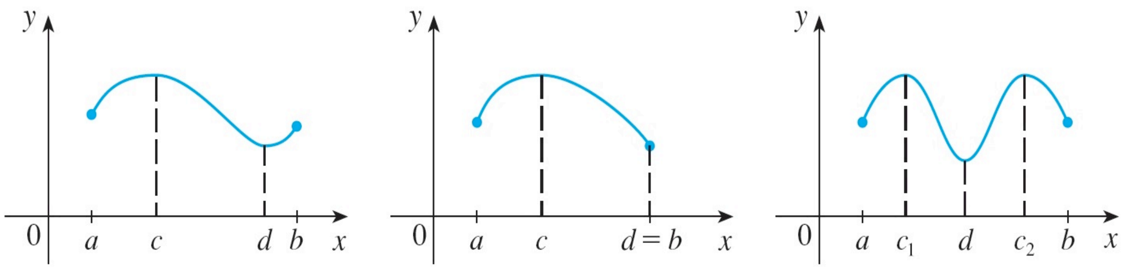
Theorem 4.2.5. *(Extreme Value Theorem) If f is continuous on $[a,b]$, then there exists a number $c \in [a,b]$ such that*

$$f(c) \geq f(x) \quad \textit{for all } x \in [a,b]$$

That is, $f(c) = \max\limits_{x\in[a,b]} f(x)$. Similarly, there exists a number $d \in [a,b]$ such that

$$f(d) \leq f(x) \quad \textit{for all } x \in [a,b]$$

That is, $f(d) = \min\limits_{x\in[a,b]} f(x)$.



Proof. Since f is continuous on $[a,b]$, it is bounded on $[a,b]$. Then the set $\{f(x) \mid x \in [a,b]\}$ is bounded and nonempty. By the least upper bound property, the set has a least upper bound, say $M = \sup\{f(x) \mid x \in [a,b]\}$.

Assume that there is no number in $[a,b]$ such that the values of f attain its maximum. Define $g(x) = \dfrac{1}{f(x) - M}$. Since f is continuous on $[a,b]$ and $f(x) \neq M$ for all $x \in [a,b]$, g is continuous on $[a,b]$, say $0 < |g(x)| < L$.

On the other hand, M is a least upper bound of $\{f(x) \mid x \in [a,b]\}$. Then there exists $x_0 \in [a,b]$ and $y_0 \in \{f(x) \mid x \in [a,b]\}$ such that $y_0 = f(x_0)$ and

$$M - \frac{1}{2L} < y_0 < M.$$

Hence,

$$|g(x_0)| = \left|\frac{1}{f(x_0) - M}\right| = \left|\frac{1}{y_0 - M}\right| > \left|\frac{1}{(M - \frac{1}{2L}) - M}\right| = \frac{1}{1/2L} = 2L > L.$$

The contradiction implies that f must attain its maximum at some number. ∎

Remark. In the theorem, the continuity and the closedness of the interval are necessary. The theorem is false if either

(i) f is not continuous on $[a, b]$, or

(ii) f is continuous on (a, b).

Exercise.

(1) Let n be an even integer and $P(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$. Then there is a number $c \in \mathbb{R}$ such that

$$P(c) \leq P(x) \quad \text{for all } x \in \mathbb{R}.$$

That is, $P(c) = \min_{x \in \mathbb{R}} P(x)$.

(2) Let n be an even integer and the equation

$$x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = \alpha.$$

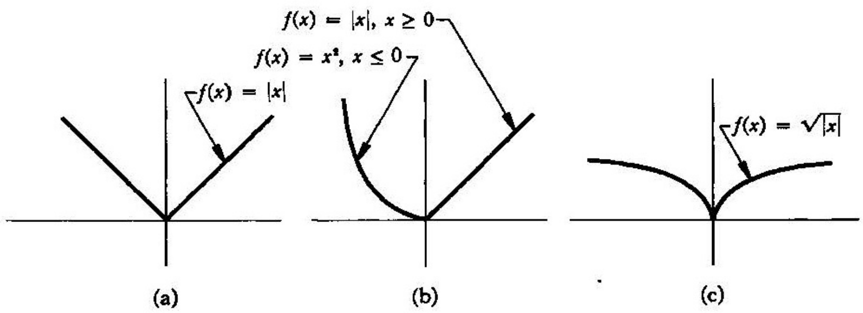
Prove that there is a number $m \in \mathbb{R}$ such that the equation has no solution for $\alpha < m$ and the equation has a solution for $\alpha \geq m$.

Chapter 5

Differentiation

5.1 Derivative of a Function

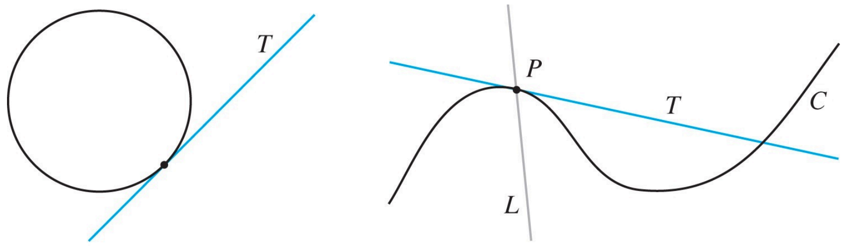
So far, we have learned the limits and continuity of functions. Some (local) information of a function can be obtained by studying its limits and continuity. But continuous functions have many different types.



The concepts in the previous chapters cannot reflect how a function changes locally. Therefore, we will discuss the "*rate of change*" of a function. Some mathematical and physical problems such as tangents and velocities involve this topic.

■ **Tangents**

The word tangent is derived from the Latin word tangens, which means "touching." How to make the idea that "*a tangent to a curve is a line that touches the curve*" precise?

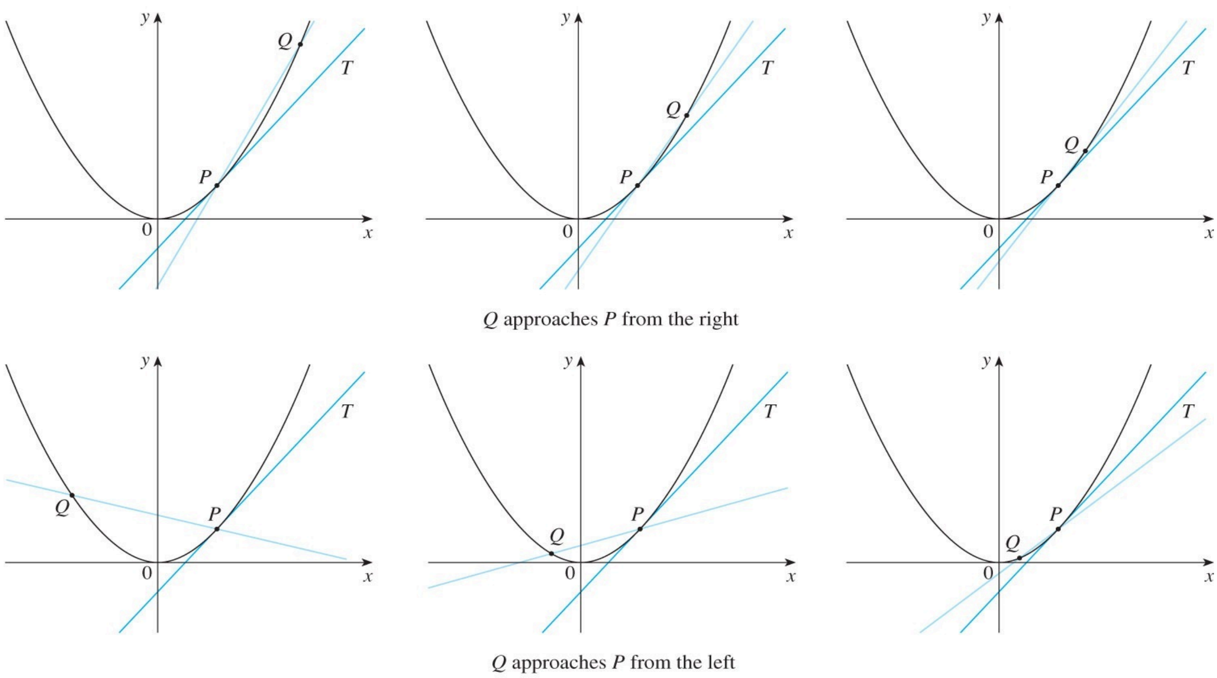


(a) (b)

Question: How to find the tangent line of the graph of $f(x)$ at a given point? To find the slope of the tangent line.

For a curve $C : y = f(x)$ and a point $P(a, f(a))$ on C, consider the slope of the secant line PQ. Say $Q(x, f(x))$ on C. Then

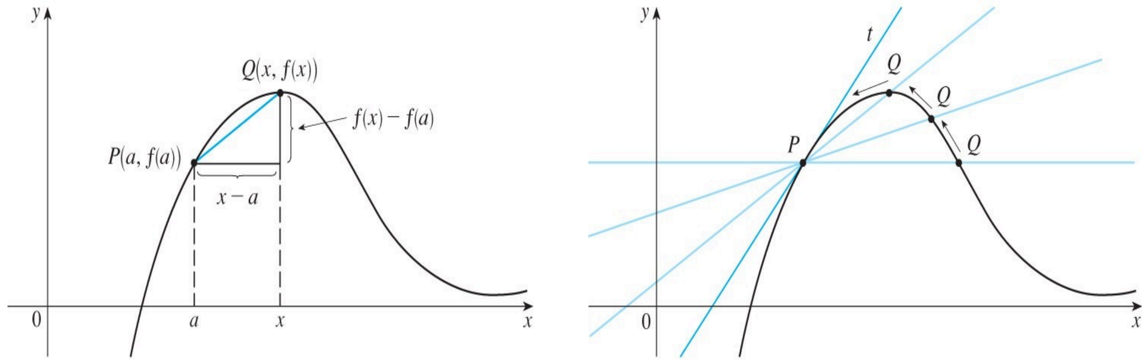
$$m_{PQ} = \frac{f(x) - f(a)}{x - a}.$$



Q approaches P from the right

Q approaches P from the left

Let Q approach P along the curve C by letting x approach a. If m_{PQ} approaches a number m, then we define tha tangent T to be the line through P with slope m.



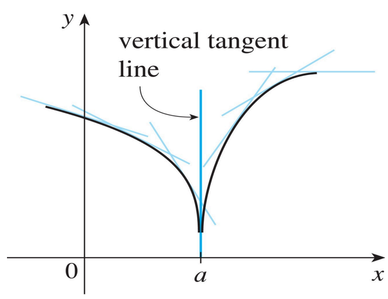
Definition 5.1.1. The "*tangent line*" to the curve $y = f(x)$ at the point $P(a, f(a))$ is

(a) the line through P with slope

$$m = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

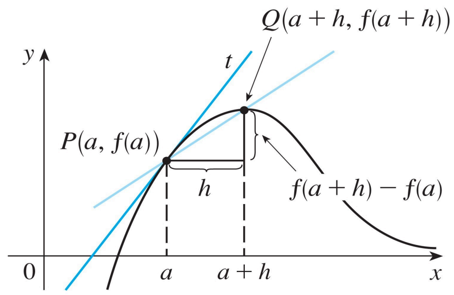
provided that this limit exists, or

(b) the (vertical) line $x = a$, if $\lim\limits_{x \to a^+} \dfrac{f(x) - f(a)}{x - a} = \pm\infty$ or $\lim\limits_{x \to a^-} \dfrac{f(x) - f(a)}{x - a} = \pm\infty$



Note. An alternating expression of the slope of the tangent line is

$$m = \lim_{h \to 0} \frac{f(a + h) - f(a)}{h}$$



Example 5.1.2. Find an equation of the tangnet line to the hyperbola $y = 3/x$ at the point $(3, 1)$.

Proof. Let $f(x) = 3/x$. The slope of the tangent line at $(3, 1)$ is

$$\lim_{x\to 3} \frac{f(x) - f(3)}{x - 3} = \lim_{x\to 3} \frac{3/x - 1}{x - 3} = \lim_{x\to 3} \frac{3 - x}{x(x - 3)} = -\frac{1}{3}.$$

Hence, the equation of the tangent of $y = f(x)$ at $(3, 1)$ is

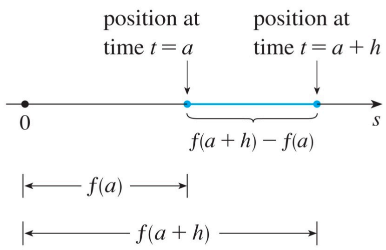
$$y - 1 = -\frac{1}{3}(x - 3)$$

or

$$x + 3y - 6 = 0$$

□

■ Velocity



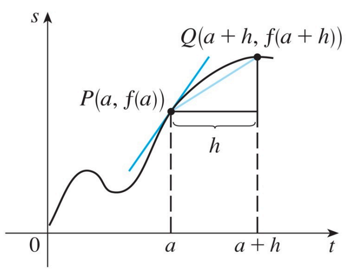
Let $f(t)$ be the position function of a particle. The average velocity from $t = a$ to $t = a + h$ is

$$\frac{f(a + h) - f(a)}{h}.$$

We define the velocity (or the instantaneous velocity) at time $t = a$ is

$$\lim_{h\to 0} \frac{f(a + h) - f(a)}{h}.$$

Remark. The value is equal to the slope of the tangent line at P.



$m_{PQ} = \dfrac{f(a+h) - f(a)}{h} =$ average velocity

❑ Derivative

We observe the "*difference quotient*" plays an important role when we study the (local) change of a function and its limit represents the "*rate of change* " of a function.

Definition 5.1.3. Let f be a function defined on D which cantains a neighborhood of a.

(a) We say that f is "*differentiable at a*" if

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

exists. The limit is denoted by $f'(a)$ and is called the "*derivative of f at a* .

(b) If $f(x)$ is differentiable at every point of a set I, we say that "*f is differnetiable on I*".

Note. If replacing x by $a + h$, we have

$$f'(a) = \lim_{x\to a} \frac{f(x) - f(a)}{x - a} = \lim_{h\to 0} \frac{f(a + h) - f(a)}{h}.$$

(c) We collect every point $x \in D$ where $f'(x)$ is defined (i.e. the limit $\lim\limits_{h\to 0}\dfrac{f(x+h)-f(x)}{h}$ exists). Then we can regard $f'(x)$ as a function and is called the "*derivative of $f(x)$*".

Note. $Dom(f') \subseteq Dom(f)$.

Example 5.1.4. Determine whether the following functions are differentiable at the given point.

(1) $f(x) = cx + d$, at $x = a$.

Proof.

$$f'(a) = \lim_{h\to 0}\frac{f(a+h)-f(a)}{h} = \lim_{h\to 0}\frac{[c(a+h)+d]-[ca+d]}{h} = c.$$

Hence, $f'(a) = c$. (independent of a) □

(2) $f(x) = x^2$ at $x = a$

Proof.

$$f'(a) = \lim_{h\to 0}\frac{f(a+h)-f(a)}{h} = \lim_{h\to 0}\frac{(a+h)^2-a^2}{h} = \lim_{h\to 0}\frac{2ah+h^2}{h} = \lim_{h\to 0}(2a+h) = 2a.$$

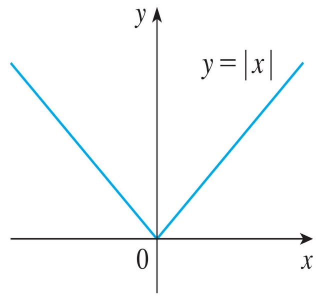
Hence, $f'(a) = 2a$. □

(3) $f(x) = |x|$ at $x = 0$.

Proof.

Consider $\lim\limits_{h\to 0}\dfrac{f(h)-f(0)}{h} = \lim\limits_{h\to 0}\dfrac{|h|-0}{h}$.

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



Hence, the limit $\lim\limits_{h\to 0}\dfrac{f(h)-f(0)}{h}$ does not exist and f is not differentiable at 0. □

(4) $f(x) = \begin{cases} 4x, & x < 1 \\ 2x^2+2, & x \geq 1 \end{cases}$ at $x = 1$.

Proof. (Exercise) □

(5) $f(x) = \sqrt{x}$ at $x > 0$.

Proof.

$$
\begin{aligned}
f'(x) &= \lim_{h\to 0}\frac{f(x+h)-f(x)}{h} = \lim_{h\to 0}\frac{\sqrt{x+h}-\sqrt{x}}{h} = \lim_{h\to 0}\left[\frac{\sqrt{x+h}-\sqrt{x}}{h}\cdot\frac{\sqrt{x+h}+\sqrt{x}}{\sqrt{x+h}+\sqrt{x}}\right] \\
&= \lim_{h\to 0}\frac{h}{h(\sqrt{x+h}+\sqrt{x})} = \frac{1}{2\sqrt{x}}.
\end{aligned}
$$

□

Note. (**Tangent line**) Suppose that $f(x)$ is differentiable at a. Then the equation of the tangent line of $y = f(x)$ at $\big(a, f(a)\big)$ is

$$y - f(a) = f'(a)(x-a)$$

or

$$y = f(a) + f'(a)(x-a).$$

■ **Continuity and Differentiability**

Theorem 5.1.5. *If $f(x)$ is differentiable at a, then $f(x)$ is continuous at a.*

Proof. Since f is differentiable at a, the derivative $f'(a) = \lim\limits_{h\to 0}\dfrac{f(a+h)-f(a)}{h}$ exists. Also, $\lim\limits_{h\to 0} h = 0$. We have

$$\lim_{h\to 0}\Big(f(a+h)-f(a)\Big) = \lim_{h\to 0}\Big(\frac{f(a+h)-f(a)}{h}\cdot h\Big) = \lim_{h\to 0}\frac{f(a+h)-f(a)}{h}\cdot\lim_{h\to 0} h = f'(a)\cdot 0 = 0.$$

Then

$$\lim_{h\to 0} f(a+h) = \lim_{h\to 0}\Big(f(a+h)-f(a)+f(a)\Big) = \lim_{h\to 0}\Big(f(a+h)-f(a)\Big) + \lim_{h\to 0} f(a) = 0 + f(a) = f(a).$$

Hence, f is continuous at a. □

Remark. The converse of this theorem is false. For example, $f(x) = |x|$.

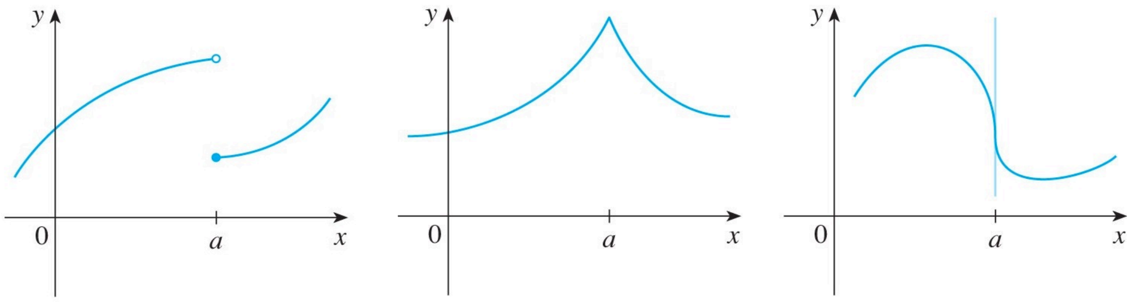
Question: How can a function fail to be differentiable?

f is not differentiable at a if

(i) f is not continuous at a;

(ii) f is continuous at a but $\lim\limits_{h\to 0}\dfrac{f(a+h)-f(a)}{h}$ DNE;

(iii) $\lim\limits_{h\to 0}\dfrac{f(a+h)-f(a)}{h} = \pm\infty$ vertical tangnet.



(a) A discontinuity (b) A corner (c) A vertical tangent

Three ways for f not to be differentiable at a

Remark. Heuristically, if the graph of a function $y = f(x)$ has a discontinuity, a corner(cusp) or a vertical tangent line at $\big(a, f(a)\big)$, then $f(x)$ is not differentiable at a. On the contrary, if the graph of $f(x)$ is smooth at $\big(a, f(a)\big)$, then f is differentiable at a.

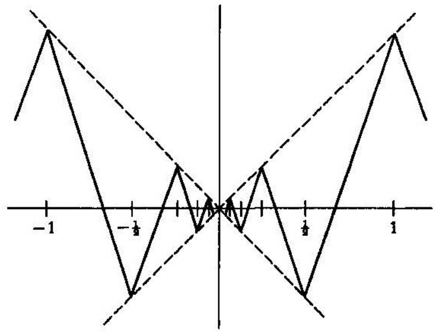
Example 5.1.6. Some continuous but nondifferentiable functions:

(1) $f(x) = |x|$;

(2) $f(x) = \begin{cases} x^2, & x \le 0 \\ x, & x > 0 \end{cases}$;

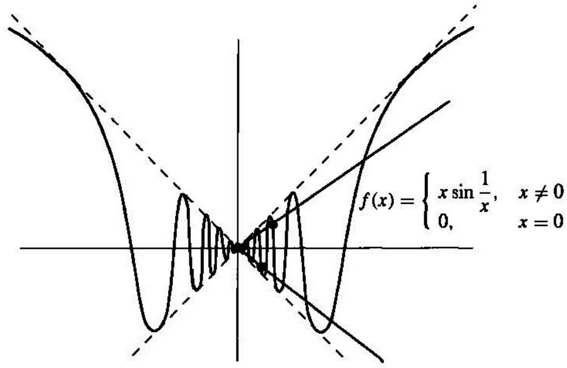
(3) $f(x) = \sqrt{|x|}$;

(4) $f(x) = \begin{cases} x \sin \dfrac{1}{x} & , x \neq 0 \\ 0, & x = 0 \end{cases}$



$\lim_{h\to 0}(h \sin \frac{1}{h} - 0)/h = \lim\limits_{h\to 0} \sin \dfrac{1}{h}$ does not exist. Hence, f is not differentiable at 0.

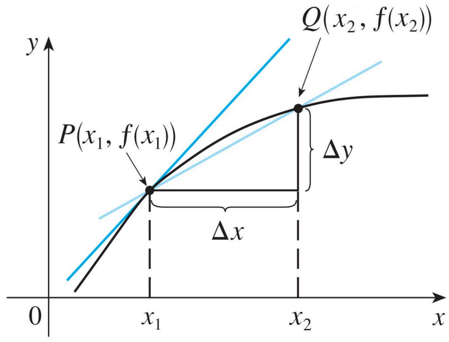
(5) $f(x) = \begin{cases} x^2 \sin \dfrac{1}{x} & , x \neq 0 \\ 0, & x = 0 \end{cases}$



$\lim_{h\to 0}(h^2 \sin \frac{1}{h} - 0)/h = \lim\limits_{h\to 0} h \sin \dfrac{1}{h} = 0$. Hence, f is differentiable at 0.

(6) There are functions which are continuous everywhere, but are differentiable nowhere.

❑ Rate of Change



average rate of change = m_{PQ}

instantaneous rate of change = slope of tangent at P

$\triangle x = x_2 - x_1$, $\triangle y = f(x_2) - f(x_1)$
The quotient differenece

$$\frac{\triangle y}{\triangle x} = \frac{f(x_2) - f(x_1)}{x_2 - x_1}$$

is called the **average rate of change of** y **with respect to** x over $[x_1, x_2]$.

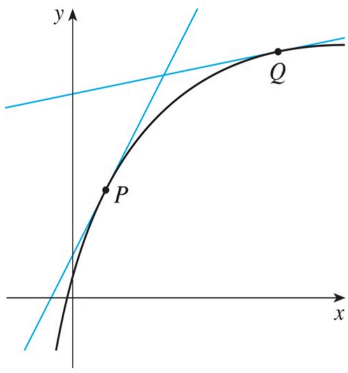
$$f'(a) = \lim_{\triangle x \to 0} \frac{\triangle y}{\triangle x}$$

is called the **instantaneous rate of change of** y **with respect to** x.

Hence, $f'(a)$ is the instantaneous rate of change of $y = f(x)$ with respect to x when $x = a$.

Remark.

If $f'(a)$ is large, it means that the curve at $x = a$ is steep. Hence, y-value changes rapidly



The y-values are changing rapidly at P and slowly at Q.

❑ Notation

Let $y = f(x)$. Some common alternative notations for thederivative are as follows:

$$f'(x) = y' = \frac{dy}{dx} = \frac{df}{dx} = \frac{d}{dx}f(x) = Df(x) = D_x f(x).$$

The symbols D and $\frac{d}{dx}$ are called "**differentiation operators**" because they indicate the operation of differentiation.

$$f'(a) = \left.\frac{dy}{dx}\right|_{x=a} = \left.\frac{df}{dx}\right|_{x=a}.$$

❑ Higher Derivatives

Recall the we can regard $f'(x)$ as a new function and consider the differentiability of f'.

$$\begin{array}{rcll}
f \xrightarrow{\frac{d}{dx}} f' & & & \text{derivative of } f \\
f' \xrightarrow{\frac{d}{dx}} (f')' & = & f'' & \text{derivative of } f' \quad (\textbf{second derivative of } f) \\
f'' \xrightarrow{\frac{d}{dx}} (f'')' & = & f''' & \text{derivative of } f'' \quad (\textbf{third derivative of } f) \\
& & \vdots & \\
& & f^{(n)} & \qquad\qquad\qquad (\textbf{nth derivative of } f)
\end{array}$$

■ **<u>Leibniz notation:</u>**

$$f \xrightarrow{\frac{d}{dx}} \frac{df}{dx} \xrightarrow{\frac{d}{dx}} \frac{d}{dx}\Big(\frac{df}{dx}\Big) = \frac{d^2 f}{d^2 x} \xrightarrow{\frac{d}{dx}} \frac{d}{dx}\Big(\frac{d^2 f}{dx^2}\Big) = \frac{d^3 f}{dx^3} \cdots \xrightarrow{\frac{d}{dx}} \frac{d^n f}{dx^n}$$

5.2 Differentiation Formulas

Using the definition to find the derivatives of functions would be tedious. We hope to study some rules to help finding the derivative without using the definition.

Theorem 5.2.1. *(Differentiation formulas) Let f and g be differentiable at a, and c be a constant. Then*

(a) $f \pm g$ is differentiable at a and $\big(f \pm g\big)'(a) = f'(a) \pm g'(a)$;

(b) cf is differentiable at a and $\big(cf\big)'(a) = cf'(a)$;

(c) fg is differentiable at a and $\big(fg\big)'(a) = f'(a)g(a) + f(a)g'(a)$ (**Product rule**)*;*

(d) $\dfrac{f}{g}$ is differentiable at a provided $g'(a) \neq 0$ and

$$\Big(\frac{f}{g}\Big)(a) = \frac{f'(a)g(a) - f(a)g'(a)}{[g(a)]^2} \qquad (\textbf{Quotient rule})$$

Proof. (Proof of product rule)

$$\begin{aligned}
& \lim_{h\to 0} \frac{(fg)(a+h) - (fg)(a)}{h} \\
= \ & \lim_{h\to 0} \frac{f(a+h)g(a+h) - f(a)g(a)}{h} \\
= \ & \lim_{h\to 0} \frac{f(a+h)g(a+h) - f(a)g(a+h) + f(a)g(a+h) - f(a)g(a)}{h} \\
= \ & \lim_{h\to 0} \frac{\big(f(a+h) - f(a)\big)g(a+h) + f(a)\big(g(a+h) - g(a)\big)}{h}
\end{aligned}$$

Since f and g are differentiable at a, they are continuous at a. Hence,

$$\lim_{h\to 0}\frac{f(a+h)-f(a)}{h}=f'(a),\ \lim_{h\to 0}\frac{g(a+h)-g(a)}{h}=g'(a),\ \lim_{h\to 0}g(a+h)=g(a).$$

Plugging the above limit, we have

$$\begin{aligned}&\lim_{h\to 0}\frac{(fg)(a+h)-(fg)(a)}{h}\\ =\ &\lim_{h\to 0}\frac{f(a+h)-f(a)}{h}\cdot\lim_{h\to 0}g(a+h)+f(a)\lim_{h\to 0}\frac{g(a+h)-g(a)}{h}\\ =\ &f'(a)g(a)+f(a)g'(a).\end{aligned}$$

□

Remark. $\big(fg\big)'\neq f'g'$ and $\left(\frac{f}{g}\right)'\neq\left(\frac{f'}{g'}\right)$.

Corollary 5.2.2. *If $f_1, f_2, \ldots, f_n$ and g are differentiable at a, then*

(a) $f_1+f_2+\cdots+f_n$ *are differentiable at a and*

$$\big(f_1+f_2\cdots+f_n\big)'(a)=f_1'(a)+f_2'(a)+\cdots+f_n'(a);$$

(b) $f_1f_2\cdots f_n$ *are differentiable at a and*

$$\big(f_1f_2\cdots f_n\big)'(a)=f_1'(a)f_2(a)\cdots f_n(a)+f_1(a)f_2'(a)\cdots f_n(a)+\cdots f_1(a)f_2(a)\cdots f_n'(a).$$

If $f_i=f$ for all $i=1,2,\ldots,n$, then part (b) can be rewritten as

$$\big(f^n\big)'(a)=nf^{n-1}(a)f'(a).$$

Note: This law is true for all $n\in\mathbb{R}$. We will discuss in the future.

(c) $\left(\frac{1}{g}\right)'(a)=-\frac{g'(a)}{[g(a)]^2}$ *provided $g(a)\neq 0$.*

Proof. Exercise □

Exercise.

(i) (constant function) $f(x)=c$ is a constant function. Then $f'(x)=0$.

(ii) (power function) Let $f(x)=x^n$ for $n\in\mathbb{N}$. Then $f'(x)=nx^n$.

(iii) $f(x)=x^{-n}$ for $n\in\mathbb{N}$. Then $f'(x)=-nx^{-n-1}$.

(iv) $f(x)=x^n$ for $n\in\mathbb{Q}$ (that is, $n=\frac{p}{q}$ where $p,q\in\mathbb{Z}$). Then $f'(x)=nx^{n-1}$.

(v) (polynomial function) Let $P(x)=a_nx^n+a_{n-1}x^{n-1}+\cdots+a_1x+a_0$ be a polynomial. Then

$$P'(x)=na_nx^{n-1}+(n-1)a_{n-1}x^{n-2}+\cdots+2a_2x+a_1=\sum_{k=1}^{n}ka_kx^{k-1}.$$

Proof. (iii)

$$\frac{d}{dx}(x^{-n}) = \frac{d}{dx}\Big(\frac{1}{x^n}\Big) = -\frac{\frac{d}{dx}(x^n)}{(x^n)^2} = -\frac{nx^{n-1}}{x^{2n}} = -nx^{-n-1}.$$

Problem (i), (ii), (iv), (v) are left as exercise. □

Remark. From (ii), (iii), (iv), we have the power rule that for $n \in \mathbb{Q}$

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

In fact, the rule is true for all $n \in \mathbb{R}$ and we will discuss it in the future.

Exercise.

(i) $\frac{d}{dx}(x^{20}) = 20x^{19}$.

(ii) $f(x) = (3x^2)(5x^4)$.
$f'(x) = (3x^2)'(5x^4) + (3x^2)(5x^4)' = 6x \cdot 5x^4 + 3x^2 \cdot 20x^3 = 30x^5 + 60x^5 = 90x^5$.
In fact, $f(x) = 15x^6$ and thus $f'(x) = 90x^5$.

(iii) $f(x) = \dfrac{5x^3 + 2x - 3}{3x^2 + 1}$.

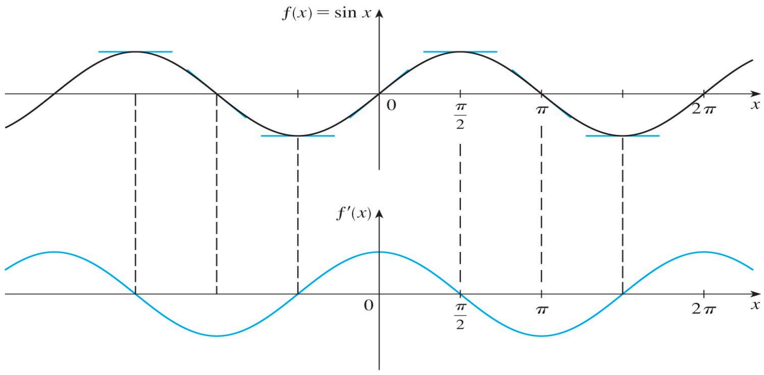
$$\begin{aligned} f'(x) &= \frac{(5x^3 + 2x - 3)(3x^2 + 1) - (5x^3 + 2x - 3)(3x^2 + 1)'}{(3x^2 + 1)^2} \\ &= \frac{(15x^2 + 2)(3x^2 + 1) - (5x^3 + 2x - 3)(6x)}{(3x^2 + 1)^2}. \end{aligned}$$

❑ Derivatives of Trigonometric Functions

- $f(x) = \sin x$

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{f(x+h)f(x)}{h} = \lim_{h\to 0} \frac{\sin(x+h) - \sin x}{h} \\ &= \lim_{h\to 0} \frac{\sin x \cos h + \cos x \sin h - \sin x}{h} \\ &= \lim_{h\to 0} \frac{\sin x(\cos h - 1) + \cos x \sin h}{h} \\ &= \lim_{h\to 0} \sin x \cdot \lim_{h\to 0} \frac{\cos h - 1}{h} + \lim_{h\to 0} \cos x \cdot \lim_{h\to 0} \frac{\sin h}{h} \\ &= \cos x \end{aligned}$$

The last equality follows the fact that $\lim_{h\to 0} \frac{\cos h - 1}{h} = 0$ and $\lim_{h\to 0} \frac{\sin h}{h} = 1$.



- $f(x) = \cos x$. Then $f'(x) = -\sin x$. (Exercise)

- $f(x) = \tan x = \dfrac{\sin x}{\cos x}$. By the quotient rule,

$$f'(x) = \frac{(\sin x)' \cos x - \sin x(\cos x)'}{\cos^2 x} = \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = \frac{1}{\cos^2 x} = \sec^2 x$$

- $f(x) = \cot x$, $f'(x) = -\csc^2 x$.

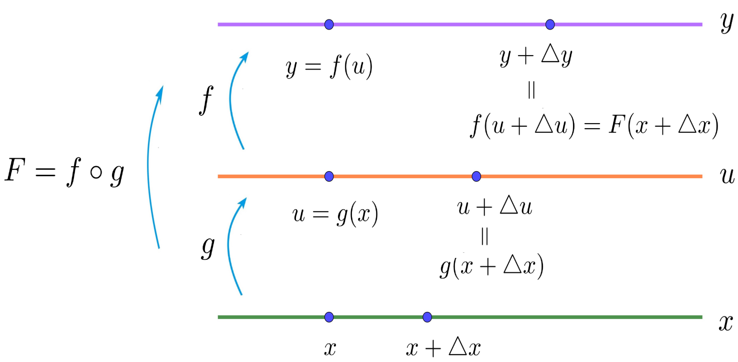
- $f(x) = \sec x$, $f'(x) = \sec x \tan x$.

- $f(x) = \csc x$, $f'(x) = -\csc x \cot x$.

5.3 The Chain Rule

So far, we cannot use the differentiation formulas in the previous section to find the derivative of $F(x) = \sqrt{x^2 + 1}$. Let $f(x) = \sqrt{x}$ and $g(x) = x^2 + 1$. then $F(x) = f\big(g(x)\big)$.

In general, we want to deal with the differentiation of a composite function. Let $F(x) = \big(f \circ g\big)(x) = f\big(g(x)\big)$. What is $F'(x)$?



Heuristic idea: Let $u = g(x)$ and $y = f(u)$ Then $y = f\big(g(x)\big) = F(x)$.

$$F'(x) = \lim_{\triangle x\to 0} \frac{F(x+\triangle x) - F(x)}{\triangle x} = \lim_{\triangle x\to 0} \frac{\triangle y}{\triangle x}.$$

Consider

$$\begin{aligned}
\frac{\triangle y}{\triangle u} &= \frac{f(u+\triangle u)}{\triangle u} &\Rightarrow \quad \lim_{\triangle u\to 0}\frac{\triangle y}{\triangle u} = \lim_{\triangle u\to 0}\frac{f(u+\triangle u) - f(u)}{\triangle u} = f'(u)\\
\frac{\triangle u}{\triangle x} &= \frac{g(x+\triangle x)}{\triangle x} &\Rightarrow \quad \lim_{\triangle x\to 0}\frac{\triangle u}{\triangle x} = \lim_{\triangle x\to 0}\frac{g(x+\triangle x) - g(x)}{\triangle x} = g'(x)
\end{aligned}$$

Hence,

$$\begin{aligned}
\frac{dy}{dx} = \lim_{\triangle x\to 0}\frac{\triangle y}{\triangle x} &= \lim_{\triangle x\to 0}\Big(\frac{\triangle y}{\triangle u}\cdot\frac{\triangle u}{\triangle x}\Big) = \Big(\lim_{\triangle x\to 0}\frac{\triangle y}{\triangle u}\Big)\Big(\lim_{\triangle x\to 0}\frac{\triangle u}{\triangle x}\Big)\\
&= \Big(\lim_{\triangle u\to 0}\frac{\triangle y}{\triangle u}\Big)\Big(\lim_{\triangle x\to 0}\frac{\triangle u}{\triangle x}\Big) = f'(u)g'(x)\\
&= \frac{dy}{du}\frac{du}{dx} = f'\big(g(x)\big)g'(x)
\end{aligned}$$

Theorem 5.3.1. *(Chain Rule) Suppose that S and T are open intervals in $\mathbb{R}$, $g(x) : S \to T$, $f(u) : T \to \mathbb{R}$ and $a \in S$. If g is differentiable at a and f is differentiable at $g(a)$ then $f \circ g$ is differentiable at a and*

$$\big(f\circ g\big)'(a) = f'\big(g(a)\big)g'(a). \tag{5.1}$$

Proof. Since f is differentiable at $g(a)$,

$$\lim_{\triangle u\to 0}\frac{f\big(g(a)+\triangle u\big) - f\big(g(a)\big)}{\triangle u} = f'\big(g(a)\big).$$

Define a new function

$$\varepsilon(\triangle u) = \frac{f\big(g(a)+\triangle u\big) - f\big(g(a)\big)}{\triangle u} - f'\big(g(a)\big). \tag{5.2}$$

Then

$$f\big(g(a)+\triangle u\big) - f\big(g(a)\big) = \triangle u f'\big(g(a)\big) + \triangle u\varepsilon(\triangle u)$$

and (5.1) implies that $\varepsilon(\triangle u) \to 0$ as $\triangle u \to 0$. On the other hand, let $\triangle u = g(a+\triangle x) - g(a)$. Since g is differentiable at a, it is continuous there. Then $\triangle u \to 0$ as $\triangle x \to 0$. This implies that

$$\varepsilon(\triangle u) \to 0 \quad \text{as } \triangle x \to 0$$

Also, by (5.2)

$$f\big(g(a+\triangle x) - g(a)\big) = \big[g(a+\triangle x) - g(a)\big]f'\big(g(a)\big) + \big[g(a+\triangle x) - g(a)\big]\varepsilon(\triangle u).$$

Dividing by $\triangle x$,

$$\frac{f\big(g(a+\triangle x) - g(a)\big)}{\triangle x} = \frac{\big[g(a+\triangle x) - g(a)\big]}{\triangle x}f'\big(g(a)\big) + \frac{\big[g(a+\triangle x) - g(a)\big]}{\triangle x}\varepsilon(\triangle u)$$

Let $\triangle x \to 0$. We have

$$\begin{aligned}
&\lim_{\triangle x\to 0} \frac{f\big(g(a+\triangle x)-g(a)\big)}{\triangle x} \\
= &\lim_{\triangle x\to 0} \frac{\big[g(a+\triangle x)-g(a)\big]}{\triangle x} f'\big(g(a)\big) + \lim_{\triangle x\to 0} \frac{\big[g(a+\triangle x)-g(a)\big]}{\triangle x} \varepsilon(\triangle u) \\
= &f'\big(g(a)\big)g'(a) + g'(a)\cdot 0.
\end{aligned}$$

□

Note.

(1) Strictly speaking, the two functions $\big(f\circ g\big)'(x)$ and $f'\big(g(x)\big)g'(x)$ may not be equal. For example, $f(x)\equiv 0$ and $g(x)=|x|$. The domain of $\big(f\circ g\big)'(x)$ is $\mathbb{R}$ and the domain of $f'\big(g(x)\big)g'(x)$ is $\mathbb{R}\backslash\{0\}$. We don't worry about this if adding some conditions.

(2) Suppose that S and T are open intervals in $\mathbb{R}$, $g: S\to T$, $f: T\to\mathbb{R}$. If g is differentiable on S and f is differentiable on T, then

$$\big(f\circ g\big)'(x) = f'(g(x))g'(x).$$

Example 5.3.2.

(1) $h(x)=\sin x^2$.
Let $f(x)=\sin x$ and $g(x)=x^2$, then $h(x)=f\big(g(x)\big)$. Since $f'(x)=cosx$ and $g'(x)=2x$, we have

$$h'(x) = f'\big(g(x)\big)g'(x) = \cos x^2\cdot 2x$$

(2) $h(x)=\sin^2 x$.
Let $f(x)=x^2$ and $g(x)=\sin x$, then $h(x)=f\big(g(x)\big)$. Since $f'(x)=2x$ and $g'(x)=\cos x$,

$$h'(x) = f'\big(g(x)\big)g'(x) = 2\sin x\cos x.$$

(3) $h(x)=\begin{cases} x^2\sin\dfrac{1}{x}, & \text{for } x\neq 0\\ 0, & \text{for } x=0\end{cases}$

For $x\neq 0$, let $f(x)=\sin x$ and $g(x)=\dfrac{1}{x}$, then $\sin\dfrac{1}{x}=(f\circ g)(x)$. Since $f'(x)=\cos x$ and $g'(x)=-\dfrac{1}{x^2}$, we have $(\sin\dfrac{1}{x})'=(\cos\dfrac{1}{x})(-\dfrac{1}{x^2})$. Then

$$h'(x) = 2x\sin\frac{1}{x} + x^2\cos\frac{1}{x}\cdot(-\frac{1}{x^2}) = 2x\sin\frac{1}{x} - \cos\frac{1}{x}$$

At $x=0$

$$\lim_{k\to 0}\frac{h(k)-h(0)}{k} = \lim_{k\to 0}\frac{k^2\sin\frac{1}{k}-0}{k} = 0.$$

Thus, $h'(0)=0$.

■ The power rule combined with the chain rule

Theorem 5.3.3. *Suppose that g is differentiable at a and $n \in \mathbb{Q}$. If $f(x) = [g(x)]^n$, then*

$$f'(a) = n[g(a)]^{n-1} g'(a).$$

Proof. Let $h(x) = x^n$, then $f(x) = h\big(g(x)\big)$. Since $h(x)$ is differentiable everywhere and $h'(x) = nx^{n-1}$, we have

$$f'(a) = h'\big(g(a)\big) g'(a) = n[g(a)]^{n-1} g'(a).$$

□

Example 5.3.4. (1) Let $y = (x^3 + 2x + 1)^{30}$, then $\dfrac{dy}{dx} = 30(x^3 + 2x + 1)^{29}(3x^2 + 2)$.

(2) Let $f(x) = \dfrac{1}{\sqrt[3]{x^2 + x + 1}} = (x^2 + x + 1)^{-1/3}$, then $f'(x) = -\frac{1}{3}(x^2 + x + 1)^{-\frac{4}{3}} \cdot (2x + 1)$.

(3) Let $g(t) = \left(\dfrac{t-2}{2t+1}\right)^9$, then $g'(t) = 9\left(\dfrac{t-2}{2t+1}\right)^8 \cdot \dfrac{(2t+1) - 2(t-2)}{(2t+1)^2}$.

Corollary 5.3.5. *Suppose that $h(x)$ is differentiable at a, $g(x)$ is differentiable at $h(a)$ and $f(x)$ is differentiable at $g\big(h(a)\big)$. If $k(x) = \big(f \circ g \circ h\big)(x)$, then*

$$k'(a) = f'\big(g(h(a))\big) g'\big(h(a)\big) h'(a).$$

In Leibniz notation, $y = f(u)$, $u = g(w)$ and $w = h(x)$,

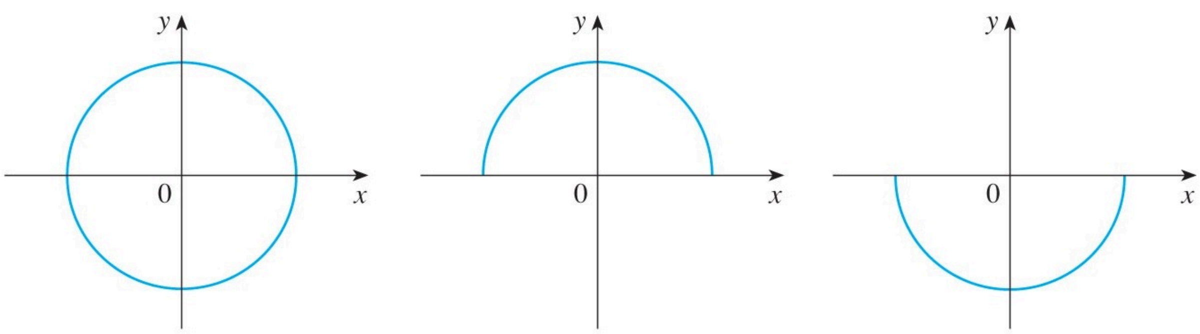
$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dw}\frac{dw}{dx}.$$

Example 5.3.6. Let $f(x) = \sin^2(2x^2 + 1) = \big[\sin(2x^2 + 1)\big]$, then

$$f'(x) = 2\sin(2x^2 + 1) \cdot \cos(2x^2 + 1) \cdot 4x.$$

5.4 Implicit Differentiation

Some functions have explicit forms. For example, $y = \sqrt{x}$, $y = \sin x^2$ etc. But not all functions can be described by expressing one variable explicitly in terms of another variable. Some functions may have relations(equations) between x and y. For example, $x^2 + y^2 = 1$.



(a) $x^2 + y^2 = 25$ (b) $f(x) = \sqrt{25 - x^2}$ (c) $g(x) = -\sqrt{25 - x^2}$

Moreover, for some equations such as $x^3y^2 + \sin(xy^2) + \dfrac{y}{x^2+1} = 1$, it is difficult to express y as a function of x (locally).

Question: How to find the $\dfrac{dy}{dx}$ at a given point?

■ **Implicit Differentiation**

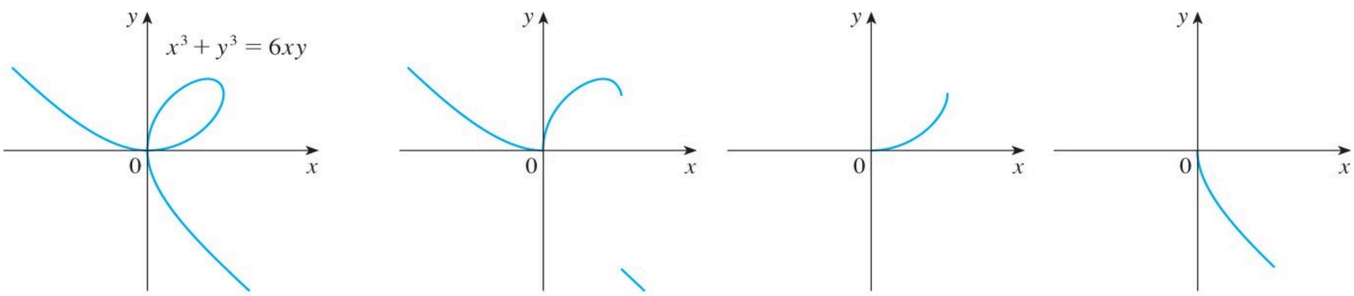
If x and y have a "*relation*" (satisfy an equation), we can regard y as a function of x (locally). Take "$\frac{d}{dx}$" on the both sides of the equation.

Example 5.4.1. Let $x^3 + y^3 = 6xy$.

(a) Find $\dfrac{dy}{dx}$.

(b) Find the equation of the tangent line of the curve at $(3,3)$.

(c) Find the points(s) on the curve such that $\dfrac{dy}{dx} = 0$.



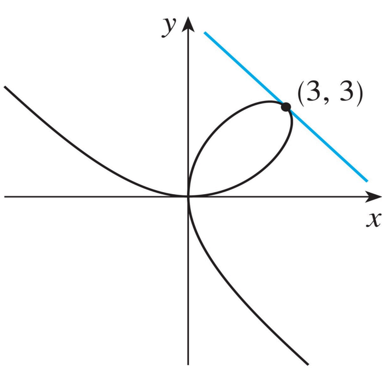
The folium of Descartes

Graphs of three functions defined by the folium of Descartes

Proof. (a)

$$
\begin{aligned}
& \frac{d}{dx}\left(x^3+y^3\right) = \frac{d}{dx}(6xy) \\
\Rightarrow\quad & 3x^2 + 3y^2\frac{dy}{dx} = 6y + 6x\frac{dy}{dx} \\
\Rightarrow\quad & \frac{dy}{dx} = \frac{3x^2-6y}{6x-3y^2} = \frac{x^2-2y}{2x-y^2}
\end{aligned}
$$

(b)



At $(3,3)$, $\left.\dfrac{dy}{dx}\right|_{(x,y)=(3,3)} = \dfrac{3}{-3} = -1$. The equation of the tangent line is

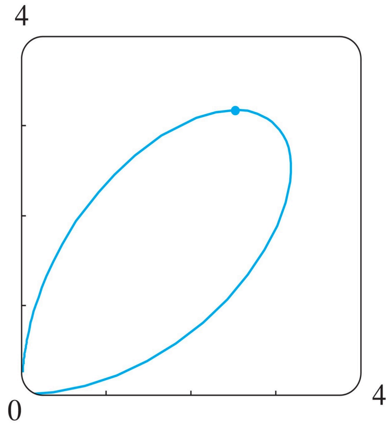
$$y - 3 = -(x-3).$$

(c) $\dfrac{dy}{dx} = \dfrac{x^2 - 2y}{2x - y^2} = 0 \quad \Rightarrow x^2 - 2y = 0 \quad \Rightarrow y = \dfrac{x^2}{2}.$

On the curve,

$$x^3 + \frac{x^6}{8} = 3x^2 \Rightarrow x^3(\frac{1}{8}x^3 - 2) = 0 \Rightarrow x = 0 \text{ or } 2^{\frac{4}{3}}.$$

At the point $(2^{4/3}, 2^{5/3})$, the curve has horizontal tangent line.
Question: How about at $(0, 0)$? There are two tangent lines.

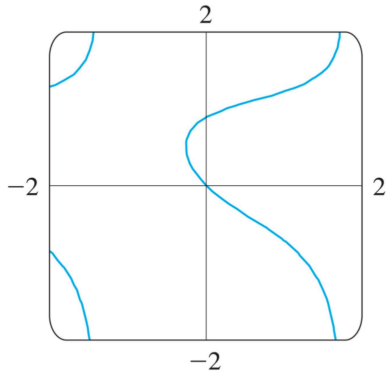


□

Example 5.4.2. $\sin(x + y) = y^2 \cos x$. Find $\dfrac{dy}{dx}$.

Proof.

$$\begin{aligned}
& \frac{d}{dx}\left[\sin(x+y)\right] = \frac{d}{dx}\left[y^2 \cos x\right] \\
\Rightarrow \quad & \cos(x+y) \cdot \left(1 + \frac{dy}{dx}\right) = 2y\frac{dy}{dx}\cos x - y^2 \sin x \\
\Rightarrow \quad & \left(\cos(x+y) - 2y\cos x\right)\frac{dy}{dx} = -\cos(x+y) - y^2 \sin x \\
\Rightarrow \quad & \frac{dy}{dx} = -\frac{\cos(x+y) + y^2 \sin x}{\cos(x+y) - 2y\cos x}
\end{aligned}$$



□

Example 5.4.3. $x^4 + y^4 = 16$. Find $\dfrac{dy}{dx}$ and $\dfrac{d^2y}{dx^2}$.

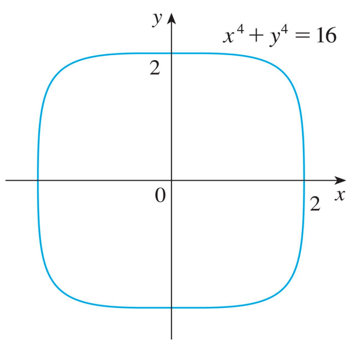
Proof.

$$\frac{d}{dx}\left[x^4 + y^4\right] = \frac{d}{dx}(16) \Rightarrow 4x^3 + 4y^3\frac{dy}{dx} = 0 \Rightarrow \frac{dy}{dx} = -\frac{x^3}{y^3}.$$

To find $\dfrac{d^2y}{dx^2}$.

Method 1:

$$\begin{aligned}
\frac{d}{dx}\left(\frac{dy}{dx}\right) &= \frac{d}{dx}\left(-\frac{x^3}{y^3}\right) = -\frac{3x^2y^3 - 3x^3y^2\frac{dy}{dx}}{(y^3)^2} = -\frac{3x^2y^3 + \frac{3x^6}{y}}{y^6} \\
&= -\frac{3x^2y^4 + 3x^6}{y^7} = \frac{3x^2(y^4 + x^4)}{y^7} = -\frac{48x^2}{y^7}
\end{aligned}$$



Method 2:

$$\begin{aligned}
& \frac{d}{dx}\left(4x^3 + 4y^3\frac{dy}{dx}\right) = \frac{d}{dx}(0) \\
\Rightarrow \quad & 12x^2 + 12y^2\left(\frac{dy}{dx}\right)\left(\frac{dy}{dx}\right) + 4y^3\frac{d^2y}{dx^2} = 0 \Rightarrow \frac{d^2y}{dx^2} = -\frac{48x^2}{y^7}.
\end{aligned}$$

□

Example 5.4.4. $y = x^{\frac{p}{q}}$ where $p, q \in \mathbb{Z}$. Find $\dfrac{dy}{dx}$.

Proof. Consider $y^q = x^p$.

$$\begin{aligned} & \frac{d}{dx}(y^q) = \frac{d}{dx}(x^p) \\ \Rightarrow \quad & qy^{q-1}\frac{dy}{dx} = px^{p-1} \\ \Rightarrow \quad & \frac{dy}{dx} = \frac{p}{q}\,\frac{x^{p-1}}{y^{q-1}} = \frac{p}{q}\,\frac{x^{p-1}}{x^{p-p/q}} = \frac{p}{q}\,x^{\frac{p}{q}-1} \end{aligned}$$

□

Remark. In advanced calculus, we will study the Implicit Function Theorem.

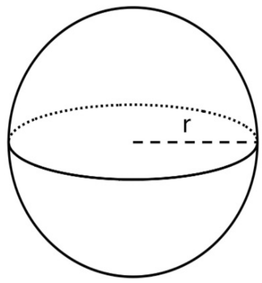
5.5 Related Rates*

Recall: (Chain Rule) Let $y = y(x)$ and $x = x(t)$. Then

$$\frac{dy}{dt} = \frac{dy}{dx}\frac{dx}{dt} \quad \Rightarrow \quad \frac{dy}{dx} = \frac{dy}{dt}\Big/\frac{dx}{dt} \quad \text{and} \quad \frac{dx}{dt} = \frac{dy}{dt}\Big/\frac{dy}{dx}.$$

Example 5.5.1.

Air is being pumped into a spherical balloon. The radius is increasing at the rate of 2 cm per second (2cm/s). What rate is the volume increasing when the radius is 5 cm?



Proof. Let V be the volume of the balloon with radius r. Then

$$V(r) = \frac{4}{3}\pi r^3.$$

Our goal is to find $\left.\dfrac{dV}{dt}\right|_{r=5}$ under the condition $\dfrac{dr}{dt} = 2$. By the volume formula,

$$\frac{dV}{dr} = 4\pi r^2.$$

Then

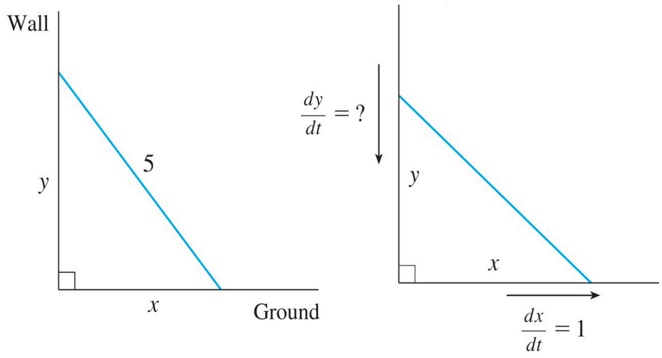
$$\frac{dV}{dt} = \frac{dV}{dr}\frac{dr}{dt} = 2 \cdot 4\pi r^2 = 8\pi r^2.$$

Hence, $\left.\dfrac{dV}{dt}\right|_{r=5} = 200\pi\ (cm^3/s)$. □

*The reference and examples in this section are from *Calculus, J. Stewart 8th Ed.*

Example 5.5.2.

A ladder 5 m long rests against a vertical wall. If the bottom of the ladder slides away from the wall at a rate of 1 m/s, how fast is the top of the ladder sliding down the wall when the bottom of the ladder is 3 m from the wall?



Proof. Let $x(t)$ be the distance from the bottom of the ladder to the wall. Let $y(t)$ be the distance from the top of the ladder to the ground. Then

$$x^2 + y^2 = 25.$$

By the implicit differentiation,

$$\frac{d}{dx}(x^2 + y^2) = \frac{d}{dx}(25) \quad \Rightarrow \quad 2x + 2y\frac{dy}{dx} = 0 \quad \Rightarrow \quad \frac{dy}{dx} = -\frac{x}{y}.$$

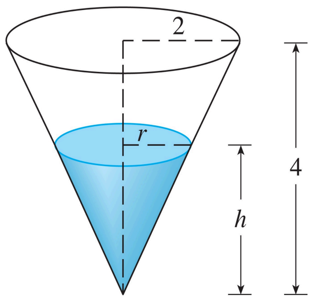
Let $x = 3$, then $y = 4$ and hence $\left.\frac{dy}{dx}\right|_{x=3} = -\frac{3}{4}$. Therefore, $\frac{dy}{dt} = \frac{dy}{dx}\frac{dx}{dt}$. The top of the ladder is sliding down the wall at a rate

$$\left.\frac{dy}{dt}\right|_{x=3} = \left.\frac{dy}{dx}\right|_{x=3} \cdot \left.\frac{dx}{dt}\right|_{x=3} = -\frac{3}{4} \cdot 1 = -\frac{3}{4}\ (m/s).$$

□

Example 5.5.3.

The water is being pumped into the tank at a rate of 2 m^3/min. find the rate at which the water level is rising when the water is 3 m deep.



Proof. Let h be the height of water level and r be the radius of the surface of the water at time t. Let $V(r, h)$ be the volume of the water when the water level is h. From the similar triangle argument, $r = \frac{1}{2}h$. Then

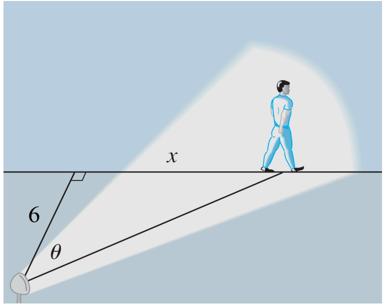
$$V(r, h) = V(h) = \frac{1}{3}\pi r^2 h = \frac{1}{3}\pi(\frac{1}{2}h)^2 h = \frac{1}{12}\pi h^3.$$

Hence, $\frac{dV}{dh} = \frac{1}{4}\pi h^2$ and $\frac{dV}{dt} = \frac{dV}{dh} \cdot \frac{dh}{dt}$. We have

$$\frac{dh}{dt} = \frac{dV}{dt} \Big/ \frac{dV}{dh} = \frac{2}{\frac{1}{4}\pi h^2} = \frac{8}{\pi h^2}.$$

The rate at which the water level is rising when $h = 3$ is $\left.\frac{dh}{dt}\right|_{h=3} = \frac{8}{9\pi}$ (m/min). □

Example 5.5.4.
A man walks along a straight path at a speed of 1.5 m/s. A searchlight is located on the ground 6 cm from the path and is kept focused on the man. At what rate is the searchlight rotating when the man is 8 m from the point on the path closest to the searchlight?



Proof. Let x be the distance from the man to the point on the path closest to the searchlight. Then $\dfrac{dx}{dt} = 1.5$. Let θ be the angle between the beam of the searchlight and the perpendicular to the path. Then

$$x = 6\tan\theta \quad \Rightarrow \quad \frac{dx}{d\theta} = 6\sec^2\theta.$$

When $x = 8$, $\cos\theta = \dfrac{6}{10} = \dfrac{3}{5}$ and

$$\frac{dx}{dt} = \frac{dx}{d\theta}\frac{d\theta}{dt} \quad \Rightarrow \frac{d\theta}{dt} = \frac{dx}{dt}\Big/\frac{dx}{d\theta} = \frac{1.5}{6\sec^2\theta} = \frac{1}{4}\cos^2\theta.$$

Hence, the rate of the searchlingt rotating is $\left.\dfrac{d\theta}{dt}\right|_{x=8} = \dfrac{1}{4}\cdot\left(\dfrac{3}{5}\right)^2 = \dfrac{9}{100}\ (rad/s)$. □

■ **Strategy**

(i) Read the problem carefully.

(ii) Draw a diagram.

(iii) Introduce notation.

(iv) Express the given information and the required rate in terms of derivatives.

(v) Write an equation that relates the various quantities of the problem.

(vi) Use the chain rule.

(vii) Substitute the given information into the resulting equation and solve for the unknown rate.

5.6 Linear Approximation and Differentials†

Motivation: A curve lies very very close to its tangnet line near the point of tangency. To evaluate the value of a function f near a point a, it is sometimes difficult to compute directly. Then we may use the tangent line at $\big(a, f(a)\big)$ as an approximation to the curve $y = f(x)$.

❑ **Linear Approximation**

†The reference and examples in this section are from *Calculus, J. Stewart 8th Ed.*

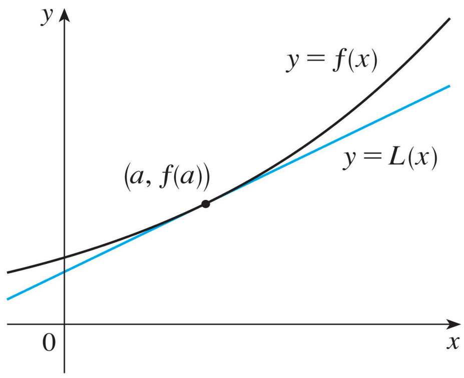
The equation of the tangent line of $y = f(x)$ at $(a, f(a))$ is

$$y = f(a) + f'(a)(x - a)$$

The approximation is

$$f(x) \approx f(a) + f'(a)(x - a)$$

if x is close to a.



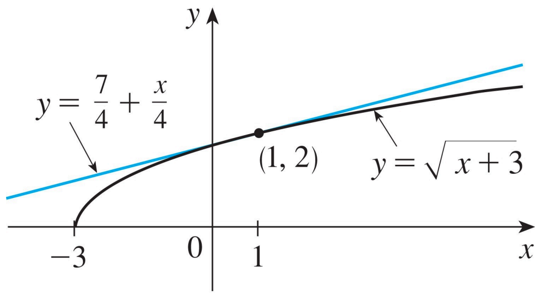
Remark. This approximation is called the "*linear approximation*" or "*tangent line approximation*" of f at a. The linear function

$$L(x) = f(a) + f'(a)(x - a)$$

is called the "*linearization*" of f at a.

Example 5.6.1.

Find the linearization of the function $f(x) = \sqrt{x+3}$ at $a = 1$ and use it to approximate the number $\sqrt{3.98}$ and $\sqrt{4.05}$.



Proof. Since $f'(x) = \dfrac{1}{2\sqrt{x+3}}$, the linearization of f at $a = 1$ is

$$L(x) = f(1) + f'(1)(x - 1) = 2 + \frac{1}{4}(x - 1) = \frac{7}{4} + \frac{x}{4}.$$

Then

$$\begin{aligned} \sqrt{3.98} &= f(0.98) \approx L(0.98) = 2 + \frac{1}{4}(0.98 - 1) = 1.995 \\ \sqrt{4.05} &= f(1.05) \approx L(1.05) = 2 + \frac{1}{4}(1.05 - 1) = 3.0125 \end{aligned}$$

□

Question: How good is the approximation?

Example 5.6.2.

For what values of x is the linear approximation

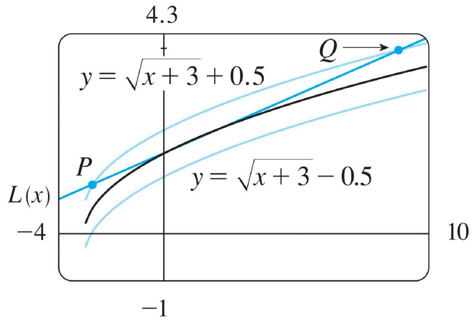
$$\sqrt{x+3} \approx \frac{7}{4} + \frac{x}{4}.$$

accurate to within 0.5?

Proof. Consider

$$\left|\sqrt{x+3}-\left(\frac{7}{4}+\frac{x}{4}\right)\right|<0.5 \quad\iff\quad \sqrt{x+3}-0.5<\frac{7}{4}+\frac{x}{4}<\sqrt{x+3}+0.5$$

See the graph to compute the points P and Q which intersect the curves $y=\sqrt{x+3}\pm 0.5$.



□

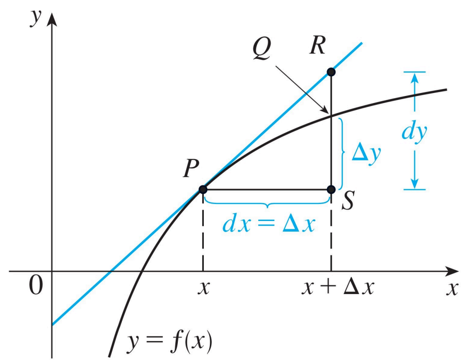
❑ Differentials

The idea behind linear approximation is formulated in the terminology and notation of "*differentials*".

Let f be a differentiable function and $y = f(x)$. Consider the change of x, $\triangle x$ and the corresponding change of y, $\triangle y$. We have

$$\triangle y = f(x+\triangle x)-f(x)\approx f'(x)\triangle x.$$

Let $\triangle x \to 0$, then $dy = f'(x)dx$. We regard the "*differential*" dx as an independent variable and "*differential*" dy as a dependent variable.



Let $dx = \triangle x$. As $\triangle x$ is sufficiently small, $dy \approx \triangle y$. Then

$$f(a+\triangle x) = f(a+dx) = f(a)+\triangle y \approx f(a)+dy = f(a)+f'(a)dx.$$

Example 5.6.3. Let $f(x) = x^3+x^2-2x+1$. Find $\triangle y$ and dy when x changes (a) from 2 to 2.05; (b) form 2 to 2.01.

Proof. $f'(x) = 3x^2+2x-2$ and $f'(2) = 144$.

(a) $\triangle x = dx = 2.05-2 = 0.05$

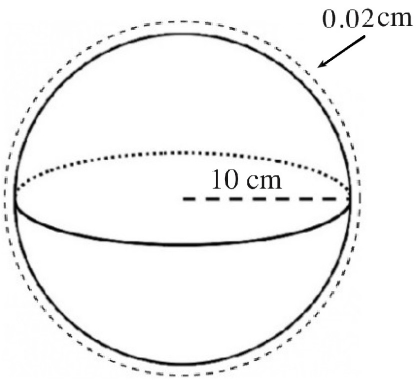
$$\triangle y = f(2.05)-f(2) = 0.717625 \quad\text{and}\quad dy = f'(2)dx = 14\cdot 0.05 = 0.7$$

(b) $\triangle x = dx = 2.01-2 = 0.01$

$$\triangle y = f(2.01)-f(2) = 0.140701 \quad\text{and}\quad dy = f'(2)dx = 14\cdot 0.01 = 0.14$$

□

Example 5.6.4.



A metal sphere with a radius of 10 *cm* is to be covered by a 0.02 *cm* coating of silver. Approximately how much silver will be required?

Proof. Let $V(r)$ be the volume of the sphere with radius r. Then $V(r) = \dfrac{4}{3}\pi r^3$ and $dV = 4\pi r^2 dr$.

$V(10.02) - V(10) = \triangle V \approx dV = 4\pi \cdot (10)^2 \cdot 0.02 = 8\pi\,(cm^3)$.

□

Remark. (Relative Error) Dividing the error by the total volume

$$\frac{\triangle V}{V} \approx \frac{dV}{V} = \frac{4\pi r^2 dr}{\frac{4}{3}\pi r^3} = 3\frac{dr}{r}$$

Note. The relative error in the volume is about three times the relative error in the radius.

Chapter 6

Applications of Differentiation

Some of the most important applications of differential calculus are optimal problems. In order to obtain the optimization, we can usually reduce these problems to finding the maximum or minimum values of a function.

6.1 Maximum and Minimum Values

Recall: In Chapter 1, we have introduced the maximum (or minimum) value and the least upper bound (or greatest lower bound) of a function. In this section, we will study the extreme values of a function more deeply.

Definition 6.1.1. Let $f : D \to \mathbb{R}$ and $x_0 \in D$. We say that

(a) the number x_0 is an "*absolute maximum number (or point)*" for f on D if

$$f(x_0) \geq f(x) \quad \text{for all } x \in D$$

and the value of f at x_0 is called the "*absolute*" maximum value of f on D.

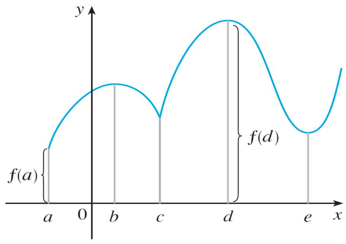
(b) the number x_0 is a "*local maximu number (or point)*" for f on D if there exists $\delta > 0$ such that

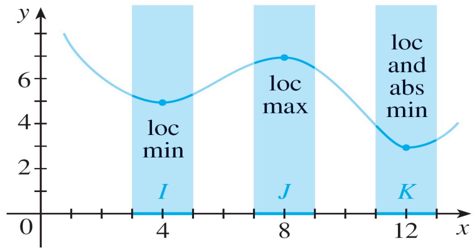
$$f(x_0) \geq f(x) \quad \text{for all } x \in D \cap (x_0 - \delta, x_0 + \delta)$$

i.e. x_0 is a maximum number for f on $D \cap (x_0 - \delta, x_0 + \delta)$. The value of f at x_0 is called the "*local*" maximum value of f on D.

(c) We can also define the "*absolute (or lcoal) minimun number*" and the "*absolute (or local) minimum value*" of f on D by replacing the inequalty "$\geq$" by "$\leq$".

(d) The maximum and minimum values of f are called the "*extreme values*" of f on D.





Remark.

(i) An absolute maximum or minimum value is sometimes called a "*global*" maximum or mumimum value.

(ii) An absolute maximum point for f on D is also a local maximum point for f on D.

(iii) The absolute maximum (or minimum) value of f on D is unique. In constast, a function may have several or infinitely many maximum (or minimum) points.

Exercise. If $A \subseteq B$, and $\max\limits_{x\in A} f(x)$ and $\max\limits_{x\in B} f(x)$ exist, then

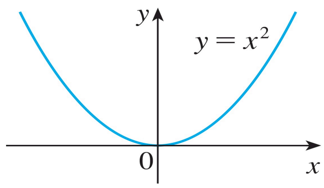
$$\max_{x\in A} f(x) \le \max_{x\in B} f(x)$$

Similarly, $\min\limits_{x\in A} f(x) \ge \min\limits_{x\in B} f(x)$, $\sup\limits_{x\in A} f(x) \le \sup\limits_{x\in B} f(x)$, $\inf\limits_{x\in A} f(x) \ge \inf\limits_{x\in B} f(x)$ if the above quantities exist.

Example 6.1.2.

$$f(x) = x^2 \quad \text{on} \quad \mathbb{R}.$$

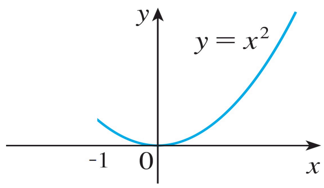
The local and absolute minimum value is 0 but there is no maximum value.



Remark. The existence of absolute (or local) maximum (or minimum) values not only depends on functions, but also depends on the domains.

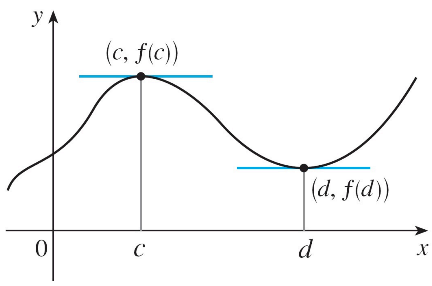
$$f(x) = x^2 \quad \text{on} \quad [-1, 3].$$

The local and absolute minimum value is 0 and the local minimum is $f(-1) = 1$ and $f(3) = 9$.



■ Extreme Points and Derivatives

We can observe that the tangent lines (if they exist) of the graph of f at the maximum and minimum points are horizontal. Hence the slopes are 0.



Theorem 6.1.3. *(Fermat's Theorem) Let f be a function defined on (a, b). If x_0 is a local maximum or local minimum point for f on (a, b) and f is differentiable at x_0, then $f'(x_0) = 0$.*

Proof. W.L.O.G, we assume that x_0 is a local maximum point for f on (a, b). Then there exsits $\delta > 0$ such that $(x_0 - \delta, x_0 + \delta) \subseteq (a, b)$ and

$$f(x_0) \geq f(x) \quad \text{for all } x \in (x_0 - \delta, x_0 + \delta).$$

For $0 < h < \delta$, $f(x_0 + h) - f(x_0) \leq 0$. Hence,

$$\frac{f(x_0 + h) - f(x_0)}{h} \leq 0. \tag{6.1}$$

Similarly, for $-\delta < h < 0$, we have $f(x_0 + h) - f(x_0) \leq 0$ and hence

$$\frac{f(x_0 + h) - f(x_0)}{h} \geq 0 \quad \text{(since } h \text{ is negative.)} \tag{6.2}$$

Since f is differentiable at x_0,

$$f'(x_0) = \lim_{h\to 0} \frac{f(x_0 + h) - f(x_0)}{h} = \lim_{h\to 0^+} \frac{f(x_0 + h) - f(x_0)}{h} = \lim_{h\to 0^-} \frac{f(x_0 + h) - f(x_0)}{h}$$

By (6.1) and (6.2),

$$f'(x_0) = \lim_{h\to 0^+} \frac{f(x_0 + h) - f(x_0)}{h} \leq 0 \quad \text{and} \quad f'(x_0) = \lim_{h\to 0^-} \frac{f(x_0 + h) - f(x_0)}{h} \geq 0.$$

Therefore,

$$f'(x_0) = \lim_{h\to 0} \frac{f(x_0 + h) - f(x_0)}{h} = 0.$$

□

Remark. The converse of the theorem is false. That is, it is possible that there exists a function f with $f'(x_0) = 0$ but f has not maximum nor minimum at x_0. For example, $f(x) = x^3$ at 0.

Corollary 6.1.4. *Let f be a function defined on (a, b) and x_0 is an extreme point. Then either $f'(x_0) = 0$ or f is not differentiable at x_0.*

Definition 6.1.5. Let f be a function defined on (a, b). We say that the point $x_0 \in (a, b)$ is a critical number (point) of f if either $f'(x_0) = 0$ or f is not differentiable at x_0. We call $f(x_0)$ a critical value of f.

Remark. If f has a (local) maximum or minimum at x_0, then x_0 is a critical number of f. But not every critical number gives rise to a maximum or minimum. For example $f(x) = x^3$ at $x = 0$.

■ Global Extreme Values for f on $[a, b]$

Recall: Theorem 4.2.5 says that a continuous function defined on $[a, b]$ must have global maximum and minimum. The Fermat's Theorem gives a method to find the extreme values of a continuous function.

• The closed interval method

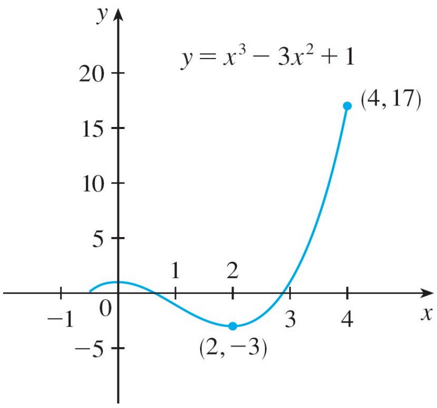
(1) Find all critical numbers of f in (a, b).

(2) Find the values of f at those ciritical points and endpoints.

(3) The largest value in the above steps is the absolute maximum value and the smallest value is the absolute minimum value.

Example 6.1.6.

Let $f(x) = x^3 - 3x^2 + 1$ defined on $[-\frac{1}{2}, 4]$. Find the absolute maximum and minimum values of f.



Proof. The derivative of f is $f'(x) = 3x(x-2)$. Since f is a polynomial, it is differentiable everywhere and the critical numbers of f are 0 and 2. The values of f at critical numbers and endpoints are

$$f(0) = 1, \quad f(2) = -3, \quad f(-\frac{1}{2}) = \frac{1}{8}, \quad f(4) = 17$$

Hence, f has the absolute maximum value $f(4) = 17$ and the absolute minimum value $f(2) = -3$. □

Theorem 6.1.7. *If f is differentiable at x_0 and $f'(x_0) > 0$, then there exists a number $\delta > 0$ such that*

$$f(x) < f(x_0) \quad \textit{for all } x \in (x_0 - \delta, x_0)$$

and

$$f(x) > f(x_0) \quad \textit{for all } x \in (x_0, x_0 + \delta)$$

Proof. Since $f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} > 0$, for $\varepsilon = \frac{1}{2} f'(x_0)$, there exists $\delta > 0$ such that

$$\left| \frac{f(x) - f(x_0)}{x - x_0} - f'(x_0) \right| < \varepsilon = \frac{1}{2} f'(x_0)$$

whenever $0 < |x - x_0| < \delta$. Then

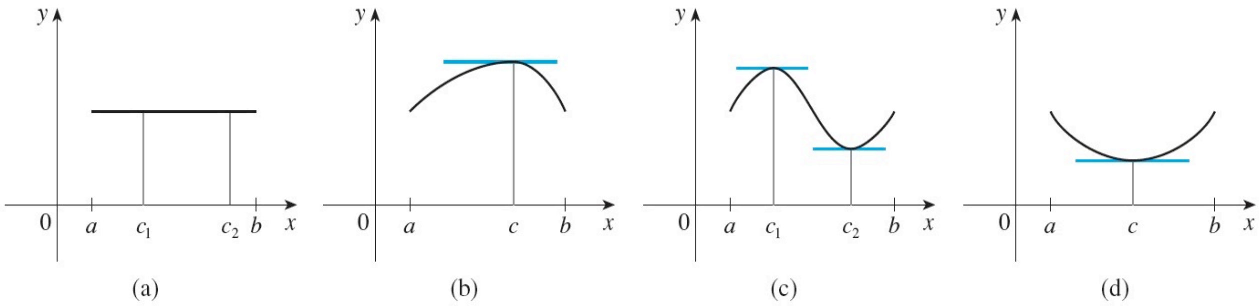
$$\frac{f(x) - f(x_0)}{x - x_0} > f'(x_0) - \frac{1}{2} f'(x_0) = \frac{1}{2} f'(x_0) > 0$$

If $x \in (x_0 - \delta, x_0)$, then $x - x_0 < 0$ and

$$f(x) - f(x_0) < \frac{1}{2} f'(x_0)(x - x_0) < 0.$$

Similarly, if $x \in (x_0, x_0 + \delta)$, then $x - x_0 > 0$ and $f(x) - f(x_0) > \frac{1}{2} f'(x_0)(x - x_0) > 0$. □

6.2 Rolle's Theorem and Mean Value Theorem



Observe a continuous function defined on $[a, b]$ with $f(a) = f(b)$. There must be a number $c \in (a, b)$ which is either a maximum point or a minimum point. Then $f'(c) = 0$.

Theorem 6.2.1. *(Rolle's Theorem) Let f be a function that satisfies*

(i) f is continuous on $[a, b]$,

(ii) f is differentiable on (a, b) and

(iii) $f(a) = f(b)$.

Then there is a number $c \in (a, b)$ such that $f'(c) = 0$.

Proof. If f is a constant function on $[a, b]$ (i.e. $f(x) = f(a) = f(b)$ for all $x \in [a, b]$), then $f'(x) = 0$ for all $x \in (a, b)$.

If $f(x)$ is not a constant function on $[a, b]$, then there exists a number $x_0 \in (a, b)$ such that $f(x_0) \neq f(a)$. W.L.O.G, say $f(x_0) > f(a)$. Since f is continuous on $[a, b]$, by the extreme value theorem, there exists $c \in [a, b]$ such that $f(c) = \max_{x\in[a,b]} f(x)$. Then $f(c) \geq f(x_0) > f(a) = f(b)$ and hence $c \neq a$ and $c \neq b$ (i.e. $c \in (a, b)$).

Since c is a maximum point of f on (a, b) and f is differentiable at c, by Fermat's theorem, $f'(c) = 0$. □

Example 6.2.2. If $s = f(t)$ is a differentiable function which represents the position of an object. Suppose that the object locates at the same position at time a and b. That is, $f(a) = f(b)$. Then three exists some time $c \in (a, b)$ such that the velocity is 0 at time c (i.e. $f'(c) = 0$).

Example 6.2.3. Prove that $3x^3 + 2x - 1 = 0$ has exactly one solution.

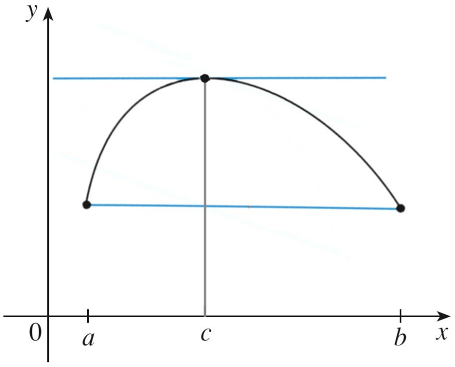
Proof. Let $f(x) = 3x^3 + 2x - 1$. To prove that there exists exactly one number c such that $f(c) = 0$.

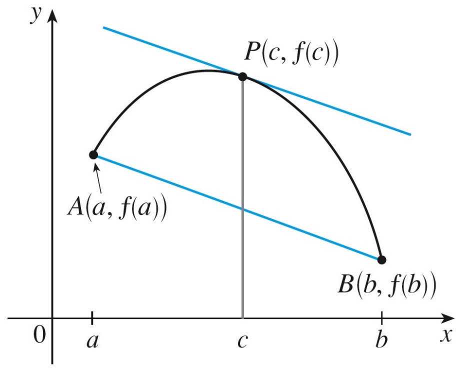
(i) (Existence: at least one root) By Intermediate Value Theorem (exercise!)

(ii) (Uniqueness: at most one root)
Assuem that there are two distinct numbers a and b such that $f(a) = f(b) = 0$. Since $f(x)$ is a polynomial function, it is continuous on $[a, b]$ and differentiable on (a, b). By the Rolle's theorem, there exists a number $c \in (a, b)$ such that $f'(c) = 0$. But $f'(x) = 3x^2 + 1$ for every x. It contradicts the conclusion of Rolle's theorem. Hence, f cannot have two or more roots.

□

The following theorem is a generalized result of the Rolle's theorem.





Theorem 6.2.4. *(Mean Value Theorem) Let f be a function that satisfies*

(i) f is continuous on $[a, b]$ and

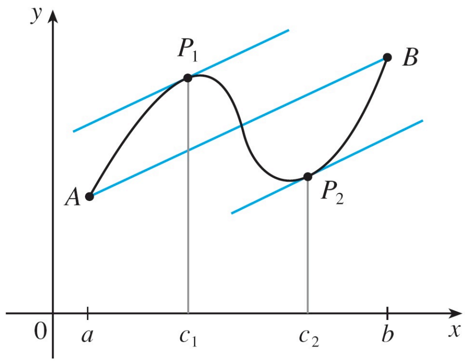
(ii) f is differentiable on (a, b).

Then there exists a number $c \in (a, b)$ such that

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

or

$$f(b) - f(a) = f'(c)(b - a)$$



Proof. Let

$$h(x) = f(x) - \left(f(a) + \frac{f(b) - f(a)}{b - a}(x - a)\right).$$

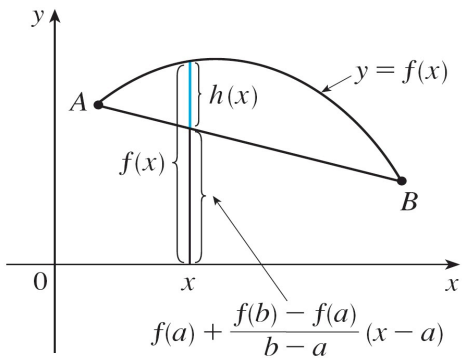
Since f is continuous on $[a, b]$ and differentiable on (a, b), so is $h(x)$.

Moreover, $h(a) = 0 = h(b)$ and

$$h'(x) = f'(x) - \frac{f(b) - f(a)}{b - a}.$$

Then, by the Rolle's theorem, there exists a number $c \in (a, b)$ such that $h'(c) = 0$. Hence,

$$f'(c) = h'(c) + \frac{f(b) - f(a)}{b - a} = \frac{f(b) - f(a)}{b - a}.$$



□

Corollary 6.2.5. *If f is continuous on $[a, b]$ and $f'(x) = 0$ for all $x \in (a, b)$, then f is a constant function on $[a, b]$.*

Proof. Let c and d be any two points in $[a, b]$. It sufficies to show that $f(c) = f(d)$.

Since f is continuous on $[a, b]$ and differentiable on (a, b), it also continuous on $[c, d]$ and differentiable on (c, d). Then, by the mean value theorem, there exists $\alpha \in (c, d)$ such that $f(c) - f(d) = f'(\alpha)(c - d) = 0$ since $f'(\alpha) = 0$. Therefore, $f(c) = f(d)$.

That c and d are arbitrary two points in $[a, b]$ implies $f(x)$ is a constant function on $[a, b]$. □

Corollary 6.2.6. *If f is differentiable on (a, b) and $f'(x) = 0$ for all $x \in (a, b)$, then f is a constant function on (a, b).*

Proof. (Skip) □

Note. The difference condition from the above corollary is that f is no longer continuous on $[a, b]$.

Corollary 6.2.7. *If f and g are continuous on $[a, b]$ and $f'(x) = g'(x)$ for all $x \in (a, b)$, then*

$$f(x) = g(x) + C \quad \textit{for some constant } C \textit{ and for all } x \in [a, b].$$

Proof. Let $h(x) = f(x) - g(x)$. Since f and g are continuous on $[a, b]$ and $f'(x) = g'(x)$ for all $x \in (a, b)$, h is continuous on $[a, b]$ and $h'(x) = 0$ for all $x \in (a, b)$. Then $h(x)$ is a constant function on $[a, b]$. Hence,

$$h(x) = h(a) = f(a) - g(a).$$

Then

$$f(x) = g(x) + \big(f(a) - g(a)\big).$$

□

Note. $f(x) = \dfrac{x}{|x|} = \begin{cases} 1 & \text{if } x > 0 \\ -1 & \text{if } x < 0 \end{cases}$ is not a constant function. But $f'(x) = 0$ for all $x \in$ *Dom*(f). It is because *Dom*(f) is not an interval. However, f is constant on $(-\infty, 0)$ and also on $(0, \infty)$.

Remark.

(i) If f is constant on (a, b), then $f'(x) = 0$ for all $x \in (a, b)$.

(ii) If $f(x) = g(x) + C$ on (a, b), and either $f'(x)$ or $g'(x)$ exists for all $x \in (a, b)$, then

$$f'(x) = g'(x) \quad \text{for all } x \in (a, b)$$

Theorem 6.2.8. *("Cauchy's Mean Value Theorem" or "Generalized Mean Value Theorem") Let f and g be two functions which are continuous on $[a, b]$ and differentiable on (a, b). Then there exists a number $c \in (a, b)$ such that*

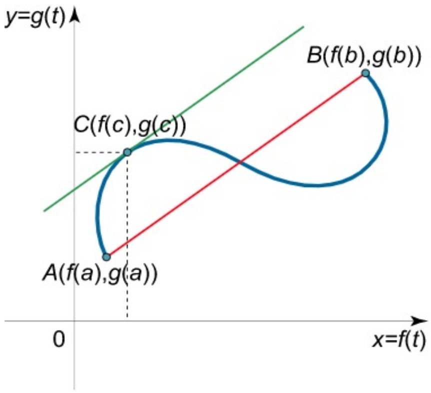
$$f'(c)\big[g(b) - g(a)\big] = g'(c)\big[f(b) - f(a)\big]. \tag{6.3}$$

Proof. Define $k(x) = f(x)\big[g(b) - g(a)\big] - g(x)\big[f(b) - f(a)\big]$. Then k is continuous on $[a, b]$ and differentiable on (a, b). Moreover, $k(a) = f(a)g(b) - f(b)g(a) = k(b)$. By the Rolle's Theorem, there exists a number $c \in (a, b)$ such that $k'(c) = 0$ and this implies (6.3). □

Remark. The generalized mean value theorem has a geometrical interpretation similar to that of the mean value theorem. Suppose that a smooth curve C can be represented as a parametric equation $(f(t), g(t))$ for $a \leq t \leq b$. There exists a tangnet line at $t = c$ whose slope is equal to the secant line connecting $(f(a), g(a))$ and $(f(b), g(b))$. The slope of the tangent line is

$$m = \frac{f(b) - f(a)}{g(b) - g(a)}.$$

Furthermore, the slope of the tangent line to the curve at any point $t = t_0$ is $\dfrac{f'(t_0)}{g'(t_0)}$.



6.3 How Derivatives Affect the Shape of a Graph

❑ Increasing and Decreasing

Definition 6.3.1. We say that

(a) a function $f(x)$ is "*(strictly) increasing*" on an interval I if

$$f(x) < f(y) \qquad \text{whenever } x, y \in I \text{ with } x < y.$$

(b) a function $f(x)$ is "*(strictly) decreasing*" on an interval I if

$$f(x) > f(y) \qquad \text{whenever } x, y \in I \text{ with } x < y.$$

(c) a function $f(x)$ is "*nondecreasing*" on I if

$$f(x) \leq f(y) \qquad \text{whenever } x, y \in I \text{ with } x < y.$$

(d) a function $f(x)$ is "*nonincreasing*" on I if

$$f(x) \geq f(y) \qquad \text{whenever } x, y \in I \text{ with } x < y.$$

(e) a function $f(x)$ is called "*monotonic*" on I if it is either nondecreasing (increasing) or non-increasing (decreasing) on I.

Theorem 6.3.2.

(a) If $f'(x) > 0$ for all $x \in (a,b)$, then f is increasing on (a,b).

(b) If $f'(x) < 0$ for all $x \in (a,b)$, then f is decreasing on (a,b).

Proof. We will prove (a) here and the proof of (b) is similar.

Let c and d be two numbers in (a,b) with $c < d$. It sufficies to show that if $f(c) < f(d)$. Since $f'(x) > 0$ for all $x \in (a,b)$, f is continuous on $[c,d]$ and differentiable on (c,d). By the mean value theorem, there exists $\alpha \in (c,d)$ such that

$$\frac{f(c) - f(d)}{c - d} = f'(\alpha)$$

Then

$$f(c) - f(d) = f'(\alpha)(c - d) < 0$$

because $f'(\alpha) > 0$ and $(c - d) < 0$. We have $f(c) < f(d)$. Since c and d are arbitrary two numbers in (a,b), f is increasing on (a,b). □

Corollary 6.3.3. *If $f'(x) > 0$ for all $x \in (a,b)$ and f is continuous on $[a,b]$, then f is increasing on $[a,b]$.*

Proof. (Exercise) □

Remark.

(i) The converse of the above theorem may be false. That is, even if f is differentiable and increasing on (a,b), it cannot imply that f' is always positive. For example, $f(x) = x^3$ is increasing but $f'(0) = 0$.

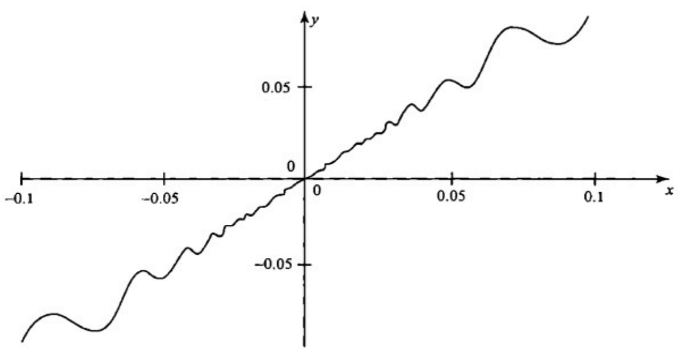
(ii) Suppose that the derivative of a function is positive at one point but not on an interval. It cannot guarantee that f must be increasing on this interval. For example

$$f(x) = \begin{cases} x + 2x^2 \sin \dfrac{1}{x} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

Then

$$f'(x) = \begin{cases} 1 + 4x \sin \dfrac{1}{x} - 2\cos \dfrac{1}{x} & \text{if } x \neq 0 \\ 1 & \text{if } x = 0 \end{cases}$$

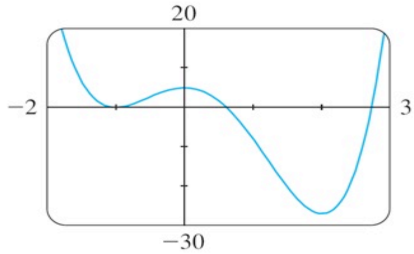
$f'(0) = 1 > 0$ but $f'(x)$ is not positive on any neighborhood of 0. The function f is not increasing on any neighborhood of 0



Example 6.3.4. Find where the function $f(x) = 3x^4 - 4x^3 - 12x^2 + 5$ is increasing and where it is decreasing.

Proof. $f'(x) = 12x(x-2)(x+1)$. To find where $f'(x) > 0$ and where $f'(x) < 0$.
f is increasing on $(-1, 0) \cup (2, \infty)$ and decreasing on $(-\infty, -1) \cup (0, 2)$. (See the table.)

Interval	$12x$	$x-2$	$x+1$	$f'(x)$	f
$x < -1$	$-$	$-$	$-$	$-$	decreasing on $(-\infty, -1)$
$-1 < x < 0$	$-$	$-$	$+$	$+$	increasing on $(-1, 0)$
$0 < x < 2$	$+$	$-$	$+$	$-$	decreasing on $(0, 2)$
$x > 2$	$+$	$+$	$+$	$+$	increasing on $(2, \infty)$



□

Remark 6.1 says that maximum or minimum value must occur at a critical point. But not every critical point gives rise to a maximum or minimum. How to determine whether a critical point gives an extreme value?

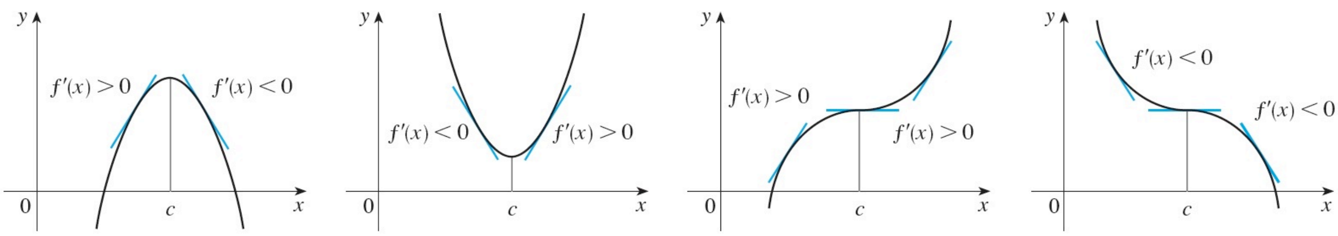
❑ First Derivative Test

Suppose that c is a critical number of a continuous function f.

(a) If $f'(x)$ changes from positive to negative at c, then f has a local maximum at c.

(b) If $f'(x)$ changes from negative to positive at c, then f has a local minimum at c.

(c) If f does not change sign at c, then f has no local maximum or minimum at c.



(a) Local maximum (b) Local minimum (c) No maximum or minimum (d) No maximum or minimum

Note. To apply the first derivative test, it only needs that f is continuous at c and f' exists near c. That $f'(c) = 0$ is not necessary.

Example 6.3.5. Find the local minimum and maximum values of the function $f(x) = 3x^4 - 4x^3 - 12x + 5$.

Proof. The derivative of f is $f'(x) = 12x(x-2)(x+1)$ and hence the critical points of $f(x)$ are -1, 0 and 2.

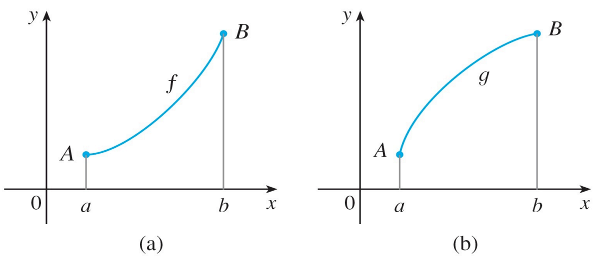
Since $f'(x) > 0$ on $(-1,0) \cup (2,\infty)$ and $f'(x) < 0$ on $(-\infty,-1) \cup (0,2)$, $f'(x)$ changes from negative to positive at -1 and 2 and from positive to negative at 0.

By the first derivative test, $f(-1) = 0$ is a local minimum, $f(0) = 5$ is a local maximum and $f(2) = -27$ is a local minimum. □

❏ Convexity and Concavity

Observe that two increasing functions may have different shapes.

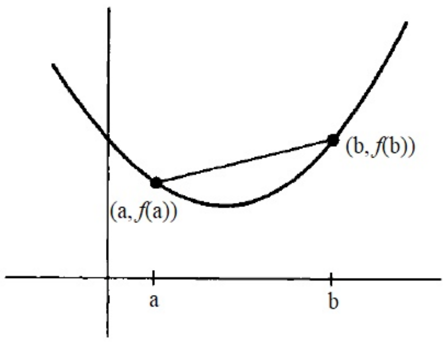
Question: How to distinguish them?



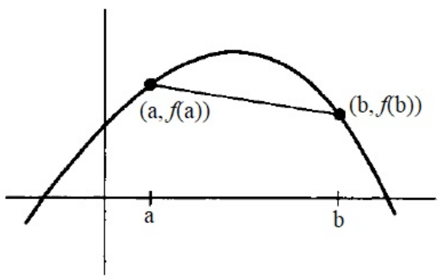
(a) (b)

Definition 6.3.6.

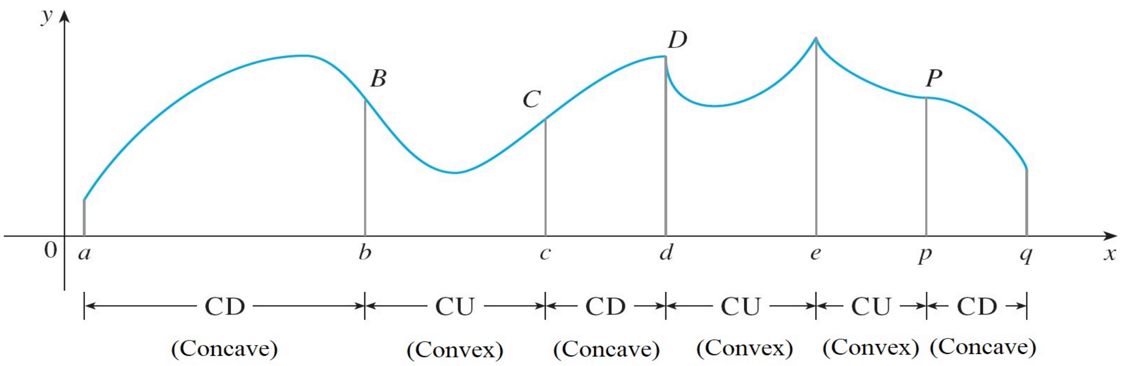
(1) A function $f(x)$ is "*concave upward*" (or "*convex*") on an interval I if for any $a, b \in I$, the segment joining $\big(a, f(a)\big)$ and $\big(b, f(b)\big)$ lies above the graph of f.



(2) A function $f(x)$ is "*concave downward*" (or "*concave*") on an interval I if for any $a, b \in I$, the segment joining $\big(a, f(a)\big)$ and $\big(b, f(b)\big)$ lies below the graph of f.



Example 6.3.7.



Example 6.3.8. $f(x) = x^2$ is convex.

Remark. An equivalent statement of the convexity is that f is convex on I if for every $a, b \in I$ and every $x \in (a, b)$

$$f(a) + \frac{f(b) - f(a)}{b - a}(x - a) \geq f(x). \tag{6.4}$$

The alternating statement of concavity of f is by replacing the inequality "$\geq$" in (6.4) by "$\leq$".

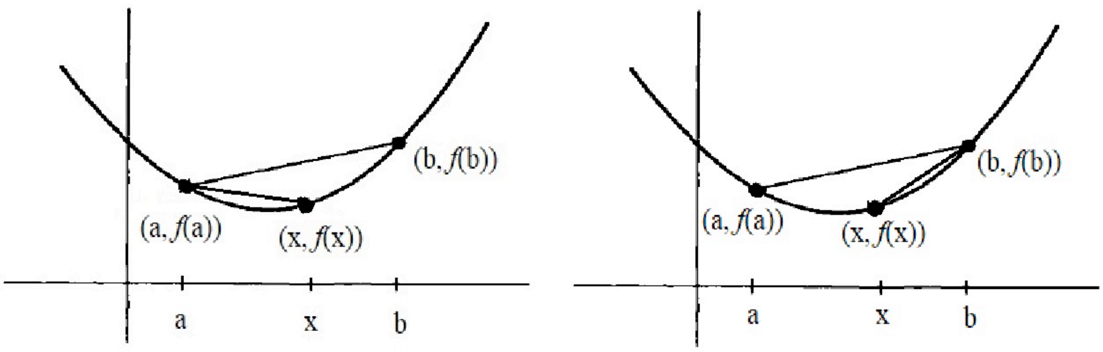
Remark.

(i) If f is convex, we can rewrite (6.4) by

$$\frac{f(x) - f(a)}{x - a} \leq \frac{f(b) - f(a)}{b - a} \leq \frac{f(x) - f(b)}{x - b} \tag{6.5}$$

for every $a, b \in I$ and $x \in (a, b)$. This implies that the slope of the segment joining $\big(a, f(a)\big)$ and $\big(b, f(b)\big)$ is greater than or equal to the slope of the segment joining $\big(a, f(a)\big)$ and $\big(x, f(x)\big)$ for every $x \in (a, b)$.

(ii) Similarly, the slope of the segment joining $\big(a, f(a)\big)$ and $\big(b, f(a)\big)$ is less than or equal to the slope of the segment joining $\big(x, f(x)\big)$ and $\big(b, f(b)\big)$.



(iii) For any $x \in (a, b)$, there exists $\lambda \in (0, 1)$ such that $x = \lambda a + (1 - \lambda)b$. Then (6.4) impliles that

$$\lambda f(a) + (1 - \lambda) f(b) = f(a) + \frac{f(b) - f(a)}{b - a}\big((1 - \lambda)(b - a)\big) \geq f\big(\lambda a + (1 - \lambda)b\big).$$

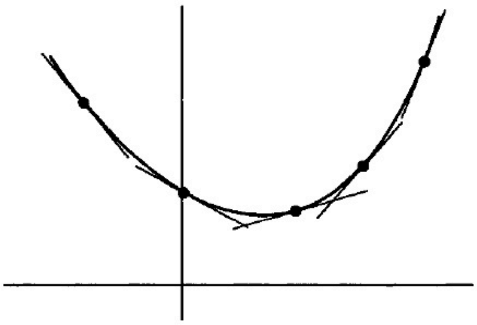
(iv) If replacing "$\geq$" by "$>$" we call the function f "*strictly convex*".

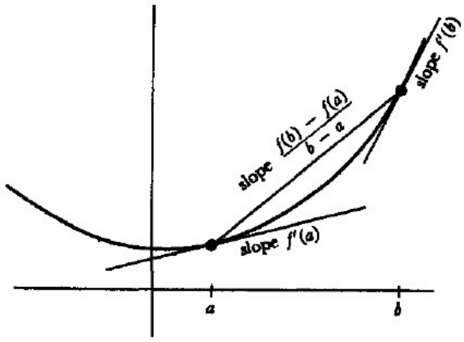
Exercise. If f is convex on I, then $-f$ is concave on I.

Theorem 6.3.9. *Let f be convex and differentiable at a.*

(a) The graph of f lies above the tangent line through $\big(a, f(a)\big)$, except at $\big(a, f(a)\big)$ itself.

(b) If $a < b$ and f is also differentiable at b, then $f'(a) < f'(b)$.





Proof.

(a) Define
$$F(h) = \frac{f(a + h) - f(a)}{h} \quad \text{for every } h \neq 0.$$

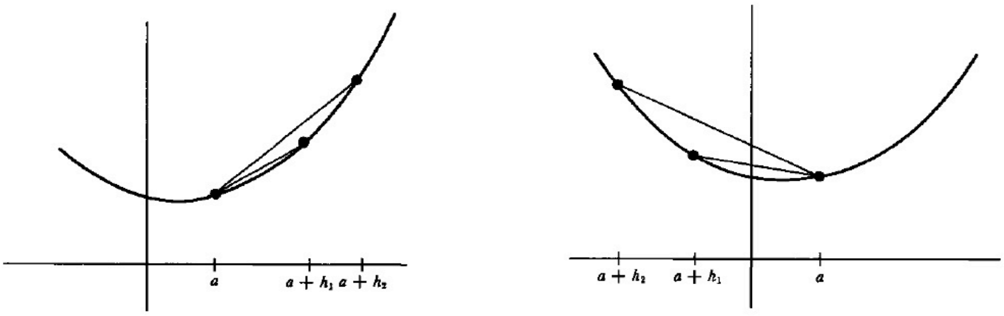
The inequality (6.5) says that F is a nondecreasing function and $\lim_{h\to 0} F(h) = f'(a)$ since f is differentiable at a. Note that $F(h)$ equals the slope of the secant line connecting $\big(a, f(a)\big)$ and $\big(a + h, f(a + h)\big)$.

Fix $h > 0$ and for $0 < h_1 < h$, since F is nondecreasing,

$$f'(a) = \lim_{h_1 \to 0^+} F(h_1) \leq F(h).$$

Hence, when $h > 0$, the slope of the secant line connecting $\big(a, f(a)\big)$ and $\big(a + h, f(a + h)\big)$ is greater than the slope of the the tangent line through $\big(a, f(a)\big)$. Then the point $\big(a + h, f(a + h)\big)$ on the graph of f is above the point $\big(a + h, L(a + h)\big)$ on the tangent line (where $y = L(x)$ is the equation of the tangent line throuth $\big(a, f(a)\big)$.)

On the contrary, we can use similar argument to show that $F(h) \leq f'(a)$ when $h < 0$. This also implies that the graph of f is above its tangent line.



The slopes of the secant lines are nondecreasing.

(b) From (6.5), we have

$$\frac{f(x)-f(a)}{x-a} \le \frac{f(b)-f(a)}{b-a} \le \frac{f(y)-f(b)}{y-b} \quad \text{for every } x, y \in (a, b).$$

Let $x \to a^+$ and $y \to b^-$. Since f is differentiable at a and b,

$$f'(a) \le \frac{f(b)-f(a)}{b-a} \le f'(b).$$

□

Note. If f is differentiable and convex on an interval I, then $f'(x)$ is nondecreasing.

Question: Is the converse of the statement true?

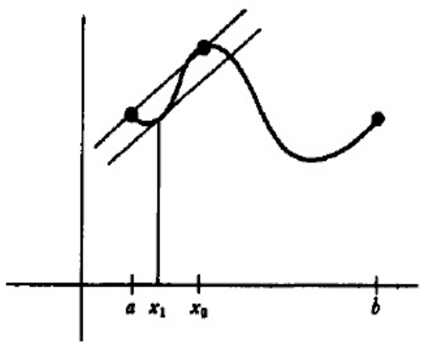
Lemma 6.3.10. *Suppose f is differentiable on I and f' is nondecreasing. If $(a, b) \subset I$ and $f(a) = f(b)$, then $f(x) \le f(a) = f(b)$ for every $x \in (a, b)$.*

Proof.

Assume that there exists some number $x \in (a, b)$ such that $f(x) > f(a) = f(b)$. Since f is continuous on $[a, b]$, by the extreme value theorem, there exists a number $x_0 \in (a, b)$ such that $f(x_0) = \max_{x\in[a,b]} f(x)$. Hence, $f'(x_0) = 0$. By the mean value theorem to the interval $[a, x_0]$, there is $x_1 \in (a, x_0)$ such that

$$f'(x_1) = \frac{f(x_0)-f(a)}{x_0-a} > 0 = f'(x_0).$$

It contradicts the fact that f' is nondecreasing and the lemma is proved.



Notice that if the hypothesis "nondecreasing" is replaced by "increasing", the inequality "≤" in the conclusion will be replaced by "<". Otherwise, there exists a number $c \in (a, b)$ such that $f(c) = f(a) = f(b)$. Then, by the mean value theorem, there is $x_0 \in (a, c)$ such that $f'(x_0) = 0$. Similar contradiction as above will be obtained.

□

Theorem 6.3.11. *If f is differentiable on I and f' is nondecreasing (increasing), then f is convex (strictly convex).*

Proof. Let $a, b \in I$ and $a < b$. Define

$$g(x) = f(x) - \frac{f(b) - f(a)}{b - a}(x - a).$$

Then $g(a) = g(b) = f(a)$ and $g'(x) = f'(x) - \dfrac{f(b) - f(a)}{b - a}$ is nondecreasing since $f'(x)$ is nondecreasing. By Lemma 6.3.10, for every $x \in (a, b)$,

$$g(x) \leq g(a) = g(b) = f(a).$$

Hence,

$$f(x) \leq f(a) + \frac{f(b) - f(a)}{b - a}(x - a)$$

and thus f is convex.

□

❑ Convexity and Concavity v.s. Differentiation

Question: How does f'' affect the graph of f?

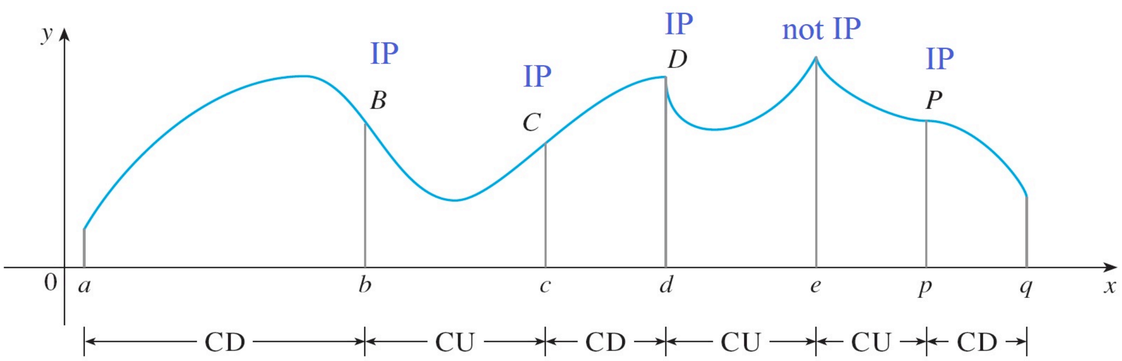
Theorem 6.3.12. *(Concavity Test)*

(a) If f has second derivative on I and $f''(x) > 0$ for all $x \in I$, then the graph of f is concave upward on I.

(b) If f has second derivative on I and $f''(x) < 0$ for all $x \in I$, then the graph of f is concave downward on I.

Proof. (Skip) □

Definition 6.3.13. We call a point P on a curve $y = f(x)$ an "*inflection point*" (or "*point of inflection*") if f is continuous there and the curve changes concavity there (i.e. either changes from CU to CD or from CD to CU).



Remark.

(i) An inflection point is a point on the curve $y = f(x)$.

(ii) The definition of an inflection point is NOT $f''(c) = 0$. For example $f(x) = x^4$, but $(0, 0)$ is not an inflection point.

(iii) If $\big(c, f(c)\big)$ is an inflection point for $f(x)$, then "f' " has an local extreme value at c. It is becasue f' is either from increase to decrease, or from decrease to increase. The conclusion is followed by the first derivative test on f'.

Theorem 6.3.14. *If the point $\big(c, f(c)\big)$ is an inflection point for f, then either $f''(c) = 0$ or $f''(c)$ does not exist.*

Proof. If $f''(c)$ does not exist, the proof is done. We may assume that $f''(c)$ exists.

Since $\big(c, f(c)\big)$ is an inflection point for f, the graph of $y = f(x)$ changes concavity at $\big(c, f(c)\big)$. Hence, $f'(x)$ has a local maximum or a local minimum at c. Since $f''(c)$ exists, by the Fermat's theorem, $f''(c) = 0$. □

■ **Stategy of finding Inflection Points**

(i) Find all points where $f''(x) = 0$ or $f''(x)$ does not exist.

(ii) determine whether the concavity changes at those points.

Example 6.3.15. Find all inflection point for $f(x) = x^3 - 6x^2 + 9x + 1$.

Proof. Since f is a polynomail function, $f''(x)$ exists everywhere. Hence, the possible inflection points happen when $f''(x) = 0$.

The second derivative of f is $f''(x) = 6(x - 2)$. We have $f''(x) < 0$ for $x \in (-\infty, 2)$ and $f''(x) > 0$ for $(2, \infty)$. Hence, the graph of f is concave downward on $(-\infty, 2)$ and concave upward on $(2, \infty)$. It implies that the graph of f changes concavity at $x = 2$. Therefore, the point $(2, 3)$ is an inflection point for f. □

Example 6.3.16. Find all inflection points for $f(x) = 3x^{5/3} - 5x$.

Proof. The first and second derivatives of f are $f'(x) = 5x^{2/3} - 5$ and $f''(x) = \frac{10}{3}x^{-1/3}$. Hence, $f''(x)$ does not exist at $x = 0$.

Since $f''(x)$ is negative for $x < 0$ and positive for $x > 0$, the graph of f is concave upward on $(-\infty, 0)$ and concave downward on $(0, \infty)$. It implies that the graph of f changes concavity at $x = 0$. Also, f is continuous at 0. Then $(0, 0)$ is an in flection point for f.

□

Question: What does f'' say about f?

❑ Second Derivative Test

Suppose $f''(x)$ is continuous near c.

(a) If $f'(c) = 0$ and $f''(c) > 0$, then f has a local minimum at c.

(b) If $f'(c) = 0$ and $f''(c) < 0$, then f has a local maximum at c.

(c) If $f''(c) = 0$, the second derivative test is inconclusive. There might be a maximum ($f(x) = -x^4$), a minimum ($f(x) = x^4$) or neither ($f(x) = x^3$).

Proof. We will prove part(a) here and the proof of part(b) is similar. By the definition of second derivative,

$$0 < f''(c) = \lim_{h \to 0} \frac{f'(c+h) - f'(c)}{h} = \lim_{h \to 0} \frac{f'(c+h)}{h}.$$

Hence, there exist $\delta > 0$ such that for $0 < |h| < \delta$,

$$\frac{f'(c+h)}{h} > 0.$$

Then $f'(c+h) < 0$ when $-\delta < h < 0$ and $f'(c+h) > 0$ when $0 < h < \delta$. Also, f is continuous at c since $f'(c)$ exists. By the first derivative test, $f(x)$ has a local minimum at c. $\square$

Remark. To find the extreme values of a function, the first deriveative test is usually easier than the second derivative test. The former can deal with more general functions than the latter. For example, the second derivative test cannot apply when $f''(c) = 0$, $f''(c)$ DNE, or $f'(c)$ DNE.

Example 6.3.17. Find all extreme values of the function $f(x) = 3x^4 - 4x^3 - 12x^2 + 5$.

Proof. The first and second derivatives of f is $f'(x) = 12(x^3 - x^2 - 2x) = 12x(x+1)(x-2)$ and $f''(x) = 12(3x^2 - 2x - 2)$. Then f has critical points -1, 0 and 2.

Check the values of $f''(x)$ at those critical points. $f''(-1) = 36 > 0$, $f''(0) = -24 < 0$ and $f''(2) = 72 > 0$. By the second derivative test, $f(-1) = 0$ and $f(2) = -27$ are local minimum of f, and $f(0) = 5$ is a local maximum of f. $\square$

The next theorem gives some result of the converse of second derivative test.

Theorem 6.3.18. *Suppose $f''(x)$ is continuous near c.*

(a) If f has a local minimum at c, then $f''(c) \geq 0$.

(b) If f has a local maximum at c, then $f''(c) \leq 0$.

Proof. We will prove part(a) by a contradiction and the proof of part(b) is similar. Assume that $f''(c) < 0$. Since $f''(c)$ exists , so does $f'(c)$. By the Fermat's theorem, $f'(c) = 0$ since f has a local minimum at c.
We apply the second derivative test, f has a local maximum at c. Hence, f is constant near c. This implies $f''(c) = 0$ and it contradicts the hypothesis $f''(c) < 0$. Therefore, $f''(c) \geq 0$. $\square$

Remark. For Theorem 6.3.18, we cannot get ">" or "<" respectively. For example, $f(x) = x^4$ or $f(x) = -x^4$ respectively.

6.4 Sketch the Graph

So far, we have learned some topics of curve sketching, for exampe domains, ranges, symmetry, limits, continuity, asymptotes, tangents, extreme values, intervals of increase and decrease, concavity, point of inflection etc. Now, we may try to drawing the graphs of functions without using graphing devices.

■ Guidelines for Sketching a Curve

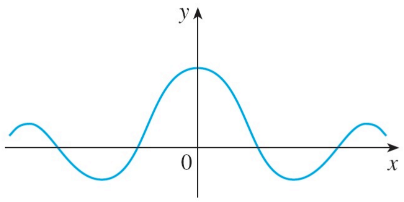
(i) **Domain**

(ii) **Intercepts**
x-intercepts: find x such that $f(x) = 0$.
y-intercepts: if $0 \in Dom(f)$, y-intercept is $f(0)$.

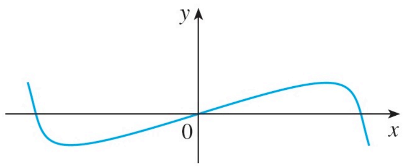
(iii) **Symmetry**

(i) *Even function:* Check the domain is symmetric about 0 and $f(-x) = f(x)$. If f is even, the graph of $f(x)$ is symmetric about the y-axis.

(ii) *Odd function:* Check the domain is symmetric about 0 and $f(-x) = -f(x)$. If f is odd, the graph of $f(x)$ is symmetric about the origin

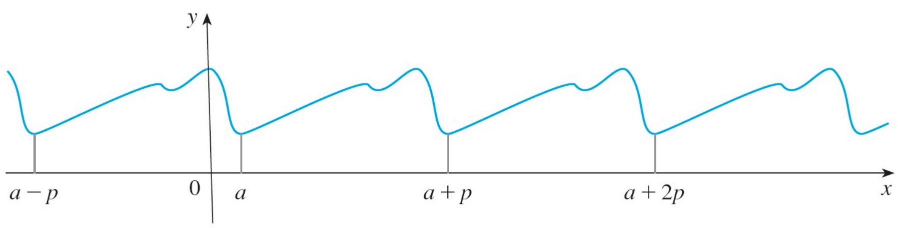


Even function: reflectional symmetry



Odd function: rotational symmetry

(iii) *Periodic function:* If there is a positive number p such that $f(x + p) = f(x)$ for all $x \in Dom(f)$.



(iv) **Asymptotes**

(i) *Horizontal asymptotes:* If $\lim_{x\to\pm\infty} f(x) = L$, the line $y = L$ is a horizontal asymptote of $y = f(x)$. If $\lim_{x\to\pm\infty} f(x) = \pm\infty$, there is no horizontal asymptote.

(ii) *Vertical asymptotes:* If $\lim_{x\to a^\pm} f(x) = \pm\infty$, then the line $x = a$ is a vertical asymptote of $y = f(x)$.

(iii) *Oblique (slant) asymptotes:* If $\lim_{x\to\pm\infty} \left[f(x) - (ax+b)\right] = 0$, then the line $y = ax + b$ is an oblique asymptote of $y = f(x)$.

(v) **Intervals of Increase or Decrease** Check the intervals of x where $f'(x) > 0$ and $f'(x) < 0$.

(vi) **Local Maximum and Minimum Values**

(i) Find all critical numbers.

(ii) Use the First Derivative Test or the Second Derivative Test to find the local maximum and local minimum values.

(vii) **Concavity and Points of Inflection**

(i) Compute $f''(x)$. The graph of $y = f(x)$ is concave upward if $f''(x) > 0$, and the graph is concave downward if $f''(x) < 0$.

(ii) *Points of Inflection:* the point where the concavity changes.

(viii) **Sketch the Curve**

■ **Examples**

Example 6.4.1. Sketch the curve $y = \dfrac{2x^2}{x^2-1} = f(x)$.

Proof. (i) **Domain:** The domain of f is $\left\{x \mid x \neq \pm 1\right\}$.

(ii) **Intercepts:** $y = \dfrac{2x^2}{x^2-1} = 0$ if $x = 0$, then 0 is a x-intercept. Taking $x = 0$, then $y = f(0) = 0$ is y-intercept.

(iii) **Symmetry:** $f(-x) = \dfrac{2(-x)^2}{(-x)^2-1} = \dfrac{2x^2}{x^2-1} = f(x)$. Thus, f is an even function and the graph of $y = f(x)$ is symmetric about the y-axis. On the other hand, the function is not periodic.

(iv) **Asymptotes:**

(i) *Horizontal Asymptote:* $\lim_{x\to\infty} \dfrac{2x^2}{x^2-1} = 2$ and $\lim_{x\to-\infty} \dfrac{2x^2}{x^2-1} = 2$. The graph $y = f(x)$ has only one horizontal asymptote $y = 2$.

(ii) *Vertical Asymptote:* $\lim_{x\to 1^+} \dfrac{2x^2}{x^2-1} = \infty$, $\lim_{x\to 1^-} \dfrac{2x^2}{x^2-1} = -\infty$, $\lim_{x\to(-1)^+} \dfrac{2x^2}{x^2-1} = -\infty$ and $\lim_{x\to(-1)^-} \dfrac{2x^2}{x^2-1} = \infty$. Thus the graph of $y = f(x)$ has two vertical asymptotes $x = 1$ and $x = -1$.

(v) **Intervals of Increase or Decrease:** The derivative of f is $f'(x) = \dfrac{-4x}{(x^2-1)^2}$. Then

$f'(x) > 0$ when $x < 0$ $(x \neq -1)$ and it implies f is increasing on $(-\infty, -1) \cup (-1, 0)$
$f'(x) < 0$ when $x > 0$ $(x \neq 1)$ and it implies f is decreasing on $(0, 1) \cup (1, \infty)$

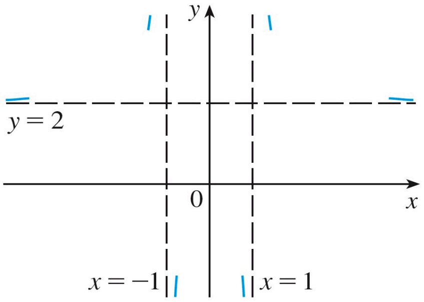
(vi) **Local maximum and minimum values** Compute $f'(x) = \dfrac{-4x}{(x^2-1)^2} = 0$ when $x = 0$. The critical number of f is $x = 0$. Since $f'(x)$ changes from positive to negative at $x = 0$, hence, $f(0) = 0$ is a local maximum by the First Derivative Test.

(vii) **Concavity and Points of Inflection** Compute $f''(x) = \dfrac{12x^2+4}{(x^2-1)^3}$. Then $f''(x) > 0$ when $x^2 - 1 > 0$ ($x > 1$ or $x < -1$), and $f''(x) < 0$ when $x^2 - 1 < 0$ ($-1 < x < 1$). We have the graph of f is concave upward on $(-\infty, -1) \cup (-1, 0)$ and concave downward on $(-1, 1)$.

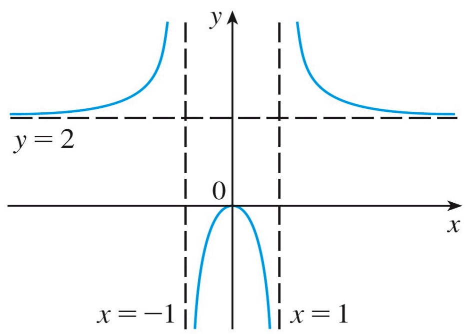
On the other hand, the graph of f does not have any point of inflection since the concavity changes at $x = \pm 1$. But ± 1 are not in the domain.

(viii) **Sketch the graph**

x	$f'(x)$	$f''(x)$	$f(x)$
$-\infty < x < -1$	+	+	increasing and CU
$-1 < x < 0$	+	-	increasing and CD
$0 < x < 1$	-	-	decreasing and CD
$1 < x < \infty$	-	+	decreasing and CU



Begin with the curve near the asymptotes and point out all special points



Connect points by following the above characters

□

Example 6.4.2. Sketch the graph of $f(x) = \dfrac{\cos x}{2 + \sin x}$.

Proof. (i) **Domain:** $Dom(f) = \mathbb{R}$.

(ii) **Intercepts:** $f(x) = \dfrac{\cos x}{2 + \sin x} = 0$ when $x = \frac{\pi}{2} + n\pi$ for every $n \in \mathbb{Z}$ and those numbers are x-intercepts. Since $f(0) = \dfrac{1}{2}$, y-intercept is $\dfrac{1}{2}$.

(iii) **Symmetry:** The domain of f is symmetric about 0. Consider $f(-x) = \dfrac{\cos(-x)}{2+\sin(-x)} = \dfrac{\cos x}{2-\sin x}$. Hence, f is neither even nor odd. On the other hand,

$$f(x+2\pi) = f(x) \quad \text{for all } x.$$

Thus, f is a periodic function with period 2π.

(iv) **Asymptotes:** Since f is periodic and continuous on $\mathbb{R}$, there is no veritcal or horizontal asymptote.

(v) **Intervals of Increase of Decrease:** It only suffices to discuss the case on $[0, 2\pi]$. Consider $f'(x) = -\dfrac{2\sin x + 1}{(2+\sin x)^2}$. $f'(x) > 0$ when $2\sin x + 1 < 0$ and hence $\dfrac{7\pi}{6} < x < \dfrac{11\pi}{6}$, and $f'(x) < 0$ when $2\sin x + 1 > 0$ and hence $0 < x < \dfrac{7\pi}{6}$ or $\dfrac{11\pi}{6} < x < 2\pi$. Then f is increasing on $(\dfrac{7\pi}{6}, \dfrac{11\pi}{6})$ and decreasing on $(0, \dfrac{7\pi}{6}) \cup (\dfrac{11\pi}{6}, 2\pi)$.

(vi) **Local maximum and minimum values:** To find all critical numbers (in $[0, 2\pi]$). $f'(x) = 0$ when $x = \frac{7\pi}{6}$ and $\frac{11\pi}{6}$. Since $f'(x)$ changes from postive to negative at $\dfrac{7\pi}{6}$, $f(\dfrac{7\pi}{6}) = -\dfrac{1}{\sqrt{3}}$ is a local maximum value. Also, since $f'(x)$ changes from negative to positive at $\frac{11\pi}{6}$, $f(\dfrac{11\pi}{6}) = \dfrac{1}{\sqrt{3}}$ is a local minimum value.

(vii) **Concavity and point of inflection on** $[0, 2\pi]$ Consider

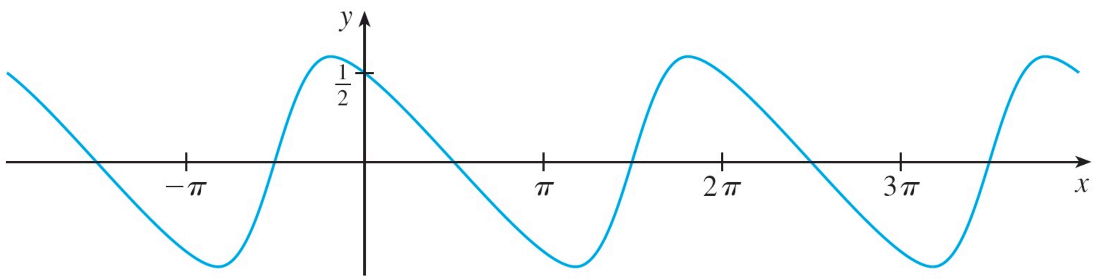
$$f''(x) = -\frac{2\cos x(1-\sin x)}{(2+\sin x)^3}.$$

Since $2 + \sin x > 0$ and $1 - \sin x \geq 0$ for all x, it suffices to consider the sign of $\cos x$. $f''(x) > 0$ when $\cos x < 0$ (hence $\dfrac{\pi}{2} < x < \dfrac{3\pi}{2}$) and $f''(x) < 0$ when $\cos x > 0$ (hence $0 < x < \frac{\pi}{2}$ or $\frac{3\pi}{2} < x < 2\pi$) We have the graph of $y = f(x)$ is concave upward on $(\dfrac{\pi}{2}, \dfrac{3\pi}{2})$ and concave downward on $(0, \dfrac{\pi}{2}) \cup (\dfrac{3\pi}{2}, 2\pi)$.

The concavity changes at $(\dfrac{\pi}{2}, 0)$ and $(\dfrac{3\pi}{2}, 0)$ and they are points of inflection.

(viii) **Sketch the graph:** It suffices to draw the graph on $[0, 2\pi]$ and then extend to whole real line since f is periodic with period 2π.

x	$f'(x)$	$f''(x)$	$f(x)$
$0 < x < \pi/2$	-	-	decreasing and CD
$\pi/2 < x < 7\pi/6$	-	+	decreasing and CU
$7\pi/6 < x < 3\pi/2$	+	+	increasing and CU
$3\pi/2 < x < 11\pi/6$	+	-	increasing and CD
$11\pi/6 < x < 2\pi$	-	-	decreasing and CD



□

Example 6.4.3. Sketch the graph of $f(x) = \dfrac{x^3}{x^2+1}$.

(i) **Domain:** $Dom(f) = \mathbb{R}$

(ii) **Intercepts:** $f(x) = \dfrac{x^3}{x^2+1} = 0$ when $x = 0$ and thus x-intercept is 0. Since $f(0) = 0$, y-intercept is 0.

(iii) **Symmetry:** Since $f(-x) = \dfrac{(-x)^3}{(-x)^2+1} = -\dfrac{x^3}{x^2+1} = -f(x)$, f is an odd function and the graph is symmetric about the origin.

(iv) **Asymptote:** Since the denominator $x^2 + 1 \neq 0$ for all x, the graph of $y = f(x)$ has no vertical asymptote. But $f(x) = x - \dfrac{x}{x^2+1}$. Hence, $f(x) - x = -\dfrac{x}{x^2+1} \to 0$ as $x \to \pm\infty$. Therefore, the line $y = x$ is an oblique asymptote.

(v) **Intervals of Increase or Decrease:**

$$f'(x) = \frac{x^2(x^2+3)}{(x^2+1)^2} \geq 0 \quad \text{for all } x.$$

Thus f is increasing on $(-\infty, \infty)$.

(vi) **Local maximum and minimum values** Since $f'(x)$ exists and $f'(x) = 0$ when $x = 0$, the only critical number is $x = 0$. Also, $f'(x) \geq 0$ for all x. Thus, f' does not change sign at $x = 0$. There is no local maximum or minimum.

(vii) **Concavity and points of inflection** The second derivative of f is

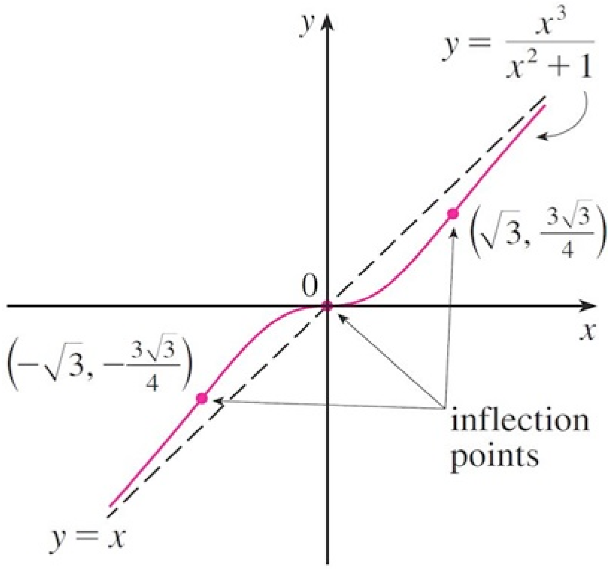
$$f''(x) = \frac{2x(3-x^2)}{(x^2+1)^3}.$$

$f''(x) > 0$ when $x < -\sqrt{3}$ or $0 < x < \sqrt{3}$ and $f''(x) < 0$ when $-\sqrt{3} < x < 0$ or $x > \sqrt{3}$. The graph is concave upward on $(-\infty, -\sqrt{3}) \cup (0, \sqrt{3})$ and is concave downward on $(-\sqrt{3}, 0) \cup (\sqrt{3}, \infty)$.

The concavity changes at $(-\sqrt{3}, -\dfrac{3\sqrt{3}}{4})$, $(0, 0)$ and $(\sqrt{3}, \dfrac{3\sqrt{3}}{4})$. They are points of inflection.

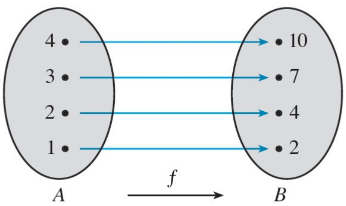
(viii) **Sketch the curve**

Interval	x	$3 - x^2$	$(x^2+1)^3$	$f''(x)$	f
$x < -\sqrt{3}$	$-$	$-$	$+$	$+$	CU on $(-\infty, -\sqrt{3})$
$-\sqrt{3} < x < 0$	$-$	$+$	$+$	$-$	CD on $(-\sqrt{3}, 0)$
$0 < x < \sqrt{3}$	$+$	$+$	$+$	$+$	CU on $(0, \sqrt{3})$
$x > \sqrt{3}$	$+$	$-$	$+$	$-$	CD on $(\sqrt{3}, \infty)$

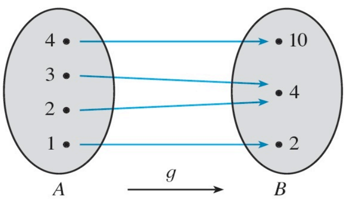


6.5 Inverse Functions

Recall: Functions



f never takes on the same value twice. If $x_1 \neq x_2$ then $f(x_1) \neq f(x_2)$.



$f(2) = 4 = f(3)$. f takes two different numbers to the same value.

Question: For every number $b \in B$, can we find a "$a \in A$ such that a function assigns a to b? In general, it is impossible since two different numbers may be assigned to one number.

■ **One-to-one and Onto Functions**

Definition 6.5.1. Let $f : A \to B$. We say that

(1) f is "one-to-one (or 1-1, or injective)" if $f(x) \neq f(y)$ whenever $x \neq y$

(2) f is "onto (or surjective)" if for every $b \in B$, there exists $a \in A$ such that $f(a) = b$.

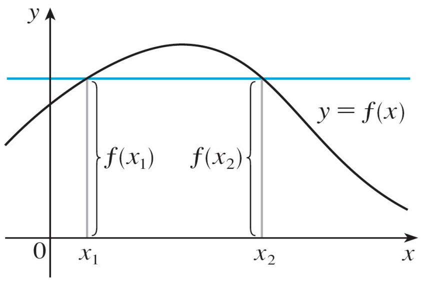
Remark.

(i) An equivalent definition of one-to-one function is that if $f(x_1) = f(x_2)$ then $x_1 = x_2$.

(ii) A function which is both one-to-one and onto is called "bijective".

(iii) If f is increasing or decreasing, then f is one-to-one.

■ **Horizontal Line Test**



A function is one-to-one if and only if no horizontal line intersects its graph more than once.

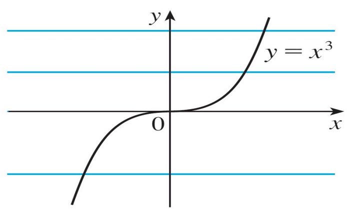
Example 6.5.2.

(1) $f(x) = x^3$ is one-to-one.

If $f(x_1) = f(x_2)$, then

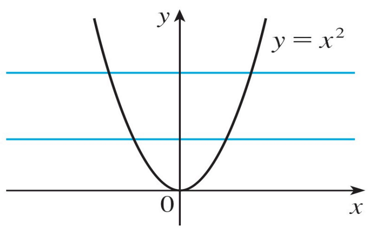
$$0 = x_1^3 - x_2^3 = (x_1 - x_2)(x_1^2 + x_1x_2 + x_2^2).$$

Since $x_1^2 + x_1x_2 + x_2^2 > 0$ if x_1 or x_2 is nonzero, $x_1 = x_2$. Hence, f is one-to-one.



(2) $g(x) = x^2$ is not one-to-one.

$g(1) = 1 = g(-1)$.



■ **Inverse of a Function**

Definition 6.5.3. Let f be a function with domain A and range B. The "*inverse*" of f, denoted by f^{-1}, is a rule that assigns each element in B to a set in A which reverses f. More precisely, for $b \in B$

$$f^{-1}(b) = \left\{a \in A \mid f(a) = b\right\}.$$

The set $f^{-1}(b) \subseteq A$ is usually called the preimage of f at b.

Note. In general, the inverse, f^{-1}, of a function f may not be a function. It sends every point b in the range of f to a set, $f^{-1}(b)$, in the domain. Every number in the set is assigned to b by f.

Definition 6.5.4. A function f is "*invertible*" if its inverse is a function.

Heuristically, if f^{-1} is a function, by the vertical line test, no vertical line will intersect the graph of f^{-1} more than one point. This implies that no horizontal line intersects the graph of f more than one point. Hence, f is ono-to-one.

Theorem 6.5.5. *A function f is invertible if and only if f is one-to-one.*

Proof. ($\Longrightarrow$) Suppose that $f(x_1) = f(x_2) = z$ for some $x_1, x_2 \in Dom(f)$. Since f is invertible, f^{-1} is a function. A function assigns an element in the domain to a unique number. Hence,

$$x_1 = f^{-1}(z) = x_2.$$

($\Longleftarrow$) If f is one-to-one, for every element b in the range of f, there is a unique element a in the domain of f such that $f(a) = b$. Hence, a is also the unique element which is assigned by f^{-1} from b. Then, f^{-1} is a function. □

Remark. If we call f^{-1} the "*inverse function of f*", then it is automatically regarded as a function. We have

(i)

$$Dom(f^{-1}) = Range(f) \qquad \text{and} \qquad Range(f^{-1}) = Dom(f).$$

(ii)

$$f^{-1}(y) = x \quad \Longleftrightarrow \quad f(x) = y \quad \text{for any } y \in Range(f).$$

Example: $f(x) = x^3$ and $f^{-1}(x) = x^{1/3}$. If $y = x^3$ then

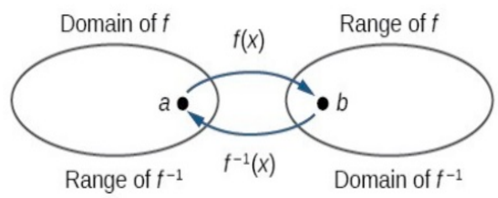
$$f^{-1}(y) = f^{-1}(x^3) = (x^3)^{1/3} = x.$$

(iii) *Caution:* Do not mistake the -1 in f^{-1} for an exponent. That is, $f^{-1} \neq \dfrac{1}{f}$.

(iv)

$$\begin{aligned} f^{-1}\big(f(x)\big) &= x \quad \text{for every } x \in Dom(f) \\ f\big(f^{-1}(y)\big) &= y \quad \text{for every } y \in Dom(f^{-1}) \end{aligned}$$

Hence, $(f^{-1})^{-1} = f$.

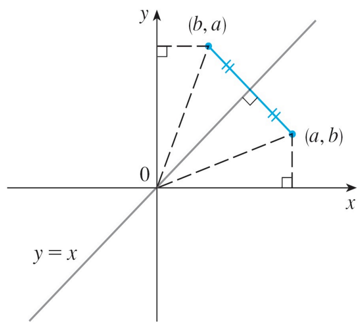


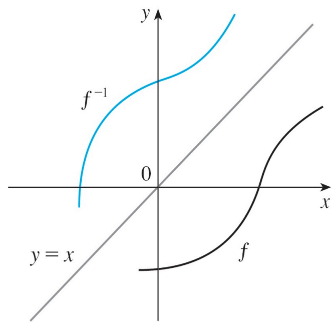
Domain and range of a function and its inverse

■ **Graph of f^{-1}**

$$\begin{aligned} Graph\ of\ f &= \big\{\big(a, f(a)\big) \mid a \in Dom(f)\big\} \\ Graph\ of\ f^{-1} &= \big\{\big(b, f^{-1}(b)\big) \mid b \in Range(f)\big\} = \big\{(f(a), a) \mid a \in Dom(f)\big\}. \end{aligned}$$

The graph of f^{-1} is obtained by reflecting the graph of f about the line $y = x$.



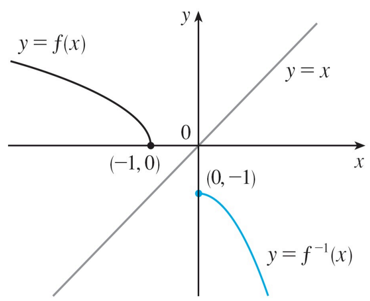


Example 6.5.6. Sketch the graph of $f(x) = \sqrt{-1-x}$ and its inverse function using the same coordinate axes.

Proof.

(i) Sketch $y = \sqrt{-1-x}$.

(ii) Reflecting the graph about the line $y = x$.



□

■ How to Find the Inverse Function of a One-to-one Function f?

1 Write $y = f(x)$.

2 Solve the equation for x in terms of y (if possible)

3 Interchange x and y to express f^{-1} as a function of x.

Example 6.5.7.

(1) Find the inverse function of $f(x) = x^3 + 2$.

Proof. Let $y = x^3 + 2$. Then $x = \sqrt[3]{y-2}$. Consider

$$y = \sqrt[3]{x-2} \quad \text{(interchange } x \text{ and } y.)$$

Then $f^{-1}(x) = \sqrt[3]{x-2}$. □

(2) Find the inverse of $f(x) = \dfrac{4x-1}{2x+3}$.

Proof. Let $y = \dfrac{4x-1}{2x+3}$. Then $4x - 1 = 2xy + 3y$ and hence $x = \dfrac{-3y-1}{2y-4}$. Interchange x and y and we have

$$y = \frac{-3x-1}{2x-4}.$$

Thus, $f^{-1}(x) = \dfrac{-3x-1}{2x-4}$. □

❑ The Calculus of Inverse Functions

■ Continuity of f^{-1}

Lemma 6.5.8. *If $f : [a, b] \to \mathbb{R}$ be continuous, one-to-one and $f(a) < f(b)$, then*

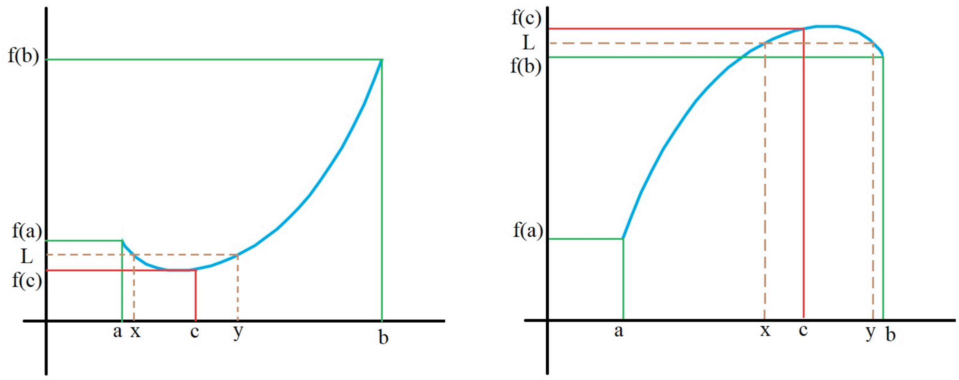
(a) $f(a) < f(c) < f(b)$ for all $c \in (a, b)$.

(b) f is increasing on $[a, b]$.

Proof. (a) If false, there is a number $c \in (a, b)$ such that either $f(c) \leq f(a)$ or $f(c) \geq f(b)$. Since f is one-to-one, the equality will not occur.

Suppose that $f(c) < f(a)$. Choose $L \in \mathbb{R}$ such that $f(c) < L < f(a) < f(b)$. By the intermediate value theorem, there exist $x, y \in (a, b)$ where $a < x < c < y < b$ such that $f(x) = L = f(y)$. It contradicts the hypothesis that f is one-to-one.

Similarly, if $f(c) > f(b)$, we can obtain a contradiction to the one-to-one hypothesis by using the intermediate value theorem.



(b) Assume that f is not increasing on $[a, b]$. There exist $a < x < y < b$ such that $f(x) > f(y)$. Since f is continuous and one-to-one on $[a, y]$ and also by part(a), $f(a) < f(y)$. Then for $x \in (a, y)$, $f(x) < f(y)$. It contradicts the assumption.

□

Theorem 6.5.9. *Let I be an interval and $f : I \to \mathbb{R}$ is continuous and one-to-one, then f is either increasing or decreasing on I.*

Proof. By Lemma 6.5.8, since f is continuous and one-to-one, it is either increasing or decreasing on any closed and bounded subinterval of I.

Assume that f is not increasing or decreasing on I. There are $a, b, c, d \in I$ with $a < b$ and $c < d$ such that $f(a) < f(b)$ and $f(c) > f(d)$.

Let $\alpha = \min(a, b, c, d)$ and $\beta = \max(a, b, c, d)$. Then $a, b, c, d \in [\alpha, \beta]$ and f is either increasing or decreasing on $[\alpha, \beta]$. Hence, one of the arguments that $f(a) < f(b)$ and $f(c) > f(d)$ is false. Thus, f is either increasing or decreasing on I. □

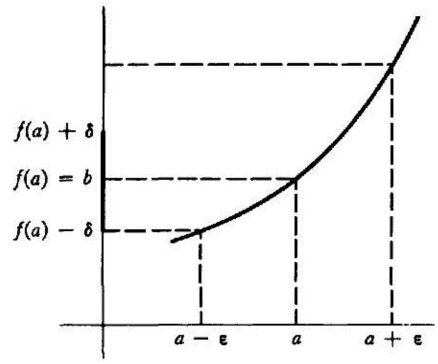
Corollary 6.5.10. *If f is continuous and one-to-one on (a, b), then the range of f is an open interval.*

Proof. (Exercise) □

Heuristically, if the graph of a continuous function has no break in it, the graph of f^{-1} also has no break in it (since it is obtained from the graph of f by reflecting about the line $y = x$). Hence, we expect that f^{-1} is also continuous.

Theorem 6.5.11. *If f is continuous and one-to-one on an interval, then f^{-1} is also continuous.*

Proof. By Theorem 6.5.9, we may assume that f is increasing. For $b \in Dom(f^{-1})$, there exists $a \in Dom(f)$ such that $f(a) = b$ ($f^{-1}(b) = a$). We will prove that f^{-1} is continuous at b.



Given $\varepsilon > 0$, we will find $\delta > 0$ such that if $|y - b| < \delta$, then $\left|f^{-1}(y) - a\right| < \varepsilon$.

Since f is increasing, $f(a - \varepsilon) < f(a) = b < f(a + \varepsilon)$. Let

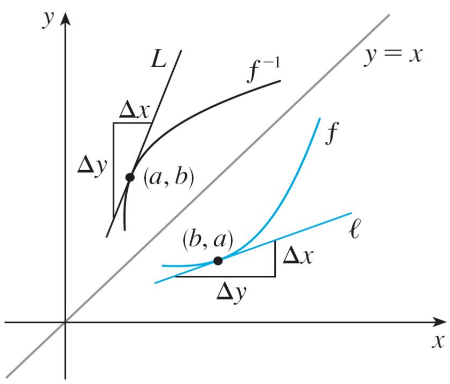
$$\delta = \min(\left|f(a - \varepsilon) - f(a)\right|, \left|f(a + \varepsilon) - f(a)\right|).$$

For $|y - b| < \delta$, $f(a - \varepsilon) < y < f(a + \varepsilon)$. By I.V.T for f, there exists $x \in (a - \varepsilon, a + \varepsilon)$ such that $f(x) = y$. Such x is the unique number satisfying $f(x) = y$ becasue f is one-to-one. This implies that

$$\left|f^{-1}(y) - a\right| < \varepsilon \quad \text{for every } |y - b| < \delta.$$

Hence, f^{-1} is continuous at b. Since b is an arbitrary number in $Dom(f^{-1})$, f^{-1} is continuous. □

■ Differentiability of f^{-1}



Let $y = f^{-1}(x)$ and $b = f^{-1}(a) \Leftrightarrow f(b) = a$. Heuristically, $(f^{-1})'(a)$ is the slope of the tangent line L of f^{-1} at (a, b). The tangent line L is obtained by reflecting the tangent line ℓ of f at (b, a). Hence

$$(f^{-1})'(a) = \frac{\Delta y}{\Delta x} = \frac{1}{\Delta x/\Delta y} = \frac{1}{f'(b)}.$$

Theorem 6.5.12. *Let f be continuous and one-to-one on an open interval I. If $f'\left(f^{-1}(a)\right) = 0$, then f^{-1} is not differentiable at a.*

Proof. Assume that $\left(f^{-1}\right)'(a)$ exists. Since $f\left(f^{-1}(x)\right) = x$, by the chain rule,

$$f'\left(f^{-1}(a)\right) \cdot \left(f^{-1}\right)'(a) = 1.$$

Then $0 \cdot \left(f^{-1}\right)'(a) = 1$. It is impossible to obtain this equality. □

Theorem 6.5.13. *Let f be continuous and one-to-one on an open interval I and $f(a) = b$. Suppose that f is differentiable at $f^{-1}(b)$ with $f'\big(f^{-1}(b)\big) \neq 0$. Then f^{-1} is differentiable at b and*

$$\big(f^{-1}\big)'(b) = \frac{1}{f'\big(f^{-1}(b)\big)} = \frac{1}{f'(a)}.$$

Proof. Since f is continuous and one-to-one, for $0 \leq |h| \ll 1$, there exists a corresponding $k = k(h)$ such that $b + h = f(a + k)$. Then $h = f(a + k) - f(a)$. Also, by Theorem 6.5.11, f^{-1} is continuous. Thus $h \to 0$ if and only if $k \to 0$.

Consider

$$\begin{aligned}
\big(f^{-1}\big)'(b) &= \lim_{h\to 0} \frac{f^{-1}(b+h) - f^{-1}(b)}{h} = \lim_{h\to 0} \frac{f^{-1}(b+h) - a}{h} \\
&= \lim_{h\to 0} \frac{(a+k) - a}{f(a+k) - f(a)} = \lim_{h\to 0} \frac{k}{f(a+k) - f(a)} \\
&= \lim_{k\to 0} \frac{1}{\frac{f(a+k)-f(a)}{k}} \\
&= \frac{1}{f'(a)} = \frac{1}{f'\big(f^{-1}(b)\big)}.
\end{aligned}$$

The equality in the last line is because f is differentiable at a and $f'(a) \neq 0$. □

Example 6.5.14. Find the derivatives of $f(x) = x^{1/n}$ where n is odd.

Proof. Let $g(y) = y^n$. Then $f = g^{-1}$ and $g'(y) = ny^{n-1}$.

$$f'(x) = \frac{1}{g'\big(g^{-1}(x)\big)} = \frac{1}{n(x^{1/n})^{n-1}} = \frac{1}{n} x^{1/n-1}.$$

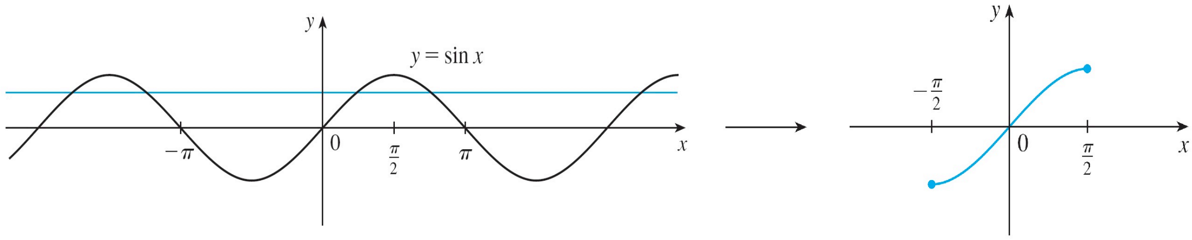
□

6.6 Inverse Trigonometric Functions

Note that the only functions that have inverse functions are one-to-one fucntions. But the trigonometric functions are not one-to-one. For example $f(x) = \sin x$ is not one-to-one (by the horizontal line test). Thus, to discuss the inverse of trigonometric functions, we should restrict those trigonometric functions on certain domain such that they are one-to-one there.

■ $f(x) = \sin^{-1} x$

The function $f(x) = \sin x$, $-\frac{\pi}{2} \leq x \leq \frac{\pi}{2}$, is one-to-one. The inverse function of this restricted sine function f exists and denoted by "$\sin^{-1}$" or "arcsin". It is called the "*inverse sine function*" or the "*arcsine function*".



Note.

(1) $\sin^{-1} x = y \Leftrightarrow \sin y = x$ for $-\dfrac{\pi}{2} \le y \le \dfrac{\pi}{2}$.

(2) $\sin^{-1} x \neq \dfrac{1}{\sin x}$.

(3) The domain of $\sin^{-1} x$ is $[-1, 1]$ and its range is $[-\dfrac{\pi}{2}, \dfrac{\pi}{2}]$.

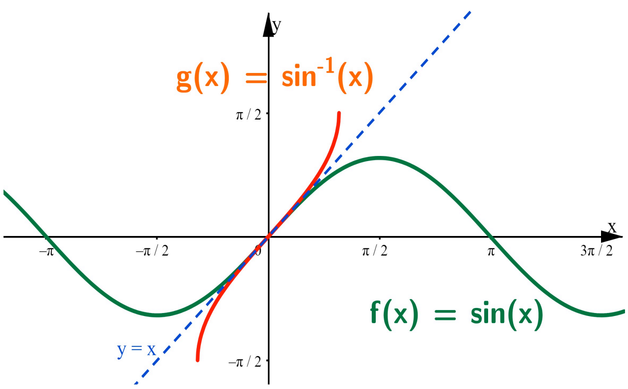
(4) For $x \in [-1, 1]$, $\sin^{-1} x$ is the number(angle) between $-\dfrac{\pi}{2}$ and $\dfrac{\pi}{2}$ whose sine is x.

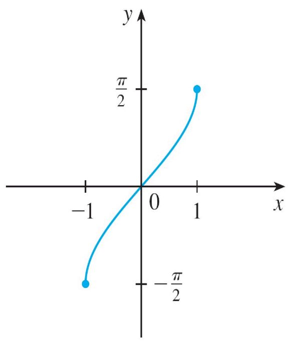
(5) $\sin\left(\sin^{-1} x\right) = x$ for $x \in [-1, 1]$ and $\sin^{-1}\left(\sin x\right) = x$ for $x \in [-\dfrac{\pi}{2}, \dfrac{\pi}{2}]$.

Example 6.6.1.

$$\sin^{-1}(\frac{1}{2}) = \frac{\pi}{6}, \qquad \sin^{-1}(-1) = -\frac{\pi}{2}, \qquad \sin^{-1}(-\frac{\sqrt{3}}{2}) = -\frac{\pi}{3}$$
$$\tan\left(\sin^{-1}(\frac{1}{2})\right) = \frac{1}{\sqrt{3}}, \quad \cos\left(\sin^{-1}(\frac{\sqrt{3}}{2})\right) = \frac{1}{2}.$$

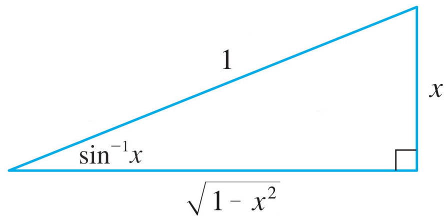
• **Graph of $\sin^{-1}$ x**





$y = \sin^{-1} x$

• **Derivative of $\sin^{-1}$ x**



$f(x) = \sin^{-1} x$. Let $g(y) = \sin y$. Then $f = g^{-1}$ and $g'(y) = \cos y$.

$$f'(x) = \frac{1}{g'\left(g^{-1}(x)\right)} = \frac{1}{\cos\left(\sin^{-1}(x)\right)} = \frac{1}{\sqrt{1-x^2}}.$$

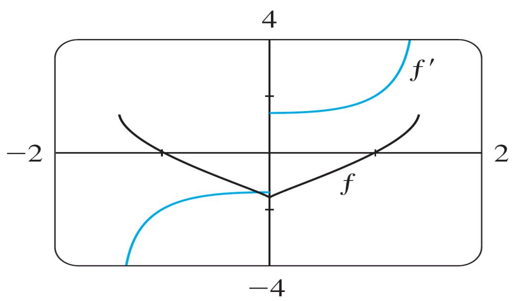
Consider the implicit differentiation, $y = \sin^{-1} x$. Then $\sin y = x$.

$$\frac{d}{dx}(\sin y) = \frac{d}{dx}(x) \quad \Rightarrow \quad \cos y \cdot \frac{dy}{dx} = 1 \quad \Rightarrow \quad \frac{dy}{dx} = \frac{1}{\cos y} = \frac{1}{\sqrt{1-x^2}}.$$

Example 6.6.2. $f(x) = \sin^{-1}(x^2 - 1)$

(a) $Dom(f) = \left\{x \,\middle|\, -1 \le x^2 - 1 \le 1\right\} = \left\{x \,\middle|\, 0 \le x^2 \le 2\right\} = \left\{x \,\middle|\, -\sqrt{2} \le x \le \sqrt{x}\right\}$.

(b) $f'(x) = \dfrac{1}{\sqrt{1-(x^2-1)^2}} \cdot 2x = \dfrac{2x}{\sqrt{2x^2 - x^4}}$ for $x \in (-\sqrt{2}, 0) \cup (0, \sqrt{2})$.



■ $f(x) = \cos^{-1} x$

Similar as the arcsine function, we should determine a region where cosine function is one-to-one there. The function $f(x) = \cos x$, $0 \le x \le \pi$, is one-to-one. The inverse function of this restricted sine function f exists and denoted by "$\cos^{-1}$" or "arccos". It is called the "*inverse cosine function*" or the "*arccosine function*".

Note.

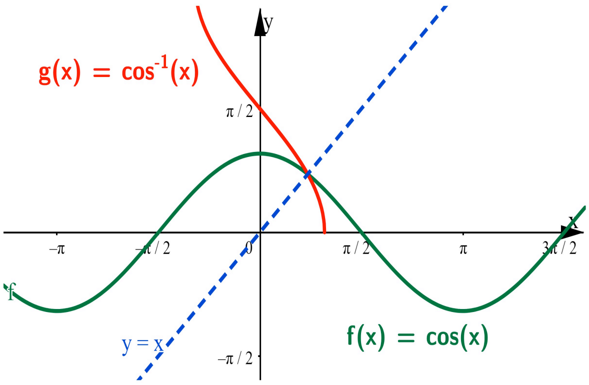
(1) $\cos^{-1} x = y \Leftrightarrow \cos y = x$ for $0 \le y \le \pi$.

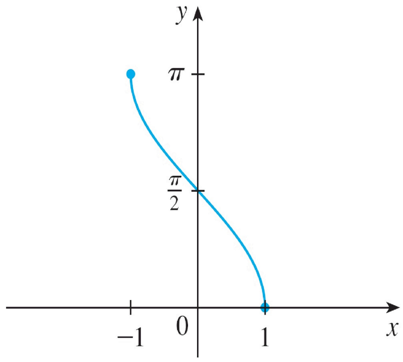
(2) The domain of $\cos^{-1} x$ is $[-1, 1]$ and its range is $[0, \pi]$.

(3) For $x \in [-1, 1]$, $\cos^{-1} x$ is the number(angle) between 0 and π whose cosine is x.

(4) $\cos\left(\cos^{-1} x\right) = x$ for $x \in [-1, 1]$ and $\cos^{-1}\left(\cos x\right) = x$ for $x \in [0, \pi]$.

• **Graph of $\cos^{-1}$ x**





$y = \cos^{-1} x$

• **Derivative of $\cos^{-1}$ x**

Following the similar steps as the derivative of arcsine, it is easy to find

$$\frac{d}{dx}\left(\cos^{-1} x\right) = -\frac{1}{\sqrt{1-x^2}} \quad \text{for} \quad -1 < x < 1.$$

Remark. Another point of view to look at the derivative of arccosine is that

$$\cos^{-1} x = \frac{\pi}{2} - \sin^{-1} x \quad \Rightarrow \quad \frac{d}{dx}\left(\cos^{-1} x\right) = -\frac{d}{dx}\left(\sin^{-1} x\right) = -\frac{1}{\sqrt{1-x^2}}.$$

■ $f(x) = \tan^{-1} x$

The function $f(x) = \tan x$, $-\frac{\pi}{2} \le x \le \frac{\pi}{2}$, is one-to-one. The inverse function of this restricted sine function f exists and denoted by "$\tan^{-1}$" or "arctan". It is called the "*inverse tangnet function*" or the "*arctangent function*".

Note.

(1) $\tan^{-1} x = y \Leftrightarrow \tan y = x$ for $-\frac{\pi}{2} < y < \frac{\pi}{2}$.

(2) The domain of $\tan^{-1} x$ is $(-\infty, \infty)$ and its range is $(-\frac{\pi}{2}, \frac{\pi}{2})$.

(3) For any x, $\tan^{-1} x$ is the number(angle) between $-\frac{\pi}{2}$ and $\frac{\pi}{2}$ whose tangent is x.

(4) $\tan\left(\tan^{-1} x\right) = x$ for $x \in (-\infty, \infty)$ and $\tan^{-1}\left(\tan x\right) = x$ for $x \in (-\frac{\pi}{2}, \frac{\pi}{2})$.

(5) $\lim\limits_{x\to-\infty} \tan^{-1} x = -\frac{\pi}{2}$ and $\lim\limits_{x\to\infty} \tan^{-1} x = \frac{\pi}{2}$.

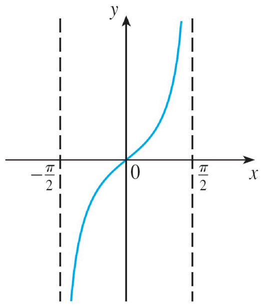
Example 6.6.3. Evaluate $\cos\left(\tan^{-1} x\right)$.

Proof. Let $y = \tan^{-1} x$. Then

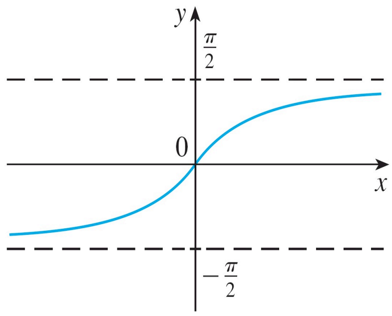
$$\tan y = x \Rightarrow \sec^2 y = 1 + \tan^2 y = 1 + x^2 \Rightarrow \cos\left(\tan^{-1} x\right) = \cos y = \frac{1}{\sec y} = \frac{1}{\sqrt{1 + x^2}}.$$

□

• **Graph of $\tan^{-1}$ x**



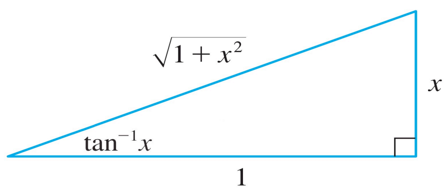
$y = \tan x$



$y = \tan^{-1} x$

• **Derivative of $\tan^{-1}$ x**

$f(x) = \tan^{-1} x$. Let $g(y) = \tan y$. Then $f = g^{-1}$ and $g'(y) = \sec^2 y$.

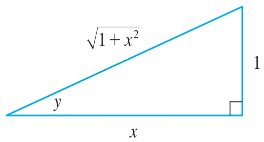
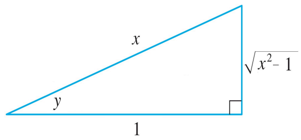
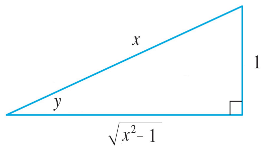


$$f'(x) = \frac{1}{g'\left(g^{-1}(x)\right)} = \frac{1}{\sec^2\left(\tan^{-1}(x)\right)} = \frac{1}{1 + x^2}.$$

Consider the implicit differentiation, $y = \tan^{-1} x$. Then $\tan y = x$.

$$\frac{d}{dx}(\tan y) = \frac{d}{dx}(x) \quad \Rightarrow \quad \sec^2 y \cdot \frac{dy}{dx} = 1 \quad \Rightarrow \quad \frac{dy}{dx} = \frac{1}{\sec^2 y} = \frac{1}{1 + x^2}.$$

• **Other inverse of trigonometric functions**

Functions	Definition	Domain	Range	
$y = \cot^{-1} x$	$x = \cot y$	$(-\infty, \infty)$	$(0, \pi)$	
$y = \sec^{-1} x$	$x = \sec y$	$(-\infty, -1) \cup (1, \infty)$	$(0, \frac{\pi}{2}) \cup (\frac{\pi}{2}, \pi)$	
$y = \csc^{-1} x$	$x = \csc y$	$(-\infty, -1) \cup (1, \infty)$	$(-\frac{\pi}{2}, 0) \cup (0, \frac{\pi}{2})$	

Remark.

(i) The ranges of $\sec^{-1} x$ and $\csc^{-1} x$ are not universally agreed.

(ii) The following is the derivatives of all inverse of trigonometric functions

- $\dfrac{d}{dx}\left(\sin^{-1} x\right) = \dfrac{1}{\sqrt{1-x^2}}$
- $\dfrac{d}{dx}\left(\cos^{-1} x\right) = -\dfrac{1}{\sqrt{1-x^2}}$
- $\dfrac{d}{dx}\left(\tan^{-1} x\right) = \dfrac{1}{1+x^2}$
- $\dfrac{d}{dx}\left(\cot^{-1} x\right) = -\dfrac{1}{1+x^2}$
- $\dfrac{d}{dx}\left(\sec^{-1} x\right) = \dfrac{1}{x\sqrt{x^2-1}}$
- $\dfrac{d}{dx}\left(\csc^{-1} x\right) = -\dfrac{1}{x\sqrt{x^2-1}}$

Example 6.6.4. Let $f(x) = x \tan^{-1}(\sqrt{x})$. Then

$$f'(x) = \tan^{-1}(\sqrt{x}) + x \cdot \frac{1}{1+(\sqrt{x})^2} \cdot \frac{1}{2\sqrt{x}} = \tan^{-1}(\sqrt{x}) + \frac{\sqrt{x}}{2(1+x)}.$$

Example 6.6.5. Prove that $\tan^{-1} x + \cot^{-1} x = \dfrac{\pi}{2}$.

Proof. Let $f(x) = \tan^{-1} x + \cot^{-1} x$ and compute $f'(x) = 0$. □

6.7 Optimization Problems*

The goal of this section is to solve some practical optimization problems.

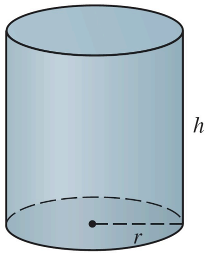
■ Strategy

*All the materials and examples in this section are from *Calculus, J. Stewart 8th Ed.*

1. Understand the problem
2. Draw a diagram
3. Introduce notation
4. Express the quantity in Step3, say Q, in terms of other symbols.
5. Express Q as a function of a single variable.
6. Use the techniques discussed in this chapter to find the extreme values.

Example 6.7.1.

A cyclindrical can is to be made to hold 1L of oil. find the dimensions that will minimized the cost tof the metal to manufacture the can.

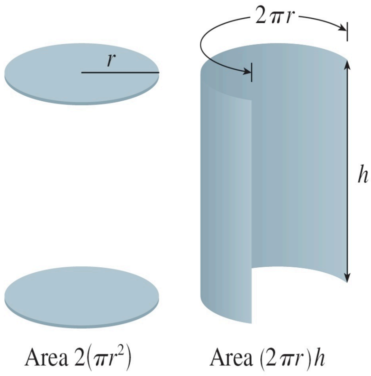


Proof.

Let r be the radius of the top of the can (in cm) and h be the height of the can. Then the surface area of the can is

$$A = 2\pi r^2 + 2\pi rh$$

The volume of the can gives rise to $\pi r^2 h = 1000$. We have $h = \dfrac{1000}{\pi r^2}$ and the surface area can be expressed as

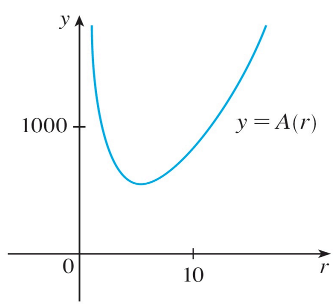


$$A(r) = 2\pi r^2 + 2\pi r \cdot \frac{1000}{\pi r^2} = 2\pi r^2 + \frac{2000}{r}.$$

To find the minimum of A, we compute the critical numbers of A.

$$A'(r) = 4\pi r - \frac{2000}{r^2} = \frac{4(\pi r^3 - 500)}{r^2}.$$

Then $A'(r) = 0$ when $r = \sqrt[3]{\frac{500}{\pi}}$. Since $A'(r) < 0$ on $(0, \sqrt[3]{\frac{500}{\pi}})$ and $A'(r) > 0$ on $(\sqrt[3]{\frac{500}{\pi}}, \infty)$, $A(r)$ decreases on $(0, \sqrt[3]{\frac{500}{\pi}})$ and increases on $(\sqrt[3]{\frac{500}{\pi}}, \infty)$.



Hence, $A(r)$ has an absolute minimum at $r = \sqrt[3]{\dfrac{500}{\pi}}$ and then $h = 2\sqrt[3]{\dfrac{500}{\pi}}$.

• <u>**Alternating Method**</u>

Since $A = A(r,h) = 2\pi r^2 + 2\pi rh$ and $\pi r^2 h = 1000$, we can compute the derivative of A and find the critical numbers of A by using the implicit differentiation and solving

$$\frac{d}{dr}A = 4\pi r + 2\pi h + 2\pi r\frac{dh}{dr} = 0 \quad \text{and} \quad 2\pi rh + \pi r^2\frac{dh}{dr} = 0.$$

Then we obtain that A has critical number when $h = 2r$. Plugging into $\pi r^2 h = 1000$, we have $r = \sqrt[3]{\dfrac{500}{\pi}}$. Following the same analysis, we get the same result as above. □

From the above example, we use the first derivtive test to check that the minimum value occurs at the critical point. But the first derivative test is to use to verify the "local" extremem values. Why can we use it here?

■ **<u>First Derivative Test for Absolute Extreme Values</u>**

Let f be a continuous function on an interval and c be a critical number of f.

(a) If $f'(x) > 0$ for $x < c$ and $f'(x) < 0$ for $x > c$, then $f(c)$ is the absolute maximum value of f.

(b) If $f'(x) < 0$ for $x < c$ and $f'(x) > 0$ for $x > c$, then $f(c)$ is the absolute minimum value of f.

Proof. Skip □

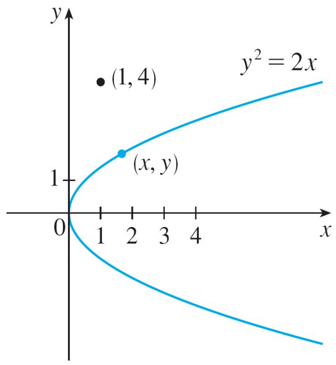
Example 6.7.2. Find the point on the parabola $y^2 = 2x$ that is closest to the point $(1,4)$.

Proof.

Let (x,y) be a point on the parabola. Then $y^2 = 2x$. The distance from (x,y) to $(1,4)$ is

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

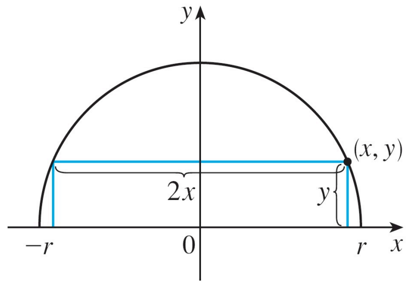
To find the minimum of $\sqrt{(\frac{1}{2}y^2 - 1)^2 + (y-4)^2}$ is equalivent to find the minumum of $(\frac{1}{2}y^2 - 1)^2 + (y-4)^2$.



Let $f(y) = (\frac{1}{2}y^2 - 1)^2 + (y-4)^2$. Then $f'(y) = 2(\frac{1}{2}y^2 - 1)\cdot y + 2(y-4) = y^3 - 8$. Hence, $f'(y) = 0$ when $y = 2$ which is the critical number of f. Since $f'(y) < 0$ when $y < 2$ and $f'(y) > 0$ when $y > 2$, $f(y)$ has an absolute minimum at $y = 2$ and thus $x = \frac{1}{2}\cdot 2^2 = 2$. Therefore, the point on the curve $y^2 = 2x$ that is closest to $(1,4)$ is $(2,2)$. □

Example 6.7.3. Find the largest rectangle that can be inscribed in a semicircle of radius r.

Proof.



Consider the semicircle to the upper half circle $x^2 + y^2 = r^2$. Let (x, y) be the point on the semicircle which is the vertex of the rectangle in the first quadrant. Hence, $x^2 + y^2 = r^2$. The area of the inscribed rectangle is

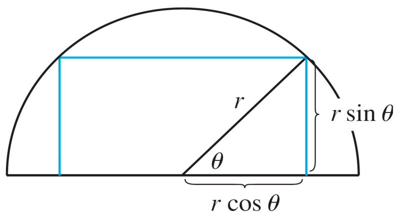
$$A = 2xy = 2x\sqrt{r^2 - x^2} \quad \text{for } 0 \le x \le r.$$

To find the absolute maximum of A, we evaluate

$$A'(x) = \frac{2(r^2 - 2x^2)}{\sqrt{r^2 - x^2}}.$$

We have $A'(x) = 0$ when $x = \dfrac{r}{\sqrt{2}}$. Since $A'(x) > 0$ when $0 < x < \frac{r}{\sqrt{2}}$ and $A'(x) < 0$ when $\frac{r}{\sqrt{2}} < x < r$. Then $A(x)$ has an absolute maximum at $x = \dfrac{r}{\sqrt{2}}$. The area of the largest inscribed rectangle is $A(r/\sqrt{2}) = r^2$.

• **Alternating Method**



We express the vertex of the rectangle in the first quadrant as $(x, y) = (r\cos\theta, r\sin\theta)$. Then the area of the inscribed rectangle is

$$A(\theta) = (2r\cos\theta)(r\sin\theta) = r^2\sin(2\theta) \quad \text{for } 0 \le \theta \le \frac{\pi}{2}.$$

Since $\sin(2\theta)$ has a maximum value when $\theta = \frac{\pi}{4}$, the area of the largest inscribed rectangle is $A(\pi/4) = r^2$.

□

Chapter 7

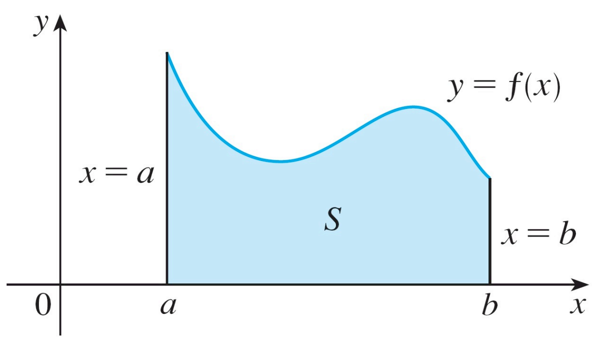
Integrals

7.1 Areas

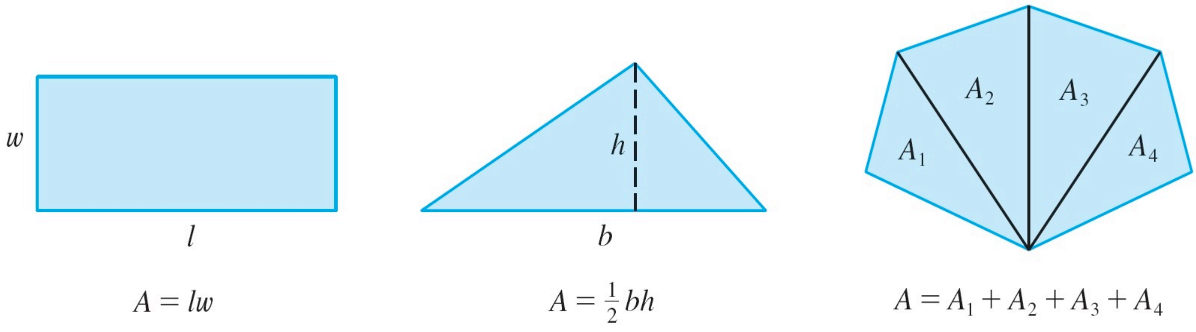
In this section, we will try to find the area under a curve.

❏ The Area Problem

Let $f(x) \geq 0$ and S be the region that lies under the curve $y = f(x)$ from a to b. We try to find the area of S.



Recall that the areas of some special regions are easily obtained. For example, the polygons.



Question: How about the areas of the general regions?

We try to approximate the area of the region S by rectangles and take the limit of the areas of these rectangles as we increase the number of rectangles.

If the function f is nonnegative on $[a, b]$, we indicate the region bounded by the x-axis, the vertical lines $x = a$ and $x = b$, and the cruve $y = f(x)$ by $R(f, a, b)$ and want to evaluate the area of $R(f, a, b)$. In order to study the integrals of more general functions, we no longer assume that f is nonnegative. The area problems will be discussed later.

Definition 7.1.1.

(a) Let P be a finite collection of points $\{x_k\}_{k=0}^{n}$ that satisfies $a = x_0 < x_1 < x_2 < \cdots < x_n = b$. We call such P a "*partition of* $[a, b]$ ".

(b) The norm (mesh) of a partition P is defined by

$$|P| = \max_{1 \le k \le n} \triangle x_k \quad \text{where } \triangle x_k = x_k - x_{k-1}.$$

(c) If P_1 and P_2 are two partitions of $[a, b]$ and $P_1 \subseteq P_2$, we say that P_2 is a refinement of P_1.

(d) If P_1 and P_2 are two partitions of $[a, b]$, then $P = P_1 \cup P_2$ is called a common refinement of P_1 and P_2.

■ Upper sums and Lower sums

Definition 7.1.2. Suppose that f is bounded on $[a, b]$ and $P = \{x_0, x_1, \cdots, x_n\}$ is a partition of $[a, b]$. Denote $m_i = \inf\limits_{x \in [x_{i-1}, x_i]} f(x)$ and $M_i = \sup\limits_{x \in [x_{i-1}, x_i]} f(x)$. We say that

$$L(P, f) = \sum_{k=1}^{n} m_k (x_k - x_{k-1}) = \sum_{k=1}^{n} m_k \triangle x_k$$

is "*the lower (Darboux) sum of f for P*" and

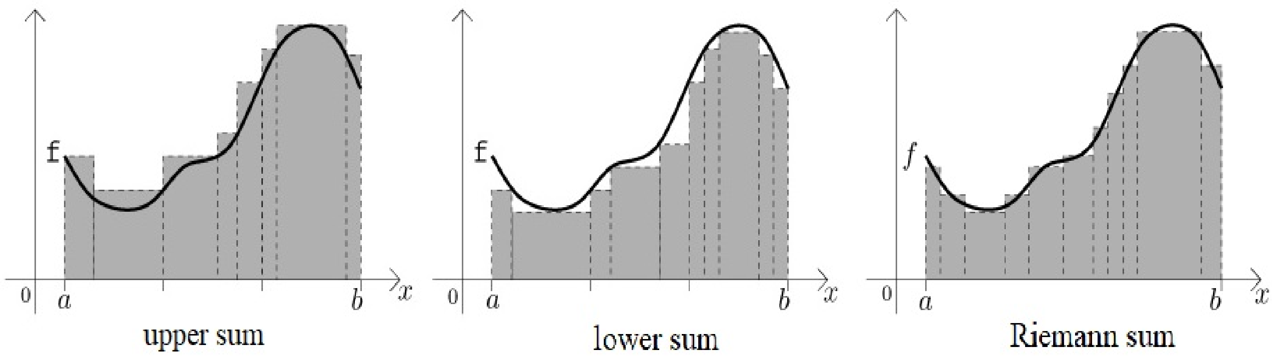
$$U(P, f) = \sum_{k=1}^{n} M_k (x_k - x_{k-1}) \sum_{k=1}^{n} M_k \triangle x_k$$

is "*the upper (Darboux) sum of f for P*".

A "*Riemann sum*" of f for P is defined by

$$S(P, f) = \sum_{k=1}^{n} f(c_k)\triangle x_k$$

with any $c_k \in [x_{k-1}, x_k]$.



Remark.

(i) The lower sum and uppoer sum of a bounded function f for P is well-defined.

(ii) Let $m = \inf_{x\in[a,b]} f(x)$ and $M = \sup_{x\in[a,b]} f(x)$. For any partition P of $[a, b]$,

$$m(b-a) \leq L(P, f) \leq S(P, f) \leq U(P, f) \leq M(b-a).$$

Example 7.1.3. Let $f(x) = x^2$ on $[0, 1]$ and a particular partition P given by

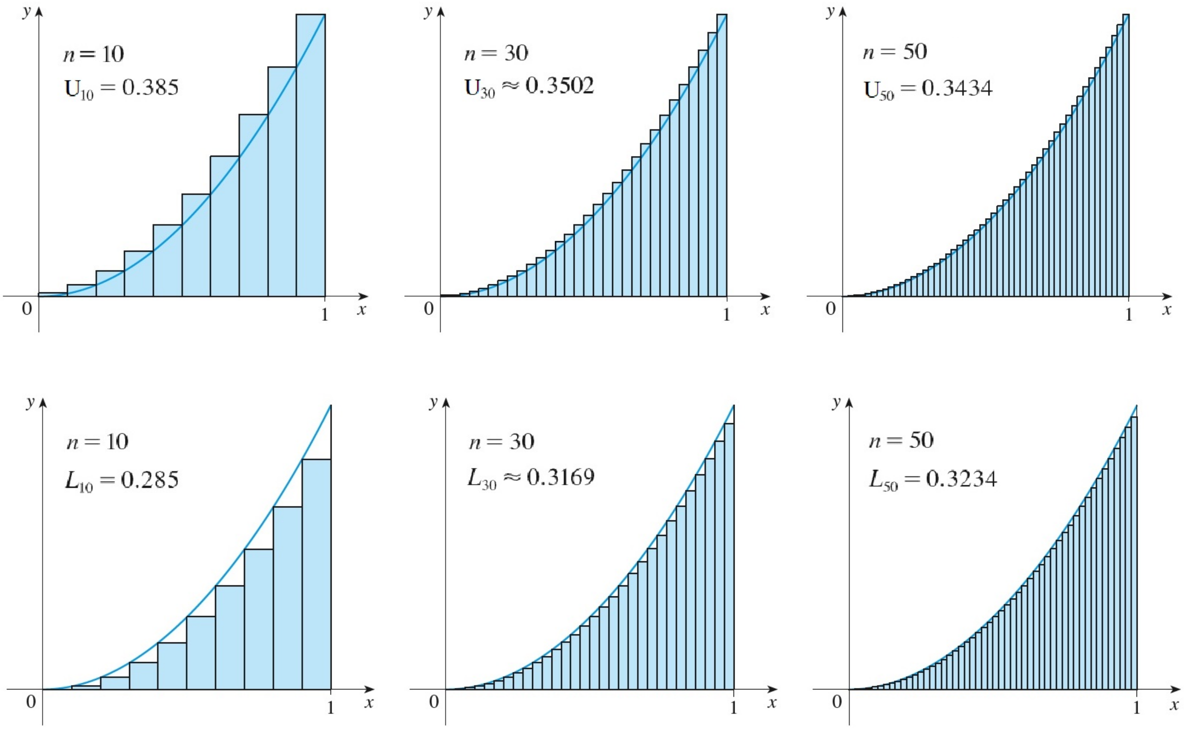
$$P = \Big\{0, \frac{1}{n}, \frac{2}{n}, \cdots, 1 - \frac{1}{n}, 1\Big\}.$$

For the partition P, $x_k = \dfrac{k}{n}$ and $\triangle x_k = \dfrac{1}{n}$ where $k = 0, 1, \cdots, n$. Since f is increasing, $f(x_k)$ is the maximum value for f and $f(x_{k-1})$ is the minimum value for f on each interval $[x_{k-1}, x_k]$. Then

$$\begin{aligned} U(P, f) &= \sum_{k=1}^{n} M_k \triangle x_k = \sum_{k=1}^{n} \Big(\frac{k}{n}\Big)^2 \cdot \frac{1}{n} = \frac{1}{n^3}\sum_{k=1}^{n} k^2 \\ &= \frac{1}{n^3} \cdot \frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6n^2} \end{aligned}$$

and

$$\begin{aligned} L(P, f) &= \sum_{k=1}^{n} m_k \triangle x_k = \sum_{k=1}^{n} \Big(\frac{k-1}{n}\Big)^2 \cdot \frac{1}{n} = \frac{1}{n^3}\sum_{k=1}^{n} (k-1)^2 \\ &= \frac{1}{n^3} \cdot \frac{(n-1)n(2n-1)}{6} = \frac{(n-1)(2n-1)}{6n^2}. \end{aligned}$$



Lemma 7.1.4. *Suppose that f is a bounded function on $[a,b]$. If P_1 and P_2 are two partitions of $[a,b]$ and P_2 is a refinement of P_1, then*

$$L(P_1,f) \le L(P_2,f) \quad \textit{and} \quad U(P_2,f) \le U(P_1,f).$$

Proof. It suffices to show that $L(P_1,f) \le L(P_2,f)$ and the proof of $U(P_2,f) \le U(P_1,f)$ is similar.

Step1: Suppose that P_2 contains only one point more than P_1. Say

$$\begin{aligned} P_1 &= \{x_0, x_1, \cdots, x_n\} \quad \text{where } x_0 < x_1 < \cdots < x_n \\ P_2 &= \{x_0, x_1, \cdots, x_{k-1}, x^*, x_k, \cdots, x_n\} \quad \text{where } x_0 < \cdots < x_k < x^* < x_{k+1} < \cdots < x_n \end{aligned}$$

Let $w_1 = \inf\limits_{x\in[x_{k-1},x^*]} f(x)$ and $w_2 = \inf\limits_{x\in[x^*,x_k]} f(x)$. Then $m_k \le w_1$ and $m_k \le w_2$. We have

$$\begin{aligned} L(P_1,f) &= \sum_{i=1}^{k-1} m_i(x_i - x_{i-1}) + m_k(x_k - x_{k-1}) + \sum_{i=k+1}^{n} m_i(x_i - x_{i-1}) \\ &= \sum_{i=1}^{k-1} m_i \triangle x_i + m_k[(x^* - x_{k-1}) + (x_k - x^*)] + \sum_{i=k+1}^{n} m_i(x_i - x_{i-1}) \\ &\le \sum_{i=1}^{k-1} m_i \triangle x_i + w_1(x^* - x_{k-1}) + w_2(x_k - x^*) + \sum_{i=k+1}^{n} m_i(x_i - x_{i-1}) \\ &= L(P_2,f). \end{aligned}$$

Step2: If P_2 contains m points more than P_1, we can repeat the procedure of Step1 m times and the lemma is proved. $\square$

Lemma 7.1.5. *Let P_1 and P_2 be partitions of $[a,b]$ and f is bounded on $[a,b]$. Then*

$$L(P_1, f) \le U(P_2, f)$$

Proof. Let $P^* = P_1 \cup P_2$. Then P is a common refinement of P_1 and P_2. By Lemma 7.1.4,

$$L(P_1, f) \le L(P, f) \le U(P, f) \le U(P_2, f).$$

□

Remark. Lemma 7.1.4 says that any upper sum $U(Q, f)$ is an upper bound of all upper sum $L(P, f)$ and any lower sum $L(Q, f)$ is a lower bound of all upper sum $U(P, f)$. Hence, any upper sum is greater than $\sup_P U(P, f)$. That is, let Q be a partition of $[a,b]$

$$\sup_P L(P, f) \le U(Q, f) \quad \text{and} \quad L(Q, f) \le \inf_P U(P, f).$$

Then

$$\sup_P L(P, f) \le \inf_P U(P, f).$$

Remark.

(i) If $\sup_P L(P, f) = \inf_P U(P, f) = c$, then c is the unique number which is greater than all lower sums and is less than all upper sums.

(ii) It is possible that $\sup_P L(P, f) < \inf_P U(P, f)$. For example, $f(x) = \begin{cases} 1, & x \in [0,1] \cap \mathbb{Q} \\ 0, & x \in [0,1] \backslash \mathbb{Q} \end{cases}$.
Then $m_k = \inf_{x \in [x_{k-1}, x_k]} f(x) = 0$ and $M_k = \sup_{x \in [x_{k-1}, x_k]} f(x) = 1$ for every $k = 1, 2, \cdots, n$. Thus,

$$\begin{aligned} L(P, f) &= \sum_{k=1}^{n} m_k \triangle x_k = 0 \\ U(P, f) &= \sum_{k=1}^{n} M_k \triangle x_k = 1 \end{aligned}$$

Since P is an arbitrary partition of $[0,1]$,

$$\sup_P L(P, f) = 0 < 1 = \inf_P U(P, f).$$

(iii) If $P_1 \subseteq P_2$, then

$$U(P_1, f) - L(P_1, f) \ge U(P_2, f) - L(P_2, f).$$

Definition 7.1.6. We write

$$\underline{\int_a^b} f(x)\, dx = \sup_P L(P, f) \quad \text{and} \quad \overline{\int_a^b} f(x)\, dx = \inf_P U(P, f)$$

which are called the "*lower integral*" and "*upper ingegral*" of f over $[a,b]$, respectively.

Remark.

(i) Suppose that f is a bounded function on $[a,b]$ and P is any partition of $[a,b]$, then

$$L(P,f) \leq \underline{\int_a^b} f(x)\,dx \leq \overline{\int_a^b} f(x)\,dx \leq U(P,f).$$

(ii) We usually write the lower and upper integral as

$$\underline{\int_a^b} f, \ \underline{\int_a^b} f\,dx, \ \text{and} \ \overline{\int_a^b} f, \ \overline{\int_a^b} f\,dx.$$

Definition 7.1.7. Let f be a bounded function on $[a,b]$. We say that f is "*(Riemann) integrable*" on $[a,b]$ if

$$\underline{\int_a^b} f(x)\,dx = \overline{\int_a^b} f(x)\,dx = A.$$

We call this number the "*definite integral*" of f on $[a,b]$ and denote

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

Remark.

(i) The symbol $\int$ is called an integral sign and $f(x)\,dx$ is called an integrand. The procedure of calculating an integral is called "*integration*".

(ii) If f is integrable on $[a,b]$ if and only if $\sup_P\ L(P,f) = \inf_P\ U(P,f) = A$.

(iii) A definite integral $\displaystyle\int_a^b f(x)\,dx$ is a number. The variable x in the preceding is a "*dummy variable*". That is, we can write

$$\int_a^b f(x)\,dx = \int_a^b f(t)\,dt = \int_a^b f(r)\,dr.$$

(iv) When $f(x) \geq 0$ on $[a,b]$, the integral of f on $[a,b]$, $\int_a^b f(x)\,dx$ is the area of $R(f,a,b)$. That is, the area of the region $R(f,a,b)$ that lies under the graph of f between a and b is the limit of the areas of approximating rectangles.

Example 7.1.8. Let $f(x) = c$ on $[a,b]$ and $P = \{a = x_0 < x_1 < \cdots < x_{n-1} < x_n\}$ be a partition of $[a,b]$. Then $m_i = c = M_i$ for $i = 1, 2, \cdots, n$. Thus

$$L(P,f) = \sum_{i=1}^n m_i(x_i - x_{i-1}) = c\sum_{i=1}^n (x_i - x_{i-1}) = c(b-a) = \sum_{i=1}^n M_i(x_i - x_{i-1}) = U(P,f).$$

Since P is an arbitrary partition of $[a,b]$,

$$\sup_P\ L(P,f) = c(b-a) = \inf_P\ U(P,f).$$

Then f is integrable on $[a,b]$ and $\displaystyle\int_a^b f(x)\,dx = c(b-a)$.

Example 7.1.9. If $f(x) = \begin{cases} 1, & x \in [a,b] \cap \mathbb{Q} \\ 0, & x \in [a,b] \backslash \mathbb{Q} \end{cases}$. Then, for any partition P of $[a,b]$,

$$L(P, f) = 0 < (b-a) = U(P, f).$$

Hence, $\sup_P L(P, f) = 0 < (b-a) = \inf_P U(P, f)$ and f is not integrable on $[a,b]$.

Example 7.1.10. Let $f(x) = x$ on $[a,b]$. Find $\int_a^b f(x)\,dx$.

Proof. Let $\triangle x = \frac{b-a}{n}$ and $P_n = \left\{ x_i \mid x_i = a + i\triangle x \text{ for } i = 0, 1, 2, \cdots, n \right\}$. Since f is an increasing function on $[a,b]$, on each subinterval $[x_{i-1}, x_i]$, $M_i = f(x_i) = a + i\triangle x = a + \frac{i(b-a)}{n}$ and $m_i = f(x_{i-1}) = a + (i-1)\triangle x = a + \frac{(i-1)(b-a)}{n}$. Then

$$\begin{aligned}
U(P_n, f) &= \sum_{i=1}^{n} \Big(a + \frac{i(b-a)}{n} \Big) \cdot \frac{b-a}{n} \\
&= a(b-a) + \Big(\frac{b-a}{n} \Big)^2 \sum_{i=1}^{n} i = \frac{1}{2}(b^2 - a^2) + \frac{(b-a)^2}{2n} \\
L(P_n, f) &= \sum_{i=1}^{n} \Big(a + \frac{((i-1)(b-a)}{n} \Big) \cdot \frac{b-a}{n} \\
&= a(b-a) + \Big(\frac{b-a}{n} \Big)^2 \sum_{i=1}^{n} (i-1) = \frac{1}{2}(b^2 - a^2) - \frac{(b-a)^2}{2n}.
\end{aligned}$$

We have

$$\frac{1}{2}(b^2 - a^2) - \frac{(b-a)^2}{2n} \le \int_a^b f(x)\,dx \le \frac{1}{2}(b^2 - a^2) + \frac{(b-a)^2}{2n}$$

for every $n \in \mathbb{N}$. The unique number makes the above inequality hold is $\frac{1}{2}(b^2 - a^2)$. Hence,

$$\int_a^b f(x)\,dx = \frac{1}{2}(b^2 - a^2).$$

□

■ <u>Areas</u>

Heuristically, for $f(x) \ge 0$ on $[a,b]$, the area of $R(f, a, b)$ under the graph $y = f(x)$ for $a \le x \le b$ is greater than any lower sum and less than any upper sum. That is, for any partition P of $[a,b]$,

$$L(P, f) \le \text{the area of } R(f, a, b) \le U(P, f).$$

Then

$$\sup_P L(P, f) \le \text{the area of } R(f, a, b) \le \inf_P U(P, f).$$

Hence, if $\sup_P L(P, f) = \inf_P U(P, f)$, we have

$$\text{the area of } R(f, a, b) = \sup_P L(P, f) = \inf_P U(P, f).$$

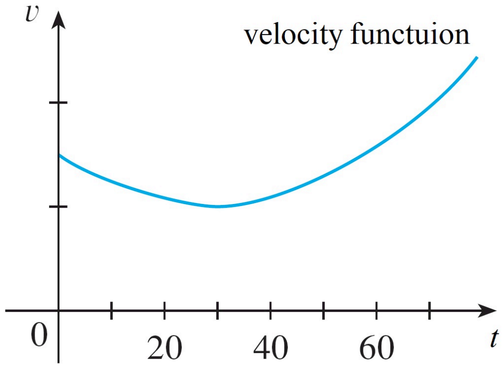
But, if $\sup L(P, f) < \inf U(P, f)$, we cannot determine the area of $R(f, a, b)$.

■ **Distances**

If the velocity of an object remains constant, then

$$\text{Distance} = \text{velocity} \times \text{time}.$$

But if the velocity varies, it is not easy to find the traveled distance.



To estimate the distance over the time period $[a, b]$, we may divide $[a, b]$ into n subinterval with width $\triangle t = \dfrac{b-a}{n}$. The distrance traveled over $[t_{i-1}, t_i]$ is approximated by

$$v(t_i^*)\triangle t.$$

Hence, the total distance is the limit of the sum of approximating rectangle

$$\text{distance} = \lim_{n\to\infty} \sum_{i=1}^{n} v(t_i^*)\triangle t = \lim_{n\to\infty} \frac{b-a}{n} \sum_{i=1}^{n} v(t_i^*).$$

7.2 Integrable Functions

Question: For a given function f on $[a, b]$, how to determine whether f is integrable?

Question: If f is integrable on $[a, b]$, how to find $\displaystyle\int_a^b f(x)\,dx$?

Theorem 7.2.1. *If f is bounded on $[a, b]$, then f is integrable on $[a, b]$ if and only if for every $\varepsilon > 0$, there is a partition P of $[a, b]$ such that*

$$U(P, f) - L(P, f) < \varepsilon.$$

Proof. ($\Longrightarrow$) Since f is bounded and integrable on $[a, b]$, by the defintion,

$$\sup_P L(P, f) = \inf_P U(P, f) = A < \infty.$$

Given $\varepsilon > 0$, there exist two partitions of $[a, b]$, P_1 and P_2, such that

$$L(P_1, f) > A - \frac{\varepsilon}{2} \quad \text{and} \quad U(P_2, f) < A + \frac{\varepsilon}{2}.$$

Let $P = P_1 \cup P_2$ be a common refinement of P_1 and P_2. Then

$$\begin{aligned} A - \frac{\varepsilon}{2} &< L(P_1, f) \le L(P, f) \text{ and} \\ A + \frac{\varepsilon}{2} &> U(P_2, f) \ge U(P, f). \end{aligned}$$

Therefore,

$$0 \le U(P, f) - L(P, f) < (A + \frac{\varepsilon}{2}) - (A - \frac{\varepsilon}{2}) = \varepsilon.$$

($\Longleftarrow$) Given $\varepsilon > 0$, let P_ε be a partition of $[a, b]$ such that

$$U(P_\varepsilon, f) - L(P_\varepsilon, f) < \varepsilon.$$

Then

$$\inf_P U(P, f) - \sup_P L(P, f) \le U(P_\varepsilon, f) - L(P_\varepsilon, f) < \varepsilon.$$

Since ε is an arbitrary positive number, $\inf_P U(P, f) = \sup_P L(P, f)$ and hence f is integrable on $[a, b]$. □

Theorem 7.2.2. *If a function f is monotonic on $[a, b]$, then it is integrable on $[a, b]$.*

Proof. W.L.O.G, we may assume that f is increasing on $[a, b]$. Then f is a bounded function with $f(a) = \min_{x \in [a,b]} f(x)$ and $f(b) = \max_{x \in [a,b]} f(x)$.
Given $\varepsilon > 0$, we wan to choose a partition P of $[a, b]$ such that $U(P, f) - L(P, f) < \varepsilon$. Let $\triangle x = \dfrac{b - a}{n}$ and $P = \{x_i = a + i\triangle x \mid i = 0, 1, 2, \cdots, n\}$ where n will be determined later. On each subinterval $[x_{i-1}, x_i]$, $M_i = f(x_i)$ and $m_i = f(x_{i-1})$. We have

$$\begin{aligned} U(P, f) - L(P, f) &= \sum_{i=1}^{n} M_i \triangle x - \sum_{i=1}^{n} m_i \triangle x = \sum_{i=1}^{n} (M_i - m_i)\triangle x \\ &= \frac{b - a}{n} \sum_{i=1}^{n} [f(x_i) - f(x_{i-1})] = \frac{b - a}{n} [f(b) - f(a)]. \end{aligned}$$

We can choose n sufficiently large such that $\dfrac{(b - a)[f(b) - f(a)]}{n} < \varepsilon$ and thus $U(P, f) - L(P, f) < \varepsilon$. Since ε is an arbitrarily positive number, f is integrable on $[a, b]$. □

❑ Uniform Continuity

Review: A function f is continuous on $[a, b]$ if for every $x \in [a, b]$ and given $\varepsilon > 0$, there exists $\delta = \delta(\varepsilon, x) > 0$ such that

$$|f(x) - f(y)| < \varepsilon$$

for every $y \in [a, b]$ and $|x - y| < \delta$.

Note that the number δ depends not only on ε but also on the point x. It could be different when x varies. For example, $f(x) = x^2$ is continuous on $\mathbb{R}$. We can show that

$$\left|x^2 - a^2\right| < \varepsilon \quad \text{whenever } |x - a| < \delta = \min\left(1, \frac{\varepsilon}{1 + 2|a|}\right).$$

On the other hand, we consider $f(x) = x^2$ on $[-2, 2]$. For every point $a \in [-2, 2]$, we can choose $\delta = \min\left(1, \frac{\varepsilon}{5}\right)$ which is independent of the point in $[-2, 2]$. Then $|x^2 - a^2| < \varepsilon$ whenever $|x - y| < \delta$. Hence, the function $f(x) = x^2$ may have different behaviors if the domain varies.

Definition 7.2.3. Let f be a function defined on $D \subseteq \mathbb{R}$ and $E \subseteq D$. We say that f is "*uniformly continuous*" on E if for any $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon) > 0$ such that

$$|f(x) - f(y)| < \varepsilon$$

for all $x, y \in E$ satisfying $|x - y| < \delta$. If f is uniformly continuous on its domain D, we simply say that f is uniformly continuous.

Remark.

(i) If f is uniformly continuous on E, then it is also continuous on E.

(ii) If f is not uniformly continuous on E, there exists $\varepsilon > 0$ such that for every $\delta > 0$ there exist $x, y \in E$ with $|x - y| < \delta$ satisfying $|f(x) - f(y)| \geq \varepsilon$.

This statement is equivalent that "there exists $\varepsilon > 0$ such that for every $n \in \mathbb{N}$ we can choose a sequence of pairs of points $x_n, y_n \in E$ such that $|x_n - y_n| \to 0$ as $n \to \infty$, but $|f(x_n) - f(y_n)| \geq \varepsilon$."

Example 7.2.4. Prove that $f(x) = x^2$ is not uniformly continuous.

Proof. Let $\varepsilon = 1$. We want to find x_n, y_n such that $|x_n - y_n| < \dfrac{1}{n}$ and $|x_n^2 - y_n^2| \geq 1$ for every $n \in \mathbb{N}$. Consider

$$|x_n^2 - y_n^2| = |x_n + y_n||x_n - y_n|.$$

Choose $x_n = n$ and $y_n = n + \dfrac{1}{n}$. Then

$$\left|x_n^2 - y_n^2\right| = \frac{1}{n} \cdot (2n + \frac{1}{n}) = \left|2 + \frac{1}{n^2}\right| > 1.$$

Hence, $f(x) = x^2$ is not uniformly continuous on $\mathbb{R}$. □

Exercise.

(i) Prove that $f(x) = \dfrac{1}{x}$ is uniformly continuous on $(\dfrac{1}{n}, \infty)$ for every $n \in \mathbb{N}$ but not uniformly continuous on $(0, \infty)$.

(ii) Prove that $f(x) = \sin x$ is uniformly continuous on $\mathbb{R}$.

Theorem 7.2.5. *If f is continuous on $[a, b]$, then f is uniformly continuous on $[a, b]$.*

Proof. We will prove this theorem by a contradition. Assume that f is not uniformly continuous on $[a, b]$. There exists $\varepsilon > 0$ and pair of sequences $\{s_n\}_{n=1}^{\infty}$, $\{t_n\}_{n=1}^{\infty} \subset [a, b]$ such that $|s_n - t_n| < \dfrac{1}{n}$ and $\left|f(s_n) - f(t_n)\right| > \epsilon$. We want to prove that $[a, b]$ cannot contain such pair of sequences. Define

$$A = \left\{x \in [a, b] \,\middle|\, [a, x] \text{ contains infinitely many } s_n\right\}.$$

Step1: To show that $a \in A$. (That is, $\{s_n\}$ contains at most finitely many "a".)

Since f is continuous at a, for the givne ε, there exists $\delta > 0$ such that for $x \in [a, b]$ and $|x - a| < \delta$ then $\left|f(x) - f(a)\right| < \varepsilon$. Hence, if $s_n = a$ and $|t_n - s_n| = |t_n - a| < \frac{1}{n}$ such that

$\left|f(t_n) - f(s_n)\right| = \left|f(t_n) - f(a)\right| > \varepsilon$, we have $1/n$ cannot be less than δ. Hence, $n \le 1/\delta$.

Since A is nonempty and b is an upper bound for A, by the least upper bound property, there exists $c \in [a, b]$ such that $c = \sup A$.
Step 2: To show that $c = b$.

Assume that $c < b$. Since f is continuous at c. For the given ε, there exists $\delta_1 > 0$ such that for every $y \in [a, b]$ and $|y - c| < \delta_1$ then

$$\left|f(y) - f(c)\right| < \frac{\varepsilon}{2}.$$

Note that we can choose δ_1 sufficiently small such that $(c - \delta_1, c + \delta_1) \subset (a, b)$. We claim that $(c - \frac{\delta_1}{2}, c + \frac{\delta_1}{2})$ contains infinitely many s_n. Suppose that the claim is false (that is, it contains finitely many s_n). Since $c = \sup A$, $[a, c - \frac{\delta_1}{4})$ contains finitely many $\{s_n\}$. Combinig the above two results, the interval $[a, c+\frac{\delta_1}{2})$ also contains finitely many s_n. This implies $\sup A \ge c+\frac{\delta_1}{2} > c$.

Now, we will show that the inequality $c < b$ is false. Since $(c - \frac{\delta_1}{2}, c + \frac{\delta_1}{2})$ contains infinitely many s_n, we can choose sufficiently large $N \in \mathbb{N}$ such that $\dfrac{1}{N} < \dfrac{\delta_1}{2}$ and $s_N \in (c - \frac{\delta_1}{2}, c + \frac{\delta_1}{2})$. Since $|t_N - s_N| < \dfrac{1}{N} < \dfrac{\delta_1}{2}$, we have $|t_N - c| < \delta_1$. Then

$$\left|f(t_N) - f(s_N)\right| \le \left|f(t_N) - f(c)\right| + \left|f(c) - f(s_N)\right| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

It contradicts the choice of $\{s_n\}$ and $\{t_n\}$. Hence $c = b$.
Step 3: To show that $A = [a, b]$. It means that $[a, b]$ contains only finitely many $\{s_n\}$ and hence such choice of pair of sequence $\{s_n\}$ and $\{t_n\}$ does not exist.

For Stpe 2, $b = \sup A$. We will show that $b \in A$. The process is similar as the proof of Step 1. Since f is continuous at b, for the given ε there exists $\delta_2 > 0$ such that for $y \in [a, b]$ and $|y - b| < \delta_2$, then $\left|f(y) - f(b)\right| < \frac{\varepsilon}{2}$. Also, since $b = \sup A$, we can show that $(b - \frac{\delta_2}{2}, b]$ contains infinitely many s_n. Choose a sufficiently large $M \in \mathbb{N}$ with $\frac{1}{M} < \frac{\delta_2}{2}$ such that $s_M \in (b - \frac{\delta_2}{2}, b]$. Then $|t_M - b| \le |t_M - s_M| + |s_M - b| < \delta_2$ and

$$\left|f(s_M) - f(t_M)\right| \le \left|f(s_M) - f(b)\right| + \left|f(b) - f(t_M)\right| \le \frac{\varepsilon}{2} + \frac{\varepsilon}{2}.$$

It contradicts that choice of $\{s_n\}$ and $\{t_n\}$. We have $b \in A$ and the theorem is proved.

□

Remark. In Theorem 7.2.5, the hypothesis that closedness and boundedness of the interval are necessary. For example, $f(x) = 1/x$ on $(0, 1)$ and $g(x) = x^2$ on $(1, \infty)$ are not uniformly continuous.

Theorem 7.2.6. *If a function f is continuous on $[a, b]$, then it is integrable on $[a, b]$.*

Proof. Since f is continuous on $[a, b]$, it is also uniformly continuous on $[a, b]$. Given $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\left|f(x) - f(y)\right| < \frac{\varepsilon}{b - a}$$

whenever $|x - y| < \delta$. Choose a partition $P = \left\{x_0, x_1, \cdots, x_n\right\}$ of $[a, b]$ with $|P| = \max\limits_{1 \le i \le n} \triangle x_i < \delta$. On each subinterval $[x_{i-1}, x_i]$, by the extreme value theorem, there exists $s_i, t_i \in [x_{i-1}, x_i]$ such that $M_i = f(s_i)$ and $m_i = f(t_i)$. Since $|s_i - t_i| < \delta$, we have $M_i - m_i < \dfrac{\varepsilon}{b-a}$ for $i = 1, 2, \cdots, n$. Then

$$\begin{aligned} U(P,f) - L(P,f) &= \sum_{i=1}^{n}(M_i - m_i)(x_i - x_{i-1}) \\ &< \frac{\varepsilon}{b-a}\sum_{i=1}^{n}(x_i - x_{i-1}) \\ &= \frac{\varepsilon}{b-a}\cdot(b-a) = \varepsilon. \end{aligned}$$

Hence, f is integrable on $[a, b]$. □

Theorem 7.2.7. *If f is bounded on $[a, b]$ and is continuous on $[a, b]$ except at one point, then f is integrable on $[a, b]$. Moreover,*

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$$

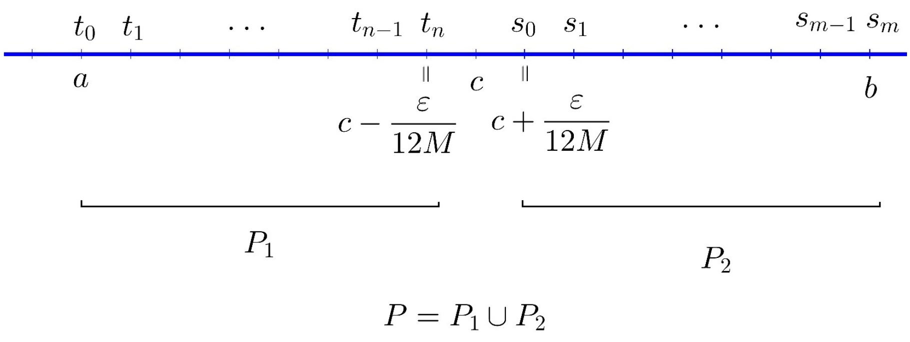
Proof. Since f is bounded on $[a, b]$, there exists $M > 0$ such that $|f(x)| < M$ for every $x \in [a, b]$. For given $\varepsilon > 0$, we will prove that there exists a partition P of $[a, b]$ such that

$$U(P,f) - L(P,f) < \varepsilon.$$

Suppose that f is discontinuous at only one point $c \in [a, b]$. Then f is continuous on $[a, c - \frac{\varepsilon}{12M}] \cup [c + \frac{\varepsilon}{12M},$ Hence, f is integrable on $[a, c - \frac{\varepsilon}{12M}]$ and on $[c + \frac{\varepsilon}{12M}, b]$. There exists a partition $P_1 = \{t_0, t_1, \cdots, t_n\}$ of $[a, c - \frac{\varepsilon}{12M}]$ and a partition $P_2 = \{s_0, s_1, \cdots, s_m\}$ of $[c + \frac{\varepsilon}{12M}, b]$ such that

$$U(P_1,f) - L(P_1,f) < \frac{\varepsilon}{3} \quad \text{and} \quad U(P_2,f) - L(P_2,f) < \frac{\varepsilon}{3}.$$

Note that, in the partitions P_1 and P_2, $t_0 = a$, $t_n = c - \frac{\varepsilon}{12M}$, $s_0 = c + \frac{\varepsilon}{12M}$ and $s_m = b$. Let $P = P_1 \cup P_2$. Then P is a partition of $[a, b]$.



Define

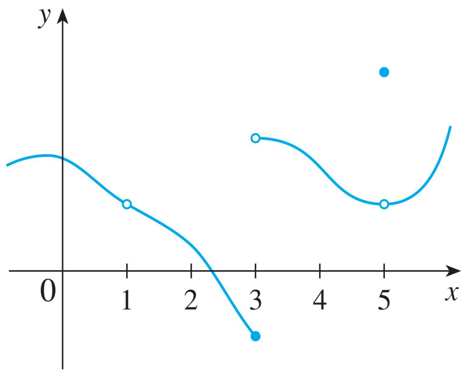
$$\begin{aligned} M_i &= \sup_{x\in[t_{i-1},t_i]} f(x) \quad &\text{and}\quad m_i &= \inf_{x\in[t_{i-1},t_i]} f(x) \\ M'_j &= \sup_{x\in[s_{j-1},s_j]} f(x) \quad &\text{and}\quad m'_j &= \inf_{x\in[s_{j-1},s_j]} f(x) \end{aligned}$$

Then

$$\begin{aligned} U(P,f) - L(P,f) &= \sum_{i=1}^{n} M_i(t_i - t_{i-1}) + \Big(\sup_{x\in[t_n,s_0]} f(x)\Big)(s_0 - t_n) + \sum_{j=1}^{m} M_j'(s_j - s_{j-1}) \\ &\quad - \sum_{i=1}^{n} m_i(t_i - t_{i-1}) + \Big(\inf_{x\in[t_n,s_0]} f(x)\Big)(s_0 - t_n) + \sum_{j=1}^{m} m_j'(s_j - s_{j-1}) \\ &= \Big[U(P_1,f) - L(P_1,f)\Big] + \Big[U(P_2,f) - L(P_2,f)\Big] \\ &\quad + \Big[\sup_{x\in[t_n,s_0]} f(x) - \inf_{x\in[t_n,s_0]} f(x)\Big](s_0 - t_n) \\ &< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + 2M \cdot \frac{\varepsilon}{6M} = \varepsilon. \end{aligned}$$

Hence, f is integrable on $[a,b]$. □

Definition 7.2.8. A function f is called "*piecewise continuous*" on an interval I if there exists finitely many points $x_1, x_2, \cdots, x_n$ in I such that f is continuous on I except at $x_1, x_2, \cdots, x_n$ and f has removable or jump discontinuities at $x_1, x_2, \cdots, x_n$.



Corollary 7.2.9. *If f is piecewise continuous on $[a,b]$, then f is integrable on $[a,b]$.*

Proof. Exercise! □

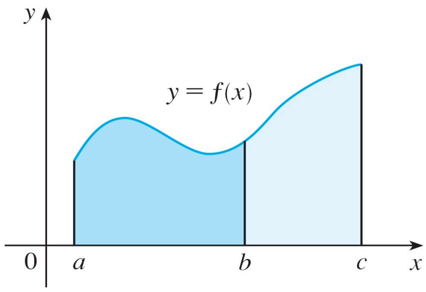
7.3 Properties of the Integrals

Theorem 7.3.1.

(a) *If f is integrable on $[a,b]$ and $[c,d] \subseteq [a,b]$, then f is integrable on $[c,d]$.*

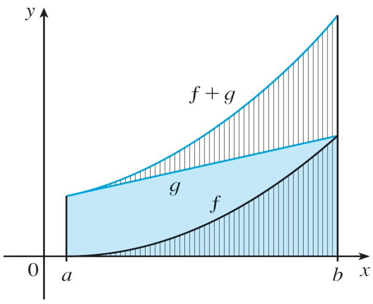
(b) *For $a < b < c$, if f is integrable on $[a,b]$ and is integrable on $[b,c]$, then f is integrable on $[a,c]$. Moreover,*

$$\int_a^c f(x)\,dx = \int_a^b f(x)\,dx + \int_b^c f(x)\,dx.$$



(c) If f and g are integrable on $[a, b]$, then $f \pm g$ is integrable on $[a, b]$. Moreover,

$$\int_a^b (f \pm g)(x)\, dx = \int_a^b f(x)\, dx \pm \int_a^b g(x)\, dx.$$



(d) If f is integrable on $[a, b]$ and $\alpha \in \mathbb{R}$, then αf is integrable on $[a, b]$. Moreover

$$\int_a^b (\alpha f)(x)\, dx = \alpha \int_a^b f(x)\, dx.$$

(e) $\displaystyle\int_a^a f(x)\, dx = 0.$

Proof. We will prove (c) here and the proofs of other parts are left to the readers.
Since f and g are integrable on $[a, b]$, for given $\varepsilon > 0$, there exist partitions, P_1 and P_2, of $[a, b]$ such that

$$U(P_1, f) - L(P_1, f) < \frac{\varepsilon}{2} \quad \text{and} \quad U(P_2, g) - L(P_2, g) < \frac{\varepsilon}{2}.$$

Let $P = P_1 \cup P_2 = \left\{a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b\right\}$ be the common refinement of P_1 and P_2. Then

$$U(P, f) - L(P, f) < \frac{\varepsilon}{2} \quad \text{and} \quad U(P, g) - L(P, g), \frac{\varepsilon}{2}.$$

Define

$$M_i = \sup_{x \in [x_{i-1}, x_i]} (f + g)(x), \; M_i' = \sup_{x \in [x_{i-1}, x_i]} f(x) \text{ and } M_i'' = \sup_{x \in [x_{i-1}, x_i]} g(x)$$
$$m_i = \inf_{x \in [x_{i-1}, x_i]} (f + g)(x), \; m_i' = \inf_{x \in [x_{i-1}, x_i]} f(x) \text{ and } m_i'' = \inf_{x \in [x_{i-1}, x_i]} g(x).$$

Then $M_i \le M_i' + M_i''$ and $m_i \ge m_i' + m_i''$. We have

$$\begin{aligned}
U(P, f+g) - L(P, f+g) &= \sum_{i=1}^{n} M_i(x_i - x_{i-1}) - \sum_{i=1}^{n} m_i(x_i - x_{i-1}) \\
&\le \sum_{i=1}^{n} (M_i' + M_i'')(x_i - x_{i-1}) - \sum_{i=1}^{n} (m_i' + m_i'')(x_i - x_{i-1}) \\
&= U(P, f) - L(P, f) + U(P, g) - L(P, g) \\
&< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.
\end{aligned}$$

Hence, $f + g$ is integrable on $[a, b]$. Moreover, since

$$\begin{aligned}
L(P, f) \le \int_a^b f(x)\,dx \le U(P, f) \quad &\text{with} \quad U(P, f) - L(P, f) < \frac{\varepsilon}{2} \ \text{ and} \\
L(P, g) \le \int_a^b g(x)\,dx \le U(P, g) \quad &\text{with} \quad U(P, g) - L(P, g) < \frac{\varepsilon}{2},
\end{aligned}$$

we have

$$L(P, f) + L(P, g) \le \int_a^b f(x)\,dx + \int_a^b g(x)\,dx \le U(P, f) + U(P, g).$$

Also,

$$L(P, f) + L(P, g) \le L(P, f+g) \le \int_a^b (f+g)(x)\,dx \le U(P, f+g) \le U(P, f) + U(P, g).$$

Hnece,

$$\left| \int_a^b (f+g)(x)\,dx - \Big(\int_a^b f(x)\,dx + \int_a^b g(x)\,dx \Big) \right| \le \left| \Big(U(P, f) + U(P, g) \Big) - \Big(L(P, f) + L(P, g) \Big) \right| < \varepsilon.$$

Thus,

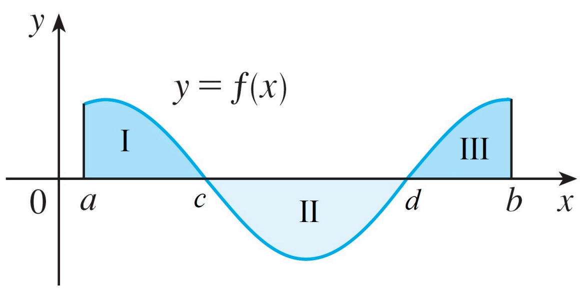
$$\int_a^b (f+g)(x)\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx.$$

□

Exercise. If f is integrable on $[a, b]$, prove that $|f(x)|$ is also integrable on $[a, b]$.

Example 7.3.2. Let $a < c < d < b$. If f is integrable on $[a, b]$, $f(x) \ge 0$ on $[a, c] \cup [d, b]$ and $f(x) \le 0$ on $[c, d]$ as the figure. Then

$$\begin{aligned}
&\int_a^b f(x)\,dx \\
= &\int_a^c f(x)\,dx + \int_c^d f(x)\,dx + \int_d^b f(x)\,dx \\
= &\int_a^c f(x)\,dx - \int_c^d (-1)f(x)\,dx + \int_d^b f(x)\,dx \\
= &\ \text{Area of } I - \text{ Area of } II + \text{ Area of } III.
\end{aligned}$$



Moreover, the area between the graph of f and x-axis is

$$\begin{aligned}\int_a^b |f(x)|\,dx &= \int_a^c |f(x)|\,dx + \int_c^d |f(x)|\,dx + \int_d^b |f(x)|\,dx \\ &= \text{Area of } I + \text{ Area of } II + \text{ Area of } III.\end{aligned}$$

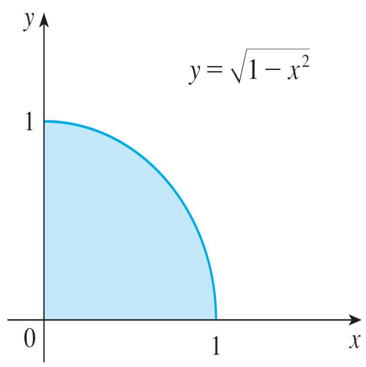
Example 7.3.3. Evaluate $\displaystyle\int_0^1 \sqrt{1-x^2}\,dx$.

Proof.

If we try to find the upper sum or lower sum, it is not easy to evaluate the form

$$\sum_{i=1}^{n} \sqrt{1-x_i}\Delta x_i.$$

Hence, we try to think that $\displaystyle\int_0^1 \sqrt{1-x^2}\,dx$ represents the area under the curve $y = \sqrt{1-x^2}$ from 0 to 1. Hence,



$$\int_0^1 \sqrt{1-x^2}\,dx = \frac{1}{4}\pi \cdot 1^2 = \frac{\pi}{4}.$$

□

Remark. So far, we only consider the integral $\displaystyle\int_a^b f(x)\,dx$ for $a \le b$. Can we define $\displaystyle\int_b^a f(x)\,dx$ for $a < b$? If the integral $\displaystyle\int_b^a f(x)\,dx$ is well-defined, by the part(2) of Theorem 7.3.1,

$$\int_a^b f(x)\,dx + \int_b^a f(x)\,dx = \int_a^a f(x)\,dx = 0.$$

Then $\int_b^a f(x)\,dx = -\int_a^b f(x)\,dx$. Therefore, for any $a, b \in \mathbb{R}$, we define

$$\int_b^a f(x)\,dx = -\int_a^b f(x)\,dx.$$

■ Comparison Properties of the Integral

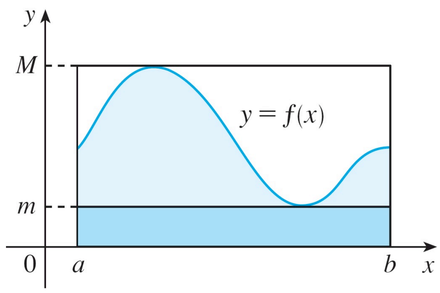
(f) If $f(x) \ge 0$ on $[a, b]$, then $\displaystyle\int_a^b f(x)\,dx \ge 0$.

(g) If f and g are integrable on $[a, b]$ and $f(x) \le g(x)$ for all $x \in [a, b]$, then

$$\int_a^b f(x)\,dx \le \int_a^b g(x)\,dx.$$

(h) If f is integrable on $[a, b]$ and $m \leq f(x) \leq M$ for all $x \in [a, b]$, then

$$m(b-a) \leq \int_a^b f(x)\,dx \leq M(b-a).$$

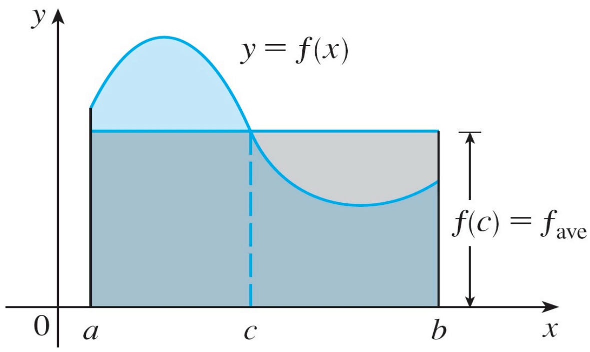


Corollary 7.3.4. *(Mean Value Theorem of Integarls) If f is continuous on $[a, b]$, then there exists a number $c \in [a, b]$ such that*

$$f(c) = f_{ave} = \frac{1}{b-a}\int_a^b f(x)\,dx,$$

that is,

$$\int_a^b f(x)\,dx = f(c)(b-a)$$



Proof. Since f is continuous on $[a, b]$, by the extreme value theorem, there exist $\alpha, \beta \in [a, b]$ such that $M = f(\alpha) = \max_{x\in[a,b]} f(x)$ and $m = f(\beta) = \min_{x\in[a,b]} f(x)$. By the property(h),

$$m \leq \frac{1}{b-a}\int_a^b f(x)\,dx \leq M.$$

That is, the value $\frac{1}{b-a}\int_a^b f(x)\,dx$ is a number between $f(\alpha)$ and $f(\beta)$. By the mean value theorem for continuous functions, there exists a number c between α and β such that

$$f(c) = \frac{1}{b-a}\int_a^b f(x)\,dx.$$

□

7.4 Riemann Sums

Suppose that $P = \{x_0, x_1, \cdots, x_n\}$ is a partition of $[a, b]$, and that for each i we choose some point x_i^* in $[x_{i-1}, x_i]$. Then we have

$$L(P, f) \le \sum_{i=1}^{n} f(x_i^*)(x_i - x_{i-1}) \le U(P, f).$$

Any sum $\displaystyle\sum_{i=1}^{n} f(x_i^*)(x_i - x_{i-1})$ is called a "*Riemann sum*" of f for P.

Theorem 7.4.1. *Suppose that f is integrable on $[a, b]$. Then for every $\varepsilon > 0$, there exists $\delta > 0$ such that if $P = \{x_0, \cdots, x_n\}$ is any partition of $[a, b]$ with $\|P\| < \delta$ then*

$$\Big| \sum_{i=1}^{n} f(x_i^*)(x_i - x_{i-1}) - \int_a^b f(x)\, dx \Big| < \varepsilon$$

for any $x_i^* \in [x_{i-1}, x_i]$.

Proof. Firstly, we consider that f is continuous on $[a, b]$. Hence, it is integrable and uniformly continuous on $[a, b]$. For given $\varepsilon > 0$, there exists $\delta > 0$ such that if $|x - y| < \delta$, then

$$\big| f(x) - f(y) \big| < \frac{\varepsilon}{b - a}.$$

Let $P = \{x_0, x_1, \cdots, x_n\}$ be a partition of $[a, b]$ with $\|P\| < \delta$. Then $M_i - m_i < \dfrac{\varepsilon}{b - a}$ for $i = 1, \cdots, n$. We have

$$U(P, f) - L(P, f) = \sum_{i=1}^{n} (M_i - m_i)(x_i - x_{i-1}) < \frac{\varepsilon}{b - a} \sum_{i=1}^{n} (x_i - x_{i-1}) = \varepsilon.$$

Also, since $L(P, f) \le \displaystyle\int_a^b f(x)\, dx \le U(P, f)$ and $L(P, f) \le \displaystyle\sum_{i=1}^{n} f(x_i^*)(x_i - x_{i-1}) \le U(P, f)$, we have

$$\Big| \sum_{i=1}^{n} f(x_i^*)(x_i - x_{i-1}) - \int_a^b f(x)\, dx \Big| < \varepsilon.$$

Moreover, for a general integrable function f, we will use a known fact that there exist continuous functions g and h on $[a, b]$ satisfying $g \le f \le h$ and

$$\int_a^b h(x)\, dx - \int_a^b g(x)\, dx < \varepsilon.$$

We have

$$\sum_{i=1}^{n} g(x_i^*)(x_i - x_{i-1}) \le \sum_{i=1}^{n} f(x_i^*)(x_i - x_{i-1}) \le \sum_{i=1}^{n} h(x_i^*)(x_i - x_{i-1})$$

Since g and h are both uniformly continuous on $[a, b]$, we can choose $\delta > 0$ such that for a partition P with $\|P\| < \delta$,

$$\Big|\sum_{i=1}^{n} g(x_i^*)(x_i - x_{i-1}) - \int_a^b g(x)\,dx\Big| < \varepsilon \quad \text{and} \quad \Big|\sum_{i=1}^{n} h(x_i^*)(x_i - x_{i-1}) - \int_a^b h(x)\,dx\Big| < \varepsilon$$

Hence,

$$\begin{aligned}
&\Big|\sum_{i=1}^{n} g(x_i^*)(x_i - x_{i-1}) - \sum_{i=1}^{n} h(x_i^*)(x_i - x_{i-1})\Big| \\
\leq\ &\Big|\sum_{i=1}^{n} g(x_i^*)(x_i - x_{i-1}) - \int_a^b g(x)\,dx\Big| + \Big|\int_a^b g(x)\,dx - \int_a^b h(x)\,dx\Big| \\
&+\Big| - \int_a^b h(x)\,dx - \sum_{i=1}^{n} h(x_i^*)(x_i - x_{i-1})\Big| < 3\varepsilon.
\end{aligned}$$

Hence,

$$\begin{aligned}
\Big|\sum_{i=1}^{n} f(x_i^*)(x_i - x_{i-1}) - \int_a^b f(x)\,dx\Big| &\leq \Big|\sum_{i=1}^{n} f(x_i^*)(x_i - x_{i-1}) - \sum_{i=1}^{n} g(x_i^*)(x_i - x_{i-1})\Big| \\
&+\Big|\sum_{i=1}^{n} g(x_i^*)(x_i - x_{i-1}) - \int_a^b g(x)\,dx\Big| + \Big|\int_a^b g(x)\,dx - \int_a^b f(x)\,dx\Big| \\
&\leq \Big|\sum_{i=1}^{n} g(x_i^*)(x_i - x_{i-1}) - \sum_{i=1}^{n} h(x_i^*)(x_i - x_{i-1})\Big| + 2\varepsilon \\
&< 5\varepsilon.
\end{aligned}$$

□

Remark. If f is integrable on $[a, b]$, then

$$\lim_{\|P\|\to 0} \sum_{i=1}^{n} f(x_i^*)(x_i - x_{i-1}) = \int_a^b f(x)\,dx.$$

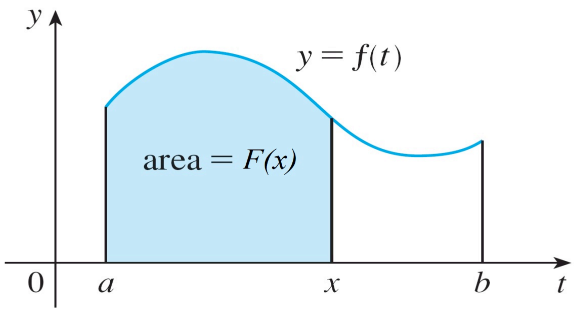
The above sample point x_i^* and $\triangle x_i$ vary as the partitions change.

7.5 The Fundamental Theorem of Calculus

Suppose that f is integrable function on $[a, b]$. For every number x in $[a, b]$, we can define a new function by

$$F(x) := \int_a^x f(t)\,dt.$$

If $f(t) \geq 0$ on $[a, b]$, then the function $F(x)$ represents the area of the region under the graph $y = f(t)$ from $t = a$ to $t = x$

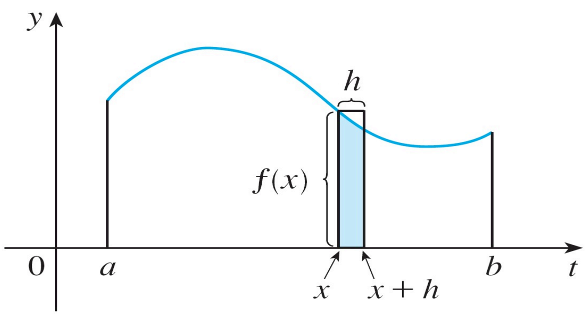


In fact, the function $F(x)$ is continuous on $[a, b]$ and the proof is left to the readers.

Exercise. If f is integrable on $[a, b]$ and $F(x)$ is defined as above, then $F(x)$ is continuous on $[a, b]$.

Question: Is $F(x)$ differentiable?

In order to compute $F'(x)$ from the definition of a derivative, we first observe that, for $h > 0$, $F(x + h) - F(x)$ is obtained by subtracting areas, so it is the area under the graph of f from x to $x + h$.



For small h, the area is approximately equal to the area of the rectangle with height $f(x)$ and width h.

$$F(x + h) - F(x) \approx hf(x)$$

an so

$$\frac{F(x + h) - F(x)}{h} \approx f(x).$$

Intuitively, we expect that

$$F'(x) = \lim_{h \to 0} \frac{F(x + h) - F(x)}{h} = f(x).$$

Theorem 7.5.1. *(Fundamental Theorem of Calculus, Part I) If f is integrable on $[a, b]$ and is continuous at c for some $c \in [a, b]$ then*

$$F(x) = \int_a^x f(t)dt$$

is differentiable and $F'(c) = f(c)$.

Proof. We will prove the theorem for $c \in (a, b)$ and the proof is similar for $c = a$ or $c = b$. Since f is continuous at c, for given $\varepsilon > 0$, there exists $\delta > 0$ such that for all $|t - c| < \delta$,

$$|f(t) - f(c)| < \varepsilon.$$

This implies that

$$f(c) - \varepsilon < f(t) < f(c) + \varepsilon \quad \text{for } t \in (c - \delta, c + \delta).$$

For $0 < h < \delta$,

$$F(c + h) - F(c) = \int_a^{c+h} f(t)\,dt - \int_a^c f(t)\,dt = \int_c^{c+h} f(t)\,dt.$$

Hence,

$$\big(f(c) - \varepsilon\big)h < F(c + h) - F(c) < \big(f(c) + \varepsilon\big)h.$$

We have

$$f(c) - \varepsilon < \frac{F((c + h) - F(c)}{h} < f(c) + \varepsilon. \tag{7.1}$$

Similarly, for $-\delta < h < 0$,

$$(-h)\big(-f(c)-\varepsilon\big) < F(c+h)-F(c) = \int_c^{c+h} f(t)\,dt = -\int_{c+h}^c f(t)\,dt = \int_{c+h}^c -f(t)\,dt < (-h)\big(-f(c)+\varepsilon\big).$$

Then

$$-f(c) - \varepsilon < \frac{F(c + h) - F(c)}{-h} < -f(c) + \varepsilon.$$

We have

$$f(c) - \varepsilon < \frac{F(c + h) - F(c)}{h} < f(c) + \varepsilon. \tag{7.2}$$

By (7.1) and (7.2), for $0 < |h| < \delta$,

$$\left|\frac{F(c + h) - F(c)}{h} - f(c)\right| < \varepsilon.$$

Hence,

$$F'(c) = \lim_{h\to 0} \frac{F(c + h) - F(c)}{h} = f(c).$$

□

Remark. Suppose that f is integrable on $[a, b]$ and is continuous at c.

(i) For $x \in [a, b]$, let

$$\begin{aligned}
F_1(x) &= \int_x^b f(t)\,dt = \int_a^b f(t)\,dt - \int_a^x f(t)\,dt = \int_a^b f(t)\,dt - F(x).\\
F_2(x) &= \int_x^a f(t)\,dt = -\int_a^x f(t)\,dt = -F(x).\\
F_3(x) &= \int_x^b f(t)\,dt = -\int_x^b f(t)\,dt = -F_1(x).
\end{aligned}$$

Then

$$F_1'(c) = -F'(c) = -f(c), \quad F_2'(c) = -F'(c) = -f(c) \quad \text{and} \quad F_3'(c) = -F_1'(c) = f(c).$$

(ii) If f is continuous on $[a, b]$, then $F(x)$ is differentiable on $[a, b]$ and $F'(x) = f(x)$.

Corollary 7.5.2. *If f is continuous on $[a, b]$ and $f = g'$ for some function g, then*

$$\int_a^x f(t)\,dt = g(x) - g(a) \quad \text{and} \quad \int_a^b f(x)\,dx = g(b) - g(a).$$

Proof. Let $F(x) = \int_a^x f(t)\,dt$. Then $F'(x) = f(x) = g'(x)$. Hence, $F(x) = g(x) + C$ for some constant C. We have $0 = F(a) = g(a) + C$ and then $C = -g(a)$. Thus

$$\int_a^x f(t)\,dt = F(x) = g(x) - g(a) \quad \text{and} \quad \int_a^b f(x)\,dx = g(b) - g(a).$$

□

Remark. This corollary seems to be useless. If $g(x) = \displaystyle\int_a^x f(t)\,dt$ then $g'(x) = f(x)$ and $g(a) = 0$. We have

$$\int_a^b f(t)\,dt = g(b) = g(b) - g(a).$$

But the useful point is that for any g satisfying $g' = f$, then

$$\int_a^b f(t)\,dt = g(b) - g(a).$$

Example 7.5.3. Let $g(x) = \frac{1}{3}x^3$ and $f(x) = x^2$. Then $g'(x) = f(x)$ and

$$\int_a^b f(x)\,dx = \int_a^b x^2\,dt = \frac{1}{2}b^3 - \frac{1}{3}a^3 = g(b) - g(a).$$

On the other hand, let $g_1(x) = \frac{1}{3}x^3 + 5$ then $g_1'(x) = f(x)$ and

$$\int_a^b f(x)\,dx = \int_a^b x^2\,dx = g_1(b) - g_1(a).$$

Theorem 7.5.4. *(Fundamental Theorem of Calculus, Part II) If f is integrable on $[a, b]$ and $f = g'$ for some function g, then*

$$\int_a^b f(x)\,dx = g(b) - g(a) \quad (\text{denoted by } g(x)\Big|_a^b).$$

Proof. Let $P = \{x_0, x_1, \cdots, x_n\}$ be a partition of $[a, b]$. By the mean value theorem, for each $i = 1, 2, \cdot 2, n$, there exists $c_i \in (x_{i-1}, x_i)$ such that

$$g(x_i) - g(x_{i-1}) = g'(c_i)(x_i - x_{i-1}) = f(c_i)(x_i - x_{i-1}).$$

Since $m_i \leq f(c_i) \leq M_i$,

$$L(P, f) \leq \sum_{i=1}^n m_i(x_i - x_{i-1}) \leq \underbrace{\sum_{i=1}^n f(c_i)(x_i - x_{i-1})}_{=\sum_{i=1}^n g(x_i) - g(x_{i-1})} \leq \sum_{i=1}^n M_i(x_i - x_{i-1}) = U(P, f)$$

Hence, $L(P, f) \leq g(b) - g(a) \leq U(P, f)$ for any partition P. Since f is integrable, we have

$$\int_a^b f(x)\,dx = g(b) - g(a).$$

$\square$

Example 7.5.5.

(1) Let $f(x) = \sin x$ and define $F(x) = \displaystyle\int_a^x \sin t\,dt$. Then $F'(x) = \sin x$.

(2) Suppose that f is continuous on $[a, b]$ and $x^2 \in [a, b]$. Let $F(x) = \displaystyle\int_a^{x^2} f(t)\,dt$. Find $F'(x)$.

Proof. Define $G(x) = \displaystyle\int_a^x f(t)\,dt$. Then $F(x) = \displaystyle\int_a^{x^2} f(t)\,dt = G(x^2)$. By the Fundamental Theorem of Calculus, $G'(x) = f(x)$ and by the chain rule,

$$F'(x) = \frac{d}{dx}\big(G(x^2)\big) = G'(x^2) \cdot 2x = f(x^2) \cdot 2x.$$

$\square$

(3) Suppose that f is continuous on $[a, b]$ and $h(x) \in [a, b]$ is differentiable. Let $F(x) = \displaystyle\int_a^{h(x)} f(t)\,dt$. Find $F'(x)$.

Proof. Define $G(x) = \displaystyle\int_a^x f(t)\,dt$ and then $F(x) = G\big(h(x)\big)$. Hence

$$F(x) = G'\big(h(x)\big) \cdot h'(x) = f\big(h(x)\big) \cdot h'(x).$$

$\square$

(4) Suppose that f is continuous on $[a, b]$ and $g(x), h(x) \in [a, b]$ are differentiable. Let $F(x) = \displaystyle\int_{g(x)}^{h(x)} f(t)\,dt$. Then

$$\begin{aligned} F(x) = \int_{g(x)}^{h(x)} f(t)\,dt &= \int_a^{h(x)} f(t)\,dt + \int_{g(x)}^{a} f(t)\,dt \\ &= \int_a^{h(x)} f(t)\,dt - \int_a^{g(x)} f(t)\,dt. \end{aligned}$$

Therefore,

$$F'(x) = f\big(h(x)\big) \cdot h'(x) - f\big(g(x)\big) \cdot g'(x).$$

(5) Let $F(x) = \displaystyle\int_a^{x^3} \frac{1}{1 + \sin^2 t}\,dt$. Then

$$F'(x) = \frac{1}{1 + \sin^2(x^3)} \cdot 3x^2.$$

(6) Let $G(x) = \displaystyle\int_{2x}^{x^3} \frac{1}{1+\sin^2 t}\,dt$. Then

$$G(x) = \int_0^{x^3} \frac{1}{1+\sin^2 t}\,dt - \int_{2x}^{0} \frac{1}{1+\sin^2 t}\,dt = \int_0^{x^3} \frac{1}{1+\sin^2 t}\,dt - \int_0^{2x} \frac{1}{1+\sin^2 t}\,dt$$

and

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

■ **Differentiation and Integration as inverse process**

By the Fundamental Theorem of Calculus,

$$\frac{d}{dx}\int_a^x f(t)]\,dt = f(x), \qquad f \xrightarrow{\int dt} \int_a^x f(t)\,dt \xrightarrow{\frac{d}{dx}} f$$

$$\int_a^b F'(x)\,dx = F(b) - F(a), \qquad F \xrightarrow{\frac{d}{dx}} f \xrightarrow{\int_a^b dt} F(b) - F(a).$$

7.6 Antiderivatives

$$\begin{array}{llll} p(t) : \text{position function} & \xrightarrow{\frac{d}{dx}} & v(t) : \text{velocity function} & v(t) = p'(t) \\ v(t) : \text{velocity function} & \xrightarrow{??} & p(t) : \text{position function} & ? \end{array}$$

Question: For a given function f, can we find a function F such that $F'(x) = f(x)$?

Definition 7.6.1. A function F is called an "*antiderivative*" of f on $[a, b]$ if $F'(x) = f(x)$ for all $x \in [a, b]$.

Example 7.6.2.

(1) Let $f(x) = x$ and $F(x) = \frac{1}{2}x^2$. Then F is an antiderivative of f since $F'(x) = f(x)$.

(2) Let $f(x) = \sin x$. Then $F(x) = -\cos x$ is an antiderivative of f. In fact, $G(x) = -\cos x + 5$ and $H(x) = -\cos x - 100$ are also antiderivatives of f.

We recall that if $G(x) = F(x) + C$ then $G'(x) = F'(x)$. Hence, if F is an antiderivative of f, then G is also an antiderivative of f where $G(x) = F(x) + C$ for any constant number C.

Theorem 7.6.3. *If $F(x)$ is an antiderivative of $f(x)$, then $F(x) + C$ is also an antiderivative for any constant C.*

Proof. $\big(F(x) + C\big)' = F'(x) = f(x)$. □

Example 7.6.4. Find the most general antiderivative of the given functions.

(1) $f(x) = \cos x$, then $F(x) = \sin x + C$.

(2) $f(x) = x^n$, $n \in \mathbb{Q}$ and $n \geq 0$, then $F(x) = \dfrac{x^{n+1}}{n+1} + C$.

(3) $f(x) = x^{-3}$, then $F(x) = -\frac{1}{2x^2}$ is an antiderivative of f. But $f(x) = x^{-3}$ is not defined at $x = 0$. The general antiderivative of f is $-\frac{1}{2x^2} + C$ on each interval that does not contain 0. Hence the general antiderivative of $f(x) = x^{-3}$ is

$$F(x) = \begin{cases} -\dfrac{1}{2x^2} + C_1 & \text{if } x > 0 \\ -\dfrac{1}{2x^2} + C_2 & \text{if } x < 0 \end{cases}$$

Remark. If $F(x)$ and $G(x)$ are antiderivative of $f(x)$ and $g(x)$ respectively on an interval, and a and b are two constants, then

$$\big(aF(x) + bG(x)\big)' = af(x) + bg(x).$$

Hence, the general antiderivative of $af(x) + bg(x)$ is $aF(x) + bG(x) + C$ where C is an arbitrary constant.

• **Antiderivative Formulas**

Functions	Antiderivative		Functions	Antiderivatives
$cf(x)$	$cF(x)$		$\cos x$	$\sin x$
$f(x) + g(x)$	$F(x) + G(x)$		$\sin x$	$-\cos x$
x^n ($n \in \mathbb{Q}, n \neq 1$)	$\dfrac{x^{n+1}}{n+1}$		$\sec^2 x$	$\tan x$
$\dfrac{1}{\sqrt{1-x^2}}$	$\sin^{-1} x$		$\csc^2 x$	$-\cot x$
$\dfrac{1}{1+x^2}$	$\tan^{-1} x$		$\sec x \tan x$	$\sec x$
			$\csc x \cot x$	$-\csc x$

Example 7.6.5.

(1) Find all functions g such that $g'(x) = 4\sin x + \dfrac{2x^5 - \sqrt{x}}{x}$.

Proof.

$$\begin{aligned} g'(x) &= 4\sin x + 2x^4 - x^{-1/2} \\ g(x) &= 4(-\cos x) + 2 \cdot \frac{1}{5}x^5 - \frac{x^{1/2}}{1/2} + C = -4\cos x + \frac{2x^5}{5} - 2\sqrt{x} + C \end{aligned}$$

□

(2) Find f if $f'(x) = x\sqrt{x}$ and $f(1) = 2$.

Proof. The general antiderivative of $f'(x)$ is $f(x) = \frac{2}{5}x^{5/2} + C$. Since $2 = f(1) = \frac{2}{5} + C$, we have $C = \frac{8}{5}$ and

$$f(x) = \frac{2}{5}x^{5/2} + \frac{8}{5}.$$

□

Note. The definite integral $\displaystyle\int_a^b f(x)\,dx$ was defined by a complicated procedure. If we know an antiderivative F of f, the definite integral can be found by the values of $F(x)$ at only two points, a and b.

Example 7.6.6.

(1) Evaluate $\displaystyle\int_{-2}^{1} x^3\,dx$.

Proof. Let $f(x) = x^3$. Then $F(x) = \frac{1}{4}x^4$ is an antiderivative of f. We have

$$\int_{-2}^{1} x^3\,dx = F(1) - F(-2) = -\frac{15}{4}.$$

□

(2) Find the area under the cosine curve from 0 to b where $0 \le b \le \dfrac{\pi}{2}$.

Proof. Let $f(x) = \cos x$. Then $F(x) = \sin x$ is an antiderivative of f. The area is

$$A = \int_0^b \cos x\,dx = \sin x\Big|_0^b = \sin b - \sin 0 = \sin b.$$

□

Caution: The following computation is wrong.

$$\int_{-1}^{3} \frac{1}{x^2}\,dx = -\frac{1}{x}\Big|_{-1}^{3} = -\frac{1}{3} - \left(\frac{1}{-1}\right) = -\frac{4}{3}.$$

The integrand $f(x) = \dfrac{1}{x^2}$ is nonnegative but the definite integral is negative. What's wrong with this?
It is because that $f(x) = \dfrac{1}{x^2}$ is not continuous on $[-1, 3]$. The Fundamental Theorem of Calculus is not applied.

7.7 Indefinite Integrals

From the Fundamental Theorem of Calculus, we observe that there are connections between integration and differentiation(antiderivatives). We need good notation for antiderivative.

From the Fundamental Theorem of Calculus, $\int_a^x f(t)\,dt$ is an antiderivative of f and $\int_a^b f(t)\,dt = F(b) - F$ where F is an antiderivative of f. Therefore, the symbol "$\int f(x)\,dx$" denotes the antiderivative of f and is called "*indefinite integral*" of f. Thus,

$$\int f(x)\,dx = F(x) + C \quad \text{means} \quad F'(x) = f(x).$$

Remark.

(i) We regard an indefinite integral as representing an entire family of functions (one antiderivative for each value of the constant C).

(ii) A definite integral $\int_a^b f(x)\,dx$ is a number and an indefinite integral $\int f(x)\,dx$ is a function (or a family of functions).

TABLE OF INDEFINITE INTEGRALS

$\int [f(x) + g(x)]\,dx = \int f(x)\,dx + \int g(x)\,dx$	$\int cf(x)\,dx = c\int f(x)\,dx$
$\int x^n\,dx = \dfrac{x^{n+1}}{n+1} + C \quad (n \neq -1)$	$\int \dfrac{1}{x}\,dx = \ln\lvert x\rvert + C$
$\int e^x\,dx = e^x + C$	$\int a^x\,dx = \dfrac{a^x}{\ln a} + C$
$\int \sin x\,dx = -\cos x + C$	$\int \cos x\,dx = \sin x + C$
$\int \sec^2 x\,dx = \tan x + C$	$\int \csc^2 x\,dx = -\cot x + C$
$\int \sec x \tan x\,dx = \sec x + C$	$\int \csc x \cot x\,dx = -\csc x + C$
$\int \dfrac{1}{x^2+1}\,dx = \tan^{-1} x + C$	$\int \dfrac{1}{\sqrt{1-x^2}}\,dx = \sin^{-1} x + C$
$\int \sinh x\,dx = \cosh x + C$	$\int \cosh x\,dx = \sinh x + C$

Example 7.7.1. Find the general indefinite integral

$$\int x^4 - 2\sec x \tan x\,dx = \int x^4\,ds - 2\int \sec x \tan x\,dx = \frac{1}{5}x^5 - 2\sec x + C.$$

7.8 The Logarithmic and Exponential Functions

In the present section, we will discuss two families of functions, the logarithimic and exponential functions. They are very useful not only in mathematics but also on physics and other fileds.

7.8.1 The Logarithmic Function

Consider the number

$$10^n.$$

(i) For $n = 0$,

$$10^0 = 1.$$

(ii) For $n \in \mathbb{N}$,

$$10^n = \overbrace{10 \times 10 \times \cdots \times 10}^{n}.$$

(iii) For $m, n \in \mathbb{N}$,

$$\begin{aligned} 10^{m+n} &= 10^m \times 10^n \\ 10^{mn} &= \left(10^m\right)^n. \end{aligned}$$

(iv) For $n \in \mathbb{N}$, $10^{-n} \times 10^n = 10^{-n+n} = 10^0 = 1$. Then

$$10^{-n} = \frac{1}{10^n}. \tag{7.3}$$

(v) For $n \in \mathbb{N}$, $\left(10^{\frac{1}{n}}\right)^n = 10^{\frac{1}{n}\cdot n} = 10^1 = 10$. Then

$$10^{\frac{1}{n}} = \sqrt[n]{10}.$$

Hence $10^{\frac{1}{n}}$ is defined for $n \in \mathbb{N}$. Moreover, by (7.3), $10^{\frac{1}{n}}$ is defined for $n \in \mathbb{Z}\backslash\{0\}$.

(vi) For $m, n \in \mathbb{Z}$,

$$10^{\frac{m}{n}} = \left(10^{\frac{1}{n}}\right)^m = \left(\sqrt[n]{10}\right)^m.$$

Hence, 10^k is defined for $k \in \mathbb{Q}$.

Question: Can we define 10^k for $k \in \mathbb{R}\backslash\mathbb{Q}$?

If yes, we can consider the function $f(x) = 10^x$ on $\mathbb{R}$. Suppose that the function $f(x) = 10^x$ is defined. We hope that it satisfies

$$f(x+y) = 10^{x+y} = 10^x \cdot 10^y = f(x)f(y) \quad \text{for every } x, y \in \mathbb{R}.$$

Moreover, assume that such a function is differentiable

$$f'(x) = \lim_{h\to 0}\frac{f(x+h)-f(x)}{h} = \lim_{h\to 0}\frac{f(x)f(h)-f(x)}{h} = f(x)\Big(\lim_{h\to 0}\frac{f(h)-1}{h}\Big) = f(x)f'(0).$$

On the other hand, assume that $f(x)$ has an inverse, say $f^{-1}(x) = \log_{10} x$ and hence $f^{-1}(1) = 0$. Then

$$\left(f^{-1}\right)'(x) = \frac{1}{f'\left(f^{-1}(x)\right)} = \frac{1}{f'(0)f\left(f^{-1}(x)\right)} = \frac{1}{f'(0)}\frac{1}{x}.$$

This gives an idea to define $f^{-1}(x) = \log_{10} x$. We expect that

$$\log_{10} x = \int_1^x (f^{-1})'(t)\,dt = \int_1^x \frac{1}{f'(0)}\frac{1}{t}\,dt = \frac{1}{f'(0)}\int_1^x \frac{1}{t}\,dt.$$

This suggests us investigate the function $\displaystyle\int_1^x \frac{1}{t}\,dt$.

Note. Another motivation to make us study the function $\displaystyle\int_1^x \frac{1}{t}\,dt$ is that $\displaystyle\int x^n\,dx = \frac{x^{n+1}}{n+1}$ for every $n \neq -1$. Hence, we want to understand the antiderivative of $\dfrac{1}{x}$.

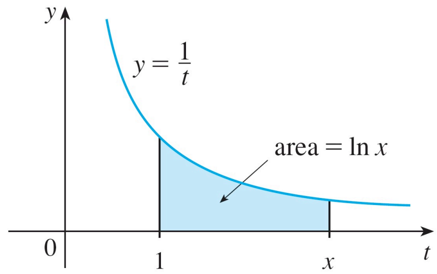
Definition 7.8.1. The "*natural logarithmic function*" is the function defined by

$$\ln x = \int_1^x \frac{1}{t}\,dt \quad x > 0.$$

Note. In some articles, the natural logarithmic function is usually denoted by "log x".

Remark.

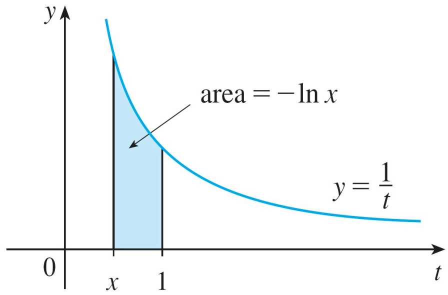
(i) If $x > 1$, $\ln x$ is the area of the region bounded by $y = \dfrac{1}{t}$, t-axis, $t = 1$ and $t = x$.



(ii) If $0 < x < 1$,

$$\ln x = \int_1^x \frac{1}{t}\,dt = -\int_x^1 \frac{1}{t}\,dt$$

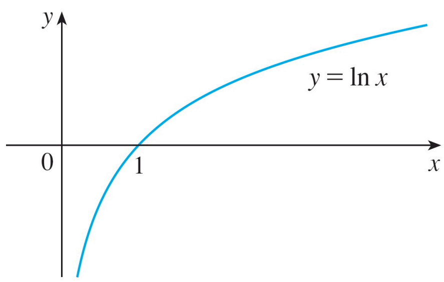
is negative area of the region bounded by $y = \dfrac{1}{t}$, t-axis, $t = x$ and $t = 1$.



(iii) By the Fundamental Theorem of Calculus, for $x > 0$,

$$\begin{aligned}\left(\ln x\right)' &= \frac{1}{x} > 0\\ \left(\ln x\right)'' &= \left(\frac{1}{x}\right)' = -\frac{1}{x^2} < 0.\end{aligned}$$

Hence, $\ln x$ is increasing and the graph of $y = \ln x$ is concave downward.



■ Laws of Logarithms

Theorem 7.8.2. *If x and y are positive nubmers and r is a rational number, then*

(a) $\ln(xy) = \ln x + \ln y$

(b) $\ln\left(\frac{x}{y}\right) = \ln x + \ln y$

(c) $\ln(x^r) = r \ln x$

Proof. (a) By the Fundamental Theorem of Calculus, $(\ln x)' = \frac{1}{x}$. Let $f(x) = \ln(xy)$. Then

$$f'(x) = \frac{1}{xy} \cdot y = \frac{1}{x} = (\ln x)'.$$

Hence, $f(x) = \ln x + C$. Since $\ln y = f(1) = \ln 1 + C = C$. Therefore,

$$\ln(xy) = f(x) = \ln x + \ln y.$$

(b) Let $x = 1/y$. Then

$$\ln \frac{1}{y} + \ln y = \ln\left(\frac{1}{y} \cdot y\right) = \ln 1 = 0.$$

Hence $\ln \frac{1}{y} = -\ln y$ and

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

(c) If $r \in \mathbb{N}$, then $\ln x^r = \ln(\overbrace{x \cdot x \cdots x}^{r}) = \overbrace{\ln x + \ln x + \cdots + \ln x}^{r} = r \ln x$.
If $r = \frac{1}{n}$, then $\ln x = \ln(x^{\frac{1}{n}})^n = n \ln x^{\frac{1}{n}}$ Hence, $\ln x^r = \ln x^{\frac{1}{n}} = \frac{1}{n} \ln x = r \ln x$.
If $r = \frac{m}{n}$. Then

$$\ln x^r = \ln(x^{\frac{1}{n}})^m = m \ln x^{\frac{1}{n}} = \frac{m}{n} \ln x = r \ln x.$$

□

Example 7.8.3.

$$\ln \frac{(x^2+5)\sin x}{x^3+1} = \ln(x^2+5) + \ln(\sin x) - \ln(x^3+1).$$

Corollary 7.8.4. $\lim_{x\to\infty} \ln x = \infty$ *and* $\lim_{x\to 0^+} \ln x = -\infty$

Proof. Consider $\ln 2 = \int_1^2 \frac{1}{x}\, dx > 0$. By the Law(3), we have $\ln 2^n = n \ln 2$ and $\ln 2^{-n} = -n \ln 2$.

Given $M > 0$, we can choose $n_0 \in \mathbb{N}$ such that $n_0 \ln 2 > M$. Also, since $\ln x$ is increasing, for $x > n_0$,

$$\ln x > \ln 2^{n_0} = n_0 \ln 2 > M.$$

Hence, $\lim_{x\to\infty} \ln x = \infty$. On the other hand, for given $N > 0$, we choose $n_1 \in \mathbb{N}$ such that $-n_1 \ln 2 < -N$. Again, since $\ln 2$ is increasing, for $0 < x < 2^{-n_1}$

$$\ln x < \ln 2^{-n_1} = -n_1 \ln 2 < -N.$$

Thus $\lim_{x\to 0^+} \ln x = -\infty$. □

❑ Derivatives of Logarithmic Functions and Logarithmic Differentiation

Recall that the Fundamental Theorem of Calculus implies that $\dfrac{d}{dx}(\ln x) = \dfrac{d}{dx}\displaystyle\int_1^x \frac{1}{t}\,dt = \frac{1}{x}$. For $f(x) > 0$, we compute the derivative of $\ln f(x)$. Let $u = f(x)$ and $y = \ln f(x) = \ln u$. By the chain rule,

$$\frac{d}{dx}\big(\ln f(x)\big) = \frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx} = \frac{1}{u}\cdot\frac{du}{dx} = \frac{f'(x)}{f(x)}.$$

Example 7.8.5.

(1) Differentiate $y = \ln(x^3 + 1)$

Proof. Let $u = x^3 + 1$, then $\dfrac{du}{dx} = 3x^2$ and $y = \ln u$. By the chain rule,

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx} = \frac{1}{u}\cdot 3x^2 = \frac{3x^2}{x^3+1}.$$

□

(2) Find $\dfrac{d}{dx}\ln(\sin x)$.

Proof.

$$\frac{d}{dx}\ln(\sin x) = \frac{1}{\sin x}\cdot\cos x = \cot x.$$

□

(3)

$$\frac{d}{dx}\ln\frac{x+1}{\sqrt{x-2}} = \frac{d}{dx}\big[\ln(x+1) - \ln\sqrt{x-2}\,\big] = \frac{d}{dx}\Big[\ln(x+1) - \frac{1}{2}\ln(x-2)\Big] = \frac{1}{x+1} - \frac{1}{2(x-2)}.$$

Example 7.8.6. Find $f'(x)$ if $f(x) = \ln|x|$.

Proof. Since

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

then

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

□

Note.

$$\frac{d}{dx}\big(\ln|x|\big) = \frac{1}{x} \quad \text{for every } x \neq 0.$$

and hence

$$\int \frac{1}{x} = \ln|x| + C$$

We have

$$\int x^n\, dx = \begin{cases} \dfrac{x^{n+1}}{n+1} + C & \text{if } n \neq -1 \\ \ln|x| + C & \text{if } n = -1. \end{cases}$$

Example 7.8.7. Differentiate $y = \dfrac{x^{3/4}\sqrt{x^2+1}}{(3x+2)^5}$.

Proof. Taking Logarithms of both sides, we have

$$\ln|y| = \frac{3}{4}\ln|x| + \frac{1}{2}\ln|x^2+1| - 5\ln|3x+2|.$$

To find $\dfrac{dy}{dx}$, we take derivatives of both sides. Then

$$\frac{1}{y}\cdot\frac{dy}{dx} = \frac{3}{4x} + \frac{2x}{2(x^2+1)} - \frac{15}{3x+2}$$

Hence,

$$\begin{aligned} \frac{dy}{dx} &= y\Big(\frac{3}{4x} + \frac{2x}{2(x^2+1)} - \frac{15}{3x+2}\Big) \\ &= \frac{x^{3/4}\sqrt{x^2+1}}{(3x+2)^5}\Big(\frac{3}{4x} + \frac{2x}{2(x^2+1)} - \frac{15}{3x+2}\Big). \end{aligned}$$

□

■ **Properties of Logarthmic Function**

(a) $\ln x = \displaystyle\int_1^x \frac{1}{t}\, dt.$

(b) $Dom(\ln x) = (0,\infty)$ and $Range(\ln x) = (-\infty,\infty)$

(c) $\ln x$ is continuous and strictly increasing.

(d) $\dfrac{d}{dx}(\ln x) = \dfrac{1}{x}$

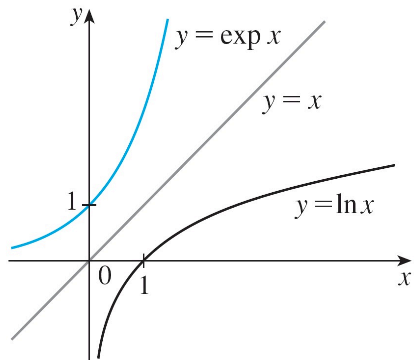
7.8.2 The Exponential Function

Recall that the natural logarithmic function $\ln x$ is one-to-one from $(0, \infty)$ onto $(-\infty, \infty)$. Hence it has an inverse function denoted by "$\exp(x)$".

Definition 7.8.8. The inverse function of $(\ln x)^{-1}$ is denoted by "$\exp(x)$". It satisfies

$$\exp(x) = y \quad \Longleftrightarrow \quad \ln y = x.$$

This function is called the "*(natural) exponential function*".



Proposition 7.8.9. *(Properties of natural exponential function)*

(a) *Domain of* $\exp(x) = (-\infty, \infty)$ *and Range of* $\exp(x) = (0, \infty)$.

(b) $\exp(\ln x) = x$ *for* $x \in (0, \infty)$ *and* $\ln\big(\exp(x)\big) = x$ *for* $x \in (-\infty, \infty)$. *In particular,*

$$\begin{aligned} \exp(0) &= 1 && \textit{since} && \ln 1 = 0 \\ \exp(1) &= e && \textit{since} && \ln e = 1. \end{aligned}$$

(c) $\lim_{x\to\infty} \exp(x) = \infty$ *and* $\lim_{x\to-\infty} \exp(x) = 0$.

(d) *Since* $\ln x$ *is differentiable and* $\frac{d}{dx}(\ln x) = \frac{1}{x} \neq 0$ *for every* $x \in (0, \infty)$*, the exponential function* $\exp(x)$ *is also differentiable everywhere. Moreover,*

$$\frac{d}{dx}\Big(\exp(x)\Big) = \frac{1}{\frac{d\ln}{dx}\big(\exp(x)\big)} = \frac{1}{1/\exp(x)} = \exp(x).$$

Notice that

$$\frac{d^{(n)}}{dx^n}\exp(x) = \exp(x) \quad \text{for every } n \in \mathbb{N}.$$

The property(d) implies that the first and second derivative of $\exp(x)$ are always positive. Hence, it is an increasing function and its graph is concave upward.

Theorem 7.8.10. *For* $x, y \in \mathbb{R}$,

$$\exp(x + y) = \exp(x) \cdot \exp(y).$$

Proof. Let $x_1 = \exp(x)$ and $y_1 = \exp(y)$. Then $\ln x_1 = x$ and $\ln y_1 = y$. Thus,

$$x + y = \ln x_1 + \ln y_1 = \ln(x_1 y_1).$$

We have

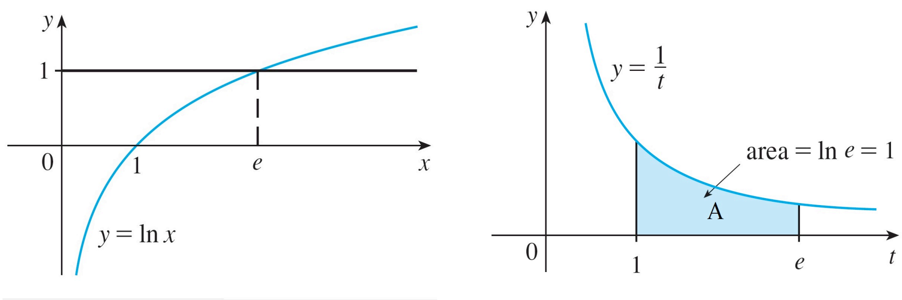
$$\exp(x + y) = x_1 y_1 = \exp(x) \cdot \exp(y).$$

□

Definition 7.8.11. We denote the number $e = \exp(1)$. This number is called "*Euler's number*".

Remark. (i) $\ln e = \ln(\exp(1)) = 1$ and e is a number such that

$$1 = \int_1^e \frac{1}{x}\, dx = \text{area of } A$$



(ii) $e^n = \overbrace{e \cdot e \cdots e}^{n} = \exp(1) \cdots \exp(1) = \exp(n)$.

(iii) $e^{-n} \cdot e^n = e^0 = 1 = \exp(0) = \exp\big(n + (-n)\big) = \exp(n) \cdot \exp(-n)$. Hence,

$$e^{-n} = \exp(-n).$$

(iv) $e^{\frac{1}{n} \cdot n} = e = \exp(1) = \exp\big(\overbrace{\frac{1}{n} + \cdots \frac{1}{n}}^{n}\big) = \big[\exp(\frac{1}{n})\big]^n$. Thus,

$$e^{\frac{1}{n}} = \exp(\frac{1}{n}).$$

(v) $e^{\frac{q}{p}} = \overbrace{e^{\frac{1}{p}} \cdots e^{\frac{1}{p}}}^{q} = \overbrace{\exp(\frac{1}{p}) \cdots \exp(\frac{1}{p})}^{q} = \exp(\frac{q}{p})$. Hence,

$$e^k = \exp(k) \quad \text{for every } k \in \mathbb{Q}.$$

Note. The function $\exp(x)$ is defined on $\mathbb{R}$ but e^x is only defined on $\mathbb{Q}$.

Question: Can we define e^x on $\mathbb{R} \backslash \mathbb{Q}$?

Definition 7.8.12. For any $x \in \mathbb{R}$, we define

$$e^x = \exp(x).$$

From Proposition 7.8.9, if $f(x) = e^x$ then

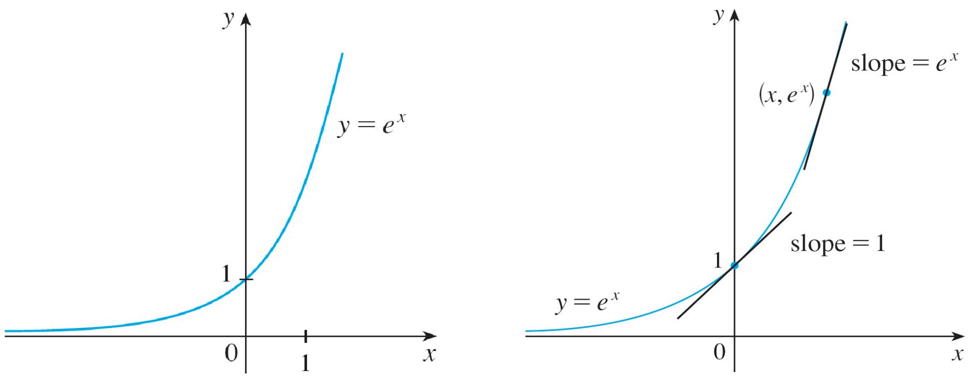
(1) $Dom(f) = \mathbb{R}$ and $Range(f) = (0, \infty)$.

(2) $e^{\ln x} = x$ for $x \in (0, \infty)$ and $\ln e^x = x$ for $x \in \mathbb{R}$.

(3) $\lim_{x\to\infty} e^x = \infty$ and $\lim_{x\to-\infty} e^x = 0$.

(4) $\dfrac{d}{dx}(e^x) = e^x$.

This means that the function $f(x) = e^x$ is its own derivative. The slope of a tangent line to the curve $y = e^x$ at any point is equal to the y-coordinate of the point The exponential curve $y = e^x$ grows very rapidly.



(5) The antiderivative of e^x is itself. That is,

$$\int e^x\, dx = e^x + C.$$

Example 7.8.13. Differentiate the function $y = e^{\tan x}$.

Proof. Let $u = \tan x$. Then $y = e^u$ and $\dfrac{du}{dx} = \sec^2 x$. Thus, by the chain rule,

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx} = e^u \cdot \sec^2 x = e^{\tan x} \sec^2 x.$$

□

7.9 General Logarithmic and Exponential Functions

In the present section, we use the natural exponential and logarithmic functions to study exponential and logarithmic functions with base $a > 0$.

7.9.1 General Exponential Functions

Question: For $a > 0$, can we define a^x for every $x \in \mathbb{R}$?

Notice that $a = \exp(\ln a) = e^{\ln a}$.

Definition 7.9.1. For $a > 0$, we define

$$a^x = \left(e^{\ln a}\right)^x = e^{x \ln a} \quad \text{for } x \in \mathbb{R}.$$

The function $f(x) = a^x$ is called the "*exponential function with base a*".

Note. For every $a > 0$,

(1) $Dom(a^x) = \mathbb{R}$ and $Range(a^x) = (0, \infty)$.

(2) $\ln a^x = \ln\left(e^{x \ln a}\right) = x \ln a \quad$ for every $x \in \mathbb{R}$.

Theorem 7.9.2. *If $a, b > 0$ and $x, y \in \mathbb{R}$, then*

(**a**) $a^{x+y} = a^x a^y$ (**b**) $a^{x-y} = \dfrac{a^x}{a^y}$ (**c**) $(a^x)^y = a^{xy}$ (**d**) $(ab)^x = a^x b^x$

Proof. Exercise □

■ **Derivative of a^x and Graph of $y = a^x$**

$$\frac{d}{dx}(a^x) = \frac{d}{dx}(e^{x \ln a}) = e^{x \ln a} \cdot \ln a = a^x \ln a.$$

Note.

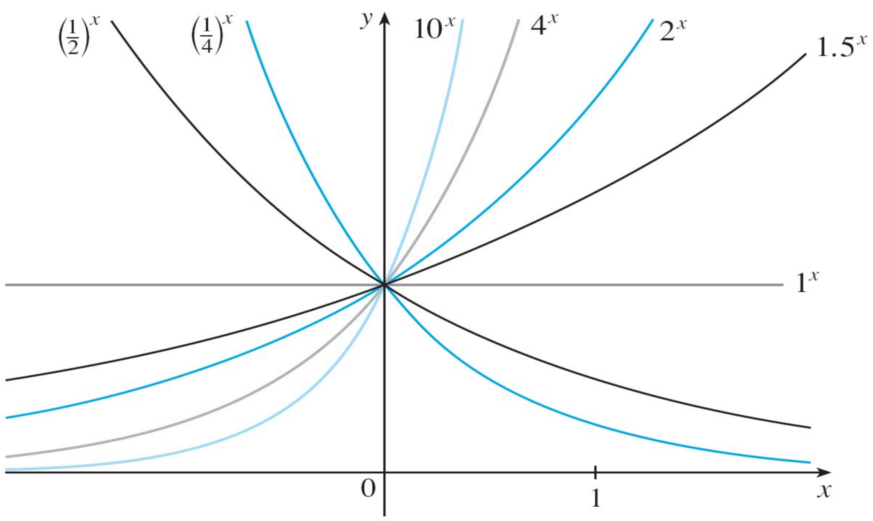
(1) If $a > 1$, then $\ln a > 0$ and hence $y = a^x$ is increasing since $\frac{d}{dx}(a^x) > 0$. Also,

$$\lim_{x\to\infty} a^x = \lim_{x\to\infty} e^{x \ln a} = \lim_{t\to\infty} e^t = \infty \quad \text{and} \quad \lim_{x\to-\infty} a^x = \lim_{x\to-\infty} e^{x \ln a} = 0.$$

(2) If $a = 1$, then $y = 1$ is a constant function.

(3) If $0 < a < 1$, then $\ln a < 0$ and hence $y = a^x$ is decreasing since $\frac{d}{dx}(a^x) < 0$. Also,

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



Since $\dfrac{d}{dx}(a^x) = a^x \ln a$, the antiderivative of a^x is $\dfrac{a^x}{\ln a}$ for $a > 0$ and $a \neq 1$. Thus,

$$\int a^x \, dx = \frac{a^x}{\ln a} + C, \quad a > 0 \text{ and } a \neq 1.$$

Example 7.9.3. $\displaystyle\int_0^5 2^x \, dx = \frac{2^x}{\ln 2}\Big|_0^5 = \frac{2^5}{\ln 2} - \frac{1}{\ln 2} = \frac{31}{\ln 2}.$

■ The Power Rule versus the Exponential Rule

Theorem 7.9.4. *If n is any real number and $f(x) = x^n$, then*

$$f'(x) = nx^{n-1}.$$

Proof. Let $y = x^n$. Then

$$\ln|y| = \ln|x|^n = n \ln|x|, \quad \text{for } x \neq 0$$

Taking differentiation of both sides,

$$\frac{y'}{y} = \frac{d}{dx}(\ln|y|) = n\frac{d}{dx}\ln|x| = \frac{n}{x}.$$

Then

$$y' = y \cdot \frac{n}{x} = x^n \cdot \frac{n}{x} = nx^{n-1}.$$

□

■ Four cases for exponents and bases

(1) $\dfrac{d}{dx}(b^n) = 0$ (constant base, constant exponent)

(2) $\dfrac{d}{dx}[f(x)]^n = n[f(x)]^{n-1} f'(x)$ (variable base, constant exponent)

(3) $\dfrac{d}{dx}[b^{g(x)}] = b^{g(x)}(\ln b)g'(x)$ (constant base, variable exponent)

(4) $\dfrac{d}{dx}[f(x)^{g(x)}] = f(x)^{g(x)}\Big(g'(x)\ln|f(x)| + g(x)\cdot\dfrac{f'(x)}{f(x)}\Big)$ (variable base, variable exponent)

Proof. We only prove (4) here. Let $y = f(x)^{g(x)}$. Then $\ln y = g(x)\ln f(x)$. Taking differentiation on both sides,

$$\frac{y'}{y} = \frac{d}{dx}\ln y = \frac{d}{dx}[g(x)\ln f(x)] = g'(x)\ln f(x) + g(x)\cdot\frac{f'(x)}{f(x)}.$$

Then

$$y' = y\Big(g'(x)\ln|f(x)| + g(x)\cdot\frac{f'(x)}{f(x)}\Big) = f(x)^{g(x)}\Big(g'(x)\ln|f(x)| + g(x)\cdot\frac{f'(x)}{f(x)}\Big).$$

Another method: $f(x)^{g(x)} = e^{\ln[f(x)^{g(x)}]} = e^{g(x)\ln f(x)}$. Then

$$\begin{aligned}\frac{d}{dx}[f(x)^{g(x)}] &= \frac{d}{dx}\Big[e^{g(x)\ln f(x)}\Big] = e^{g(x)\ln f(x)}\cdot\frac{d}{dx}\Big(g(x)\ln f(x)\Big)\\ &= e^{g(x)\ln f(x)}\Big(g'(x)\ln f(x) + g(x)\cdot\frac{f'(x)}{f(x)}\Big).\end{aligned}$$

□

Example 7.9.5. Differentiate $y = x^{\sqrt{x}}$.

Proof. Taking logarithm on both sides, $\ln y = \sqrt{x}\ln x$. Then

$$\frac{y'}{y} = \frac{d}{dx}(\ln y) = \frac{\ln x}{2\sqrt{x}} + \frac{\sqrt{x}}{x}.$$

Then

$$y' = y\Big(\frac{\ln x}{2\sqrt{x}} + \frac{\sqrt{x}}{x}\Big) = x^{\sqrt{x}}\Big(\frac{\ln x}{2\sqrt{x}} + \frac{\sqrt{x}}{x}\Big) = x^{\sqrt{x}}\Big(\frac{2+\ln x}{2\sqrt{x}}\Big).$$

□

7.9.2 General Logarithmic Functions

If $a > 0$ and $a \neq 1$, then $f(x) = a^x$ is a one-to-one function and thus its inverse function exists.

Definition 7.9.6. For $a > 0$ and $a \neq 1$, the inverse function of a^x is called the "*logarithmic function with base a*" and is denoted by $\log_a x$.

Note. For $a > 0$ and $a \neq 1$,

(1) $Dom(\log_a x) = (0, \infty)$ and $Range(\log_a x) = \mathbb{R}$.

(2) $\log_a 1 = 0$.

(3) $\log_a x = y$ if and only if $a^y = x$.

(4) $a^{\log_a x} = x$ for every $x \in (0, \infty)$ and $\log_a a^x = x$ for every $x \in \mathbb{R}$.

(5) $\log_e x = \ln x$.

(6) (Change of Base Formula) For any positive number a ($a \neq 1$), we have

$$\log_a x = \frac{\ln x}{\ln a}.$$

Proof. Let $y = \log_a x$. Then $a^y = x$. Taking natural logarithms of both sides, we obtain $y \ln a = \ln x$ and thus

$$y = \frac{\ln x}{\ln a}.$$

□

(7) For $a > 1$,

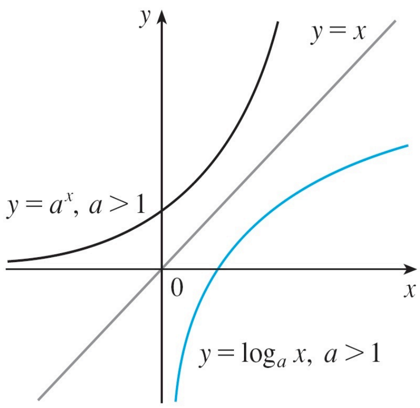
$$\lim_{x\to\infty} \log_a x = \infty \quad \text{and} \quad \lim_{x\to 0^+} \log_a x = -\infty.$$

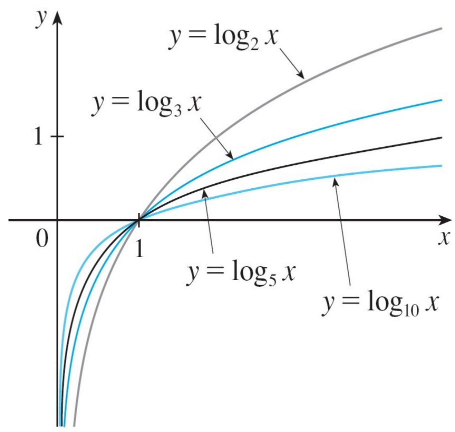
and for $0 < a < 1$,

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

■ **Graph of** $\log_a x$

Heuristically, for $a > 1$, the fact that $y = a^x$ is a very rapidly increasing function for $x > 0$ is reflected in the fact that $y = \log_a x$ is a very slowly increasing function for $x > 1$.





■ Derivative of $\log_a x$

For $a > 0$ and $a \neq 1$,

$$\frac{d}{dx}(\log_a x) = \frac{d}{dx}\left(\frac{\ln x}{\ln a}\right) = \frac{1}{x \ln a}.$$

Example 7.9.7.

$$\frac{d}{dx}\ln_{10}(2+\sin x) = \frac{1}{(2+\sin x)\ln 10}\cdot \cos x.$$

■ The Number e as a Limit

Let $f(x) = \ln x$ then $f'(x) = \dfrac{1}{x}$ and $f'(1) = 1$. By the definition of $f'(1)$,

$$\begin{aligned} f'(1) &= \lim_{h\to 0}\frac{f(1+h)-f(1)}{h} = \lim_{x\to 0}\frac{f(1+x)-f(1)}{x} \\ &= \lim_{x\to 0}\frac{\ln(1+x)-\ln 1}{x} \\ &= \lim_{x\to 0}\frac{1}{x}\ln(1+x) \\ &= \lim_{x\to 0} = \ln(1+x)^{1/x} \end{aligned}$$

Since $f'(1) = 1$, then

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

Since e^x is continuous, we have

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

Note. If we put $n = 1/x$, then $n \to \infty$ as $x \to 0^+$and we have

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

7.10 L'Hôpital's Rule

In Chapter3, we have computed some limits with special form like

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

where $f(x), g(x) \to 0$ as $x \to a$. When f and g have common factor $(x-a)$, we can evaluate the limit by dividing this common factor. We can also compute some specific limit such as $\lim\limits_{x\to 0}\dfrac{\sin x}{x} = 1$.

On the other hand, the generalized mean value theorem some ideas to compute the above limit. Recall that a curve C on the plane can be represented as $\big(f(t), g(t)\big)$ where f and g are differentiable. The slope of the secant line connected $\big(f(a), g(b)\big)$ and $\big(f(t), g(t)\big)$ is $m = \dfrac{f(t)-f(a)}{g(t)-g(a)}$ for any $t \in (a,b)$. G.M.V.T says that there exists $c_t \in (a,t)$ such that

$$\frac{f'(c_t)}{g'(c_t)} = \frac{f(t)-f(a)}{g(t)-g(a)}.$$

Suppose that $f(t), g(t) \to 0$ as $x \to 0$ and then $f(a) = g(a) = 0$. Heuristically, as $t \to a$,

$$\frac{f'(a)}{g'(a)} = \lim_{t\to a}\frac{f'(c_t)}{g'(c_t)} = \lim_{t\to a}\frac{f(t)-f(a)}{g(t)-g(a)} = \lim_{t\to a}\frac{f(t)}{g(t)}.$$

In this section, we will study the limit with some specific form.

Note.

(1) The limit $\lim\limits_{x\to a}\dfrac{f(x)}{g(x)}$ where $f(x) \to 0$ and $g(x) \to 0$ as $x \to a$ is called an "*indeterminate form of type* $\big(\frac{0}{0}\big)$".

(2) The limit $\lim\limits_{x\to a}\dfrac{f(x)}{g(x)}$ where $f(x) \to \pm\infty$ and $g(x) \to \pm\infty$ as $x \to a$ is called an "*indeterminate form of type* $\big(\frac{\infty}{\infty}\big)$".

Example: $\lim\limits_{x\to 1}\dfrac{\ln x}{x-1}$ is of type $\big(\frac{0}{0}\big)$ and $\lim\limits_{x\to\infty}\dfrac{\ln x}{x-1}$ is of type $\big(\frac{\infty}{\infty}\big)$.

Theorem 7.10.1. *(L'Hôpital's Rule) Suppose f and g are differentiable on an open interval I containing a (except possibly at a itself), and $g'(x) \neq 0$ near a. Suppose that*

$$\lim_{x\to a} f(x) = 0 = \lim_{x\to a} g(x)$$

or that

$$\lim_{x\to a} f(x) = \pm\infty \quad \textit{and} \quad \lim_{x\to a} g(x) = \pm\infty.$$

If the limit $\lim\limits_{x\to a}\dfrac{f'(x)}{g'(x)}$ exists (or equals $\pm\infty$), then

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

Proof. We will show the case $a \in \mathbb{R}$ and the case $a = \pm\infty$ will be left to the readers.

Case1: $-\infty < L < \infty$

Suppose that $\lim_{x\to a} \frac{f'(x)}{g'(x)} = L$. Then, for given $\varepsilon > 0$, there exists $\delta > 0$ such that for all $0 < |x-a| < \delta$,

$$\left|\frac{f'(x)}{g'(x)} - L\right| < \varepsilon. \tag{7.4}$$

Fix a number $s \in (a, a+\delta)$. By Cauchy Mean Value Theorem, for any $t \in (a, s)$, there exists $c_t \in (t, s) \subset (a, a+\delta)$ such that

$$\frac{f(s)-f(t)}{g(s)-g(t)} = \frac{f'(c_t)}{g'(c_t)}.$$

Then, by (7.4)

$$\left|\frac{f(s)-f(t)}{g(s)-g(t)} - L\right| = \left|\frac{f'(c_t)}{g'(c_t)} - L\right| < \varepsilon. \tag{7.5}$$

(i) For $\lim_{t\to a} f(t) = 0 = \lim_{t\to a} g(t)$, by (7.5)

$$\left|\lim_{t\to a^+} \frac{f(s)}{g(s)} - L\right| \le \varepsilon.$$

Since s is arbitrary number in $(a, a+\delta)$, we have

$$\left|\frac{f(s)}{g(s)} - L\right| \le \varepsilon$$

for every $s \in (a, a+\delta)$. This implies that $\lim_{x\to a^+} \frac{f(x)}{g(x)} = L$. Similarly, we can also prove that $\lim_{x\to a^-} \frac{f(x)}{g(x)} = L$ and thus

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

(ii) For $\lim_{x\to a} f(x) = \infty$ and $\lim_{x\to\infty} g(x) = \infty$, we can choose $\delta > 0$ such that for every $x \in (a, a+\delta)$, $g(x) > 0$ and

$$\left|\frac{f'(x)}{g'(x)} - L\right| < \varepsilon.$$

Fix $s \in (a, a+\delta)$, since $\lim_{x\to a} g(x) = \infty$, there exists $0 < \delta_1 < \delta$ such that for every $t \in (a, a+\delta_1)$, $g(t) > g(s) > 0$. By the Cauchy Mean Value Theorem, for every $t \in (a, a+\delta_1)$, there exists $c_t \in (t, s)$ such that

$$\frac{f(s)-f(t)}{g(s)-g(t)} = \frac{f'(c_t)}{g'(c_t)}.$$

Then

$$-\varepsilon\cdot\frac{g(t)-g(s)}{g(t)} \le \Big(\frac{f(t)-f(s)}{g(t)-g(s)} - L\Big)\cdot\frac{g(t)-g(s)}{g(t)} = \Big(\frac{f'(c_t)}{g'(c_t)} - L\Big)\cdot\frac{g(t)-g(s)}{g(t)} < \varepsilon\cdot\frac{g(t)-g(s)}{g(t)}.$$

Hence,

$$-\varepsilon\Big(1-\frac{g(s)}{g(t)}\Big) < \frac{f(t)}{g(t)} - \frac{f(s)}{g(t)} - L - \frac{g(s)}{g(t)} < \varepsilon\Big(1-\frac{g(s)}{g(t)}\Big)$$

for every $t \in (a, a+\delta_1)$. Since s is fiexed and $g(t) \to \infty$as $t \to a$, when t is sufficiently close to a from the right, we obtain $\left|\frac{g(s)}{g(t)}\right| < \varepsilon$ and $\left|\frac{f(s)}{g(t)}\right| < \varepsilon$. Therefore,

$$-2\varepsilon < \Big(\frac{f(t)}{g(t)} - L\Big) - \Big(\frac{f(s)}{g(t)} + \frac{g(s)}{g(t)}\Big) < 2\varepsilon.$$

Then

$$4\varepsilon \leq \frac{f(t)}{g(t)} - L \leq 4\varepsilon.$$

This implies that $\lim_{t\to a^+} \frac{f(t)}{g(t)} = L$. Similarly, we can evaluate $\lim_{t\to a^-} \frac{f(t)}{g(t)} = L$ and thus $\lim_{t\to a} \frac{f(t)}{g(t)} = L$.

Case2: $L = \pm\infty$**,** left to the readers.

□

Note.

(1) The L'Hôpital's Rule says that

(i) Check the indeterminate form $\left(\frac{0}{0}\right)$ or $\left(\frac{\infty}{\infty}\right)$ is satisfied.

(ii) Check $\lim_{x\to a} \frac{f'(x)}{g'(x)}$ exists (or equals $\pm\infty$).

(iii) the limit of a quotient of function is equal to the limit quotient of their derivatives.

(2) The rule is also valid for sided limits. That is "$\lim_{x\to a^+}$", "$\lim_{x\to a^-}$", "$\lim_{x\to -\infty}$" and "$\lim_{x\to\infty}$".

Example 7.10.2. Determine whether the limit $\lim_{x\to 1} \frac{\ln x}{x-1}$ exists.

Proof. Since $\lim_{x\to 1} \ln x = 0 = \lim_{x\to 1}(x-1)$, the limit is of type $\left(\frac{0}{0}\right)$. Consider

$$\lim_{x\to 1} \frac{\frac{d}{dx}(\ln x)}{\frac{d}{dx}(x-1)} = \lim_{x\to 1} \frac{\frac{1}{x}}{1} = 1.$$

By the L'Hôpital's rule,

$$\lim_{x\to 1} \frac{\ln x}{x-1} = \lim_{x\to 1} \frac{\frac{1}{x}}{1} = 1.$$

□

Example 7.10.3. Determine whether the limit $\lim_{x\to\infty} \frac{e^x}{x^2}$ exists.

Proof. Since $\lim_{x\to\infty} e^x = \infty = \lim_{x\to\infty} x^2$, the limit is of type $\left(\frac{\infty}{\infty}\right)$. Consider

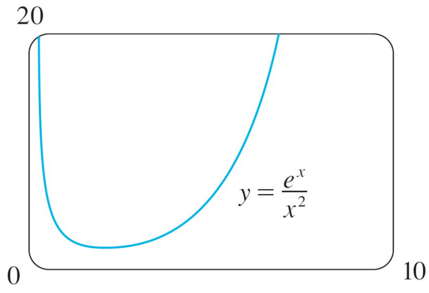
$$\lim_{x\to\infty} \frac{\frac{d}{dx}(e^x)}{\frac{d}{dx}(x^2)} = \lim \frac{e^x}{2x}.$$

Again, since $\lim_{x\to\infty} e^x = \infty = \lim_{x\to\infty} 2x$, the limit $\lim_{x\to\infty} \frac{e^x}{2x}$ is of type $\left(\frac{\infty}{\infty}\right)$. Consider

$$\lim_{x\to\infty} \frac{\frac{d}{dx}(e^x)}{\frac{d}{dx}(2x)} = \lim_{x\to\infty} \frac{e^x}{2} = \infty.$$

By the L'Hôpital's rule,

$$\lim_{x\to\infty} \frac{e^x}{x^2} = \lim_{x\to\infty} \frac{e^x}{2x} = \lim_{x\to\infty} \frac{e^x}{2} = \infty.$$



□

Remark. The exponential functions grow much more rapidly than any power functions as $x \to \infty$.

Example 7.10.4. Determine whether the limit $\lim_{x\to\infty} \frac{\ln x}{\sqrt[3]{x}}$.

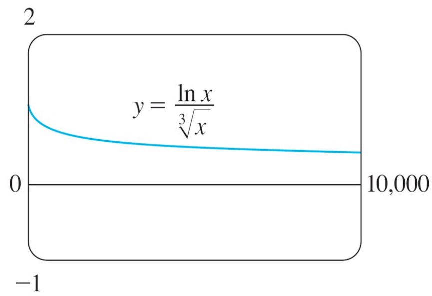
Proof. Since $\lim_{x\to\infty} \ln x = \infty = \lim_{x\to\infty} \sqrt[3]{x}$, the limit is of type $\left(\frac{\infty}{\infty}\right)$.

Consider

$$\lim_{x\to\infty} \frac{\frac{d}{dx}(\ln x)}{\frac{d}{dx}(\sqrt[3]{x})} = \lim_{x\to\infty} \frac{\frac{1}{x}}{\frac{1}{3}x^{-2/3}} = \lim_{x\to\infty} \frac{1}{3x^{1/3}} = 0.$$

By the L'Hôpital's Rule,

$$\lim_{x\to\infty} \frac{\ln x}{\sqrt[3]{x}} = \lim_{x\to\infty} \frac{1}{3x^{1/3}} = 0.$$



□

Remark. The logarithms grow slowly than power function as $x \to \infty$.

Exercise. Determine whether the limit $\lim_{x\to 0} \frac{\tan x - x}{x^3}$ exists.

Remark.

(i) If the limit $\lim_{x\to a} \frac{f(x)}{g(x)}$ is not of type $\left(\frac{0}{0}\right)$ or $\left(\frac{\infty}{\infty}\right)$, then the L'Hôpital's rule could be wrong. For example,

$$\lim_{x\to\pi^-} \frac{\sin x}{1-\cos x} = \frac{0}{2} = 0.$$

But

$$\lim_{x\to\pi^-} \frac{\frac{d}{dx}(\sin x)}{\frac{d}{dx}(1-\cos x)} = \lim_{x\to\pi^-} \frac{\cos x}{\sin x} = -\infty.$$

(ii) If the limit $\lim_{x\to a}\dfrac{f'(x)}{g'(x)}$ DNE, then the L'Hôpital's rule could be wrong. For exmaple, $\lim_{x\to\infty}\dfrac{x}{x-\sin x}=1$ but $\lim_{x\to\infty}\dfrac{1}{1-\cos x}$ DNE.

❑ Applications

The L'Hôpital's rule can be applied some speical forms of limits.

(I) **Indterminate products:** the limit $\lim_{x\to a}[f(x)g(x)]$ is of the type $0\cdot\infty$ or $\infty\cdot 0$. Either

$$f(x)\neq 0 \text{ near } a \text{ (except possibly at } a),\ f(x)\to 0 \text{ and } g(x)\to\pm\infty$$

or

$$g(x)\neq 0 \text{ near } a \text{ (except possibly at } a),\ f(x)\to \pm\infty \text{ and } g(x)\to 0$$

Express $\lim_{x\to a} f(x)g(x)$ as $\lim_{x\to a}\dfrac{f(x)}{1/g(x)}$ or $\lim_{x\to a}\dfrac{g(x)}{1/f(x)}$.

Example 7.10.5. $\lim_{x\to 0^+} x\ln x$.

Proof. Since $\lim_{x\to 0^+} x=0$ and $\lim_{x\to 0^+}\ln x=-\infty$, the limit is of type $(0\cdot\infty)$. Consider

$$\lim_{x\to 0^+}\frac{\frac{d}{dx}(\ln x)}{\frac{d}{dx}(\frac{1}{x})}=\lim_{x\to 0^+}\frac{1/x}{-1/x^2}=\lim_{x\to 0^+}(-x)=0.$$

By the L'Hôpital's rule,

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

□

Note.

(1) If we rewrite $x\ln x$ as $\dfrac{x}{1/\ln x}$, then the limit is of type $\left(\frac{0}{0}\right)$. But it is difficult to use L'Hôpital's rule to solve it.

(2) When $x\to 0^+$, the rate of power functions ($x^a,\ a>0$) which decay to 0 is more rapid than the ralte of logarithms ($\ln x$) which grow to ∞,

(II) **Indeterminate differences:** the indeterminate form is of tyep $(\infty-\infty)$ $\lim_{x\to a}[f(x)-g(x)]$ where $\lim_{x\to a} f(x)=\infty$ and $\lim_{x\to a} g(x)=\infty$.

Example 7.10.6.

$$\lim_{x\to(\frac{\pi}{2})^-}(\sec x-\tan x)=\lim_{x\to(\frac{\pi}{2})^-}\left(\frac{1}{\cos x}-\frac{\sin x}{\cos x}\right)=\lim_{x\to(\frac{\pi}{2})^-}\frac{1-\sin x}{\cos x}\quad(\frac{0}{0})\text{-type}$$

Consider

$$\lim_{x\to(\frac{\pi}{2})^-}\frac{\frac{d}{dx}(1-\sin x)}{\frac{d}{dx}(\cos x)}=\lim_{x\to(\frac{\pi}{2})^-}\frac{-\cos x}{-\sin x}=0.$$

By the L'Hôpital's rule,

$$\lim_{x\to(\frac{\pi}{2})^-}(\sec x-\tan x)=\lim_{x\to(\frac{\pi}{2})^-}\frac{-\cos x}{-\sin x}=0.$$

(III) **Indeterminate powers** the limit $\lim\limits_{x\to a}[f(x)]^{g(x)}$ is of the form 0^0, ∞^0 or 1^∞.

(a) $\lim\limits_{x\to a} f(x) = 0$ and $\lim\limits_{x\to a} g(x) = 0$ (type 0^0)

(b) $\lim\limits_{x\to a} f(x) = \infty$ and $\lim\limits_{x\to a} g(x) = 0$ (type ∞^0)

(c) $\lim\limits_{x\to a} f(x) = 1$ and $\lim\limits_{x\to a} g(x) = \pm\infty$ (type 1^∞)

Strategy:

(i) Taking "ln" on $[f(x)]^{g(x)}$ and then taking "$\lim\limits_{x\to a}$", we have

$$\lim_{x\to a} \ln[f(x)^{g(x)}] = \lim_{x\to a} g(x)\ln f(x) \begin{cases} \text{(a)} & \Rightarrow \quad \text{type } 0\cdot\infty \\ \text{(b)} & \Rightarrow \quad \text{type } 0\cdot\infty \\ \text{(c)} & \Rightarrow \quad \text{type } \infty\cdot 0 \end{cases}$$

(ii) $\lim\limits_{x\to a}[f(x)]^{g(x)} = \lim\limits_{x\to a} e^{\ln[f(x)]^{g(x)}} = e^{\lim_{x\to a}\ln[f(x)]^{g(x)}}$.

Example 7.10.7. $\lim\limits_{x\to 0^+}(1+\sin 4x)^{\cot x}$ (1^∞)

Proof. Let $y = (1+\sin 4x)^{\cot x}$. Then $\ln y = (\cot x)\ln(1+\sin 4x)$ and our goal is to compute $\lim\limits_{x\to a} y$.

$$\begin{aligned} \lim_{x\to 0^+} y &= \lim_{x\to 0^+}(\cot x)\ln(1+\sin 4x) && (\infty\cdot 0) \\ &= \lim_{x\to 0^+}\frac{(\cos x)\ln(1+\sin 4x)}{\sin x} && \left(\frac{0}{0}\right). \end{aligned}$$

Consider

$$\lim_{x\to 0^+}\frac{\frac{d}{dx}[\cos x\ln(1+\sin 4x)]}{\frac{d}{dx}(\sin x)} = \lim_{x\to 0^+}\frac{-\sin x\ln(1+\sin 4x)+\cos x\cdot\frac{1}{1+\sin 4x}\cdot\cos 4x\cdot 4}{\cos x} = 4$$

By the L'Hôpital's rule, $\lim_{x\to 0^+}\ln y = 4$. Hence,

$$\begin{aligned} \lim_{x\to 0^+}(1+\sin 4x)^{\cot x} &= \lim_{x\to 0^+} e^{\ln[(1+\sin 4x)^{\cot x}]} \\ &= e^{\lim_{x\to 0^+}\ln[(1+\sin 4x)^{\cot x}]} \\ &= e^{\lim_{x\to 0^+}\ln y} = e^4. \end{aligned}$$

□

Example 7.10.8. $\lim\limits_{x\to 0^+} x^x$ $\left(0^0\right)$.

Proof. Let $y = x^x$. Then $\ln y = x\ln x$ and our goal is to compute $\lim_{x\to 0^+} y$.

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

Consider

$$\lim_{x\to 0^+} \frac{\frac{d}{dx}(\ln x)}{\frac{d}{dx}(\frac{1}{x})} = \lim_{x\to 0^+} \frac{1/x}{-1/x^2} = \lim_{x\to 0^+}(-x) = 0.$$

By the L'Hôpital's rule,

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

Hence,

$$\lim_{x\to 0^+} x^x = \lim_{x\to 0^+} y = \lim_{x\to 0^+} e^{\ln y} = e^{\lim_{x\to 0^+} \ln y} = e^0 = 1.$$

□

Exercise. Use the L'Hôpital's rule to show

$$\lim_{h\to 0^+}(1+h)^{1/h} = e.$$

Chapter 8

Techniques of Integration

The Fundamental Theorem of Calculus says that if f is a continuous function on $[a,b]$ then the function $F(x) = \displaystyle\int_a^x f(t)\,dt$ is differentiable on $[a,b]$ and $F'(x) = f(x)$. The indefinite $\displaystyle\int f(x)\,dx$ indicates the family of antiderivative of f. The differentiation and integration are inverse operations.

Differentiation	$\longleftrightarrow$	**Integration**
Chain rule	$\longleftrightarrow$	Substitution rule
Product rule	$\longleftrightarrow$	Integration by parts

8.1 The Substitution Rule

So far, our experience does not tell the antiderivative of the function $f(x) = 2x\sqrt{1+x^2}$. By the chain rule, the derivative of $F\big(g(x)\big)$ is $F'\big(g(x)\big)g'(x)$. Hence, the antiderivative of $F'\big(g(x)\big)g'(x)$ is $F\big(g(x)\big)$.

Chain Rule:

$$F\big(g(x)\big) \xrightarrow{\frac{d}{dx}} F'\big(g(x)\big)g'(x)$$

$$F'\big(g(x)\big)g'(x) \xrightarrow{\int dx} F\big(g(x)\big)$$

For $F' = f$,

$$\int f\big(g(x)\big)g'(x)\,dx = \int F'\big(g(x)\big)g'(x)\,dx = \int \frac{d}{dx}\big[F\big(g(x)\big)\big]\,dx =$$

Let $u = g(x)$, then $\dfrac{du}{dx} = g'(x)$ and $du = g'(x)dx$. Hence,

$$\int F'\big(\underbrace{g(x)}_{u}\big)\ \underbrace{g'(x)\,dx}_{du} = \int F'\big(u\big)\ du = F(u) + C = F\big(g(x)\big) + C.$$

<u>**Substitution rule (u-substitition)**</u>

Theorem 8.1.1. *If $u = u(x)$ is a continuously differentiable function whose range is an interval I and f is continuous on I, then*

$$\int f\big(u(x)\big)u'(x)\,dx = \int f(u)\,du.$$

Moreover, if $F' = f$ then

$$\int f\big(u(x)\big)u'(x)\,dx = F(u) + C.$$

Proof. Since f is continuous and $F' = f$, by the chain rule

$$\int f\big(u(x)\big)u'(x)\,dx = \int F'\big(u(x)\big)u'(x)\,dx = \int \frac{d}{dx}\Big(F\big(u(x)\big)\Big)\,dx = F\big(u(x)\big) + C.$$

□

Example 8.1.2.

(1) Evaluate $\displaystyle\int x^3\cos(x^4+2)\,dx$

Proof. Let $u = x^4 + 2$. Then $\dfrac{du}{dx} = 4x^3$ and $du = 4x^3\,dx$. Thus $\dfrac{1}{4}du = x^3dx$. We have

$$\begin{aligned}\int x^3\cos(x^4+2)\,dx &= \int \cos u\cdot\frac{1}{4}\,du = \frac{1}{4}\int \cos u\,du\\ &= \frac{1}{4}\sin u + C = \frac{1}{4}\sin(x^4+2) + C\end{aligned}$$

□

(2) Evaluate $\displaystyle\int \sqrt{2x+1}\,dx.$

Proof. **Solution 1:** Let $u = 2x + 1$. Then $\dfrac{du}{dx} = 2$ and $du = 2dx$. Thus $du = \dfrac{1}{2}dx$. We have

$$\begin{aligned}\int \sqrt{2x+1}\,dx &= \int \sqrt{u}\cdot\frac{1}{2}du = \frac{1}{2}\int u^{1/2}\,du\\ &= = \frac{1}{2}(\frac{2}{3}u^{3/2} + C) = \frac{1}{3}u^{3/2} + C\\ &= \frac{1}{3}(2x+1)^{\frac{3}{2}} + C.\end{aligned}$$

Solution 2: Let $u = \sqrt{2x+1}$. Then $du = \dfrac{1}{\sqrt{2x+1}}\,dx$ and $dx = \sqrt{2x+1}\,du = udu$. Hence,

$$\begin{aligned}\int \sqrt{2x+1}\,dx &= \int u \cdot u\,du = \int u^2\,du \\ &= \frac{1}{3}u^3 + C = \frac{1}{3}(2x+1)^{\frac{3}{2}} + C.\end{aligned}$$

□

(3) Evaluate $\displaystyle\int \frac{x}{\sqrt[3]{2x^2+1}}\,dx$.

Proof. Let $u = 2x^2 + 1$. Then $\dfrac{du}{dx} = 4x$ and $du = 4xdx$. Thus $xdx = \dfrac{1}{4}du$. We have

$$\begin{aligned}\int \frac{x}{\sqrt[3]{2x^2+1}}\,dx &= = \int \frac{1}{\sqrt[3]{u}} \cdot \frac{1}{4}du = \frac{1}{4}\int u^{-1/3}\,du \\ &= \frac{1}{4} \cdot \frac{1}{2/3}u^{2/3} + C = \frac{3}{8}(2x^2+1)^{\frac{2}{3}} + C.\end{aligned}$$

□

(4) Evaluate $\displaystyle\int \cos 5x\,dx$.

Proof. Let $u = 5x$. Then $du = 5dx$. Hence,

$$\begin{aligned}\int \cos 5x\,dx &= \int \cos u \cdot \frac{1}{5}du = \frac{1}{5}\int \cos u\,du \\ &= \frac{1}{5}\sin u + C = \frac{1}{5}\sin 5x + C.\end{aligned}$$

□

(5) Evaluate $\displaystyle\int x^5\sqrt{1+x^2}\,dx$.

Proof. Let $u = 1 + x^2$. Then $du = 2xdx$ and $x^2 = u - 1$. Hence,

$$\begin{aligned}\int \sqrt{1+x^2} \cdot x^5\,dx &= \int \sqrt{1+x^2} \cdot (x^2)^2 \cdot xdx \\ &= \frac{1}{2}\int \sqrt{u}(u-1)^2\,du \\ &= \frac{1}{2}\int u^{5/2} - 2u^{3/2} + u^{1/2}\,du \\ &= \frac{1}{2}\Big(\frac{2}{7}u^{\frac{7}{2}} - \frac{4}{5}u^{\frac{5}{2}} + \frac{2}{3}u^{\frac{3}{2}}\Big) + C \\ &= \frac{1}{7}(1+x^2)^{\frac{7}{2}} - \frac{2}{5}(1+x^2)^{\frac{5}{2}} + \frac{1}{3}(1+x^2)^{\frac{3}{2}} + C.\end{aligned}$$

□

(6) Evaluate $\displaystyle\int \tan x\, dx$.

Proof. We observe that $\tan x = \dfrac{\sin x}{\cos x}$. Let $u = \cos x$. Then $du = -\sin x\, dx$. Hence,

$$\begin{aligned}\int \tan x\, dx &= \int \frac{\sin x}{\cos x}\, dx = -\int \frac{1}{u}\, du = -\ln|u| + C \\ &= -\ln|\cos x| + C = \ln|\sec x| + C.\end{aligned}$$

□

■ **Definite Integral**

Example 8.1.3.

$$\int_0^4 \sqrt{2x+1}\, dx = \int \sqrt{2x+1}\, dx\Big|_0^4 = \frac{1}{3}(2x+1)^{1/2}\Big|_0^4 = \frac{1}{3}(27-1) = \frac{26}{3}.$$

Theorem 8.1.4. *If f and u' are continuous, then*

$$\int_a^b f\big(u(x)\big)u'(x)\, dx = \int_{u(a)}^{u(b)} f(u)\, du.$$

Proof. Let $F(x) = \displaystyle\int_{u(a)}^{x} f(t)\, dt$, then $F'(x) = f(x)$. Hence,

$$\frac{d}{dx}\Big(F\big(u(x)\big)\Big) = F'\big(u(x)\big)u'(x) = f\big(u(x)\big)u'(x).$$

Then

$$\int_a^b f\big(u(x)\big)u'(x)\, dx = \int_a^b \frac{d}{dx}\Big(F\big(u(x)\big)\Big)\, dx = F\big(u(b)\big) - F\big(u(a)\big) = \int_{u(a)}^{u(b)} f(u)\, du.$$

□

Note. The theorem notices that the upper and lower limits of the integral will change when we take the change of variables. The readers sholud carefully deal with this.

Example 8.1.5.

(1) Evaluate $\displaystyle\int_0^{1/2} \cos^3(\pi x)\sin(\pi x)\, dx$.

Proof. Let $u = \cos(\pi x)$. Then $du = -\pi\sin(\pi x)\, dx$. Hence,

$$\int_0^{1/2} \cos^3(\pi x)\sin(\pi x)\, dx = -\frac{1}{\pi}\int_1^0 u^3\, du = -\frac{1}{\pi}\Big(\frac{1}{4}u^4\Big|_1^0\Big) = \frac{1}{4\pi}.$$

□

(2) Evaluate $\displaystyle\int_1^e \frac{\ln x}{x}\,dx$.

Proof. Let $u = \ln x$. Then $du = \frac{1}{x}\,dx$. Hence,

$$\int_1^e \frac{\ln x}{x}\,dx = \int_0^1 u\,du = \frac{1}{2}u^2\Big|_0^1 = \frac{1}{2}.$$

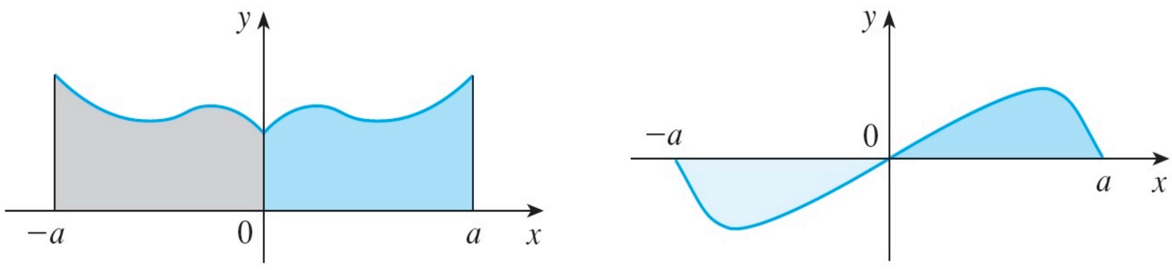
□

■ Integral of Symmetric Functions

Proposition 8.1.6. *Suppose that f is integrable on $[-a, a]$.*

(a) If f is even, then $\displaystyle\int_{-a}^{a} f(x)\,dx = 2\int_0^a f(x)\,dx.$

(b) If f is odd, then $\displaystyle\int_{-a}^{a} f(x)\,dx = 0.$



(a) f even, $\int_{-a}^{a} f(x)\,dx = 2\int_0^a f(x)\,dx$ (b) f odd, $\int_{-a}^{a} f(x)\,dx = 0$

Proof. (a)

$$\begin{aligned}
\int_{-a}^{a} f(x)\,dx &= \int_{-a}^{0} f(x)\,dx + \int_0^a f(x)\,dx \\
(\text{let } u = -x) \quad &= -\int_a^0 f(-u)\,du + \int_0^a f(x)\,dx \\
&= \int_0^a f(u)\,du + \int_0^a f(x)\,dx \\
&= 2\int_0^a f(x)\,dx.
\end{aligned}$$

(b) Skip

□

Example 8.1.7.

(1) The function $f(x) = x^6 + 1$ is an even function on $[-2, 2]$ since $f(-x) = f(x)$. Then

$$\int_{-2}^{2} x^6 + 1\, dx = 2\int_{0}^{2} x^6 + 1\, dx = 2(\frac{1}{7}x^7 + x)\Big|_0^2 = \frac{284}{7}.$$

(2) The function $f(x) = \dfrac{\tan x}{1 + x^2 + x^4}$ is an odd function on $[-1, 1]$ since $f(-x) = f(x)$. Then

$$\int_{-1}^{1} \frac{\tan x}{1 + x^2 + x^4}\, dx = 0.$$

8.2 Integration by Parts

In the present section, we will study another technique of integration which is an inverse operation of product rule of differentiation.

$$\text{Product rule} \quad \longleftrightarrow \quad \text{Integration by parts}$$

Let $u(x)$ and $v(x)$ be differentiable functions. By the product rule,

$$\frac{d}{dx}\Big(u(x)v(x)\Big) = u'(x)v(x) + u(x)v'(x).$$

By the fundamental theorem of calculus,

$$u(x)v(x) = \int \frac{d}{dx}\Big(u(x)v(x)\Big)\, dx = \int u'(x)v(x) + u(x)v'(x)\, dx + C.$$

Then

$$\int u(x)v'(x)\, dx = u(x)v(x) - \int u'(x)v(x)\, dx + C$$

The process is called the "*integration by parts*".

By using the symbols of differential, $du = u'(x)dx$ and $dv = v'(x)dv$. Then

$$\int u(x) \underbrace{v'(x)\, dx}_{dv} = u(x)v(x) - \int v(x) \underbrace{u'(x)\, dx}_{du} + C.$$

Another form of the integration by parts is

$$\int u\, dv = uv - \int v\, du.$$

Strategy:

(i) Obersve the integrand as a product of two functions

(ii) One will be differentiated and the other will be integrated

(iii) Convert the integral of $\int u\, dv$ into $\int v\, du$ and to solve the latter integral.

Example 8.2.1.

(1) Evaluate $\int xe^x\,dx$.

Proof.

Solution1:

$$\begin{aligned}\int xe^x\,dx &\overset{I.B.P}{=} xe^x - \int e^x\,dx \qquad [\text{where } x = u(x),\ e^x = v'(x)]\\ &= xe^x - e^x + C\end{aligned}$$

Solution2: Let $u = x$ and $dv = e^x dx$. Then $du = dx$ and $v = e^x$. Hence,

$$\begin{aligned}\int xe^x\,dx &= \int u\,dv \overset{I.B.P}{=} uv - \int v\,du\\ &= xe^x - \int e^x\,dx = xe^x - e^x + C.\end{aligned}$$

□

(2) Evaluate $\int x\ln x\,dx$.

Proof. Let $u = \ln x$ and $dv = xdx$. Then $du = \frac{1}{x}dx$ and $v = \frac{1}{2}x^2$. Hence,

$$\begin{aligned}\int x\ln x\,dx &= \int u\,dv = \overset{I.B.P}{=} uv - \int v\,du\\ &= (\ln x)\cdot\frac{1}{2}x^2 - \int \frac{1}{2}x^2\cdot\frac{1}{x}\,dx = \frac{1}{2}x^2\ln x - \frac{1}{2}\int x\,dx\\ &= \frac{1}{2}x^2\ln x - \frac{1}{4}x^2 + C.\end{aligned}$$

□

(3) Evaluate $\int x^2e^x\,dx$.

Proof. Use the integration by parts twice,

$$\begin{aligned}\int x^2e^x\,dx &= x^2e^x - \int 2xe^x\,dx \qquad [\text{where } u(x) = x^2,\ v'(x) = e^x]\\ &= x^2e^x - 2\int xe^x\,dx\\ &= x^2e^x - 2\Big(xe^x - \int e^x\,dx\Big) \qquad [\text{where } u(x) = x,\ v'(x) = e^x]\\ &= x^2e^2 - 2\Big(xe^x - e^x\Big) + C\\ &= x^2e^x - 2xe^x + 2e^x + C.\end{aligned}$$

□

(4) Evaluate $\int \ln x \, dx$.

Proof.

$$\int \ln x \, dx = \int \ln x \cdot 1 \, dx = x \ln x - \int \frac{1}{x} \cdot x \, dx = x \ln x - \int 1 \, dx = x \ln x - x + C.$$

□

(5) Evaluate $\int e^x \sin x \, dx$.

Proof. Use the integration by parts twice,

$$\begin{aligned}
\int e^x \sin x \, dx &= e^x \sin x - \int e^x \cos x \, dx \\
&= e^x \sin x - \left(e^x \cos x + \int e^x \sin x \, dx\right) \\
&= e^x \sin x - e^x \cos x - \int e^x \sin x \, dx
\end{aligned}$$

Then

$$\int e^x \sin x \, dx = \frac{1}{2} e^x (\sin x - \cos x) + C.$$

□

(6) Evaluate $\int x^5 \cos x^3 \, dx$.

Proof. Let $u = x^3$ then $du = 3x^2 dx$. We have

$$\begin{aligned}
\int x^5 \cos x^3 \, dx &= \int x^3 \cos x^3 \cdot x^2 \, dx = \frac{1}{3} \int u \cos u \, du \\
&= \frac{1}{3} \left[u \sin u - \int \sin u \, du\right] \\
&= \frac{1}{3} \left[u \sin u + \cos u + C\right] \\
&= \frac{1}{3} (x^3 \sin x^3 + \cos x^3) + C.
\end{aligned}$$

□

(7) Evaluate $\int (\ln x)^2 \, dx$.

Proof. Let $u = \ln x$. Then$x = e^u$ and $du = \frac{1}{x} dx$. Thus $dx = x du = e^u du$ and we have

$$\begin{aligned}
\int (\ln x)^2 \, dx &= \int u^2 e^u \, du = \cdots \qquad (I.B.P \text{ twice}) \\
&= u^2 e^u - 2ue^u + 2e^u + C \\
&= x (\ln x)^2 - 2x \ln x + 2x + C.
\end{aligned}$$

□

Example 8.2.2. For $n \in \mathbb{N}$, evaluate $\displaystyle\int \sin^n x\,dx$.

Proof.

For $n = 1$, $\displaystyle\int \sin x\,dx = -\cos x + C$.
For $n = 2$, $\displaystyle\int \sin x\,dx = \int \frac{1-\cos 2x}{2}\,dx = \frac{1}{2}(x - \frac{1}{2}\sin 2x) + C = \frac{1}{2}x - \frac{1}{4}\sin 2x + C$.
For $n \geq 3$,

$$\begin{aligned}
\int \sin^n x\,dx &= \int \sin^{n-1} x \cdot \sin x\,dx \\
&= \sin^{n-1} x \cdot (-\cos x) - \int (n-1)\sin^{n-2} x \cos x \cdot (-\cos x)\,dx \\
&= -\sin^{n-1} x \cos x + (n-1)\int \sin^{n-2} x(\cos^2 x)\,dx \\
&= -\sin^{n-1} x \cos x + (n-1)\int \sin^{n-2} x(1-\sin^2 x)\,dx \\
&= -\sin^{n-1} x \cos x + (n-1)\int \sin^{n-2} x\,dx - (n-1)\int \sin^n x\,dx.
\end{aligned}$$

Then

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

□

■ **Definite Integral**

$$\int_a^b u(x)v'(x)\,dx = u(x)v(x)\Big|_a^b - \int_a^b u'(x)v(x)\,dx.$$

Example 8.2.3.

$$\begin{aligned}
\int_0^1 \tan^{-1} x\,dx &= x\tan^{-1} x\Big|_0^1 - \int_0^1 \frac{x}{1+x^2}\,dx \\
&= x\tan^{-1} x\Big|_0^1 - \int_1^2 \frac{1}{u}\,du \qquad (\text{let } u = 1+x^2) \\
&= \frac{\pi}{4} - \frac{1}{2}\ln u\Big|_1^2 = \frac{\pi}{4} - \frac{1}{2}\ln 2.
\end{aligned}$$

8.3 Trigonometric Integrals

In the present section, we will study the integrals of combination of trigonometric functions with some specific forms.

(I) **Product of sine and cosine:**

$$\int \sin^m x \cos^n x\,dx \quad \text{for } m, n \in \mathbb{N}$$

Case1: Either m or n is odd.

For example $m = 2k + 1$, then taking $u = \cos x$.

Example 8.3.1. $\displaystyle\int \sin^3 x\, dx$.

Proof. Let $u = \cos x$. Then $du = -\sin x dx$.

$$\begin{aligned}
\int \sin^3 x\, dx &= \int \sin^2 x \sin x\, dx = \int (1 - \cos^2 x) \sin x\, dx \\
&= -\int 1 - u^2\, du = -(u - \frac{1}{3}u^3) + C \\
&= -\cos x + \frac{1}{3}\cos^3 x + C.
\end{aligned}$$

□

$$\int \sin^6 x \cos^5 x\, dx$$

Proof. Let $u = \sin x$. Then $du = \cos x dx$.

$$\begin{aligned}
\int \sin^6 \cos^5 x\, dx &= \int \sin^6 x \cos^4 x \cos x\, dx = \int \sin^6 x(1 - \cos^2 x)^2 \cos x\, dx \\
&= \int u^6(1 - u^2)^2\, du = \int u^6 - 2u^8 + u^{10}\, du \\
&= \frac{1}{7}u^7 - \frac{2}{9}u^9 + \frac{1}{11}u^{11} + C \\
&= \frac{1}{7}\sin^7 x - \frac{2}{9}\sin^9 x + \frac{1}{11}\sin^{11} x + C.
\end{aligned}$$

□

Case2: Both m and n are even.

Using the half-angle identity, either we can reduced the integrand $\sin^m x \cos^n x$ to the form of Case1, or it can be coverted into another form of Case2. Then taking the half-angle identity until it can be coverted into the form of Case1.

Example 8.3.2.

(1) $\displaystyle\int \sin^4 x\, dx$.

Solution1 : Using the integration by parts to lower down the power of sine function by 2 each time.

Solution2 :

$$
\begin{aligned}
\int \sin^4 x\,dx &= \int \Big(\frac{1-\cos 2x}{2}\Big)^2\,dx = \frac{1}{4}\int 1-2\cos 2x+\cos^2 2x\,dx\\
&= \frac{1}{4}\int 1-2\cos 2x+\frac{1+\cos 4x}{2}\,dx\\
&= \frac{1}{4}\int \frac{3}{2}-2\cos 2x+\frac{1}{2}\cos 4x\,dx\\
&= \frac{3}{8}x-\frac{1}{4}\sin 2x+\frac{1}{32}\sin 4x+C.
\end{aligned}
$$

(2)

$$
\begin{aligned}
\int \sin^4 x\cos^2 x\,dx &= \int (1-\cos^2 x)^2\cos^2 x\,dx = \int \cos^2 x-2\cos^4 x+\cos^6 x\,dx\\
&= \int \frac{1+\cos 2x}{2}-2\Big(\frac{1+\cos 2x}{2}\Big)^2+\Big(\frac{1+\cos 2x}{2}\Big)^3\,dx\\
&= \frac{1}{8}\int 1-\cos 2x-\cos^2 2x+\cos^3 2x\,dx\\
&= \frac{1}{8}\Big[\int 1-\cos 2x-\Big(\frac{1+\cos 4x}{2}\Big)\,dx+\int \cos^2 2x\cdot\cos 2x\,dx\Big]\\
&= \cdots\\
&= \frac{1}{16}x-\frac{1}{64}\sin 4x-\frac{1}{48}\sin^3 2x+C.
\end{aligned}
$$

(II) Product of tangent and secant:

$$\int \tan^m x\sec^n x\,dx \quad \text{for } m,n\in\mathbb{N}$$

Case1: ***n* is even.**
Let $u=\tan x$. Then $du=\sec^2 x dx$.

Example 8.3.3.

$$
\begin{aligned}
\int \tan^5 x\sec^6 x\,dx &= \int \tan^5 x\sec^4 x\sec^2 x\,dx\\
&= \int \tan^5 x(1+\tan^2 x)^2\sec^2 x\,dx\\
&= \int u^5(1+u^2)^2\,du \qquad (\text{let } u=\tan x)\\
&= \int u^9-2u^7+u^5\,du\\
&= \frac{1}{10}u^{10}-\frac{1}{4}u^8+\frac{1}{6}u^6+C\\
&= \frac{1}{10}\tan^{10}x-\frac{1}{4}\tan^8 x+\frac{1}{6}\tan^6 x+C.
\end{aligned}
$$

Case2: ***m* is odd.**
Let $u=\sec x$. Then $du=\tan x\sec x\,dx$.

Example 8.3.4.

(1)

$$\begin{aligned}
\int \tan^5 x \sec^6 x\,dx &= \int \tan^4 x \sec^5 x \tan x \sec x\,dx \\
&= \int (\sec^2 x - 1)^2 \sec^5 x(\tan x \sec x)\,dx \\
&= \int (u^2-1)^2 u^5\,du \qquad (\text{let } u = \sec x) \\
&= \int u^9 - 2u^7 + u^5\,du \\
&= \frac{1}{10}u^{10} - \frac{1}{4}u^8 + \frac{1}{6}u^6 + C \\
&= \frac{1}{10}\sec^{10} x - \frac{1}{4}\sec^8 x + \frac{1}{6}\sec^6 x + C.
\end{aligned}$$

(2) Recall that $\int \tan x\,dx = \ln|\sec x| + C$.

$$\begin{aligned}
\int \tan^3 x\,dx &= \int \tan x(\sec^2 x - 1)\,dx \\
&= \int \tan x \sec^2 x - \tan x\,dx \\
&= \int \sec x(\tan x \sec x)\,dx - \int \tan x\,dx \\
&= \int u\,du - \ln|\sec x| + C \qquad (\text{let } u = \sec x) \\
&= \frac{1}{2}u^2 - \ln|\sec x| + C \\
&= \frac{1}{2}\sec^2 x - \ln|\sec x| + C.
\end{aligned}$$

Case3: Others, m is even or n is odd.
Notice that if $m = 2k$, we can convert the term $\tan^{2k} x$ into $(\sec^2 x - 1)^k$. Hence, the integral

$$\int \tan^{2k} x \sec^n x\,dx = \int (\sec^2 x - 1)^k \sec^n x\,dx.$$

Suppose that we can compute $\int \sec^k x\,dx$ for any $k \in \mathbb{N}$. Then every integral in Case3 can be evaluated.

(i) ($k = 1$)

$$\begin{aligned}
\int \sec x\,dx &= \int \sec x \cdot \frac{\sec x + \tan x}{\sec x + \tan x}\,dx = \int \frac{1}{u}\,du \qquad (\text{let } u = \sec x + \tan x) \\
&= \ln|u| + C \\
&= \ln\big|\sec x + \tan x\big| + C.
\end{aligned}$$

(ii) $(k = 2)$

$$\int \sec^2 x\, dx = \tan x + C.$$

(iii) $(k \geq 3,\ \text{integer})$ By the integration by parts,

$$\int \sec^k x\, dx = \frac{1}{n-1}\tan x \sec^{k-2} x + \frac{n-2}{n-1}\int \sec^{k-2} x\, dx.$$

(III)

$$\int \sin mx \cos nx\, dx, \qquad \int \sin mx \sin nx\, dx, \qquad \int \cos mx \cos nx\, dx.$$

By the identities,

$$\begin{aligned} \sin A \cos B &= \frac{1}{2}[\sin(A-B) + \sin(A+B)] \\ \sin A \sin B &= \frac{1}{2}[\cos(A-B) - \cos(A+B)] \\ \cos A \cos B &= \frac{1}{2}[\cos(A-B) + \cos(A+B)] \end{aligned}$$

8.4 Trigonometric Substitution

Recall that the substitution method says that if $u = g(x)$ then $du = g'(x)dx$ and

$$\int \underbrace{f\big(g(x)\big)}_{f(u)} \underbrace{g'(x)\, dx}_{du} = \int f(u)\, du.$$

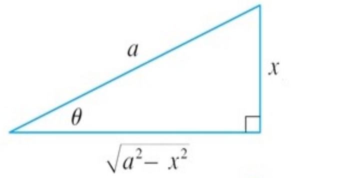
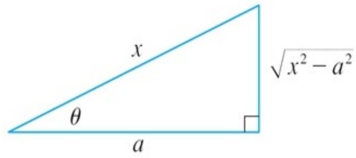
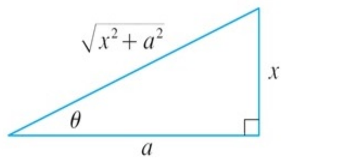
In the formula, "x" is the old variable in the left hand side and "u" is a new variable in the right hand side. The new variable u is a function of the old variable x. Conversely, assume that the old variable x is a function of a new variable t, say $x = g(t)$. Then $dx = g'(t)dt$ and we have the "*inverse substitutuion*"

$$\int f(x)\, dx = \int \underbrace{f\big(g(t)\big)}_{f(x)} \underbrace{g'(t)\, dt}_{dx}.$$

Note. The inverse substitution provides a new method to evaluate the integral $\int f(x)\, dx$. Suppose that we can find a suitable function $x = g(t)$ such that we could compute the integral $\int f\big(g(t)\big)g'(t)\, dt$ and the problem would be solved. In general, the suitable function g is not easy to find. But, it is effective for the given radical expression because of the secified trigonometric identities.

❑ Trigonometric Substitutions

In the present section, we will set $x = a\sin\theta$, $x = a\tan\theta$ and $x = a\sec\theta$ to deal with the integral with integrand $\sqrt{a^2 - x^2}$, $\sqrt{a^2 + x^2}$ and $\sqrt{x^2 - a^2}$ respectively.

Expression	Substitution	Identity	
$\sqrt{a^2 - x^2}$	$x = a \sin\theta,\quad -\frac{\pi}{2} \leqslant \theta \leqslant \frac{\pi}{2}$	$1 - \sin^2\theta = \cos^2\theta$	
$\sqrt{a^2 + x^2}$	$x = a \tan\theta,\quad -\frac{\pi}{2} < \theta < \frac{\pi}{2}$	$1 + \tan^2\theta = \sec^2\theta$	
$\sqrt{x^2 - a^2}$	$x = a \sec\theta,\quad 0 \leqslant \theta < \frac{\pi}{2}$ or $\pi \leqslant \theta < \frac{3\pi}{2}$	$\sec^2\theta - 1 = \tan^2\theta$	

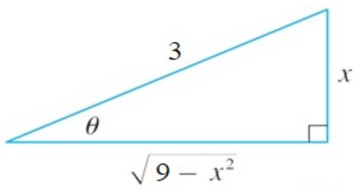
Example 8.4.1.

(1) Evaluate $\displaystyle\int \frac{\sqrt{9-x^2}}{x^2}\,dx$.

Proof.

Let $x = 3\sin\theta$, $-\frac{\pi}{2} \le \theta \le \frac{\pi}{2}$. Then $dx = 3\cos\theta\, d\theta$.

$$\begin{aligned}
\int \frac{\sqrt{9-x^2}}{x^2}\,dx &= \int \frac{3\cos\theta}{9\sin^2\theta}\cdot 3\cos\theta\, d\theta = \int \cot^2\theta\, d\theta \\
&= \int \csc^2\theta - 1\, d\theta = -\cot\theta - \theta + C \\
&= -\frac{\sqrt{9-x^2}}{x} - \sin^{-1}\left(\frac{x}{3}\right) + C.
\end{aligned}$$



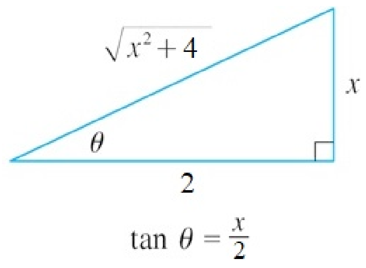
$\sin\theta = \frac{x}{3}$

□

(2) Evaluate $\displaystyle\int \frac{1}{x^2\sqrt{x^2+4}}\,dx$.

Proof.

Let $x = 2\tan\theta$, $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$. Then $dx = 2\sec^2\theta\, d\theta$.

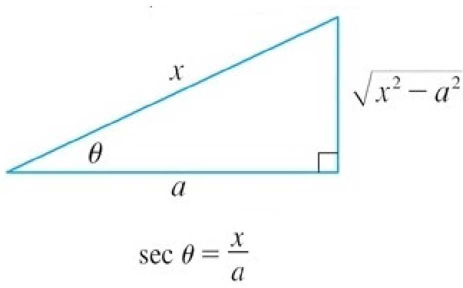


$$\begin{aligned}
\int \frac{1}{x^2\sqrt{x^2+4}}\,dx &= \int \frac{1}{4\tan^2\theta\cdot 2\sec\theta}\cdot 2\sec^2\theta\,d\theta \\
&= \frac{1}{4}\int \frac{\sec\theta}{\tan^2\theta}\,d\theta = \frac{1}{4}\int \frac{\cos\theta}{\sin^2\theta}\,d\theta \\
(\text{let } u = \sin\theta) \quad &= \frac{1}{4}\int \frac{1}{u^2}\,du = -\frac{1}{4u} + C \\
&= -\frac{1}{4\sin\theta} + C = -\frac{\sqrt{x^2+4}}{4x} + C.
\end{aligned}$$

□

(3) Evaluate $\displaystyle\int \frac{1}{\sqrt{x^2-a^2}}\,dx$, $a > 0$.

Proof. Let $x = \sec\theta$, $0 < \theta < \frac{\pi}{2}$ or $\frac{\pi}{2} < \theta < \pi$. Then $dx = a\tan\theta\sec\theta\,d\theta$.

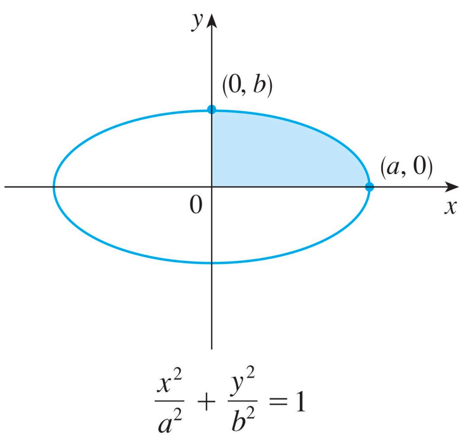


$$\begin{aligned}
\int \frac{1}{\sqrt{x^2-a^2}}\,dx &= \int \frac{1}{a\tan\theta}\cdot a\tan\theta\sec\theta\,d\theta \\
&= \int \sec\theta\,d\theta = \ln|\sec\theta + \tan\theta| + C \\
&= \ln\left|\frac{x}{a} + \frac{\sqrt{x^2-a^2}}{a}\right| + C \\
&= \ln\left|x + \sqrt{x^2-a^2}\right| - \ln a + C \\
&= \ln\left|x + \sqrt{x^2-a^2}\right| + C.
\end{aligned}$$

□

(4) Find the area enclosed by the ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$.

Proof.



$$\begin{aligned}
\text{Area} &= 4\int_0^a b\sqrt{1-\frac{x^2}{a^2}}\,dx = \frac{4b}{a}\int_0^a \sqrt{a^2-x^2}\,dx \\
&= \frac{4b}{a}\int_0^{\frac{\pi}{2}} a\cos\theta\cdot a\cos\theta\,d\theta \quad (\text{let } x = a\sin\theta) \\
&= 4ab\int_0^{\frac{\pi}{2}} \frac{1+\cos^2\theta}{2}\,d\theta \\
&= 2ab\left(\theta + \frac{1}{2}\sin 2\theta\right)\Big|_0^{\frac{\pi}{2}} = \pi ab.
\end{aligned}$$

□

(5) Evaluate $\displaystyle\int_0^{\frac{3\sqrt{3}}{2}} \frac{x^3}{(4x^2+9)^{3/2}}\,dx$.

Proof. Let $x = \frac{3}{2}\tan\theta$, then $ds = \frac{3}{2}\sec^2\theta\,d\theta$.

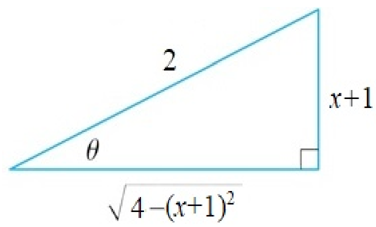
$$\begin{aligned}
\int_0^{\frac{3\sqrt{3}}{2}} \frac{x^3}{(4x^2+9)^{3/2}}\,dx &= \int_0^{\frac{\pi}{3}} \frac{\frac{27}{8}\tan^3\theta}{27\sec^3\theta}\cdot\frac{3}{2}\sec^2\theta\,d\theta = \frac{3}{16}\int_0^{\frac{\pi}{3}}\frac{\tan^3\theta}{\sec\theta}\,d\theta \\
&= \frac{3}{16}\int_0^{\frac{\pi}{3}}\frac{\sin^3\theta}{\cos^2\theta}\,d\theta = \frac{3}{16}\int_1^{\frac{1}{2}}\frac{1-u^2}{u^2}\,(-du) \\
&= \frac{3}{16}\int u^{-2}-1\,du = \frac{3}{16}\Big(-u^{-1}-u\Big)\Big|_{\frac{1}{2}}^{1} \\
&= \frac{3}{32}.
\end{aligned}$$

□

(6) Evaluate $\displaystyle\int \frac{x}{\sqrt{3-2x-x^2}}\,dx$.

Proof. Let $x+1 = 2\sin\theta$. Then $dx = 2\cos\theta\,d\theta$.

$$\begin{aligned}
\int \frac{x}{\sqrt{3-2x-x^2}}\,dx &= \int \frac{x}{\sqrt{4-(x+1)^2}} \\
&= \int \frac{2\sin\theta-1}{2\cos\theta}\cdot 2\cos\theta\,d\theta \\
&= \int 2\sin\theta - 1\,d\theta = -2\cos\theta-\theta+C \\
&= -\sqrt{4-(x+1)^2}-\sin\Big(\frac{x+1}{2}\Big)+C.
\end{aligned}$$



□

8.5 Partial Fractions

In this section, we will try to solve the integral of rational functions. Let's observe the following example that

$$\begin{aligned}
\int \frac{2}{x+1}\,dx &= 2\ln|x+1|+C \quad \text{and} \\
\int \frac{1}{x-2}\,dx &= \ln|x-2|+C.
\end{aligned}$$

Then

$$\int \frac{x-5}{x^2-x-2}\,dx = \int \frac{2}{x+1}-\frac{1}{x-2}\,dx = 2\ln|x+1|-\ln|x-2|+C.$$

Question: For a general rational function f, can we express f as sum of several fractions such that we can evaluate the integral of each fraction?

■ Breaking a rational function into several fractions

Consider

$$f(x) = \frac{P(x)}{Q(x)}$$

where

$$\begin{array}{ll} P(x) = a_n x^n + \cdots + a_1 x + a_0, & \\ Q(x) = b_m x^m + \cdots + b_1 x + b_0, & a_n, b_m \neq 0 \end{array}$$

Definition 8.5.1. If $n < m$, we call $f(x) \left(= \dfrac{P(x)}{Q(x)} \right)$ a "*proper*" rational function; if $n \geq m$, we call f a "*improper*" rational function.

Notice that in high school algebra, we can use long-divison to express a rational function as a sum of a polynomial and a proper rational function. That is,

$$\frac{P(x)}{Q(x)} = \underbrace{S(x)}_{\text{polynomial}} + \underbrace{\frac{R(x)}{Q(x)}}_{\text{proper rational function}}.$$

■ Strategy of the integration of ration functions

$$\int \frac{P(x)}{Q(x)}\, dx$$

Step1: By using the long-divison to express $\dfrac{P(x)}{Q(x)} = S(x) + \dfrac{R(x)}{Q(x)}$. Hence,

$$\int \frac{P(x)}{Q(x)}\, dx = \int S(x)\, dx + \int \frac{R(x)}{Q(x)}\, dx.$$

Step2: Factorizing the denominator $Q(x)$ as far as possible. For example,

$$\begin{aligned} Q(x) &= x^4 - 16 = (x^2+4)(x+2)(x-2) \\ Q(x) &= x^3 - 5x^2 + 7x - 2 = (x-2)(x - \frac{3+\sqrt{5}}{2})(x - \frac{3-\sqrt{5}}{2}) \\ Q(x) &= x^5 - 2x^4 + 6x + 32 = (x-2)^2(x+2)(x^2+4) \\ Q(x) &= x^3 - 5x^2 + 12x - 12 = (x-2)(x^2 - 3x + 6) \end{aligned}$$

Step3: To express $\dfrac{R(x)}{Q(x)}$ as a sum of several terms of the forms

$$\frac{A}{(ax+b)^i} \quad \text{or} \quad \frac{Ax+B}{(ax^2+bx+c)^i}$$

Note. Not all improper rational functions can be expressed as a sum of the above terms.

Step4: Take the integral of each of the above terms and use the techniques in the previous sections to solve them.

■ **Different Cases**

• **Case1:** $Q(x) = (a_1x + b_1)(a_2x + b_2)\cdots(a_kx + b_k)$ all distinct (i.e. no factor repeated). Express

$$\frac{R(x)}{Q(x)} = \frac{A_1}{(a_1x + b_1)} + \frac{A_2}{(a_2x + b_2)} + \cdots + \frac{A_k}{(a_kx + b_k)}.$$

and solve $A_1, \cdots, A_k$.

Example 8.5.2. Evaluate $\displaystyle\int \frac{x^2 + 2x - 1}{2x^3 + 3x^2 - 2x}\,dx.$

Proof. Consider the factorization $2x^3 + 3x^2 - 2x = x(2x - 1)(x + 2)$. We can express

$$\frac{x^2 + 2x - 1}{2x^3 + 3x^2 - 2x} = \frac{A}{x} + \frac{B}{2x - 1} + \frac{C}{x + 2} = \frac{1}{2}\cdot\frac{1}{x} + \frac{1}{5}\cdot\frac{1}{2x - 1} - \frac{1}{10}\cdot\frac{1}{x + 2}.$$

Hence,

$$\begin{aligned}\int \frac{x^2 + 2x - 1}{2x^3 + 3x^2 - 2x}\,dx &= \frac{1}{2}\int \frac{1}{x}\,dx + \frac{1}{5}\int \frac{1}{2x - 1}\,dx - \frac{1}{10}\int \frac{1}{x + 2}\,dx \\ &= \frac{1}{2}\ln|x| + \frac{1}{10}\ln|2x - 1| - \frac{1}{10}\ln|x + 2| + C.\end{aligned}$$

□

• **Case2:** $Q(x) = (a_1x + b_1)^{r_1}(a_2x + b_2)^{r_2}\cdots(a_kx + b_k)^{r_k}$. Express

$$\begin{aligned}\frac{R(x)}{Q(x)} &= \frac{A_{11}}{a_1x + b_1} + \frac{A_{12}}{(a_1x + b_1)^2} + \cdots + \frac{A_{1r_1}}{(a_1x + b_1)^{r_1}} \\ &+ \\ &\vdots \\ &+ \frac{A_{k1}}{a_kx + b_k} + \frac{A_{k2}}{(a_kx + b_k)^2} + \cdots + \frac{A_{kr_k}}{(a_kx + b_k)^{r_k}}.\end{aligned}$$

and solve $A_{11}, \cdots, A_{kr_k}$.

Example 8.5.3. Evaluate $\displaystyle\int \frac{4x}{x^3 - x^2 - x + 1}\,dx.$

Proof. Consider the factorization $x^3 - x^2 - x + 1 = (x - 1)^2(x + 1)$. Then

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

Hence,

$$\begin{aligned}\int \frac{4x}{x^3 - x^2 - x + 1}\,dx &= \int \frac{1}{x + 1}\,dx + 2\int \frac{1}{(x - 1)^2}\,dx - \int \frac{1}{x + 1}\,dx \\ &= \ln|x - 1| - \frac{2}{x - 1} - \ln|x + 1| + C.\end{aligned}$$

□

• **Case3:** $Q(x) = (a_1x^2 + b_1x + c_1)(a_2x^2 + b_2x + c_2)\cdots(a_kx^2 + b_kx + b_k)$. Express

$$\frac{R(x)}{Q(x)} = \frac{A_1x + B_1}{(a_1x^2 + b_1x + c_1)} + \frac{A_2x + B_2}{(a_2x^2 + b_2x + c_2)} + \cdots + \frac{A_kx + B_k}{(a_kx^2 + b_kx + c_k)}.$$

Example 8.5.4. Evaluate $\displaystyle\int \frac{2x^2 - x^4}{x^3 + 4x}\,dx.$

Proof. Consider the factorization $x^3 + 4x = x(x^2 + 4)$. Then

$$\frac{2x^2 - x + 4}{x^3 + 4x} = \frac{A}{x} + \frac{Bx + C}{x^2 + 4} = \frac{1}{x} + \frac{x - 1}{x^2 + 4}.$$

Hence,

$$\begin{aligned}
\int \frac{2x^2 - x + 4}{x^3 + 4x}\,dx &= \int \frac{1}{x}\,dx + \int \frac{x - 1}{x^2 + 4}\,dx \\
&= \int \frac{1}{x}\,dx + \frac{1}{2}\int \frac{2x}{x^2 + 4}\,dx - \int \frac{1}{x^2 + 4}\,dx \\
&= \ln|x| + \frac{1}{2}\ln|x^2 + 1| - \tan^{-1}\left(\frac{x}{2}\right) + C.
\end{aligned}$$

□

Remark.

(i) In this case, we usually use the trick

$$\int \frac{Cx + D}{x^2 + a^2}\,dx = \frac{C}{2}\int \frac{2x}{x^2 + a^2}\,dx + D\int \frac{1}{x^2 + a^2}\,dx = \frac{C}{2}\ln|x^2 + a^2| + D\tan^{-1}\left(\frac{x}{a}\right) + K.$$

(ii) As long as the denominator $ax^2 + bx + c$ cannot be factorized, $\dfrac{Ax + B}{ax^2 + bx + c}$ must can be expressed as

$$\frac{A}{2a}\cdot\frac{(2ax + b)}{ax^2 + bx + c} + \left(B - \frac{Ab}{2a}\right)\frac{1}{ax^2 + bx + c} = \frac{A}{2a}\cdot\frac{2ax + b}{ax^2 + bx + c} + \left(B - \frac{Ab}{2a}\right)\frac{1}{(\alpha x + \beta)^2 + \gamma^2}.$$

For example,

$$\begin{aligned}
\int \frac{x - 1}{4x^2 - 4x + 3}\,dx &= \frac{1}{8}\int \frac{8x - 4}{4x^2 - 4x + 3}\,dx - \frac{1}{2}\int \frac{1}{(2x - 1)^2 + 2}\,dx \\
&= \frac{1}{8}\ln\left|4x^2 - 4x + 3\right| - \frac{1}{4}\int \frac{1}{u^2 + 2}\,du \qquad (u = 2x - 1) \\
&= \frac{1}{8}\ln\left|4x^2 - 4x + 3\right| - \frac{1}{4}\tan^{-1}\left(\frac{u}{\sqrt{2}}\right) + C \\
&= \frac{1}{8}\ln\left|4x^2 - 4x + 3\right| - \frac{1}{4}\tan^{-1}\left(\frac{2x - 1}{\sqrt{2}}\right) + C.
\end{aligned}$$

- **Case4:** $Q(x) = (a_1x^2 + b_1x + c_1)^{r_1}(a_2x^2 + b_2x + c_2)^{r_2} \cdots (a_kx^2 + b_kx + b_k)^{r_k}$. Express

$$
\begin{aligned}
\frac{R(x)}{Q(x)} &= \frac{A_{11}x + B_{11}}{(a_1x^2 + b_1x + c_1)} + \frac{A_{12}x + B_{12}}{(a_1x^2 + b_1x + c_1)^2} + \cdots + \frac{A_{1r_1}x + B_{1r_1}}{(a_1x^2 + b_1x + c_1)^{r_1}} \\
&+ \\
&\qquad\vdots \\
&+ \frac{A_{k1}x + B_{k1}}{(a_kx^2 + b_kx + c_k)} + \frac{A_{k2}x + B_{k2}}{(a_kx^2 + b_kx + c_k)^2} + \cdots + \frac{A_{kr_k}x + B_{kr_k}}{(a_kx^2 + b_kx + c_k)^{r_k}}.
\end{aligned}
$$

Example 8.5.5. Evaluate $\displaystyle\int \frac{1 - x + 2x^2 - x^3}{x(x^2 + 1)^2}$.

Proof.

$$
\begin{aligned}
\int \frac{1 - x + 2x^2 - x^3}{x(x^2 + 1)^2} &= \int \frac{A}{x}\,dx + \frac{Bx + C}{x^2 + 1}\,dx + \int \frac{Dx + E}{(x^2 + 1)^2}\,dx \\
&= \int \frac{1}{x}\,dx - \int \frac{x + 1}{x^2 + 1}\,dx + \int \frac{x}{(x^2 + 1)^2}\,dx \\
&= \int \frac{1}{x}\,dx + \frac{1}{2}\int \frac{2x}{x^2 + 1}\,dx - \int \frac{1}{x^2 + 1}\,dx + \frac{1}{2}\int \frac{2x}{(x^2 + 1)^2}\,dx \\
&= \ln|x| - \frac{1}{2}\ln\left|x^2 + 1\right| - \tan^{-1} x - \frac{1}{2(x^2 + 1)} + K.
\end{aligned}
$$

□

- **Case5:** $Q(x) = (a_1x + b_1)^{r_1} \cdots (a_kx + b_k)^{r_k}(c_1x^2 + d_1x + e_1)^{s_1} \cdots (c_\ell x^2 + d_\ell x + e_\ell)^{s_\ell}$. Express

$$
\begin{aligned}
\frac{R(x)}{Q(x)} &= \frac{A_{11}}{a_1x + b_1} + \cdots + \frac{A_{1r_1}}{(a_1x + b_1)^{r_1}} \\
&+ \\
&\qquad\vdots \\
&+ \frac{A_{k1}}{a_kx + b_k} + \cdots + \frac{A_{kr_k}}{(a_kx + b_k)^{r_k}} \\
&+ \frac{C_{11}x + D_{11}}{(c_1x^2 + d_1x + e_1)} + \cdots + \frac{C_{1s_1}x + B_{1s_1}}{(c_1x^2 + d_1x + e_1)^{s_1}} \\
&+ \\
&\qquad\vdots \\
&+ \frac{C_{\ell 1}x + D_{\ell 1}}{(c_kx^2 + d_kx + e_k)} + \cdots + \frac{C_{\ell s_\ell}x + D_{\ell s_\ell}}{(c_\ell x^2 + d_\ell x + e_\ell)^{s_\ell}}.
\end{aligned}
$$

■ Rationalizing Substitutions

Example 8.5.6. Evaluate $\displaystyle\int \frac{\sqrt{x + 4}}{x}\,dx$.

Proof. Let $u = \sqrt{x+4}$. Then $x = u^2 - 4$ and $du = \dfrac{1}{2\sqrt{x+4}}dx = \dfrac{1}{2u}\,dx$.

$$\begin{aligned}
\int \frac{\sqrt{x+4}}{x}\,dx &= \int \frac{u}{u^2-4}\cdot 2u\,du = 2\int 1 + \frac{4}{u^2-4}\,du \\
&= 2u + 2\int \frac{1}{u-2} - \frac{1}{u+2}\,du \\
&= 2u + 2\ln\left|\frac{u-2}{u+2}\right| + C \\
&= 2\sqrt{x+4} + 2\ln\left|\frac{\sqrt{x+4}-2}{\sqrt{x+4}+2}\right| + C.
\end{aligned}$$

□

Remark. Since every polynomial function $Q(x)$ can be factorized into products of several 1-degree or 2-degree irreducible polynomial functions, by following above steps and cases, we can deal with the integrations of all rational functions.

8.6 Strategy for Integration

Momorized the following table

Table of Integration Formulas Constants of integration have been omitted.

1. $\int x^n\,dx = \dfrac{x^{n+1}}{n+1} \quad (n \neq -1)$

2. $\int \dfrac{1}{x}\,dx = \ln|x|$

3. $\int e^x\,dx = e^x$

4. $\int a^x\,dx = \dfrac{a^x}{\ln a}$

5. $\int \sin x\,dx = -\cos x$

6. $\int \cos x\,dx = \sin x$

7. $\int \sec^2 x\,dx = \tan x$

8. $\int \csc^2 x\,dx = -\cot x$

9. $\int \sec x \tan x\,dx = \sec x$

10. $\int \csc x \cot x\,dx = -\csc x$

11. $\int \sec x\,dx = \ln|\sec x + \tan x|$

12. $\int \csc x\,dx = \ln|\csc x - \cot x|$

13. $\int \tan x\,dx = \ln|\sec x|$

14. $\int \cot x\,dx = \ln|\sin x|$

15. $\int \sinh x\,dx = \cosh x$

16. $\int \cosh x\,dx = \sinh x$

17. $\int \dfrac{dx}{x^2 + a^2} = \dfrac{1}{a}\tan^{-1}\left(\dfrac{x}{a}\right)$

18. $\int \dfrac{dx}{\sqrt{a^2 - x^2}} = \sin^{-1}\left(\dfrac{x}{a}\right)$

*19. $\int \dfrac{dx}{x^2 - a^2} = \dfrac{1}{2a}\ln\left|\dfrac{x-a}{x+a}\right|$

*20. $\int \dfrac{dx}{\sqrt{x^2 \pm a^2}} = \ln\left|x + \sqrt{x^2 \pm a^2}\right|$

■ **Strategy**

(1) Simplify the integrand if possible.

(2) Look for an obvious substitution.

(3) Classify the integrand according to its form

(a) trigonometric function: products of powers of $\sin x, \cdots, \csc x$.

(b) rational functions: $\dfrac{P(x)}{Q(x)}$

(c) Integration by parts: $\int u(x)v'(x)\,dx = u(x)v(x) - \int u'(x)v(x)\,dx$

(d) radicals: $\sqrt{x^2 \pm a^2}$, $\sqrt{a^2 \pm x^2}$ (trigonometric substitituion); $\sqrt[n]{ax+b}$ (rationalizing substitution)

(4) Try again!

Question: Can we integrate all continuous functions?
Answer: No, the majority of elementary functions don't have elementary antiderivatives. For example, $f(x) = e^{x^2}$ has no antiderivative which is an elementary function.

8.7 Improper Integral

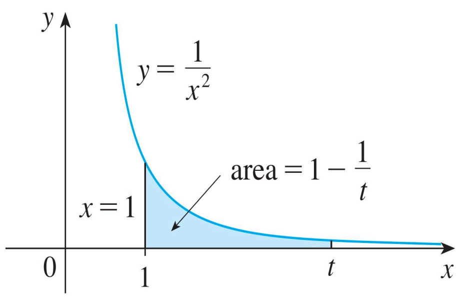
In the previous sections, we discuss the definite integral $\int_a^b f(x)\,dx$ of f under the assumptions that f is defined on a finite interval $[a, b]$ and f does not have an infinite discontinuity. In the presect section, we extend the concept of a definite integral to the case where the interval is infinite and also to the case where f has an infinite discontinuity in $[a, b]$. In either case the integral is called an "*improper integral*".

❏ Type1: Infinite Intervals

Let f be a function defined on an infinite interval such as $[a, \infty]$, $(-\infty, a]$ or $(-\infty, \infty)$.

Example 8.7.1. Let $f(x) = \dfrac{1}{x^2}$ be defined on $[1, \infty)$.

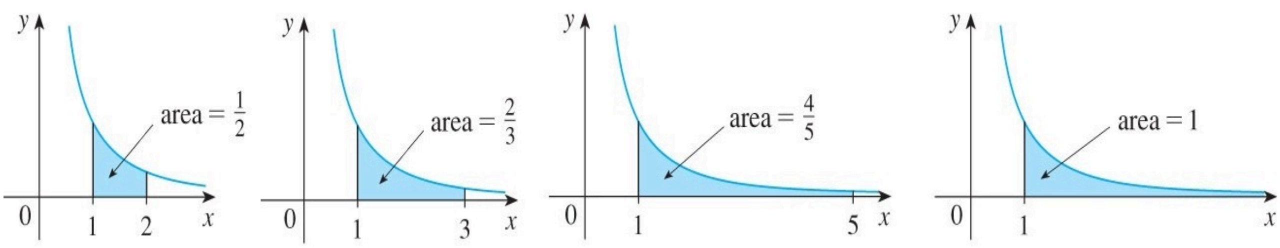
So far, we can only evaluate the integral of f on an finite interval. Fix $t > 1$, we have the area of the region bounded by $y = \dfrac{1}{x^2}$, x-axis, $x = 1$ and $x = t$



$$A(t) = \int_1^t \frac{1}{x^2}\,dx = -\frac{1}{x}\Big|_1^t = 1 - \frac{1}{t}.$$

To evaluate the area of the region bounded by $y = \dfrac{1}{x^2}$, x-axis and $x = 1$, we let t tend to infinity and consider the limit

$$\lim_{t\to\infty} A(t) = \lim_{t\to\infty} \int_1^t \frac{1}{x^2}\,dx = \lim_{t\to\infty}(1 - \frac{1}{t}) = 1.$$



Note. In the above process, the integral $\int_1^t \dfrac{1}{x^2}\,dx$ should be defined for all $t > 1$.

Definition 8.7.2. (Improper Integral of Type1)

(a) 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.

(b) If $\int_t^b f(x)\,dx$ exists for every number $t \leq b$, then

$$\int_{-\infty}^b f(x)\,dx = \lim_{t\to-\infty} f(x)\,dx$$

proveided this limit exists.

We call the above improper integrals $\displaystyle\int_a^\infty f(x)\,dx$ and $\displaystyle\int_{-\infty}^b f(x)\,dx$ "*convergent*" if the corresponding limit exists and "*divergent*" if the limit does not exists.

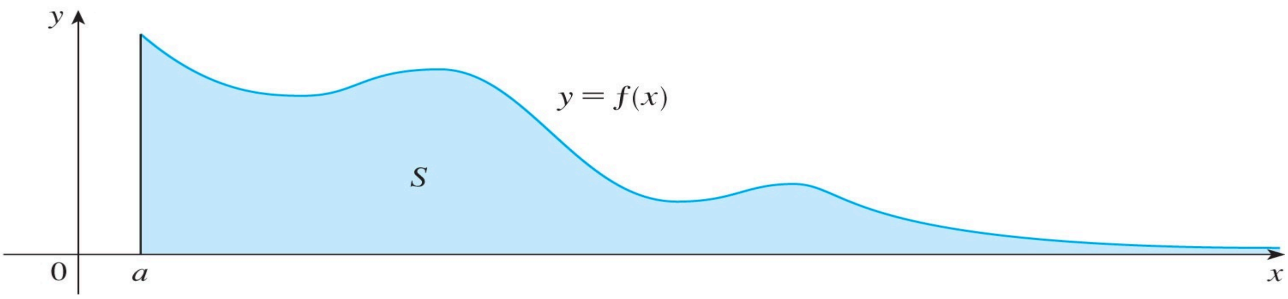
(c) If both $\int_a^\infty f(x)\,dx$ and $\int_{-\infty}^a f(x)\,dx$ are convergent, then we definte

$$\int_{-\infty}^\infty f(x)\,dx = \int_{-\infty}^a f(x)\,dx + \int_a^\infty f(x)\,dx.$$

In part (c) any real number a can be used.

Remark. If $f(x) \geq 0$ and the integral $\int_a^\infty f(x)\,dx$ is convergent, we define the area of the region $S = \{(x,y)\,|\,x \geq a, 0 \leq y \leq f(x)\}$ to be

$$A(S) = \int_a^\infty f(x)\,dx.$$



Example 8.7.3.

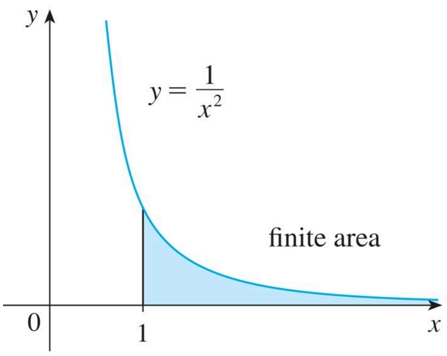
(1) Discuss for what values of p the integral $\displaystyle\int_1^\infty \frac{1}{x^p}\,dx$ is convergent or divergent.

Proof.

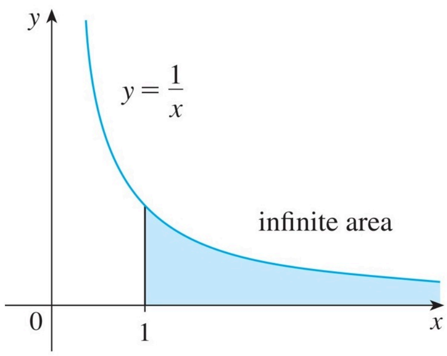
$$
\begin{aligned}
\int_1^\infty \frac{1}{x^p}\,dx &= \lim_{t\to\infty}\int_1^t \frac{1}{x^p}\,dx \\
&= \begin{cases} \lim\limits_{t\to\infty}\left(\dfrac{1}{1-p}\cdot\dfrac{1}{x^{p-1}}\right)\Big|_1^t & p\neq 1 \\[2ex] \lim\limits_{t\to\infty}\left(\ln|x|\right)\Big|_1^t & p=1 \end{cases} \\
&= \begin{cases} \dfrac{1}{1-p}\lim\limits_{t\to\infty}\left(\dfrac{1}{t^{p-1}}-1\right) & p\neq 1 \\[2ex] \lim\limits_{t\to\infty}\ln t & p=1 \end{cases} \\
&= \begin{cases} \dfrac{1}{1-p}\left(\lim\limits_{t\to\infty}\dfrac{1}{t^{p-1}}-1\right) = \begin{cases} \infty & p<1 \\[1ex] \dfrac{1}{p-1} & p>1 \end{cases} \\[4ex] \infty & p=1 \end{cases}
\end{aligned}
$$

□

Conclusion: $\displaystyle\int_1^\infty \frac{1}{x^p}\,dx$ is convergenet if $p > 1$ and divergent if $p \leq 1$.



$\int_1^\infty (1/x^2)\,dx$ converges.



$\int_1^\infty (1/x)\,dx$ diverges.

(2) Evaluate $\displaystyle\int_{-\infty}^0 xe^x\,dx$.

Proof.

$$
\begin{aligned}
\int_{-\infty}^0 xe^x\,dx &= \lim_{t\to-\infty}\int_t^0 xe^x\,dx \overset{I.B.P}{=} \lim_{t\to-\infty}\left[xe^x\Big|_t^0 - \int_t^0 e^x\,dx\right] \\
&= \lim_{t\to-\infty}\left[-te^t - e^x\Big|_t^0\right] = \lim_{r\to-\infty}\left[-te^t - 1 + e^t\right] \\
&= -1.
\end{aligned}
$$

□

(3) Evaluate $\int_{-\infty}^{\infty} \frac{1}{1+x^2}\,dx$.

Proof.

$$\int_{-\infty}^{\infty} \frac{1}{1+x^2}\,dx = \int_{-\infty}^{0} \frac{1}{1+x^2}\,dx + \int_{0}^{\infty} \frac{1}{1+x^2}\,dx.$$

Consider

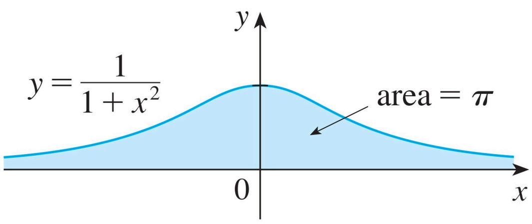
$$\begin{aligned}\int_{0}^{\infty} \frac{1}{1+x^2}\,dx &= \lim_{t\to\infty}\int_{0}^{t} \frac{1}{1+x^2}\,dx = \lim_{t\to\infty} \tan^{-1} x\Big|_0^t \\ &= \lim_{t\to\infty} \tan^{-1} t = \frac{\pi}{2}.\end{aligned}$$

$$\begin{aligned}\int_{-\infty}^{0} \frac{1}{1+x^2}\,dx &= \lim_{t\to-\infty}\int_{t}^{0} \frac{1}{1+x^2}\,dx = \lim_{t\to-\infty} \tan^{-1} x\Big|_t^0 \\ &= \lim_{t\to-\infty} (-\tan^{-1} t) = \frac{\pi}{2}.\end{aligned}$$

Hnece,

$$\int_{-\infty}^{\infty} \frac{1}{1+x^2}\,dx = \int_{-\infty}^{0} \frac{1}{1+x^2}\,dx + \int_{0}^{\infty} \frac{1}{1+x^2}\,dx = \frac{\pi}{2}+\frac{\pi}{2} = \pi.$$

Note that $f(x) = \dfrac{1}{1+x^2}$ is an even function.

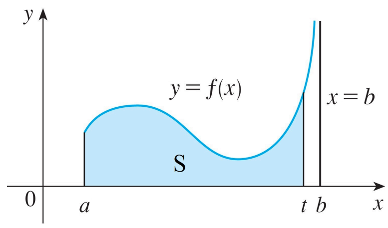


□

❏ Type2: Discontinuous Integrands

Let f be a function defined on a finite interval $[a, b)$ but has a vertical asymptote at b.

In type1 integrals, the regions extended indefinitely in a horizontal direction. In type2 integrals, the regioin is infinite in a vertical direction.



For $a \le t < b$, the area of the region S under the graph $y = f(x)$ from $x = a$ to $x = t$ is

$$A(t) = \int_a^t f(x)\,dx.$$

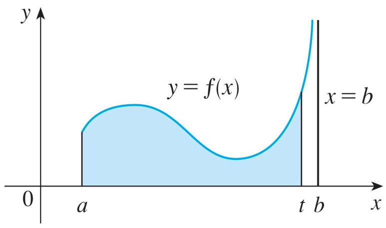
If the limit $\lim_{t\to b^-} A(t) = \lim_{t\to b^-} \int_a^t f(x)\,dx = A$ exists, we say that the area of the region S is A.

Definition 8.7.4. (Improper Integral of Type 2)

(a) If f is defined on $[a,b)$ and $\int_a^t f(x)\,dx$ exists for all $a \le t < b$, then

$$\int_a^b f(x)\,dx = \lim_{t\to b^-} \int_a^t f(x)\,dx$$

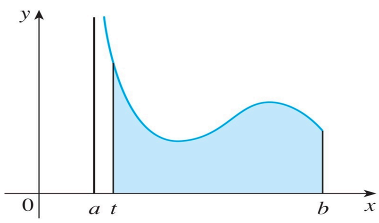
if this limit exists.



(b) If f is defined on $(a,b]$ and $\int_t^b f(x)\,dx$ exists for all $a < t \le b$, then

$$\int_a^b f(x)\,dx = \lim_{t\to a^+} \int_t^b f(x)\,dx$$

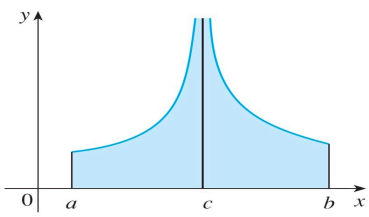
if this limit exists.



We call the improper integral $\displaystyle\int_a^b f(x)\,dx$ "*convergent*" if the corresponding limit exists and "*divergent*" if the limit does not exist.

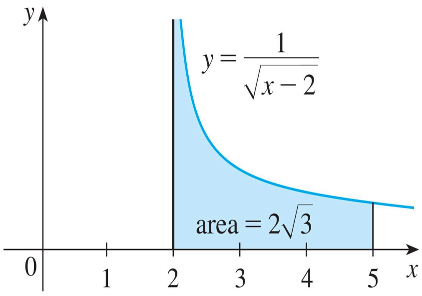
(c) For $a < c < b$, if f has an (infinite) discontinuity at c, if both $\displaystyle\int_a^c f(x)\,dx$ and $\displaystyle\int_c^b f(x)\,dx$ converge then we say that $\displaystyle\int_a^b f(x)\,dx$ converges and

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$$



Example 8.7.5.

(1) Evaluate $\displaystyle\int_2^5 \frac{1}{\sqrt{x-2}}\,dx.$

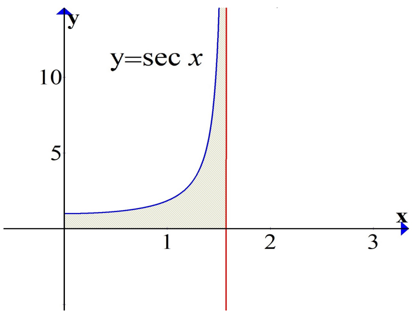


Proof. The function $f(x) = \dfrac{1}{\sqrt{x-2}}$ has the vertical asymptote $x = 2$. Thus,

$$\begin{aligned}\int_2^5 \frac{1}{\sqrt{x-2}}\,dx &= \lim_{t\to 2^+}\int_t^5 \frac{1}{\sqrt{x-2}}\,dx\\ &= \lim_{t\to 2^+} 2\sqrt{x-2}\Big|_t^5\\ &= \lim_{t\to 2^+} 2(\sqrt{3}-\sqrt{t-2}) = 2\sqrt{3}.\end{aligned}$$

□

(2) Evaluate $\displaystyle\int_0^{\frac{\pi}{2}} \sec x\,dx.$

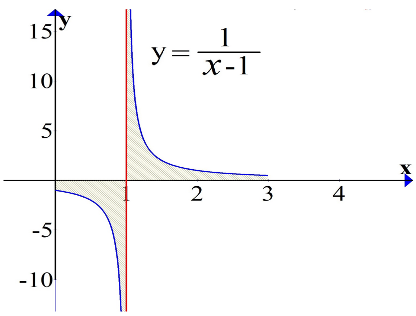


Proof. The function $f(x) = \sec x$ has the vertical asymptote $x = \frac{\pi}{2}$. Thus,

$$\begin{aligned}\int_0^{\frac{\pi}{2}} \sec x\,dx &= \lim_{t\to(\frac{\pi}{2})^-}\int_0^t \sec x\,dx\\ &= \lim_{t\to(\frac{\pi}{2})^-} \ln|\sec x+\tan x|\Big|_0^t\\ &= \lim_{t\to(\frac{\pi}{2})^-} [\ln|\sec x+\tan x| - \ln 1] = \infty.\end{aligned}$$

□

(3) Evaluate $\displaystyle\int_0^3 \frac{1}{x-1}\,dx.$



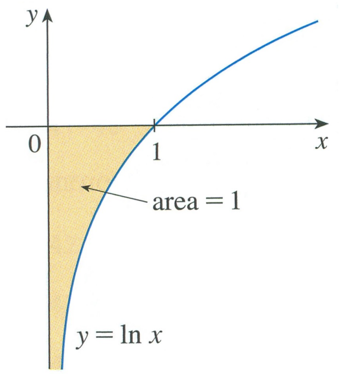
Proof. The function $f(x) = \dfrac{1}{x-1}x$ has the vertical asymptote $x = 1$. Thus,

$$\begin{aligned}\int_0^1 \frac{1}{x-1}\,dx &= \lim_{t\to 1^-}\int_0^t \frac{1}{x-1}\,dx\\ &= \lim_{t\to 1^-}(\ln|x-1|)\Big|_0^t\\ &= \lim_{t\to 1^-}\ln|t-1| = -\infty.\end{aligned}$$

Hence, $\displaystyle\int_0^3 \frac{1}{x-1}\,dx$ is divergent. □

Wrong method: $\displaystyle\int_0^3 \frac{1}{x-1}\,dx = (\ln|x-1|)\Big|_0^3 = \ln 2 - \ln 1 = \ln 2.$

(4) Evaluate $\displaystyle\int_0^1 \ln x\,dx.$



Proof. The function $f(x) = \ln x$ has the vertical asymptote $x = 0$. Thus,

$$\begin{aligned}\int_0^1 \ln x\,dx &= \lim_{t\to 0^+}\int_t^1 \ln x\,dx \\ &= \lim_{t\to 0^+}[x\ln x - x]\Big|_t^1 \\ &= \lim_{t\to 0^+}(-t\ln t - 1 + t) \overset{L.H.}{=} -1.\end{aligned}$$

□

(5) Discuss for what values of p the integral $\displaystyle\int_0^1 \frac{1}{x^p}\,dx$ is convergent or divergent.

Proof. When $p \le 0$, $f(x) = \frac{1}{x^p}$ is continuous on $[0, 1]$. Hence, the integral is convergent and $\displaystyle\int_0^1 \frac{1}{x^p}\,dx = \frac{1}{1-p}$. Consider the cases $p > 0$, then function $f(x) = \frac{1}{x^p}$ has a vertical asymptote $x = 0$. Then

$$\begin{aligned}\int_0^1 \frac{1}{x^p}\,dx &= \lim_{t\to 0^+}\int_t^1 \frac{1}{x^p}\,dx = \begin{cases} \dfrac{1}{1-p}\lim_{t\to 0^+}\dfrac{1}{x^{p-1}}\Big|_t^1 & p \ne 1 \\[2ex] \lim_{t\to 0^+}(\ln|x|)\Big|_t^1 & p = 1 \end{cases} \\ &= \begin{cases} \dfrac{1}{1-p}\lim_{t\to 0^+}(1 - t^{1-p}) = \begin{cases} \dfrac{1}{1-p} & p < 1 \\ \infty & p > 1 \end{cases} \\[2ex] \lim_{t\to 0^+}(-\ln t) = \infty & p = 1 \end{cases}\end{aligned}$$

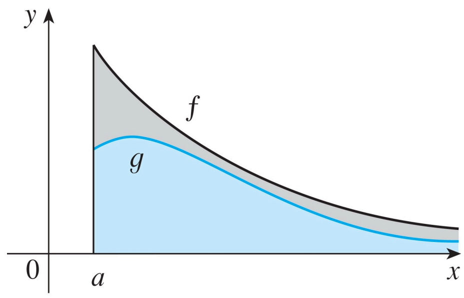
□

Conclusion: $\displaystyle\int_0^1 \frac{1}{x^p}\,dx$ is convergent if $p < 1$ and divergent if $p \ge 1$.

❑ Comparison Theorem

Note. For some definite integrals, it is impossible (difficult) to find their exact values but we can still determine whether these integrals are convergent or divergent.

Theorem 8.7.6. *(Comparison Theorem) Suppose that f and g satisfy $0 \le g(x) \le f(x)$ for every $x \ge a$.*



(a) If $\int_a^\infty f(x)\,dx$ *is convergent, then* $\int_a^\infty g(x)\,dx$ *is convergent.*

(b) If $\int_a^\infty g(x)\,dx$ *is divergent, then* $\int_a^\infty f(x)\,dx$ *is divergent.*

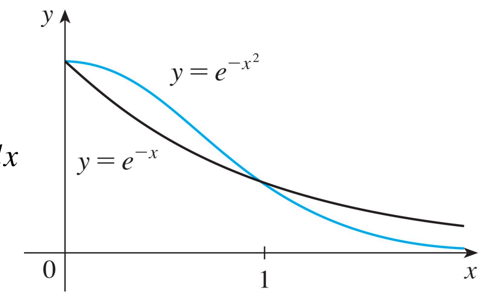
Example 8.7.7.

(1) Determine whether the improper integral $\int_0^\infty e^{-x^2}\,dx$ is convergent or divergent.

Proof.

Since $f(x) = e^{-x^2}$ is continuous on $[0,1]$, it is integrable on $[0,1]$. On the other hand, $0 \le e^{-x^2} \le e^{-x}$ for every $x \ge 1$ and

$$\int_1^\infty e^{-x}\,dx = \lim_{t\to\infty}\int_1^t e^{-x}\,dx = \lim_{t\to\infty}(-e^x)\Big|_1^t = e^{-1}.$$



By the Comparison Theorem, the improper integral $\int_1^\infty e^{-x^2}\,dx$ is convergent. Hence,

$$\int_0^\infty e^{-x^2}\,dx = \int_0^1 e^{-x^2}\,dx + \int_1^\infty e^{-x^2}\,dx$$

is also convergent. In fact, $\int_0^\infty e^{-x^2}\,dx = \dfrac{\sqrt{\pi}}{2}$.

□

(2) Determine whether the improper integral $\int_1^\infty \dfrac{1-e^{-x}}{x}\,dx$ is convergent or divergent.

Proof. Since $0 < \dfrac{1}{2x} < \dfrac{1-e^{-x}}{x}$ for every $1 \le x < \infty$ and

$$\int_1^\infty \frac{1}{2x}\,dx = \frac{1}{2}\lim_{t\to\infty}\int_1^t \frac{1}{x}\,dx = \frac{1}{2}\lim_{t\to\infty}\ln t = \infty.$$

By the Comparison Theorem, the improper integral $\int_1^\infty \dfrac{1-e^{-x}}{x}\,dx$ is divergent. □

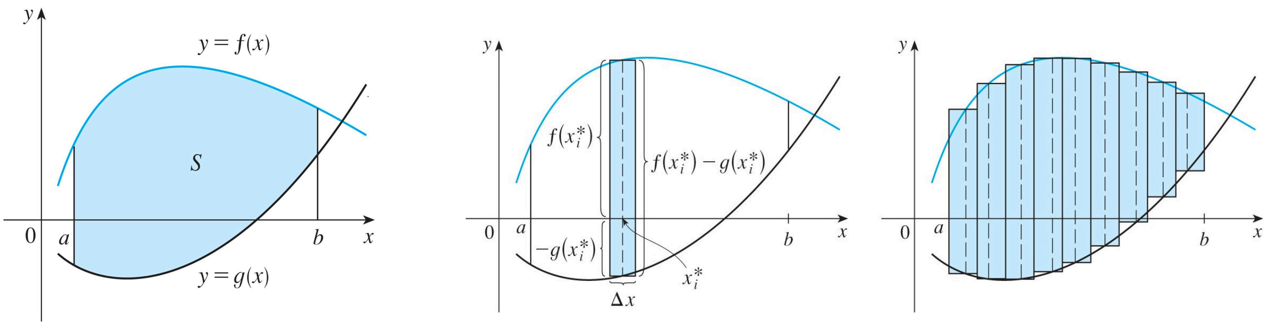
Chapter 9

Applications of Integration

9.1 Areas Between Curves

In the present section, we try to evaluate the integrals to find areas of regions that lie between the graphs of two functions.

Let f and g be two continuous functions satisfying $f(x) \geq g(x)$ for every $x \in [a, b]$. Let S be the region between the two curves $y = f(x)$ and $y = g(x)$, and the vertical lines $x = a$ and $x = b$. We use the approximating rectangles method to evaluate the area of S.



$S = \{(x, y) \mid a \leqslant x \leqslant b, g(x) \leqslant y \leqslant f(x)\}$ (a) Typical rectangle   (b) Approximating rectangles

Let P be a partition of $[a, b]$. The Riemann sum

$$\sum_{i=1}^{n} [f(x_i^*) - g(x_i^*)] \triangle x_i$$

is an approximation to the area of S. We define the area A of the region S as the limiting value of the sum of the area of these approximating rectangles

$$A = \lim_{\|P\|\to 0} \sum_{n=1}^{n} [f(x_i^*) - g(x_i^*)]\triangle x_i.$$

Theorem 9.1.1. *The area A of the region bounded by the cruve $y = f(x)$, $y = g(x)$ and the lines $x = a$ and $x = b$, where f and g are integrable and $f(x) \geq g(x)$ for all $x \in [a, b]$, is*

$$A = \int_a^b [f(x) - g(x)]\, dx$$

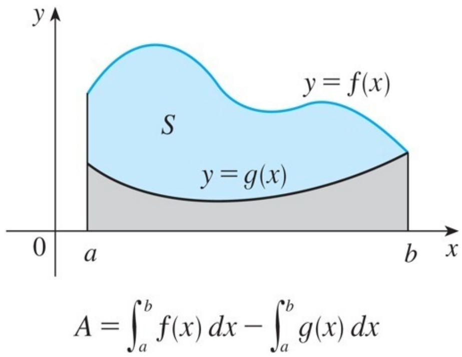
Note. (1) If $g(x)$, S is the region under the graph of f. The area of S is

$$A = \int_a^b [f(x) - 0]\, dx = \int_a^b f(x)\, dx$$

is the same as the area we discussed before.

(2) If $f(x) \geq g(x) \geq 0$ for all $x \in [a, b]$

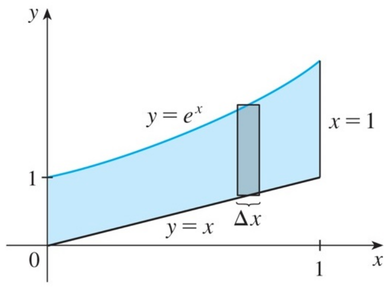
$$\begin{aligned} A &= [\text{area under } y = f(x)] - [\text{area under } y = g(x)] \\ &= \int_a^b f(x)\, dx - \int_a^b g(x)\, dx \\ &= \int_a^b [f(x) - g(x)]\, dx. \end{aligned}$$



Example 9.1.2. Find the area of the region bounded above by $y = e^x$, bounded below by $y = x$ and bounded on the sides by $x = 0$ and $x = 1$.

Proof.

$$\begin{aligned} A &= \int_0^1 [e^x - x]\, dx \\ &= e^x - \frac{1}{2}x^2 \Big|_0^1 = e - \frac{3}{2}. \end{aligned}$$

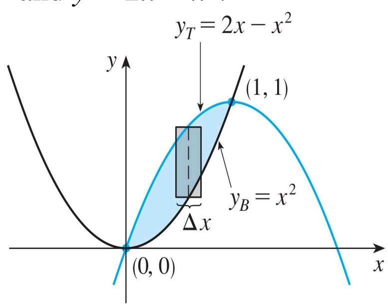


□

Example 9.1.3. Find the area of the region enclosed by the parabola $y = x^2$ and $y = 2x - x^2$.

Proof. The points of intersection of $y = x^2$ and $y = 2x - x^2$ are given by solving the equation $x^2 = x - x^2$. They are $x = 0$ and $x = 1$. The graph $y = 2x - x^2$ is above the graph of $y = x^2$ for all $x \in [0, 1]$. The area of the region is

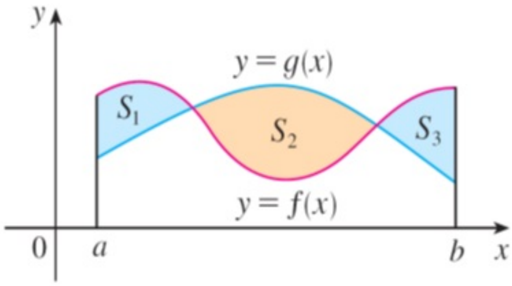
$$A = \int_0^1 [(2x - x^2) - x^2]\, dx = x^2 + x - \frac{2}{3}x^3 \Big|_0^1 = \frac{1}{3}.$$



□

To find the area between the curves $y = f(x)$ and $y = g(x)$ where $f(x) \geq g(x)$ for some values of x but $g(x) \geq f(x)$ for other values.

We splits the region S into several subregions $S_1, S_2, \cdots S_n$ with areas $A_1, A_2, \cdots A_n$. Then the area of S is



$$A = A_1 + A_2 + \cdots + A_n.$$

Since

$$|f(x) - g(x)| = \begin{cases} f(x) - g(x) & \text{when } f(x) \geq g(x) \\ g(x) - f(x) & \text{when } f(x) \leq g(x), \end{cases}$$

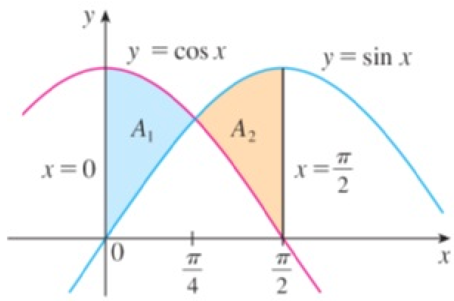
we have the following results.

Theorem 9.1.4. *The are between the curves $y = f(x)$ and $y = g(x)$ and between $x = a$ and $x = b$ is*

$$A = \int_a^b |f(x) - g(x)|\, dx.$$

Example 9.1.5. Find the area of the region bounded by the cruves $y = \sin x$, $y = \cos x$, $x = 0$ and $x = \dfrac{\pi}{2}$.

Proof. The points of intersection of two curves in $[0, \dfrac{\pi}{2}]$ is $\dfrac{\pi}{4}$. Also, $\cos x \geq \sin x$ when $0 \leq x \leq \frac{\pi}{4}$ and $\sin x \geq \cos x$ when $\frac{\pi}{4} \leq x \leq \frac{\pi}{2}$. The area of the region is

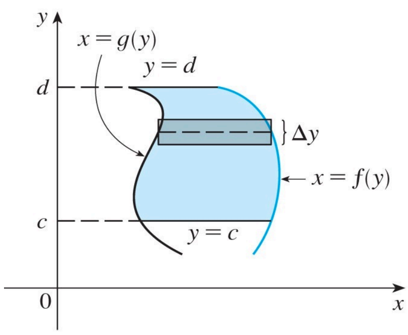


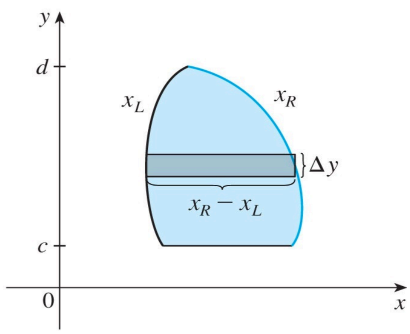
$$\begin{aligned} A &= \int_0^{\frac{\pi}{2}} |\cos x - \sin x|\, dx \\ &= \int_0^{\frac{\pi}{4}} \cos x - \sin x\, dx + \int_{\frac{\pi}{4}}^{\frac{\pi}{2}} \sin x - \cos x\, dx \\ &= 2\sqrt{2} - 2 \end{aligned}$$

□

Some regions are treated by regarding x as a function of y. Suppose that the region S is bounded by curves with equation $x = f(y)$, $x = g(y)$, $y = c$ and $y = d$ where f and g are continuous and $f(y) \geq g(y)$ for all $c \leq y \leq d$. The area of the region S is

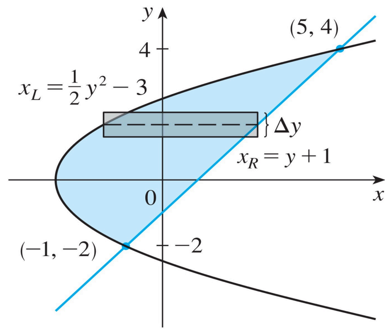
$$A = \int_c^d [f(y) - g(y)]\, dy.$$





Example 9.1.6. Find the area enclosed by the line $y = x - 1$ and the parabola $y^2 = 2x + 6$.

Proof. The points of intersection is obtained by solving $y^2 = 2y + 8$. Hence, those points are $y = 4$ and $y = 2$. The area of the enclosed region is

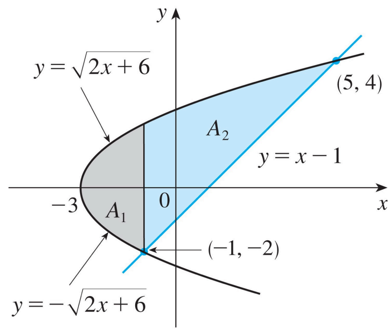


$$
\begin{aligned}
A &= \int_{-2}^{4} (y+1) - (\frac{1}{2}y^2 - 3)\, dy \\
&= \int_{-2}^{4} -\frac{1}{2}y^2 + y + 4 dy \\
&= 18
\end{aligned}
$$

□

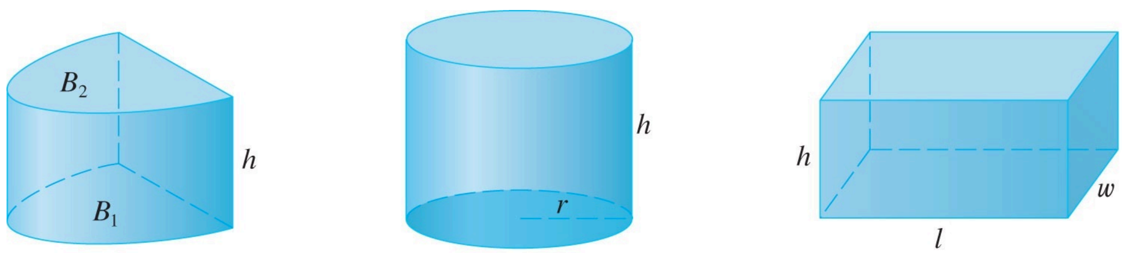
Note. We can also obtain the area of the above region by integrating with respect to x instead of y.

Splitting the region into two subregions A_1 and A_2 and computing each area and adding them up. But it is very complicated.



9.2 Volume

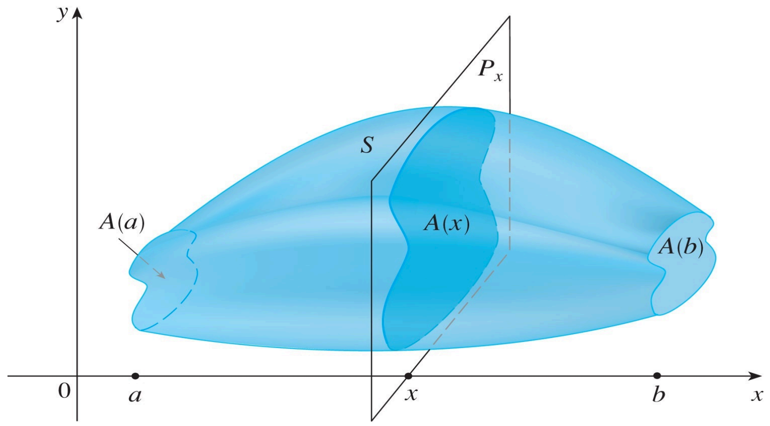
In the present section, we wnat to find the volume of a solid by using the techniques of integral to give an exact definition. We start with a simple type of solid called a "cylinder (right cylinder)".



(a) Cylinder $V = Ah$ (b) Circular cylinder $V = \pi r^2 h$ (c) Rectangular box $V = lwh$

For a general solid S (not a cylinder), we cut it into several slices and approximate each slice by regarding them as cylinders. We estimate the volume of S by adding the volume of those approximating volumes of slabs.

(i) The intersection of S with a plane and obtaining a plane region that is called a "*cross-section*" of S. Let $A(x)$ be the area of the cross-section of S in a plane P_x perpendicular to the x-axis and passing through the point x where $a \le x \le b$.



(ii) Dividing S into n "slabs" of equal width $\triangle x$ by using the planes $P_{x_1}, P_{x_2}, \cdots$ to slice the solid.

(iii) Choosing sample points x_i^* in $[x_{i-1}, x_i]$, we can approximate the ith slab S_i by a cylinder with base $A(x_i^*)$ and "height" $\triangle x_i$. The volume of this cylinder is $A(x_i^*)\triangle x_i$. Hence, the volume of S_i is

$$V_i \approx A(x_i^*)\triangle x_i.$$

(iv) Adding the volumes of these slabs, we get an approximation to the total volume of S,

$$V \approx \sum_{i=1}^{n} A(x_i^*)\triangle x_i.$$

(v) Let n tends to infinity, we define the volume of S as the limit of these sums.

Definition 9.2.1. Let S be a solid that lies between $x = a$ and $x = b$. If the cross-sectional area of S in the plane P_x through x and perpendicular to the x-axis, is $A(x)$, where A is a continuous function, then the volume of S is

$$V = \lim_{n\to\infty} \sum_{i=1}^{n} A(x_i^*)\triangle x_i = \int_a^b A(x)\, dx.$$

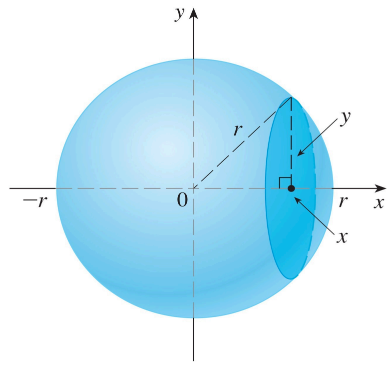
Note. For a (right) cylinder, $A(x) = A$ for all x. Then the volume is

$$V = \int_a^b A(x)\, dx = \int_a^b A\, dx = A(b-a).$$

Example 9.2.2. Find the volume of a sphere of radius r.

Proof. The plane P_x intersects the sphere in a circle whose radius is $y = \sqrt{r^2 - x^2}$. Hence, the cross-sectional area is

$$A(x) = \pi(\sqrt{r^2 - x^2})^2 = \pi(r^2 - x^2).$$

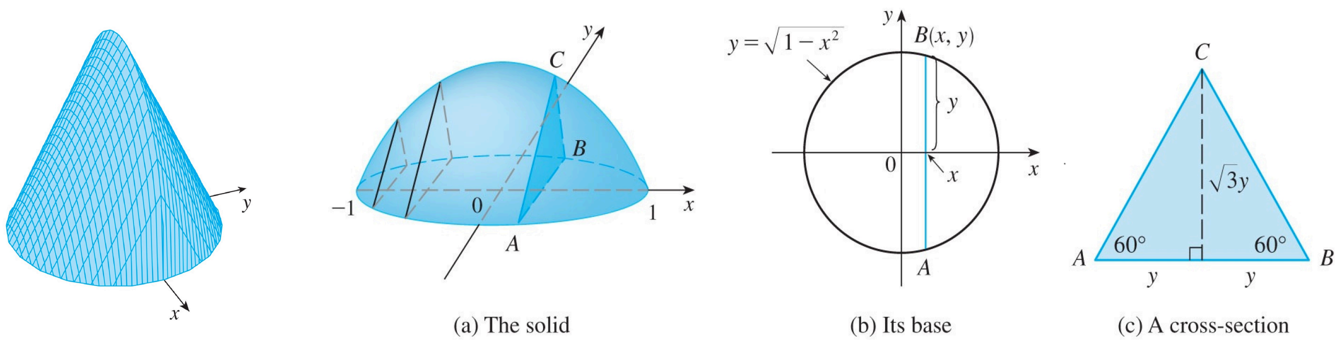


The volume of the sphere is

$$\begin{aligned} V &= \int_{-r}^{r} A(x)\,dx = \int_{-r}^{r} \pi(r^2 - x^2)\,dx \\ &= \pi(r^2 x - \frac{1}{3}x^3)\Big|_{-r}^{r} \\ &= \frac{4}{3}\pi r^3. \end{aligned}$$

□

Example 9.2.3. A solid with a circular base of radius 1. Parallel cross-sections perpendicular to the base are equilateral triangles. Find the volume of the solid.



(a) The solid (b) Its base (c) A cross-section

Proof. Each cross-section is an equilateral triangle, the base is $2y$ and the height is $\sqrt{3}y$. Hence the area of the cross-section is

$$A(x) = \sqrt{3}y^2 = \sqrt{3}(1 - x^2).$$

The volume of the solid is

$$V = \int_{-1}^{1} A(x)\,dx = \int_{-1}^{1} \sqrt{3}(1 - x^2)\,dx = \frac{4\sqrt{3}}{3}.$$

□

Example 9.2.4. A wedge is cut out of circular cylinder of radius 4 by two planes. One plane is perpendicular to the axis of the cylinder. The other intersects the first at an angle 30° along a diameter of the cylinder. Find the volume of the wedge.

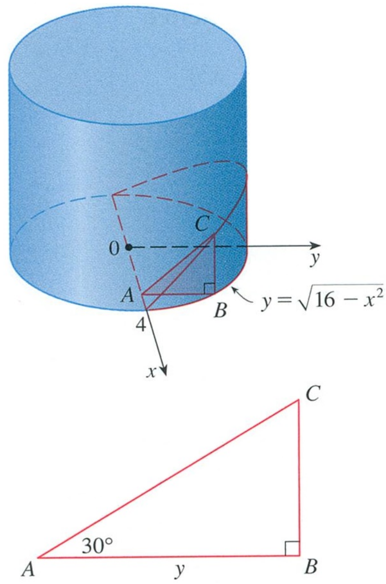
Proof. Each cross-section is a right triangle with base $y = \sqrt{16 - x^2}$. The intersection angle 30° implies that the height is $y \tan 30° = \frac{\sqrt{16-x^2}}{\sqrt{3}}$. The area of the cross-section is

$$A(x) = \frac{1}{2}\sqrt{16 - x^2} \cdot \frac{\sqrt{16 - x^2}}{\sqrt{3}} = \frac{1}{2\sqrt{3}}(16 - x^2).$$

The volume of the solid is

$$\begin{aligned} V &= \int_{-4}^{4} A(x)\,dx = \int_{-4}^{4} \frac{1}{2\sqrt{3}}(16 - x^2)\,dx \\ &= \frac{1}{2\sqrt{3}}(16x - \frac{1}{3}x^3)\Big|_{-4}^{4} = \frac{128}{3\sqrt{3}}. \end{aligned}$$

□



9.3 Solid of Revolution

In the present section, we wnat to find the volume of the solid which is obtained by rotating a region about a line. We calculate the area of cross-section. The the volume is

$$V = \int_a^b A(x)\,dx \quad \text{or} \quad V = \int_c^d A(y)\,dy.$$

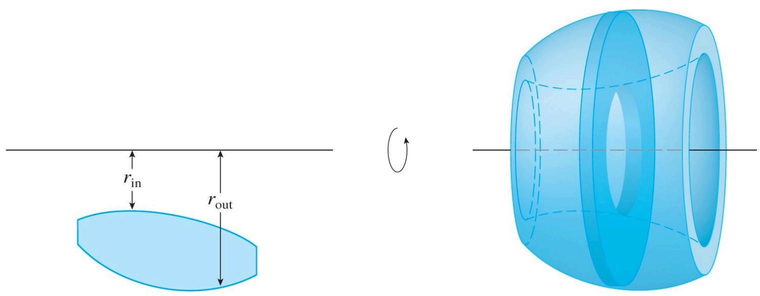
To find the area of each cross-section.

(i) If the cross-section is a dist, the area is

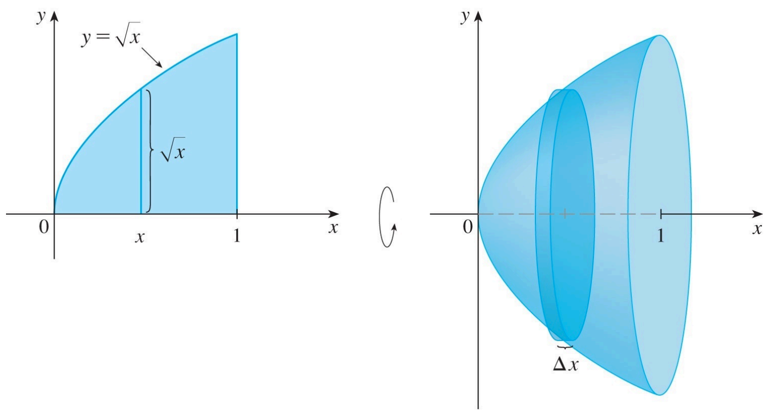
$$A = \pi(\text{radius})^2$$

(ii) If the cross-section is a washer, the area is

$$A = \pi r_{outer}^2 - \pi r_{inner}^2.$$



Example 9.3.1. Find the volume of the solid obtained by rotating about the x-axis the region under the curve $y = \sqrt{x}$ from 0 to 1.



Proof. The cross-sectional area is

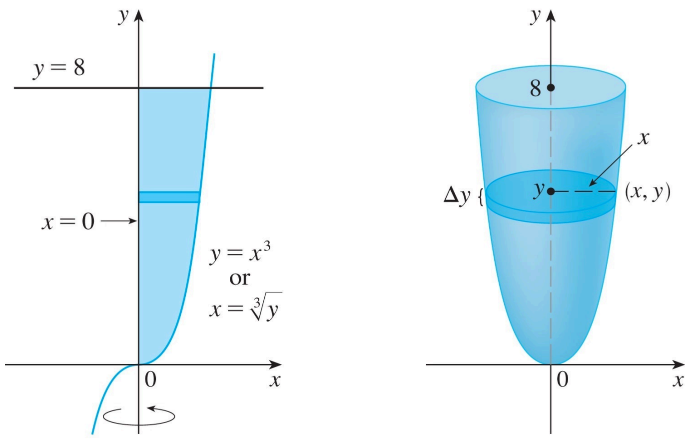
$$A(x) = \pi(\sqrt{x})^2 = \pi x.$$

The solid lies between $x = 0$ and $x = 1$ has volume

$$V = \int_0^1 A(x)\,dx = \int_0^1 \pi x\,dx = \frac{\pi x^2}{2}\bigg|_0^1 = \frac{\pi}{2}.$$

□

Example 9.3.2. Find the volume of the solid obtained by rotating the region bounded by $y = x^3$, $y = 8$ and $x = 0$ about the y-axis.



Proof. The region is rotated about y-axis. It makes to slice the solid perpendicular to the y-axis obtaining circular cross-sections. The area of a cross-section through y is

$$A(y) = \pi x^2 = \pi(\sqrt[3]{y})^2 = \pi y^{2/3}.$$

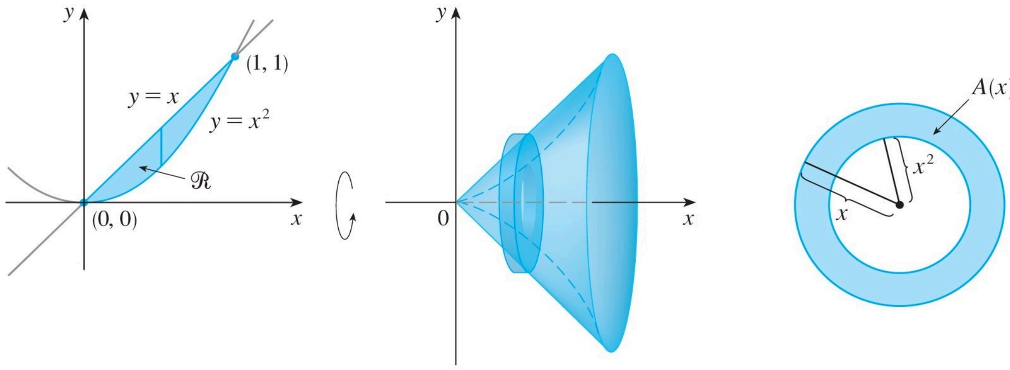
The volume of the solid is

$$V = \int_0^8 A(y)\,dy = \int_0^8 \pi y^{2/3}\,dy = \frac{3\pi}{5}y^{5/3}\bigg|_0^8 = \frac{96\pi}{5}.$$

□

❑ Washer Method (Method of Washer)

Example 9.3.3. Find the volume of the solid obtained by rotating the region which is enclosed by $y = x$ and $y = x^2$, about the x-axis.



Proof. The points of intersection is obtained by $x = x^2$ and hence those points are $x = 0$ and $x = 1$. The area of the cross-section perpendicular to x-axis is

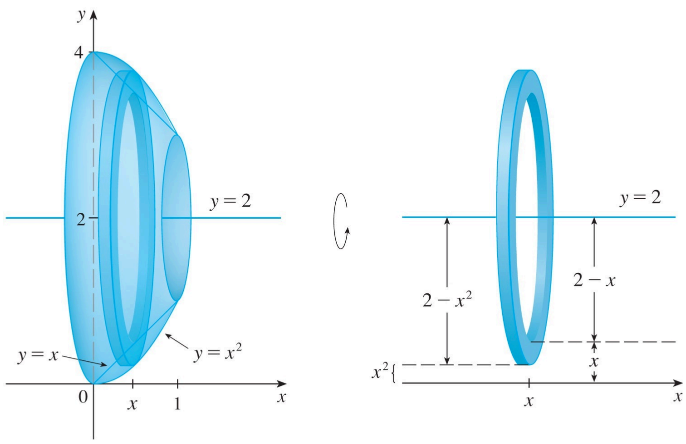
$$A(x) = \pi r_{outer}^2 - \pi r_{inner}^2 = \pi(x)^2 - \pi(x^2)^2 = \pi(x^2 - x^4).$$

The volume of the solid is

$$V = \int_0^1 A(x)\,dx = \int_0^1 \pi(x^2 - x^4)\,dx = \pi(\frac{1}{3}x^3 - \frac{1}{5}x^5)\Big|_0^1 = \frac{2\pi}{15}.$$

□

Example 9.3.4. Find the volume of the solid obtained by rotating the region which is enclosed by $y = x$ and $y = x^2$, about the line $y = 2$.



Proof. The cross-section is a washer and its area is

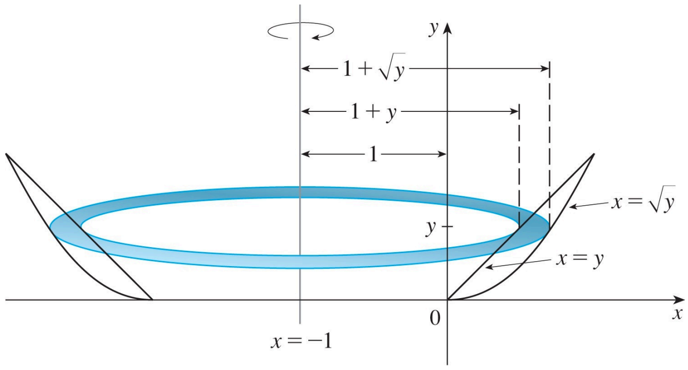
$$A(x) = \pi r_{outer}^2 - \pi r_{inner}^2 = \pi(2 - x^2)^2 - \pi(2 - x)^2 = \pi(x^4 - 5x^2 + 4x).$$

The volume of the solid is

$$V = \int_0^1 A(x)\,dx = \pi\int_0^1 x^4 - 5x^2 + 4x\,dx = \pi\left(\frac{1}{5}x^5 - \frac{5}{3}x^3 + 2x^2\right)\Big|_0^1 = \frac{8\pi}{15}.$$

□

Example 9.3.5. Find the volume of the solid obtained by rotating the region which is enclosed by $y = x$ and $y = x^2$, about the line $x = -1$.



Proof. The area of the cross-section is

$$\pi r_{outer}^2 - \pi r_{inner}^2 = \pi\left(\sqrt{y} - (-1)\right)^2 - \pi\left(y - (-1)\right)^2 = \pi(2\sqrt{y} - y - y^2).$$

The volume of the solid is

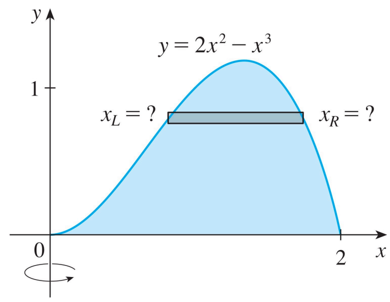
$$V = \int_0^1 \pi(2\sqrt{y} - y - y^2)\, dy = \pi\left(\frac{4}{3}y^{3/2} - \frac{1}{2}y^2 - \frac{1}{3}y^3\right)\Big|_0^1 = \frac{\pi}{2}.$$

$\square$

❑ Method of Cylindrical Shells

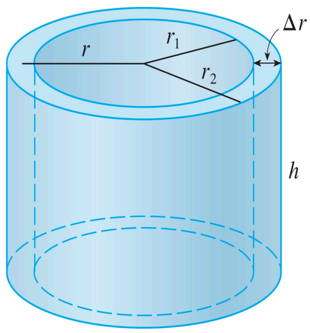
For some solids of revolution, it is difficult to find their volumes by using the washer method.

For example, the solid obtained by rotating the region which is enclosed by $y = 2x^2 - x^3$ and x-axis. If we want to use the washer method to find the volume of the solid, we have to evaluate the areas of each cross-section, $A(y)$, for every $0 \leq y \leq \frac{32}{27}$. But it is not easy to solve the equation $y = 2x^2 - x^3$.



Hence, we study a different method, called the method of "cylindrical shells", to find its volume here.

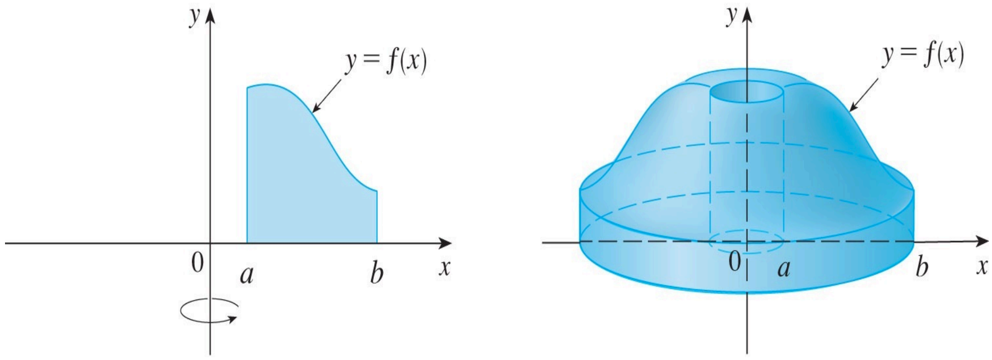
Consider a cylindrical shell with inner radius r_1, outer radius r_2 and height h. Then the thickness of the shell is $\Delta r = r_2 - r^1$. The volume of the shell is



$$\begin{aligned} V &= \pi r_2^2 h - \pi r_1^2 h = \pi(r_2^2 - r_1^2)h \\ &= \pi(r_2 + r_1)(r_2 - r_1)h = 2\pi \cdot \underbrace{\frac{r_2 + r_1}{2}}_{\approx r} h \underbrace{(r_2 - r_1)}_{=\Delta r} \\ &\approx 2\pi r h \Delta r. \end{aligned}$$

The approximating volume of the cylindrical shell is $2\pi rh\triangle r$.[†]

Let S be the solid obtained by rotating about the y-axis the region bound by $y = f(x)$, $y = 0$, $x = a$ and $x = b$ where $0 \leq a < b$.

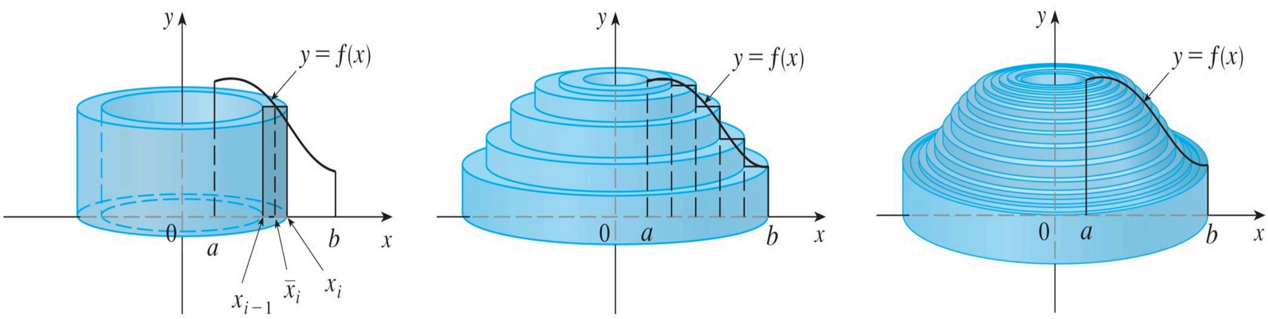


Dividing $[a, b]$ into n subintervals $[x_{i-1}, x_i]$ of equal width $\triangle x$ and choose $\bar{x}$ as the midpoint of the ith subinterval. Consider the rectangle with base $[x_{i-1}, x_i]$ and height $f(\bar{x})$. The solid whcih is obtained by rotating the above region about the y-axis has volume

$$V_i \approx (2\pi\bar{x})(f(\bar{x}_i))\triangle x.$$

The approximation to the volume of S is

$$V \approx \sum_{i=1}^{n} V_i = \sum_{i=1}^{n} 2\pi\bar{x}_i f(\bar{x}_i)\triangle x.$$



Let $n \to \infty$, the volume of the solid is,

$$\lim_{n\to\infty} \sum_{i=1}^{n} 2\pi\bar{x}_i f(\bar{x}_i)\triangle x = \int_a^b 2\pi x f(x)\, dx.$$

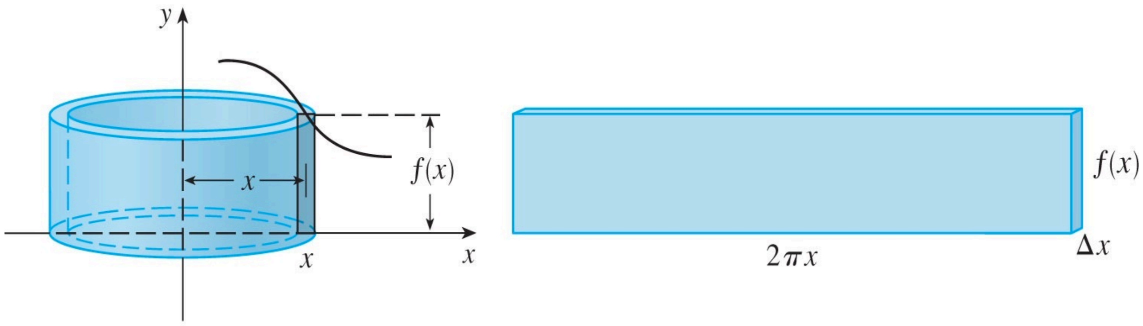
Theorem 9.3.6. *The volume of the solid obtained by rotating about the y-axis the region under the curve $y = f(x)$ from a to b is*

$$V = \int_a^b 2\pi x f(x)\, dx.$$

[†]It can be remembered as $V \approx$ [circumference][height][thickness].

Note. Flattening a cylindrical shell with radius x, circumference $2\pi x$, height $f(x)$ and thickness $\triangle x$ (or dx). Hence, the volume of S is

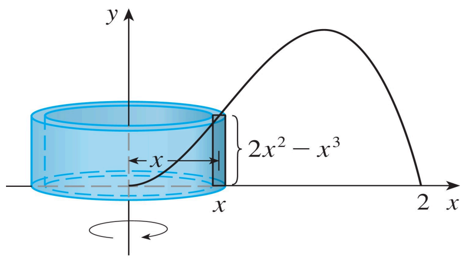
$$V - \int_a^b \underbrace{2\pi x}_{\text{circumference}} \underbrace{f(x)}_{\text{height}} \underbrace{dx}_{\text{thickness}}.$$



Example 9.3.7. Find the volume of the solid obtained by rotating about the y-axis the region bounded by $y = 2x^2 - x^3$ and $y = 0$.

Proof.

$$\begin{aligned} V &= \int_0^2 2\pi x(2x^2 - x^3)\, dx \\ &= 2\pi(\frac{1}{2}x^4 - \frac{1}{5}x^5)\Big|_0^2 \\ &= \frac{16\pi}{5}. \end{aligned}$$

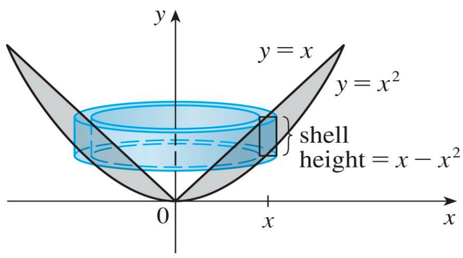


□

Example 9.3.8. Find the volume of the solid obtained by rotating about the y-axis the region bounded by $y = x$ and $y = x^2$.

Proof. The points of intersection of $y = x$ and $y = x^2$ is $(0, 0)$ and $(1, 0)$. Therefore, the volume of the solid is

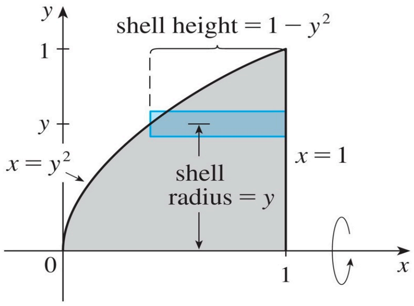
$$V = \int_0^1 2\pi x(x - x^2)\, dx = \frac{\pi}{6}.$$

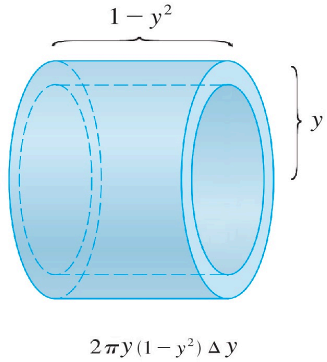


□

Example 9.3.9. Find the volume of the solid obtained by rotating about the x-axis the region under the curve $y = \sqrt{x}$ from 0 to 1.

Proof.



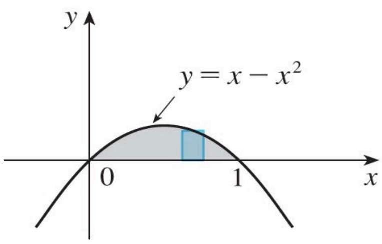


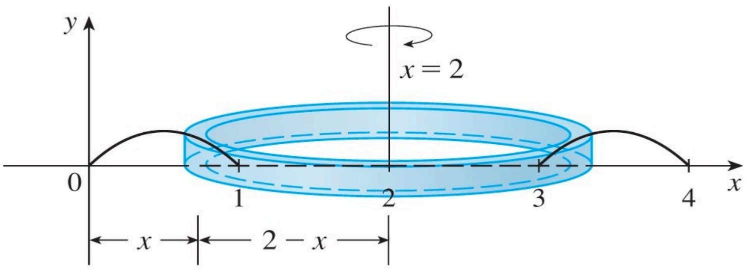
$$V = \int_0^1 2\pi y(1-y^2)\,dy = 2\pi\left(\frac{y^2}{2} - \frac{y^4}{4}\right)\Big|_0^1 = \frac{\pi}{2}.$$

□

Example 9.3.10. Find the volume of the solid obtained by rotating about the line $x = 2$ the region bounded by $y = x - x^2$ and $y = 0$.

Proof.





$$V = \int_0^1 2\pi(2-x)(x-x^2)\,dx = \frac{\pi}{2}.$$

□

9.4 Arc Length

In the present section, we want to evaluate the arc length of a curve which is the graph of a smooth function.

Question: For a given curve C, what is the length of C?
 If the curve is a polygon, it is easy to find its length.

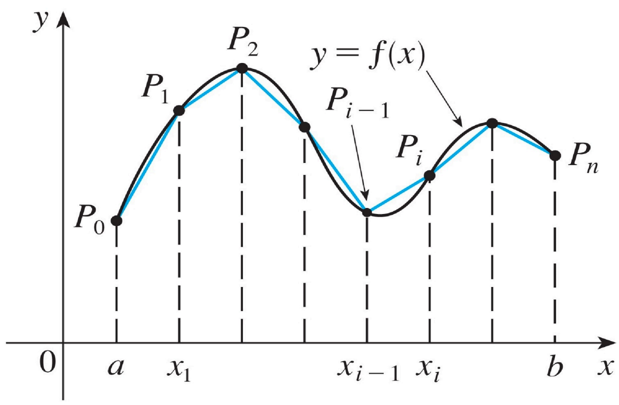
Question: How about the length of a general curve?

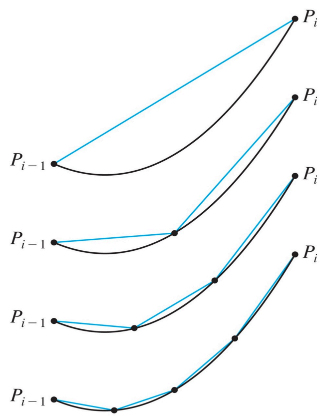
We try to approximate the length of a general curve by polygons and take a limit as the numbers of thy polygon is increased.

Suppose that f is a function defined on $[a, b]$ and C is the graph of f with equation $y = f(x)$. Let $P = \{x_0, x_1, \cdots, x_n\}$ be a partition of $[a, b]$ and the point $P_i\big(x_i, f(x_i)\big)$ are points on C. Consider the polygon with vertices $P_0, P_1, \cdots, P_n$. The number $\ell(P, f)$ represents the length of a polygonal curve inscribed in the graph of f. Then the length L of the curve C is approximately the length of the polygon

$$\ell(P, f) = \sum_{i=1}^{n} |P_{i-1}P_i| = \sum_{i=1}^{n} \sqrt{(x_i - x_{i-1})^2 + \big[f(x_i) - f(x_{i-1})\big]^2}.$$

As n increases, the approximation gets better





Exercise. Let P and Q be two partitions of $[a, b]$. If Q is a refinement of P, then

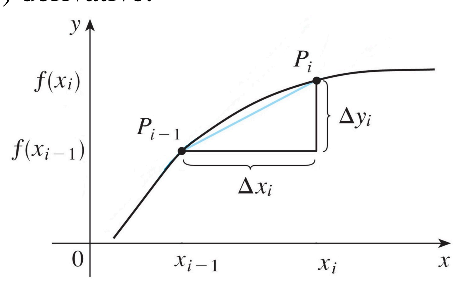
$$\ell(P, f) \le \ell(Q, f).$$

Definition 9.4.1. We define the length of f on $[a, b]$ to be the least upper bound of all $\ell(P, f)$ for all partition P (provided that the set of all such $\ell(P, f)$ is bounded above). That is, the length of f on $[a, b]$ is

$$L = \sup_P \ell(P, f).$$

Unfortunately, for a general function f, the approximating length $\ell(P, f)$ is not easy to obtain. Therefore, from now on, we assume that f has a (continuous) derivative.

The length of the segment $P_{i-1}P_i$ is



$$\begin{aligned}
& \sqrt{(\triangle x_i)^2 + (\triangle y_i)^2} \\
= \; & \sqrt{(x_i - x_{i-1})^2 + [f(x_i) - f(x_{i-1})]^2} \\
\overset{M.V.T}{=} \; & \sqrt{(x_i - x_{i-1})^2 + [f'(x_i^*)(x_i - x_{i-1})]^2} \\
= \; & \sqrt{1 + [f'(x_i^*)]^2} \triangle x_i.
\end{aligned}$$

The length of the curve C with the equation $y = f(x)$ on $[a, b]$ is

$$\sup_P \ell(P, f) = \sup_P \sum_{i=1}^{n} \sqrt{1 + [f'(x_i^*)]^2} \triangle x_i = \int_a^b \sqrt{1 + [f'(x)]^2}\, dx.$$

The last equality is followed the hypothesis that f is continuously differentiable.

■ **Arc Length Formula**

If $f'(x)$ is continuous on $[a, b]$, then the length of the curve $y = f(x)$, $a \le x \le b$, is

$$L = \int_a^b \sqrt{1 + [f'(x)]^2}\, dx.$$

The expression in Leibniz notation is

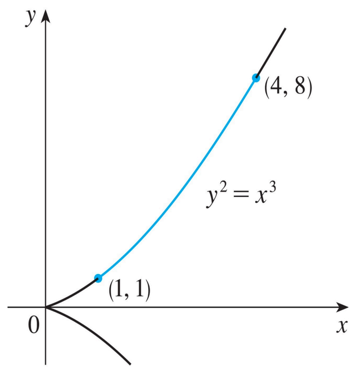
$$L = \int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx.$$

Example 9.4.2. Find the arc length of the semicubical parabola $y^2 = x^3$ between $(1, 1)$ and $(4, 8)$.

Proof. The curve between $(1, 1)$ and $(4, 8)$ satisfies the equation $y = x^{3/2}$. Then $\dfrac{dy}{dx} = \dfrac{3}{2}x^{1/2}$. The arc length of the curve is

$$L = \int_1^4 \sqrt{1 + (\frac{3}{2}x^{\frac{1}{2}})^2}\, dx = \frac{8}{27}u^{\frac{3}{2}}\Big|_{\frac{13}{4}}^{10} = \frac{1}{27}(80\sqrt{10} - 13\sqrt{13}).$$

□



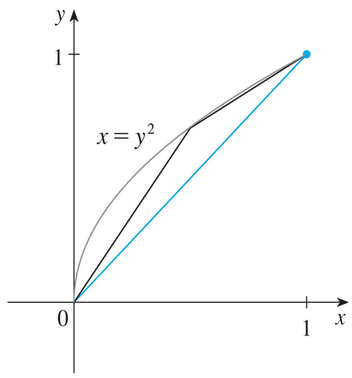
Suppose that the curve C has equation $x = g(y)$, $c \le y \le d$. Then the arc length of C is

$$L = \int_c^d \sqrt{1 + [g'(y)]^2}\, dy = \int_c^d \sqrt{1 + \left(\frac{dx}{dy}\right)^2}\, dy.$$

Example 9.4.3. Find the arc length of the curve C with the equation $y^2 = x$ from $(0, 0)$ to $(1, 1)$.

Proof. Since the curve has equation $x = y^2$, then $\dfrac{dx}{dy} = 2y$. The arc length of the curve is

$$\begin{aligned}
L &= \int_0^1 \sqrt{1 + (2y)^2}\, dy \\
&= \int_0^{\tan^{-1} 2} \sqrt{1 + \tan^2\theta} \cdot \frac{1}{2}\sec^2\theta\, d\theta \quad (y = \frac{1}{2}\tan\theta) \\
&= \frac{1}{4}\Big(\sec\theta\tan\theta + \ln|\sec\theta + \tan\theta|\Big)\Big|_0^{\tan^{-1} 2} \\
&= \frac{\sqrt{5}}{2} + \frac{1}{4}\ln(\sqrt{5} + 2).
\end{aligned}$$

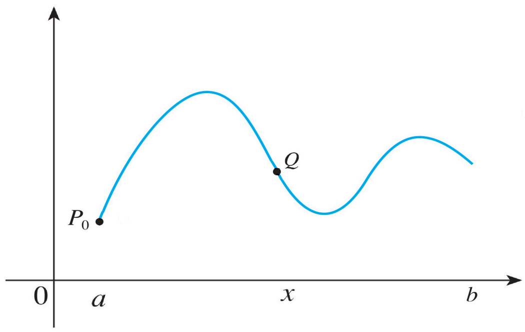


□

■ Arc Length Function

Suppose that a smooth curve C has the equation $y = f(x)$, $a \le x \le b$. Let $s(x)$ be the distance along C from the initial point $P_0\big(a, f(a)\big)$ to the point $Q\big(x, f(x)\big)$. Then s is a function, called the "*arc length function*" and

$$s(x) = \int_a^x \sqrt{1 + [f'(t)]^2}\, dt.$$



By the Fundamental Theorem of Calculus,

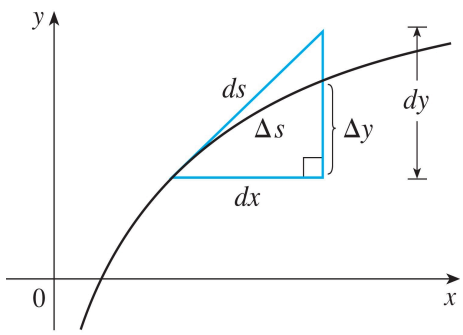
$$\frac{ds}{dx} = \sqrt{1 + [f'(x)]^2} = \sqrt{1 + \left(\frac{dy}{dx}\right)^2}.$$

This shows that the rate of change of s with respect to x is always at least 1 and is equal to 1 when $f'(x)$, the slope of the curve, is 0. The differential of arc length is

$$ds = \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx.$$

It is sometimes written in the symmetric form

$$(ds)^2 = (dx)^2 + (dy)^2.$$



Similarly,

$$ds = \sqrt{1 + \left(\frac{dx}{dy}\right)^2}\, dy.$$

Hence, the arc length along the curve C from $\big(a, f(a)\big)$ to $\big(t, f(t)\big)$ is

$$L = \int_0^t \underbrace{\sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx}_{ds} = \int 1\, ds = s(x)\Big|_a^t = s(t) - s(a) = s(t).$$

Example 9.4.4. Find the arc length function for the curve $y = x^2 - \frac{1}{8}\ln x$ taking $P_0(1, 1)$ as the starting point.

Proof. The rate of change of y with respect to x is

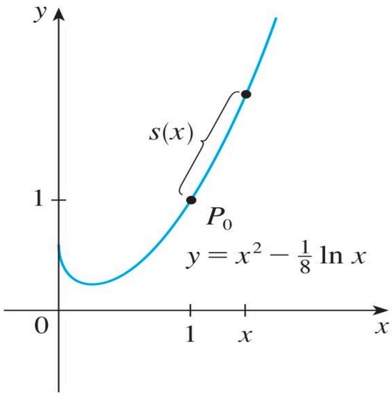
$$\frac{dy}{dx} = 2x - \frac{1}{8x}.$$

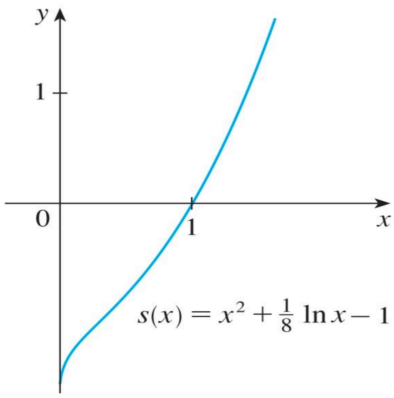
The arc length function is

$$\begin{aligned} s(x) &= \int_1^x \sqrt{1 + (2t - \frac{1}{8t})^2}\, dt = \int_1^x \sqrt{(2t + \frac{1}{8t})^2}\, dt \\ &= \int_1^x 2t + \frac{1}{8t}\, dt = x^2 + \frac{1}{8}\ln x - 1. \end{aligned}$$

The arc length from (1, 1) to (3, f(3)) is

$$s(3) = 3^2 + \frac{1}{8}\ln 3 - 1 = 8 + \frac{\ln 3}{8}.$$

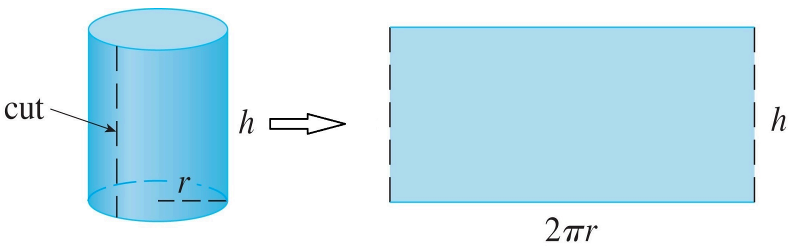




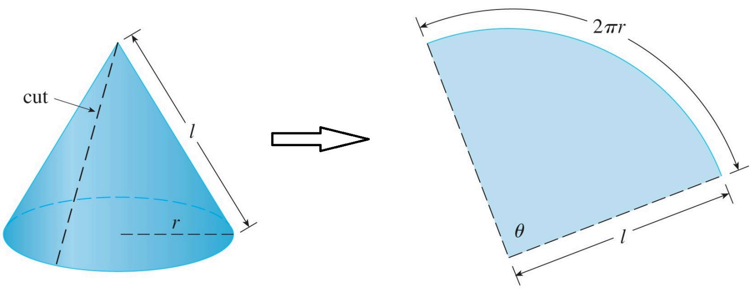
□

9.5 Area of a Surface of Revolution

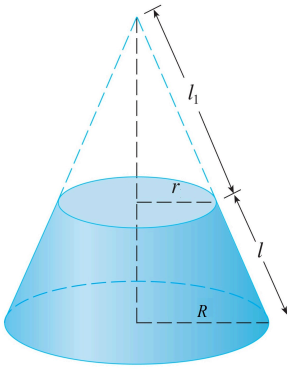
In the present section, we want to evaluate the area of a surface of revolution which is formed when a curve is rotated about a line. Let's look at some simple cases.



Area $= 2\pi rh$.

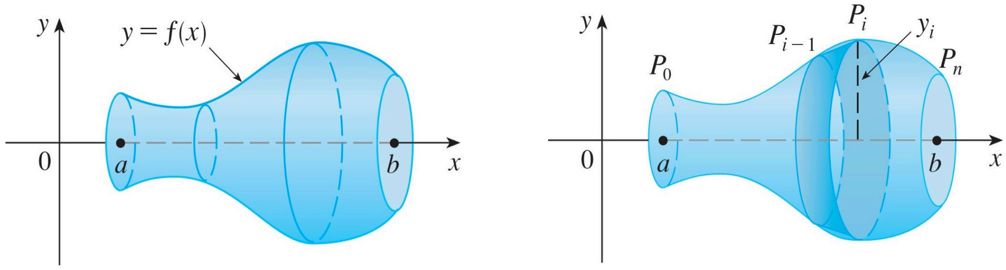


$$\begin{aligned} \theta &= \frac{2\pi r}{\ell} \\ \text{Area} &= \frac{1}{2}\ell^2\theta = \pi r\ell. \end{aligned}$$



$$\begin{aligned} \frac{r}{R} &= \frac{\ell_1}{\ell+\ell_1} \\ \Rightarrow \quad \ell_1 &= \frac{r\ell}{R-r} \end{aligned} \qquad\qquad \begin{aligned} \text{Area} &= \pi R(\ell+\ell_1) - \pi r\ell_1 \\ &= \pi(R-r)\ell_1 + \pi R\ell \\ &= \pi(R+r)\ell. \end{aligned}$$

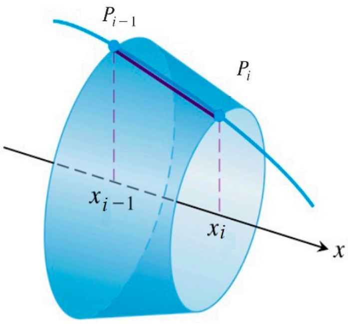
Consider the surface which is obtained by rotating the curve $y = f(x)$, $a \le x \le b$, about the x-axis where f is positive and has a continuous derivative. Let $P = \{x_0, x_1, \cdots, x_n\}$ be a partition of $[a, b]$. The points $P_0\big(x_0, f(x_0)\big), \cdots, P_n\big(x_n, f(x_n)\big)$ are on the curve $y = f(x)$.



(a) Surface of revolution

(b) Approximating band

The surface of revolution S is divided into several "belts". The surface area of one belt can be calculated in terms of its radius and its arc length.



$$\begin{aligned} \text{Area} &= \pi\big(f(x_i)+f(x_{i-1})\big)\sqrt{(x_i^2-x_{i-1}^2)^2\big[f(x_i)-f(x_{i-1})\big]^2} \\ &\overset{M.V.T}{=} \pi\big(f(x_i)+f(x_{i-1})\big)\sqrt{\triangle x_i^2\big[f'(x_i^*)\triangle x_i\big]^2} \\ &= \pi\big(f(x_i)+f(x_{i-1})\big)\sqrt{1+[f'(x_i^*)]^2}\triangle x_i \\ &\approx 2\pi f(x_i^*)\sqrt{1+[f'(x_i^*)]^2}\triangle x_i \end{aligned}$$

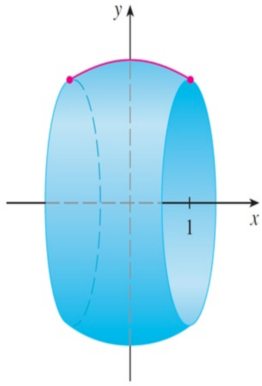
Hence, the sufrace area of the revolution is

$$\begin{aligned} S &= \lim_{n\to\infty}\sum_{i=1}^{n} 2\pi f(x_i^*)\sqrt{1+[f'(x_i^*)]^2}\triangle x_i \\ &= \int_a^b 2\pi f(x)\sqrt{1+[f'(x)]^2}\,dx \\ \text{(Leibniz notation)} \quad &= \int_a^b 2\pi y\sqrt{1+\Big(\frac{dy}{dx}\Big)^2}\,dx \\ \text{(arc length notation)} \quad &= \int_a^b 2\pi y\,ds \qquad\qquad \Big(ds = \sqrt{1+\Big(\frac{dy}{dx}\Big)^2}\,dx\ \Big) \end{aligned}$$

Example 9.5.1. The curve $y = \sqrt{4 - x^2}$, $-1 \le x \le 1$, is an arc of the circle $x^2 + y^2 = 4$. Find the area of the surface obtained by rotating this arc about the x-axis.

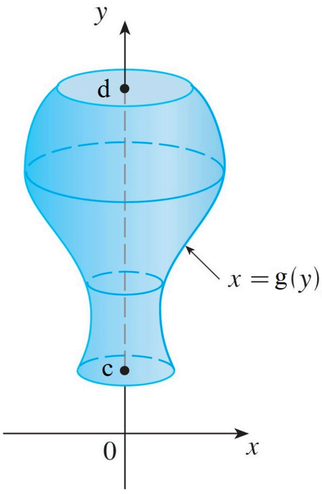
Proof. Since $y = \sqrt{4 - x^2}$, then $\dfrac{dy}{dx} = \dfrac{-x}{\sqrt{4 - x^2}}$. The surface area is

$$\begin{aligned} S &= \int_{-1}^{1} 2\pi y \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx \\ &= 2\pi \int_{-1}^{1} \sqrt{4 - x^2} \sqrt{1 + \left(\frac{-x}{\sqrt{4 - x^2}}\right)^2}\, dx \\ &= 2\pi \int_{-1}^{1} 2\, dx = 8\pi. \end{aligned}$$



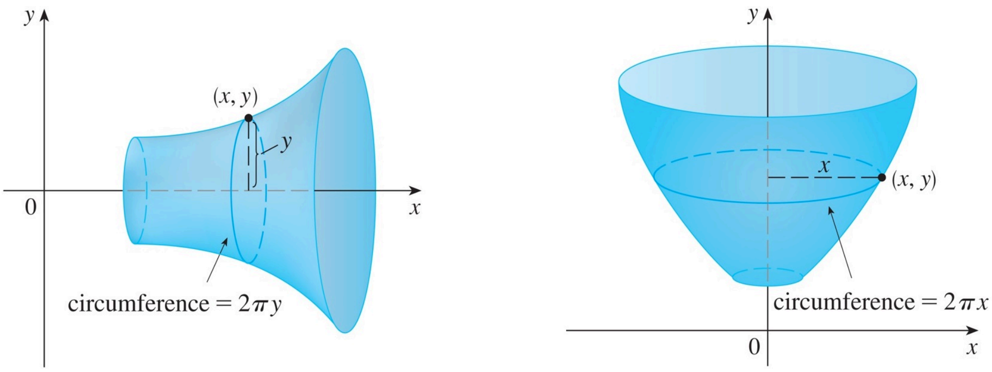
□

Similarly, the surface is obtained by rotating the curve $x = g(y)$, $c \le y \le d$, about the y-axis. The surface area is



$$\begin{aligned} \text{Area} &= \int_c^d 2\pi g(y) \sqrt{1 + [g'(y)]^2}\, dy \\ &= \int_c^d 2\pi x \sqrt{1 + \left(\frac{dx}{dy}\right)^2}\, dy \\ &= \int_c^d 2\pi x\, ds \qquad \left(ds = \sqrt{1 + \left(\frac{dx}{dy}\right)^2}\, dy\right) \end{aligned}$$

Note. Thinking of $2\pi y$ or $2\pi x$ as the circumference of a circle traced out by the point (x, y) on the curve as it is rotated about the x-axis or y-axis respectively.

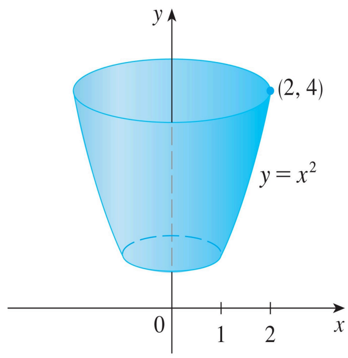


(a) Rotation about x-axis: $S = \int 2\pi y\, ds$ (b) Rotation about y-axis: $S = \int 2\pi x\, ds$

Example 9.5.2. The arc of the parabola $y = x^2$ from $(1, 1)$ to $(2, 4)$ is rotated about the y-axis. Find the area of the resulting surface.

Proof.

Method 1: Since $y = x^2$, then $\dfrac{dy}{dx} = 2x$. The surface area is



$$\begin{aligned} S &= \int 2\pi x\, dx = \int_1^2 2\pi x \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx \\ &= 2\pi \int_1^2 x\sqrt{1+4x^2}\, dx \\ &= \frac{\pi}{4}\left[\frac{2}{3}u^{\frac{3}{2}}\right]_5^{17} = \frac{\pi}{6}(17\sqrt{17} - 5\sqrt{5}). \end{aligned}$$

Method 2 : Since $x = \sqrt{y}$, then $\dfrac{dx}{dy} = \dfrac{1}{2\sqrt{y}}$. The surface area is

$$\begin{aligned} S &= \int 2\pi x\, ds = \int_1^4 2\pi \sqrt{y} \sqrt{1 + \left(\frac{dx}{dy}\right)^2}\, dy \\ &= \pi \int_1^4 \sqrt{4y+1}\, dy \\ &= \frac{\pi}{4}\int_5^{17} \sqrt{u}\, du = \frac{\pi}{6}(17\sqrt{17} - 5\sqrt{5}). \end{aligned}$$

□

Example 9.5.3. Find the area of the surface generated by rotating the curve $y = e^x$, $0 \le x \le 1$, about the x-axis.

Proof. Since $y = e^x$, then $\dfrac{dy}{dx} = e^x$. The surface area is

$$\begin{aligned} S &= \int 2\pi y\, ds = \int_0^1 2\pi e^x \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx \\ &= 2\pi \int_0^1 e^x\sqrt{1+e^{2x}}\, dx \\ (u = e^x) \quad &= 2\pi \int_1^e \sqrt{1+u^2}\, du \\ (u = \tan\theta) \quad &= 2\pi \int_{\pi/4}^{\tan^{-1} e} \sec^3\theta\, d\theta \\ &= \pi\Big[\sec\theta\tan\theta + \ln|\sec\theta + \tan\theta|\Big]_{\pi/4}^{\tan^{-1} e} \\ &= \pi\Big[e\sqrt{1+e^2} + \ln(e + \sqrt{1+e^2}) - \sqrt{2} - \ln(\sqrt{2}+1)\Big]. \end{aligned}$$

□

Chapter 10

Parametric Equations and Polar Coordinates

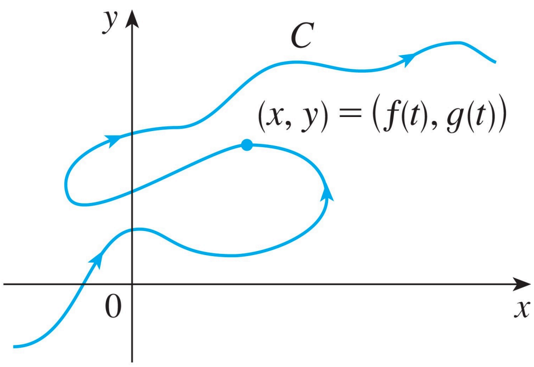
So far, we have studied the plane curves which are the graphs of explicit functions ($y = f(x)$ or $x = g(y)$) or implicit functions ($f(x, y) = 0$). In the present chapter, we will discuss those curves which are given in terms of a third variable t ($x = f(t)$ and $y = g(t)$).

10.1 Parametric Curves

When a particle moves on a plane along the curve C, in general, the path may not be described as an equation of the form $y = f(x)$ (or $x = g(y)$). Suppose that x and y are both given as functions of a third variable t (called a "*parameter*"). The equation

$$x = f(t), \quad y = g(t)$$

is called a "*parametric equation*".



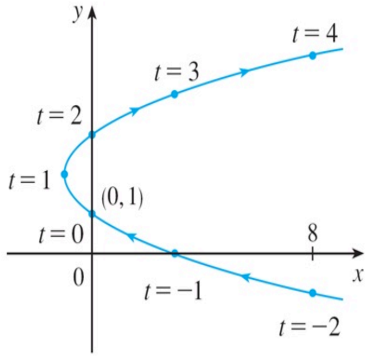
Each value of t determines a point (x, y) which we can plot in a coordinate plane. As t varies, the point $(x, y) = \big(f(t), g(t)\big)$ varies and traces out a curve C. We call the curve C : $(x, y) = \big(f(t), g(t)\big)$ a "*parametric curve*".

Example 10.1.1. Sketch and identify the curve defined by the parametric equation

$$x = t^2 - 2t \quad y = t + 1$$

$$t = y - 1 \Rightarrow x = (y - 1)^2 - 2(y - 1) = y^2 - 4y + 3 \quad \text{(Cartesian equation)}$$

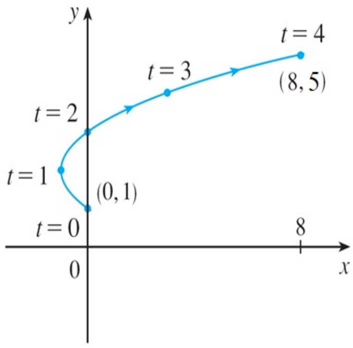
t	x	y
-2	8	-1
-1	3	0
0	0	1
1	-1	2
2	0	3
3	3	4
4	8	5



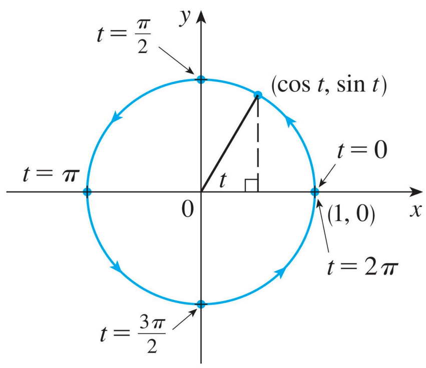
We sometimes restrict t to lie in a finite interval.

Example 10.1.2. Consider the parametric equation

$$x = t^2 - 2t \quad y = t + 1 \quad 0 \leq t \leq 4$$



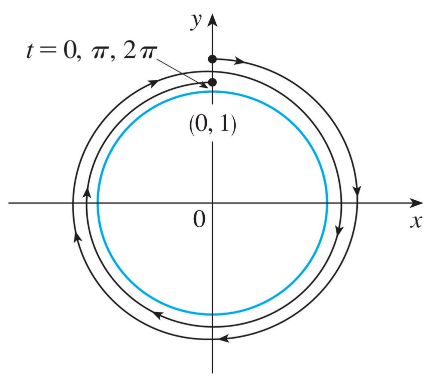
Example 10.1.3. Observe the parametric equation



$$x = \cos t \quad y = \sin t \quad 0 \leq t \leq 2\pi$$

represents the circle $x^2 + y^2 = 1$. As t increase from 0 to 2π, the point $(x, y) = (\cos t, \sin t)$ moves once around the circle in the counterclockwise direction starting from the point $(1, 0)$.

Example 10.1.4. The parametric equation

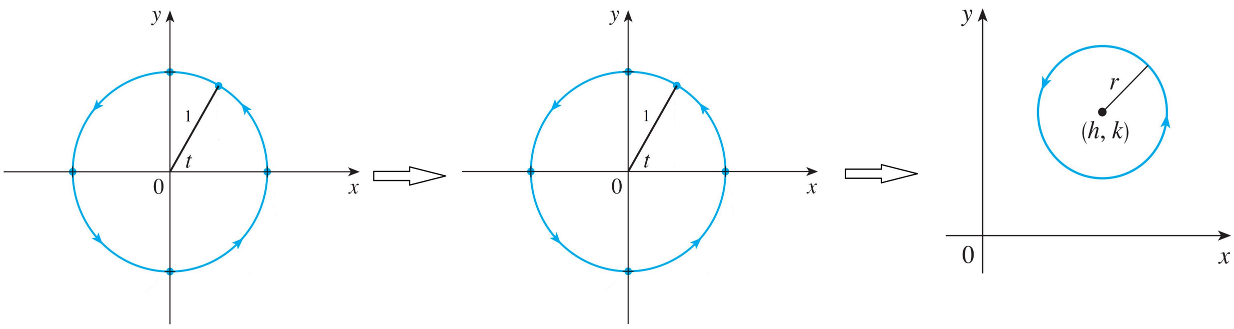


$$x = \sin 2t \quad y = \cos 2t \quad 0 \leq t \leq 2\pi$$

still represents the unit circle $x^2 + y^2 = 1$. But as t increases from 0 to 2π, the point $(x, y) = (\sin 2t, \cos 2t)$ starts at $(0, 1)$ and moves twice around the circle in the clockwise direction.

Example 10.1.5. Find parametric equations for the circle with center (h, k) and radius r.

Proof. We start from the circle $x = \cos t$, $y = \sin t$. Multiplying the expressions for x and y by r, we get $x = r\cos t$, $y = r\sin t$ and it represents a circle with radius r and center the origin traced counterclockwise. Then we shift h units in the x-direction and k units in the y-direction and obtain parametric equations of the circle with center (h, k) and raidus r.



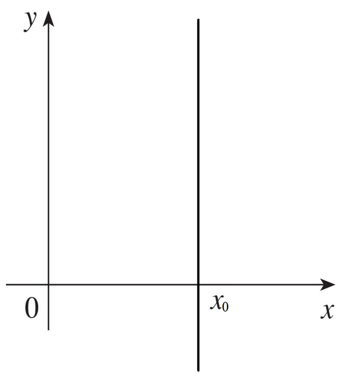
$x = \cos t$, $y = \sin t$ $\quad$ $x = r\cos t$, $y = r\sin t$ $\quad$ $x = h + r\cos t$, $y = k + r\sin t$

□

Example 10.1.6. (Straight Line)

The parametric equation of a straight line perpendicular the x-axis and passing $(x_0, 0)$ is

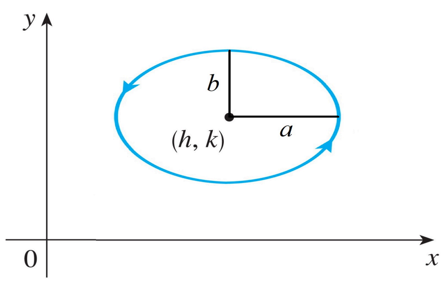
$$x = x_0 \quad y = t.$$



Example 10.1.7. (Ellipsoid)

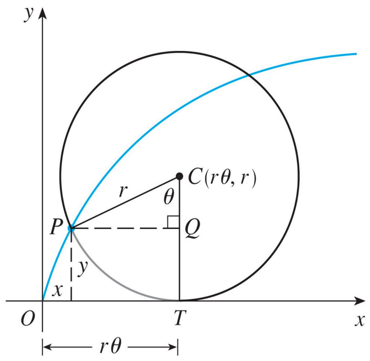
The parametric equation of an ellipsoid with center (h, k) and two axes with lengths a and b is

$$x = h + a\cos t \quad y = k + b\sin t \quad 0 \le t \le 2\pi.$$

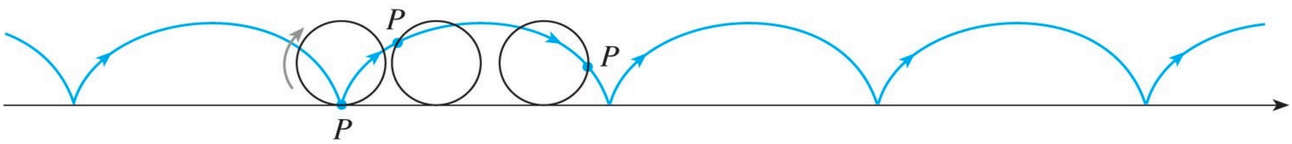


$x = h + a\cos t$, $y = k + b\sin t$

Example 10.1.8. (The Cycloid 擺線) The curve traced out by a point P on the circumference of a circle as the circle rolls along a straight line is called a "*cycloid*".



$$\begin{aligned} x &= |OT| - |PQ| = r\theta - r\sin\theta = r(\theta - \sin\theta) \\ y &= |CT| - |CQ| = r - r\cos\theta = r(1 - \cos\theta). \end{aligned}$$

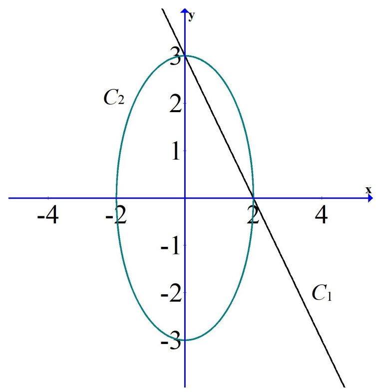


Example 10.1.9. Two particles move along the curves C_1 and C_2, respectively, with parametric equations

$$C_1 : \begin{cases} x = \dfrac{16}{3} - \dfrac{8}{3}t \\ y = 4t - 5 \end{cases}, \ t \geq 0 \qquad\qquad C_2 : \begin{cases} x = 2\sin(\dfrac{1}{2}\pi t) \\ y = -3\cos(\dfrac{1}{2}\pi t) \end{cases}, \ t \geq 0$$

(a) Do the two curves intersect?

Proof. The Cartesian equations of C_1 and C_2 are $C_1 : 3x + 2y - 6 = 0$ and $C_2 : \dfrac{x^2}{4} + \dfrac{y^2}{9} = 1$. We can solve the two equations and find the points where the the curves intersect at $(2, 0)$ and $(0, 3)$. □



(b) Do the two particles collide?

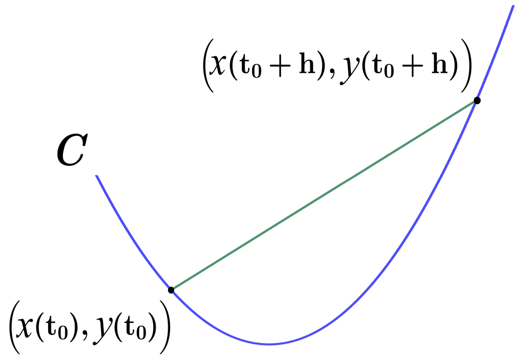
Proof. Find $t \geq 0$ such that both $\dfrac{16}{3} - \dfrac{8}{3}t = 2\sin(\dfrac{1}{2}\pi t)$ and $4t - 5 = -3\cos(\dfrac{1}{2}\pi t)$. We have $t = 2$ and the two particles collide at $(0, 3)$ when $t = 2$. □

10.2 Calculus with Parametric Curves

In the present section, we will apply the methods of calculus to the parametric curves. We will solve problems involving tangents, areas, arc length, and surface area.

❏ Tangents

Suppose that f and g are differentiable functions and C is a curve with parametric equation $x = x(t)$, $y = y(t)$. We want to find the tangent line of the curve C at a given point. In order to find the equation of the tangent line, it suffices to obtain its slope $\dfrac{dy}{dx}$.



The slope of the secant line connecting $\big(x(t_0), y(t_0)\big)$ and $\big(x(t_0+h), y(t_0+h)\big)$ is

$$\frac{y(t_0+h)-y(t_0)}{x(t_0+h)-x(t_0)} = \frac{\frac{y(t_0+h)-y(t_0)}{h}}{\frac{x(t_0+h)-x(t_0)}{h}} \xrightarrow{h\to 0} \frac{y'(t_0)}{x'(t_0)} = \frac{dy/dt}{dx/dt}\bigg|_{t=t_0}$$

By the Chain Rule,

$$\frac{dy}{dt} = \frac{dy}{dx}\cdot\frac{dx}{dt}.$$

If $\dfrac{dx}{dt} \neq 0$, we have

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt}$$

Remark.

(i) The rate of change of y with respect to x, $\dfrac{dy}{dx}$, is followed by the Chain Rule. It is not necessary to express y in terms of x.

(ii) The curve has a horizontal tangent line when $\dfrac{dy}{dt} = 0$ and $\dfrac{dx}{dt} \neq 0$.

(iii) The curve has a vertical tangent line when $\dfrac{dy}{dt} \neq 0$ and $\dfrac{dx}{dt} = 0$.

(iv) How about $\dfrac{dx}{dt} = 0 = \dfrac{dy}{dt}$? It may need further discussion.

(v) To discuss the concavity of a curve, we consider

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\left(\frac{dy}{dx}\right) = \frac{\frac{d}{dt}\left(\frac{dy}{dx}\right)}{\frac{dx}{dt}}.$$

Notice that

$$\frac{d^2y}{dt^2} \neq \frac{\frac{d^2y}{dt^2}}{\frac{d^2x}{dt^2}}.$$

Example 10.2.1. A curve C is defined by the parametric equations $x = t^2$, $y = t^3 - 3t$.

(a) Show that C has two tangents at the point $(3, 0)$ and find their equations.

Proof. Find the value(s) of t at which the curve passes $(3, 0)$. Consider

$$t^2 = 3 \implies t = \pm\sqrt{3} \quad \text{and} \quad t^3 - 3t = 0 \implies t = 0, \pm\sqrt{3}.$$

Hence, when $t = \pm\sqrt{3}$, the curve passes $(3, 0)$. Also, $\dfrac{dy}{dt} = 3t^2 - 3$ and $\dfrac{dx}{dt} = 2t$. Then

$$\frac{dy}{dt}\Big|_{t=-\sqrt{3}} = \frac{dy/dt}{dx/dt}\Big|_{t=-\sqrt{3}} = \frac{3}{2}(t - \frac{1}{t})\Big|_{t=-\sqrt{3}} = -\sqrt{3}.$$

The equation of the tangent line is $\underline{y = -\sqrt{3}(x - 3)}$. Similarly, $\dfrac{dy}{dx}\Big|_{\sqrt{3}} = \dfrac{3}{2}(t - \dfrac{1}{t})\Big|_{t=\sqrt{3}} = \sqrt{3}$.
The equation of the tangent line is $\underline{y = \sqrt{3}(x - 3)}$. □

(b) Find the points on C where the tangent is horizontal or vertical.

Proof.

(i) **Horizontal tanglent line:** Let $\dfrac{dy}{dt} = 3t^2 - 3 = 0$, then $t = \pm 1$. Also, $\dfrac{dx}{dt} = 2t \neq 0$ when $t = \pm 1$. Hence, when $t = 1$, $\big(x(1), y(1)\big) = (1, -2)$. The curve has a horizontal tangent line $\underline{y = -2}$. When $t = -1$, $\big(x(-1), y(-1)\big) = (1, 2)$. The curve has a horizontal tangent line $\underline{y = 2}$.

(ii) **Vertical tangent line:** Let $\dfrac{dx}{dt} = 2t = 0$. Then $t = 0$. Also, $\dfrac{dy}{dt} = 3t^2 - 3 \neq 0$ when $t = 0$ and $\big(x(0), y(0)\big) = (0, 0)$. The curve has a vertical tangent line $\underline{x = 0}$.

□

(c) Determine where the curve is concave upward or downward.

Proof. Consider

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\Big(\frac{dy}{dx}\Big) = \frac{\frac{d}{dt}\Big(\frac{dy/dt}{dx/dt}\Big)}{\frac{dx}{dt}} = \frac{\frac{d}{dt}[\frac{3}{2}(t - \frac{1}{t})]}{2t} = \frac{3(t^2 + 1)}{4t^3}.$$

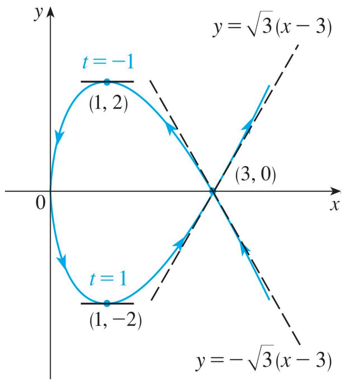
Then

$$\frac{d^2y}{dx^2} > 0 \quad \text{when} \quad t > 0 \qquad \text{and} \qquad \frac{d^2y}{dx^2} < 0 \quad \text{when} \quad t < 0.$$

The curve is concave upward when $t > 0$ and concave downward when $t < 0$. □

(d) Sketch the curve

Proof.



□

Example 10.2.2. (a) Find the tangent to the cycloid $x = r(\theta - \sin\theta)$, $y = r(1-\cos\theta)$ at the point where $\theta = \frac{\pi}{3}$.

Proof. Consider

$$\frac{dy}{dx} = \frac{dy/d\theta}{dx/d\theta} = \frac{r\sin\theta}{r(1-\cos\theta)} = \frac{\sin\theta}{1-\cos\theta}.$$

When $\theta = \frac{\pi}{3}$, $\big(x(\theta), y(\theta)\big) = \big(r(\frac{\pi}{3} - \frac{\sqrt{3}}{2}), \frac{r}{2}\big)$ and $\frac{dy}{dx}\Big|_{\theta=\frac{\pi}{3}} = \frac{\sqrt{3}/2}{1-\frac{1}{2}} = \sqrt{3}$. Therefore, when $\theta = \frac{\pi}{3}$, the tangent line is

$$y - \frac{r}{2} = \sqrt{3}\big(x - r(\frac{\pi}{3} - \frac{\sqrt{3}}{2})\big).$$

□

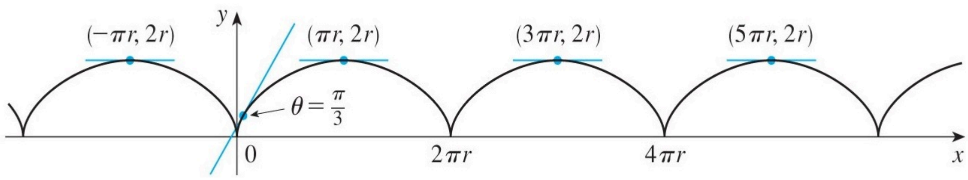
(b) At what point(s) is the tangent horizontal? Where is it vertical?

Proof. (i) When $n = 2m-1$ is odd, $\frac{dx}{d\theta} = r(1-\cos\theta) \neq 0$. The curve has horizontal tangent lines at $\big(x\big((2m-1)\pi\big), y\big((2m-1)\pi\big)\big) = \big((2m-1)\pi r, 2r\big)$, $m \in \mathbb{Z}$.

(ii) When $n = 2m$ is even. $\frac{dx}{d\theta} = 0$. Consider the limit

$$\lim_{\theta\to 2m\pi^+} \frac{dy}{dx} = \lim_{\theta\to 2m\pi^+} \frac{\sin\theta}{1-\cos\theta} \overset{L.H}{=} \lim_{\theta\to 2m\pi^+} \frac{\cos\theta}{\sin\theta} = \infty.$$

Similarly, $\lim_{\theta\to 2m\pi^-} \frac{dy}{dx} = -\infty$. The curve has vertical tangent line at $\big(x(2m\pi), y(2m\pi)\big) = (2m\pi r, 0)$, $m \in \mathbb{Z}$.



□

❏ Areas

Recall that, for a function $F(x) \geq 0$, the area under the cruve $y = F(x)$ from a to b is $A = \int_a^b F(x)\,dx$. Suppose that a curve has the parametric equation $x = f(t)$ and $y = g(t)$, $\alpha \leq t \leq \beta$, we want to calculate an area formula. Let $a = f(\alpha)$ and $b = f(\beta)$. Then the area of the region under the curve is

$$A = \int_a^b y\,dx = \int_\alpha^\beta y\,\frac{dx}{dt}\,dt = \int_\alpha^\beta g(t)f'(t)\,dt.$$

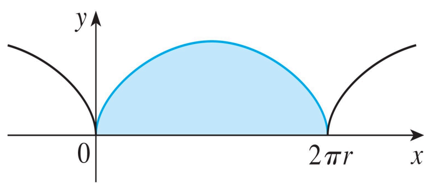
Example 10.2.3. Find the area under one arch of the cycloid

$$x = r(\theta - \sin\theta) \qquad y = r(1 - \cos\theta)$$

Proof.

Using the Substitution Rule with $y = r(1 - \cos\theta)$ and $dx = r(1 - \cos\theta)\,d\theta$, the area of one arch is

$$\begin{aligned} A &= \int_0^{2\pi r} y\,dx = \int_0^{2\pi} r(1-\cos\theta)r(1-\cos\theta)\,d\theta \\ &= r^2(\frac{3}{2}\cdot 2\pi) = 3\pi r^2. \end{aligned}$$



□

❏ Arc Length

Let C be a curve with equation $y = F(x)$, $a \leq x \leq b$. If $F'(x)$ is continuous, the arc length of C is

$$L = \int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\,dx = \int_a^b \sqrt{1 + \left(F'(x)\right)^2}\,dx.$$

We want to calculate the arc length of C with parametric equation $x = f(t)$, $y = g(t)$, $\alpha \leq t \leq \beta$.

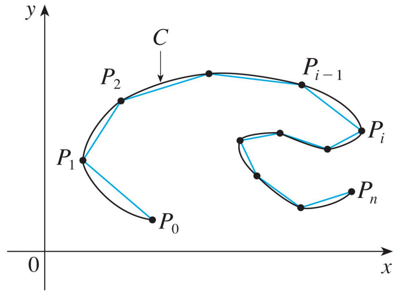
(i) If C can be expressed as the graph of a function $y = F(x)$, it is traversed once from left to right as t increases (i.e. $\frac{dx}{dt} = f'(t) > 0$). The arc length is

$$\begin{aligned} L &= \int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\,dx \\ &= \int_\alpha^\beta \sqrt{1 + \left(\frac{dy/dt}{dx/dt}\right)^2}\left(\frac{dx}{dt}\right)dt \\ &= \int_\alpha^\beta \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\,dt \end{aligned}$$

(ii) If C cannot be expressed in the form $y = F(x)$, we take a partition $P = \{t_0, t_1, \cdots, t_n\}$ of $[\alpha, \beta]$. Let $P_i\big(f(t_i), g(t_i)\big)$, $i = 1, \cdots, n$, be point on the curve C. Then the length of the segment $\overline{P_{i-1}P_i}$ is

$$\sqrt{[f(t_i) - f(t_{i-1})]^2 + [g(t_i) - g(t_{i-1})]^2}$$

By the polygonal approximateions and the mean value theorem,



$$\begin{aligned}
\sum_{i=1}^{n} |\overline{P_{i-1}P_i}| &= \sum_{i=1}^{n} \sqrt{[f(t_i) - f(t_{i-1})]^2 + [g(t_i) - g(t_{i-1})]^2} \\
&= \sum_{i=1}^{n} \sqrt{[f'(t_i^*)\Delta t_i]^2 + [g'(t_i^{**}\Delta t_i)]^2} \\
&= \sum_{i=1}^{n} \sqrt{[f'(t_i^*)]^2 + [g'(t_i^{**})]^2}\Delta t_i
\end{aligned}$$

The arc length of C is

$$\begin{aligned}
L &= \lim_{\|P\|\to 0} \sum_{i=1}^{n} \sqrt{[f'(t_i^*)]^2 + [g'(t_i^{**})]^2}\Delta t_i \\
&= \int_{\alpha}^{\beta} \sqrt{[f'(t)]^2 + [g'(t)]^2}\, dt \\
&= \int_{\alpha}^{\beta} \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dx}\right)^2}\, dt
\end{aligned}$$

Theorem 10.2.4. *If a curve C is described by the parametric equation $x = f(t)$, $y = g(t)$, $\alpha \le t \le \beta$ where f' and g' are continuous on $[\alpha, \beta]$ and C is traversed exactly once as t increases from α to β, then the arc length of C is*

$$L = \int_{\alpha}^{\beta} \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dx}\right)^2}\, dt = \int_{\alpha}^{\beta} \sqrt{[f'(t)]^2 + [g'(t)]^2}\, dt.$$

Note. The formula is consisent with the general formulas $L = \int 1\, ds$ and $(ds)^2 = (dx)^2 + (dy)^2$.

Example 10.2.5. Compute the circumference of a unit circle by expressing it as the parametric equation

$$x = \cos t \qquad y = \sin t \qquad 0 \le t \le 2\pi$$

Proof. We have $\dfrac{dx}{dt} = -\sin t$ and $\dfrac{dy}{dt} = \cos t$. Then the arc length is

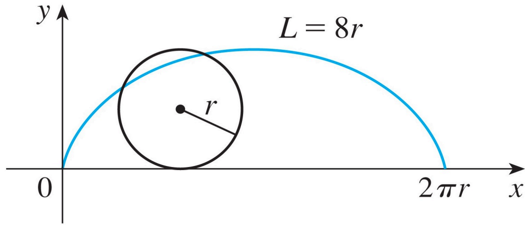
$$L = \int_{0}^{2\pi} \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt = \int_{0}^{2\pi} \sqrt{\sin^2 t + \cos^2 t}\, dt = 2\pi.$$

□

Example 10.2.6. Find the length of one arch of the cycloid $x = r(\theta - \sin\theta)$ and $y = r(1 - \cos\theta)$.

Proof. We have $\dfrac{dx}{d\theta} = r(1 - \cos\theta)$ and $\dfrac{dy}{d\theta} = r\sin\theta$. The arc length of one arch is

$$
\begin{aligned}
L &= \int_0^{2\pi} \sqrt{\left(\frac{dx}{d\theta}\right)^2 + \left(\frac{dy}{d\theta}\right)^2}\, d\theta \\
&= \int_0^{2\pi} \sqrt{r^2(1-\cos\theta)^2 + r^2\sin^2\theta}\, d\theta \\
&= r\int_0^{2\pi} \sqrt{2(1-\cos\theta)}\, d\theta \\
&= r\int_0^{2\pi} 2\sin\left(\frac{\theta}{2}\right) d\theta \\
&= 8r.
\end{aligned}
$$



□

❑ Surface Area

Recall that the surface area of the surface obtained by rotating a curve, $C : y = F(x)$ where $F(x) \geq 0$ for $a \leq x \leq b$, about x-axis is

$$S = \int_a^b 2\pi y \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx.$$

Suppose that C has the parametric equation $x = f(t)$ and $y = g(t)$, $\alpha \leq t \leq \beta$ where f' and g' are continuous and $g(t) \geq 0$. Then rotating the curve C about x-axis and the surface area is

$$
\begin{aligned}
S &= \int_a^b 2\pi y \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx \\
&= \int_\alpha^\beta 2\pi y \sqrt{1 + \left(\frac{dy/dt}{dx/dt}\right)^2} \left(\frac{dx}{dt}\right) dt \\
&= \int_\alpha^\beta 2\pi y \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt.
\end{aligned}
$$

Note. Let $s(t)$ be the arc length function. Then

$$ds = \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt.$$

The surface area formula is

$$S = \int 2\pi y\, ds$$

Example 10.2.7. Find the surface area of a sphere of radius r.

Proof. The sphere is obtained by rotating the semicircle

$$x = r\cos t \qquad y = r\sin t \qquad 0 \leq t \leq \pi$$

about x-axis. The surface area of the sphere is

$$
\begin{aligned}
S &= \int_0^\pi 2\pi r\sin t \sqrt{(-r\sin t)^2 + (r\cos t)^2}\, dt \\
&= 2\pi \int_0^\pi r\sin t \cdot r\, dt = 4\pi r^2
\end{aligned}
$$

□

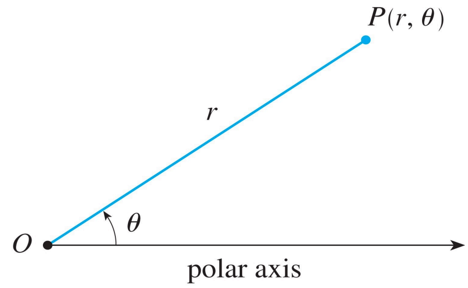
10.3 Polar Coordinates

A coordinate system represents a point in the plane by an ordered pair of numbers called coordinates. In the present section, we will study a coordinate system which is called the "*polar coordinate system*". The coordinate is established by the following steps

(i) We choose a point in the plane that is called the "*pole*" (or origin) and is labeled O.

(ii) We drwa a ray starting at O called the "*polar axis*" . It is usually horizontal to the right and corresponds to the positive x-axis in Cartesian coordinates.

(iii) If $P \neq O$ is an point in the plane, let r be the distance from O to P and let θ be then angle between the polar axis. We use the convention that an angle is positive if measured in the counterclockwise direction from the polar axis and negative in the clockwise direction.



Then the point P is represented by the ordered pair (r, θ) as well as r and θ are called "*polar coordinates*" of P.

Note. The origin $O = (0, \theta)$ for any θ.

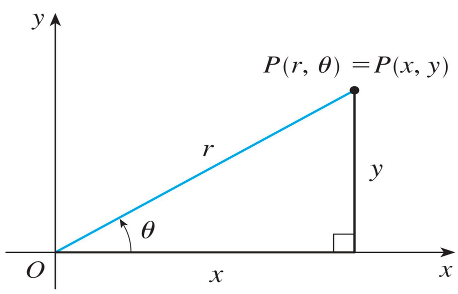
Now, we extend (r, θ) to the case that in which r is negative. The point $(-r, \theta)$ means the point which is opposite to (r, θ) about the origin. Hence, $(-r, \theta) = (r, \theta + \pi)$. Moreover, we can also extend (r, θ) to the case where $r \in \mathbb{R}$ (not only on $[0, 2\pi]$). We have

$$\begin{aligned}(r, \theta) &= (-r, \theta + \pi) = (r, \theta + 2\pi) \\ &= (-r, \theta + 3\pi) = (r, \theta + 4\pi) \\ &= \big(-r, \theta + (2k+1)\pi\big) = \big(r, \theta + 2k\pi\big) \quad \text{for every } k \in \mathbb{Z}.\end{aligned}$$

Remark. In the Cartesian coordinate system, every point has only one representation, but in the polar coordinate system, each point has infinitely many representations.

■ **The connection between polar and Cartesian coordinates**

$$\begin{cases} x = r\cos\theta \\ y = r\sin\theta \end{cases} \Longleftrightarrow \begin{cases} \cos\theta = \dfrac{x}{r} \\ \sin\theta = \dfrac{y}{r} \end{cases} \Longrightarrow \begin{cases} r^2 = x^2 + y^2 \\ \tan\theta = \dfrac{y}{x} \end{cases}$$



Example 10.3.1. (1) Convert $(2, \frac{\pi}{3})$ from polar to Cartesian coordinates.

Proof. From the above formulas, $x = 2\cos\frac{\pi}{3} = 1$ and $y = 2\sin\frac{\pi}{3} = \sqrt{3}$. Then $(x, y) = (1, \sqrt{3})$. □

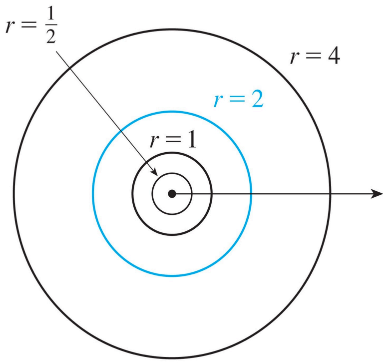
(2) Convert $(1,-1)$ from Cartesian to polar coordinates.

Proof. Again, $r = \sqrt{1^2+(-1)^2} = \sqrt{2}$ and $\tan\theta = \frac{-1}{1} = -1$. Then $\theta = \dfrac{3\pi}{4}$ or $\dfrac{7\pi}{4}$. Since $(1,-1)$ is a point in the fourth quadrant, $\theta = \dfrac{7\pi}{4}$ and $(r,\theta) = (\sqrt{2}, \dfrac{7\pi}{4})$. □

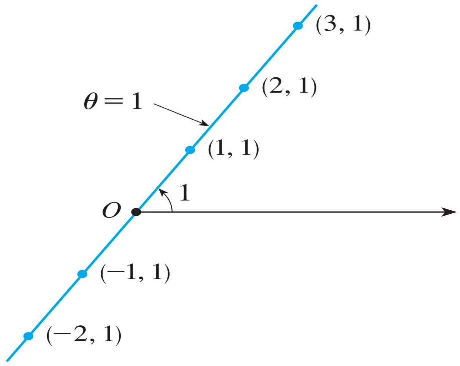
❑ Polar Curves

Definition 10.3.2. A polar curve is the graph of a polar equation, $r = f(\theta)$ or $F(r,\theta) = 0$, consists of all points P that have at least one polar representation (r,θ) whose coordinates satisfy the equation.

Example 10.3.3.



$r = 2$



$\theta = 1$

Example 10.3.4. (a) Sketch the curve with polar equation $r = 2\cos\theta$.

Proof.

θ	$r = 2\cos\theta$
0	2
$\pi/6$	$\sqrt{3}$
$\pi/4$	$\sqrt{2}$
$\pi/3$	1
$\pi/2$	0
$2\pi/3$	-1
$3\pi/4$	$-\sqrt{2}$
$5\pi/6$	$-\sqrt{3}$
π	-2

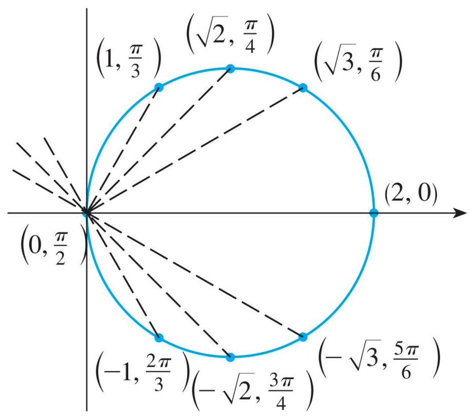


Table of values and graph of $r = 2\cos\theta$

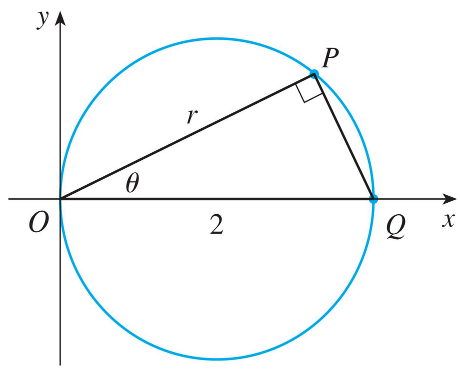
□

(b) Find a Cartesian equation for this curve.

Proof.

Consider $r = 2\cos\theta$. Then $r^2 = 2r\cos\theta$. Convert this polar equation into Cartesian equation $x^2 + y^2 = 2x$ and we have

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

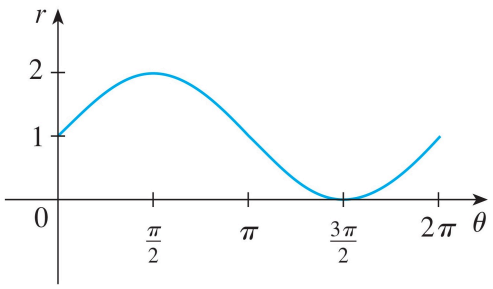


□

Example 10.3.5. Sketch the curve $r = 1 + \sin\theta$.

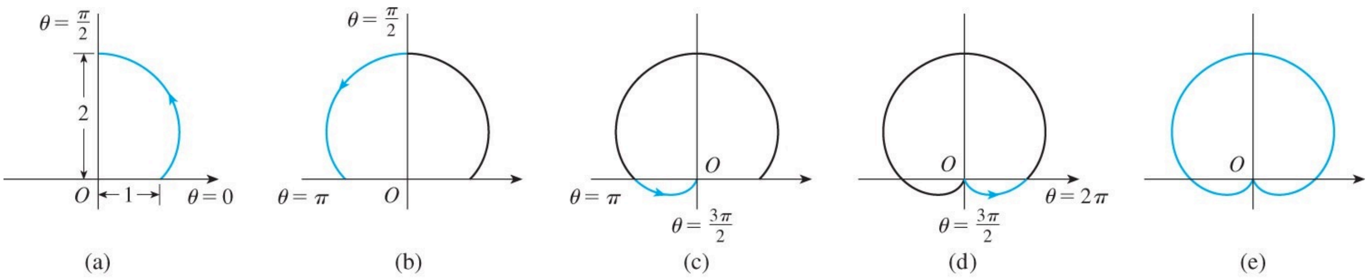
Proof.

(a) Sketch the graph of $r = 1 + \sin\theta$ in Cartesina coordinates (θ-r plane). That is a shift of the curve of sine function up by one unit.



$r = 1 + \sin\theta$ in Cartesian coordinates, $0 \leqslant \theta \leqslant 2\pi$

(b) Sketch the polar curve as θ increases $0 \to \dfrac{\pi}{2} \to \pi \to \dfrac{3\pi}{2} \to 2\pi$.
(Cardioid)

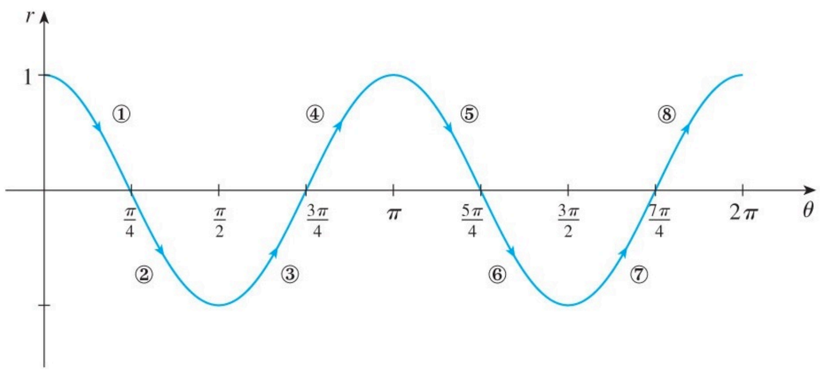


Stages in sketching the cardioid $r = 1 + \sin\theta$

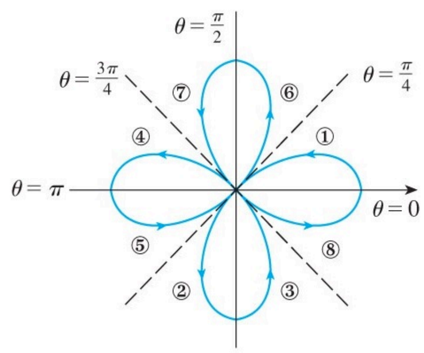
□

Example 10.3.6. Sketch the curve $r = \cos 2\theta$.

Proof. □



$r = \cos 2\theta$ in Cartesian coordinates

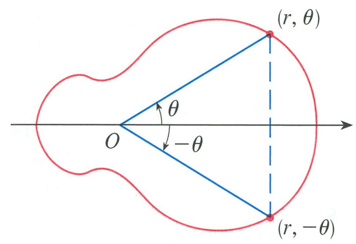


Four-leaved rose $r = \cos 2\theta$

■ Symmetry

(a)

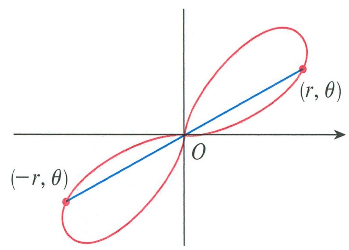
If $f(\theta) = f(-\theta)$ or $F(r, \theta) = F(r, -\theta)$, then the curve is symmetric about the polar axis.



(a)

(b)

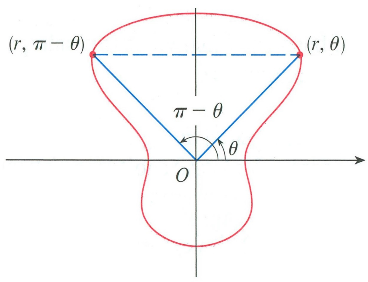
If $f(\theta) = f(\theta + \pi)$ or $F(r, \theta) = F(r, \theta + \pi)$, then the curve is symmetric about the pole.



(b)

(c)

If $f(\theta) = f(\pi - \theta)$ or $F(r, \theta) = F(r, \pi - \theta)$, then the curve is symmetric about the vertical line $\theta = \dfrac{\pi}{2}$.



(c)

❑ Tangents to Polar Curves

We want to use the techniques of finding the tangent lines of parametric curves to obtain the tangents of polar curves. Consider the curve with polar equation $r = f(\theta)$. Then

$$\begin{cases} x = r\cos\theta = f(\theta)\cos\theta \\ y = r\sin\theta = f(\theta)\sin\theta \end{cases} \implies \frac{dy}{dx} = \frac{dy/d\theta}{dx/d\theta} = \frac{\frac{dr}{d\theta}\sin\theta + r\cos\theta}{\frac{dr}{d\theta}\cos\theta - r\sin\theta}.$$

(i) **Horizontal tangent line:** When $\dfrac{dy}{d\theta} = 0$ and $\dfrac{dx}{d\theta} \neq 0$, the polar curve has a horizontal tangent line.

(ii) **Vertical tangent line:** When $\dfrac{dy}{d\theta} \neq 0$ and $\dfrac{dx}{d\theta} = 0$, the polar curve has a vertical tangent line.
(Special case: $\dfrac{dy}{d\theta} = 0 = \dfrac{dx}{d\theta}$, we should further consider the limit $\lim\limits_{\theta\to\theta_0} \dfrac{dy/d\theta}{dx/d\theta}$).

(iii) **Tangent line at pole:**

$$\frac{dy}{dx} = \frac{\frac{dr}{d\theta}\sin\theta}{\frac{dr}{d\theta}\cos\theta} = \tan\theta, \quad \text{if} \quad \frac{dr}{d\theta} \neq 0.$$

Example 10.3.7. The cardioid has polar equation $r = 1 + \sin\theta$.

(a) Find the slope of the tangent line when $\theta = \dfrac{\pi}{3}$.

Proof. Consider $\dfrac{dr}{d\theta} = \cos\theta$. Then

$$\frac{dy}{dx} = \frac{\cos\theta\sin\theta + (1+\sin\theta)\cos\theta}{\cos\theta\cos\theta - (1+\sin\theta)\sin\theta} = \frac{\cos\theta(1+2\sin\theta)}{(1+\sin\theta)(1-2\sin\theta)}.$$

Hence, the slope of the tangent line when $\theta = \dfrac{\pi}{3}$ is $\left.\dfrac{dy}{dx}\right|_{\theta=\frac{\pi}{3}} = -1$. □

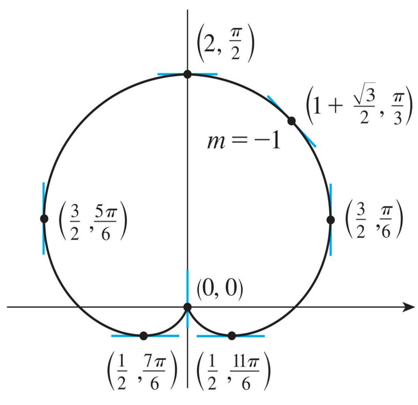
(b) Find the points on the cardioid where the tangent line is horizontal or vetical.

Proof. We have

$$\frac{dy}{d\theta} = \cos\theta(1+2\sin\theta) = 0 \implies \theta = \frac{\pi}{2}, \frac{3\pi}{2}, \frac{7\pi}{6}, \frac{11\pi}{6}.$$

$$\frac{dx}{d\theta} = (1+\sin\theta)(1-2\sin\theta) = 0 \implies \theta = \frac{3\pi}{2}, \frac{\pi}{6}, \frac{5\pi}{6}.$$

The curve has horizontal tangent lines at $(2, \pi/2)$, $(1/2, 7\pi/6)$ and $(1/2, 11\pi/6)$ and has vertical tangent lines at $(3/2, \pi/6)$, $(3/2, 5\pi/6)$.



Tangent lines for $r = 1 + \sin\theta$

For $\theta = \dfrac{3\pi}{2}$, $\dfrac{dy}{d\theta} = \dfrac{dx}{d\theta} = 0$. Consider

$$\lim_{\theta\to(3\pi/2)^-} \frac{dy}{dx} = \Big(\lim_{\theta\to(3\pi/2)^-} \frac{1+2\sin\theta}{1-2\sin\theta}\Big)\Big(\lim_{\theta\to(3\pi/2)^-} \frac{\cos\theta}{1+\sin\theta}\Big) \overset{L.H.}{=} -\frac{1}{3}\lim_{\theta\to(3\pi/2)^-} \frac{-\sin\theta}{\cos\theta} = \infty.$$

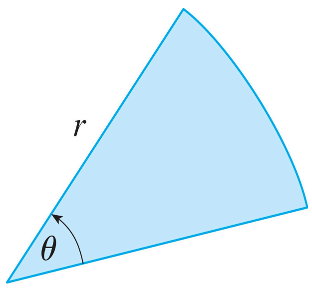
Similarly, $\displaystyle\lim_{\theta\to(3\pi/2)^+} \frac{dy}{dx} = -\infty$. Hence, the cardioid has a vertical tangent line at $(0, 3\pi/2)$.

□

10.4 Areas and Lengths in Polar Coordinates

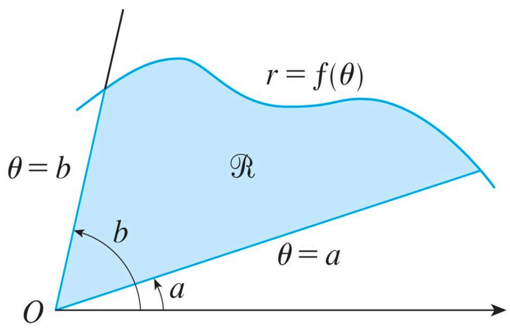
❏ Areas

We try to find the area of a region whose boundary is given by a polar equation. Let's start with an easy case that the area of an sector of a circle with radius r and central angle θ.



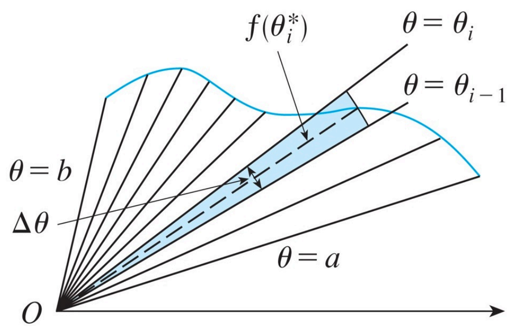
$$\text{Area } = \frac{1}{2}r^2\theta.$$

Let $\mathcal{R}$ be the region bounded by the polar curve $r = f(\theta)$ and by the rays $\theta = a$ and $\theta = b$, where f is a positive continuous function and where $0 < b - a < 2\pi$. We will use the approximating sectors to estimate the area of $\mathcal{R}$.



Let $P = \{\theta_0, \theta_1, \cdots, \theta_n\}$ be a paratition of $[a, b]$ with $\triangle\theta_i = \theta_i - \theta_{i-1}$. The region $\mathcal{R}$ is divided into n subregions by the rays $\theta = \theta_i$. The area of each subregion denotes $\triangle A_i$. Choose a sample point $\theta_i^* \in [\theta_{i-1}, \theta_i]$. Then

$$\triangle A_i \approx \frac{1}{2}[f(\theta_i^*)]^2 \triangle\theta_i.$$



Then an approximation to the total area A of $\mathcal{R}$ is

$$\text{Area} \approx \sum_{i=1}^{n} \frac{1}{2}[f(\theta_i^*)]^2 \triangle\theta_i$$

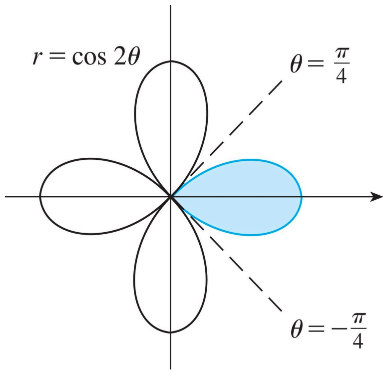
Taking $||P|| \to 0$, then

$$\begin{aligned}
\text{Area} &= \lim_{||P||\to 0} \sum_{i=1}^{n} \frac{1}{2}[f(\theta_i^*)]^2 \triangle\theta_i \\
&= \frac{1}{2}\int_a^b [f(\theta)]^2\, d\theta \\
&= \frac{1}{2}\int_a^b r^2\, d\theta \quad \text{where } r = f(\theta).
\end{aligned}$$

Note. The area formula is to compute the area of the region which area enclosed by a polar curve and two straight lines connecting the origin and their intersections of the polar curve

Example 10.4.1. Find the area enclosed by one loop of the four-leaved rose $r = \cos 2\theta$.

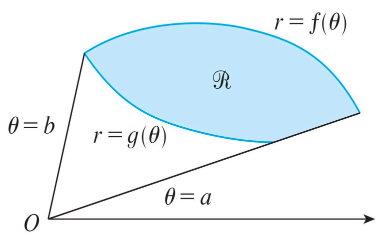
Proof.

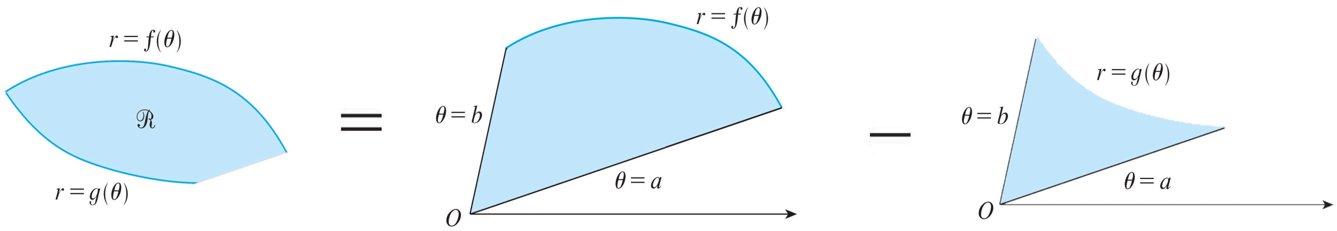


$$\begin{aligned}
\text{Area} &= \int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} \frac{1}{2} r^2\, d\theta \\
&= \frac{1}{2}\int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} \cos^2 2\theta\, d\theta \\
&= \frac{1}{2}\int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} \frac{1+\cos 4\theta}{2}\, d\theta \\
&= \frac{\pi}{8}.
\end{aligned}$$

□

■ **Region enclosed by two polar curves**

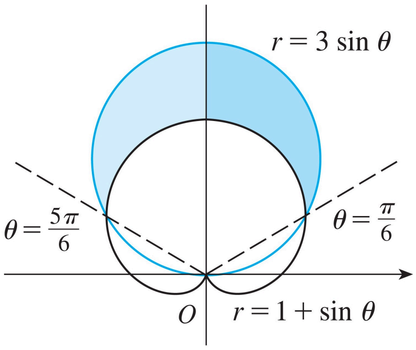




The area of $\mathcal{R}$ is

$$\int_a^b \frac{1}{2}[f(\theta)]^2\, d\theta - \int_a^b \frac{1}{2}[g(\theta)]^2\, d\theta = \frac{1}{2}\int_a^b f^2(\theta) - g^2(\theta)\, d\theta.$$

Example 10.4.2. Find the area of the region that lies inside the circle $r = 3\sin\theta$ and outside the cardioid $r = 1 + \sin\theta$.



Proof. The points of intersection of the two polar curves are obtained by solving $3\sin\theta = 1 + \sin\theta$ and hence $\theta = \frac{\pi}{6}, \frac{5\pi}{6}$. The area of the region is

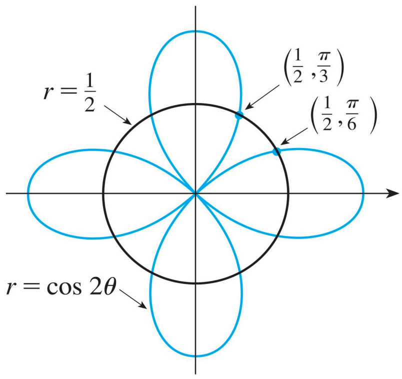
$$A = \int_{\frac{\pi}{6}}^{\frac{5\pi}{6}} \frac{1}{2}(3\sin\theta)^2 - \frac{1}{2}(1+\sin\theta)^2\, d\theta = \pi.$$

□

Note. The origin O is also a point of intersection of the two polar curves. But it cannot be obtained by solving the equation $3\sin\theta = 1 + \sin\theta$ since $r = 3\sin\theta = 0$ when $\theta = 0$ and π and $r = 1 + \sin\theta = 0$ when $\theta = \frac{3\pi}{2}$.

Remark. It is usually difficult to find the points of intersection of two polar curves since a single point may have many representations in polar coordinates. Suppose we want to find the points of intersection by solving $f_1(\theta) = r = f_2(\theta)$. The point of intersection has polar coordinate $\big(f_1(\theta_1), \theta_1\big) = \big(f_2(\theta_2), \theta_2\big)$. But, in general, the angles θ_1 may not equal θ_2.

Example 10.4.3. Find all points of intersection of the curves $r = \cos 2\theta$ and $r = \frac{1}{2}$.

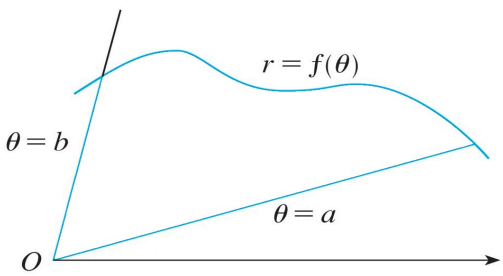


Proof. Let $\cos 2\theta = \frac{1}{2}$. Then $\theta = \frac{\pi}{6}, \frac{5\pi}{6}, \frac{7\pi}{6}, \frac{11\pi}{6}$. The points of intersection are $(\frac{1}{2}, \frac{\pi}{6})$, $(\frac{1}{2}, \frac{5\pi}{6})$, $(\frac{1}{2}, \frac{7\pi}{6})$ and $(\frac{1}{2}, \frac{11\pi}{6})$.

However, the points $(\frac{1}{2}, \frac{\pi}{3})$, $(\frac{1}{2}, \frac{2\pi}{3})$, $(\frac{1}{2}, \frac{4\pi}{3})$ and $(\frac{1}{2}, \frac{5\pi}{3})$ are also points of intersection of the two polar curves. Those points can be found by solving $\cos 2\theta = -\frac{1}{2}$. □

❑ Arc Length

To find the length of a polar curve $r = f(\theta)$, $a \le \theta \le b$, we regard θ as the parameter if we write the polar equation of the curve as

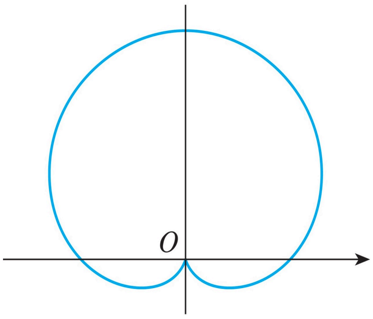


$$\begin{cases} x = r\cos\theta \\ y = r\sin\theta \end{cases} \Rightarrow \begin{cases} \dfrac{dx}{d\theta} = \dfrac{dr}{d\theta}\cos\theta - r\sin\theta \\[2mm] \dfrac{dy}{d\theta} = \dfrac{dr}{d\theta}\sin\theta + r\cos\theta \end{cases}$$

The arc length is

$$L = \int_a^b \sqrt{\Big(\frac{dx}{d\theta}\Big)^2 + \Big(\frac{dy}{d\theta}\Big)^2}\, d\theta = \int_a^b \sqrt{r^2 + \Big(\frac{dr}{d\theta}\Big)^2}\, d\theta.$$

Example 10.4.4. Find the length of the cardioid $r = 1 + \sin\theta$.



$r = 1 + \sin\,\theta$

Proof. The arc length fo the cardioid is

$$
\begin{aligned}
L &= \int_0^{2\pi} \sqrt{r^2 + \left(\frac{dr}{d\theta}\right)^2}\, d\theta = \int_0^{2\pi} \sqrt{(\cos\theta)^2 + (1+\sin\theta)^2}\, d\theta \\
&= \int_0^{2\pi} \sqrt{2 + 2\sin\theta}\, d\theta = \int_0^{2\pi} \frac{\sqrt{4 - 4\sin^2\theta}}{\sqrt{2 - 2\sin\theta}} \\
&= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{2\cos\theta}{\sqrt{2 - 2\sin\theta}}\, d\theta - \int_{\frac{\pi}{2}}^{\frac{3\pi}{2}} \frac{2\cos\theta}{\sqrt{2 - 2\sin\theta}}\, d\theta \\
&= 8.
\end{aligned}
$$

□

Chapter 11

Sequences and Series

In Chapter 2, we have learned some basic concepts of infinite sequences. In the present chapter, we will further study some important facts of sequences. Based on the ideas of infinite sequence, we will introduce an important mathematical object, "(infinite) series" in the rest sections.

11.1 Monotonic Sequences and Cauchy Sequences

Before revisit the infinite sequences, the readers should review and make familiar with all results of sequences in Chapter 2, such as "convergence and divergence of a sequence", "bounded sequence", "limit laws", "squeeze theorem", "infinite limit" etc.

❏ Monotonic Sequence

Definition 11.1.1. Let $\{a_n\}$ be a sequence.

(a) We say that $\{a_n\}$ is "*increasing* (and "*decreasing*") if for every $n, m \in \mathbb{N}$ with $n < m$ then

$$a_n < a_m \qquad (\text{and } \ a_n > a_m).$$

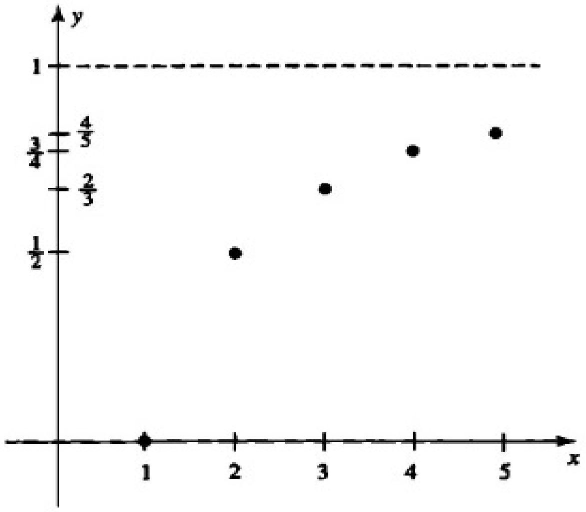
(b) We say that $\{a_n\}$ is "*nondecreasing*" (and "*nonincreasing*") if for every $n, m \in \mathbb{N}$ with $n < m$ then

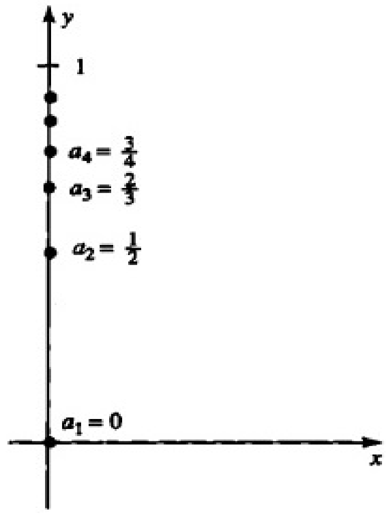
$$a_n \leq a_m \qquad (\text{and } a_n \geq a_m).$$

(c) If a sequence is either increasing (nondecreasing) or decreasing (or nonincreasing), we call it is a "*monotonic*" seqnence.

Example 11.1.2. (1) The sequence $\{1, 1, 2, 2, 3, 3, \cdots\}$ is an example of a nondecreasing sequence

(2) The sequence $\{1 - \frac{1}{n}\}$ is an increasing sequence.





Remark. For a sequence $\{a_n\}$ where $a_n > 0$ for all n, the following statements are equivalent

(i) $\{a_n\}$ is nondecreasing.

(ii) $a_n \leq a_{n+1}$ for all n.

(iii) $\dfrac{a_{n+1}}{a_n} \geq 1$ for all n.

(iv) $a_{n+1} - a_n \geq 0$ for all n.

Notice that an increasing (nondecreasing) sequence $\{a_n\}_{n=1}^{\infty}$ is bounded below by a_1. But it may not be bounded above.

Theorem 11.1.3. *Let $\{a_n\}$ be an increasing (nondecreasing) sequence.*

(a) If $\{a_n\}$ is bounded above by M, then there exists a number $L \leq M$ such that $\lim_{n\to} a_n = L$.

(b) If $\{a_n\}$ is unbounded, then $\lim_{n\to\infty} a_n = \infty$.

Proof. (Exercise) □

Note. The similar results for a decreasing (nonincreasing) sequence hold if replacing "above boundedness" by "below boundedness" in part(a), and "$\lim_{n\to\infty} a_n = \infty$" by "$\lim_{n\to\infty} a_n = -\infty$" in part(b).

Example 11.1.4. Determine whether the sequence $\{a_n\}$ with $a_n = \sqrt{n+1} - \sqrt{n}$ converges or diverges.

Proof. Consider

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

Then $\{a_n\}$ is a decreasing sequence and is bounded below by 0. The sequence is convergent. Moreover, $\lim\limits_{n\to\infty} a_n = \lim\limits_{n\to\infty} \dfrac{1}{\sqrt{n+1} + \sqrt{n}} = 0.$ □

❑ Subsequence

Example 11.1.5. The sequence $\{a_n\}_{n=1}^{\infty}$ with $a_n = (-1)^n$ is a divergent sequence. But if we restrict our attention on the sequence $\{a_{2n}\}_{n=1}^{\infty} = \{1, 1, 1, \cdots\}$, it is a convergent sequence. Conversely, for a given sequence $\{b_n\}_{n=1}^{\infty}$, if we know that $\{b_{2n}\}_{n=1}^{\infty}$ diverges, then the orginal sequence $\{b_n\}_{n=1}^{\infty}$ must be divergent which can be proved by the definition of convergence of a sequence.

It seems that a "subsequence" which comes from a certain sequence, $\{a_n\}$, may give some information of the sequence $\{a_n\}$.

Definition 11.1.6. Let $\{a_n\}_{n=1}^{\infty}$ be a sequence and we say that a sequence $\{b_n\}_{n=1}^{\infty}$ is a subsequence of $\{a_n\}$ if there exists a strictly increasing function $f : \mathbb{N} \to \mathbb{N}$ such that $b_n = a_{f(n)}$. for all $n \in \mathbb{N}$.

Note. Since $f : \mathbb{N} \to \mathbb{N}$ is a strictly increasing function, we can prove that $f(n) \geq n$.

Example 11.1.7. Let $\{a_n\} = \{1, 5, 8, 3, 6, 2, 1, 4, \cdots\}$ be a sequence.

$$\begin{array}{ccccccccc} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 & a_8 & \cdots \\ \shortparallel & \shortparallel & \shortparallel & \shortparallel & \shortparallel & \shortparallel & \shortparallel & \shortparallel & \\ 1 & 5 & 8 & 3 & 6 & 2 & 1 & 4 & \cdots \end{array}$$

If $f(1) = 3,\ f(2) = 5,\ f(3) = 8, \cdots$, then $b_1 = a_3 = 8,\ b_2 = a_5 = 6,\ b_3 = a_8 = 4, \cdots$. The sequence $\{b_n\} = \{8, 6, 4, \cdots\}$ is a subsequence of $\{a_n\}$.

Note. It is customary to write a subsequence $\{b_n\}$ obtained from the sequence $\{a_n\}$ as $b_k = a_{n_k}$, since the terms of the subsequence come from a_n. That is, $f(k) = n_k$ and we write the subsequence as $\{a_{n_k}\}_{k=1}^{\infty}$.

Remark. Let $\{a_{n_k}\}_{k=1}^{\infty}$ be a subsequence of $\{a_n\}_{n=1}^{\infty}$. Then

(i) $\{a_{n_k}\}_{k=1}^{\infty}$ is a sequence.

(ii) $\{a_{n_k} \mid k \in \mathbb{N}\}$ is a subset of $\{a_n \mid n \in \mathbb{N}\}$.

(iii) The order of a_{n_k} follows the order of a_n.

Remark. Let $\{a_n\}$ be a sequence. Then

(i) $\{a_n\}$ is a subsequence of itself .

(ii) If $\{b_n\}$ is a subsequence of $\{a_n\}$ and $\{c_n\}$ is a subsequence of $\{b_n\}$, then $\{c_n\}$ is a subsequence of $\{a_n\}$. That is, a subsequence of a subsuequence of $\{a_n\}$ is also a subsequence of $\{a_n\}$.

Lemma 11.1.8. *Any sequence $\{a_n\}$ contains a subsequence which is either nondecreasing or nonincreasing.*

Proof. We call the positive integer n "*peak point*" if

$$a_n \geq a_m \qquad \text{for all } \ m > n.$$

Case 1: $\{a_n\}$ contains infinitely many peak points, say $\{a_{n_k}\}$. By the definition of peak point,

$$a_{n_1} \geq a_{n_2} \geq a_{n_3} \geq \cdots .$$

Hence, $\{a_{n_k}\}$ is a nonincreasing subsequence of $\{a_n\}$.

Case 2: $\{a_n\}$ contains finitely many peak points. Then there exists a number $N \in \mathbb{N}$ such that for every $n \geq N$, a_n is not a peak point.

Let $a_{n_1} \geq a_N$. Since a_{n_1} is not a peak point, there exists a number $n_2 > n_1$ such that $a_{n_2} > a_{n_1}$. Also, a_{n_2} is not a peak point and hence there exists $n_3 > n_2$ such that $a_{n_3} > a_{n_2}$. Continue the process, we can obtain a sequence $\{a_{n_k}\}$ such that

$$a_{n_1} < a_{n_2} < a_{n_3} < \cdots .$$

Hence, $\{a_{n_k}\}$ is a nondecreasing subsequence of $\{a_n\}$. □

Heuristically, a subsequence $\{a_{n_k}\}$ is a portion of the sequence $\{a_n\}$. If $\{a_n\}$ has a certain property, every subsubsequence of $\{a_n\}$ is supposed to keep this property. But the converse is usually not true unless every subsequence has this property.

■ Bounded Subsequences

Recall that a convergent sequence must be bounded. But the converse is false, that is, a bounded sequence may not be convergent. For example, $\{(-1)^n\}$ is a bounded and divergent sequence. If a sequence has some nice hypothesis, the convergence of this sequence would be obtained. For example, "*a bounded and monotonic sequence is convergent*". In general, we cannot expect that a bounded will be convergent. But we can still get some results of the subsequence of a bounded sequence.

Lemma 11.1.9. *A sequence $\{a_n\}$ is bounded, say $|a_n| < M$ for all $n \in \mathbb{N}$ if and only if every subsequence $\{a_{n_k}\}$ of $\{a_n\}$ is bounded and $|a_{n_k}| < M$ for all $k \in \mathbb{N}$.*

Proof. (Exercise) □

The sequence in Example 11.1.5 is bounded and divergent but it contains one (or some) convergent subsequence. In fact, this phenomenon is true for any bounded sequence.

Theorem 11.1.10. *(Bolzano-Weierstrass Theorem) Every bounded sequence has a convergent subsequence.*

Proof. Let $\{a_n\}$ be a bounded sequence. By Lemma 11.1.8, $\{a_n\}$ contains a monotonic sequence. W.L.O.G, say $\{a_{n_k}\}$ is a nondecreasing subsequence of $\{a_n\}$.

Since $\{a_n\}$ is bounded, $\{a_{n_k}\}$ is bounded above. By the Least Upper Bound property, the set $S = \left\{a_{n_k} \,\middle|\, k \in \mathbb{N}\right\}$ has a least upper bound, say $L = \sup S$. For given $\varepsilon > 0$, there exists $k_0 \in \mathbb{N}$ such that $|a_{n_{k_0}} - L| < \varepsilon$. Since $\{a_{n_k}\}$ is nondecreasing, for every $k \geq k_0$, $|a_{n_k} - L| \leq |a_{n_{k_0}} - L| < \varepsilon$. Hence $\{a_{n_k}\}$ converges to L.

□

Theorem 11.1.11. *If $\{a_n\}$ converges to L if and only if every subsequence of $\{a_n\}$ converges to L.*

Proof. ($\Longrightarrow$) If $\{a_n\}$ converges to L, for given $\varepsilon > 0$, there exists $N_0 \in \mathbb{N}$ such that

$$|a_n - L| < \varepsilon \qquad \text{whenever} \quad n > N_0.$$

Let $\{a_{n_k}\}$ be a subsequence of $\{a_n\}$. Then $n_1 < n_2 < n_3 < \cdots$ and $n_k \geq k$ for every $k \in \mathbb{N}$. Thus there exists $K_0 \in \mathbb{N}$ such that $n_k \geq N_0$ for every $k \geq K_0$. We have

$$|a_{n_k} - L| < \varepsilon \qquad \text{whenever} \quad k \geq K_0.$$

This implies that the subsequence $\{a_{n_k}\}$ converges to L.

($\Longleftarrow$) This direction is clear since $\{a_n\}$ is a subsequence of itself.

□

Remark. There exists a divergent sequence contains convergent subsequences.

Example 11.1.12. For example $a_n = \begin{cases} 1 & \text{if } n \text{ is odd} \\ \dfrac{1}{n} & \text{if } n \text{ is even} \end{cases}$ and $a_{n_k} = a_{2k} = \dfrac{1}{2k}$. Then the sequence $\{a_n\}$ diverges but the subsequence $\{a_{n_k}\}$ converges to 0.

Remark. The importance of Theorem 11.1.11 is that if one subsequence of $\{a_n\}$ diverges, then so does $\{a_n\}$. Hence, if we can prove that two subsequences of $\{a_n\}$ converge to different values, then $\{a_n\}$ diverges.

❑ Cauchy Sequence

Theorem 11.1.13. *If $\{a_n\}$ converges, then for given $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that*

$$|a_m - a_n| < \varepsilon$$

for every $m, n > N$.

Proof. Since $\{a_n\}$ converges, say $\lim_{n\to\infty} a_n = L$, for given $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$|a_n - L| < \frac{\varepsilon}{2}$$

for every $n > N$. Hence, for $m, n > N$,

$$|a_m - a_n| \le |a_m - L| + |L - a_n| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

□

Remark. When we consider the sequence in the real-valued field $\mathbb{R}$, the converse of Theorem 11.1.13 is also true. We will prove this statement later. However, if we discuss a sequence in a different field (for example, rational number field $\mathbb{Q}$), Theorem 11.1.13 still holds but the converse will not be true.*

Definition 11.1.14. A sequence $\{a_n\}$ is called a "*Cauchy sequence*" if for every $\varepsilon > 0$ there exists a number $N \in \mathbb{N}$ such that

$$|a_m - a_n| < \varepsilon$$

for every $m, n > N$.

Lemma 11.1.15. *A Cauchy sequence is bounded.*

Proof. Let $\{a_n\}$ be a Cauchy sequence. Given $\varepsilon = 1$, there exists $N \in \mathbb{N}$ such that for every $m, n > N$,

$$|a_m - a_n| < 1.$$

Hence, for every $n > N$, $|a_n - a_{N+1}| < 1$.

Let $M = \max\big(|a_1|,\ |a_2|, \cdots,\ |a_N|,\ |a_{N+1}| + 1\big)$. Then

$$|a_n| \le M \qquad \text{for every } n \in \mathbb{N}.$$

Therefore, $\{a_n\}$ is bounded. □

Theorem 11.1.16. *A sequence $\{a_n\}$ converges (in $\mathbb{R}$) if and only if $\{a_n\}$ is a Cauchy sequence.*

Proof.

($\Longrightarrow$) This direction is proved in Theorem 11.1.13.
($\Longleftarrow$) Since $\{a_n\}$ is a Cauchy sequence, it is bounded. By Bolzano-Weierstrass Theorem, $\{a_n\}$ contains a convergent subsequence $\{a_{n_k}\}$, say $\lim_{k\to\infty} a_{n_k} = L$. We claim that the sequence $\{a_n\}$ converges to L.

Since the subsequence $\{a_{n_k}\}$ converges to L, for given $\varepsilon > 0$, there exists $K \in \mathbb{N}$ such that for every $k > K$

$$|a_{n_k} - L| < \frac{\varepsilon}{2}.$$

*If a set has the property that every Cauchy sequence converges, we called the set "*complete*" and we will discuss it in the future.

Since $\{a_n\}$ is a Cauchy sequence, there exists $N \in \mathbb{N}$ such that for every $m, n > N$

$$|a_m - a_n| < \frac{\varepsilon}{2}.$$

Choose a sufficiently large number $k_0 \in \mathbb{N}$ such that $k_0 > K$ and $n_{k_0} > N$. Then for every $n > N$,

$$|a_n - L| \le |a_n - a_{n_{k_0}}| + |a_{n_{k_0}} - L| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Hence, the sequence $\{a_n\}$ converges to L. □

11.2 Infinite Series

Every real number can be expressed as a digital number. Especially, most numbers have the expression of infinite deciamls. For example,

$$\begin{aligned} \pi &= 3.1415926\ldots \\ &= \underbrace{3}_{a_1} + \underbrace{\frac{1}{10}}_{a_2} + \underbrace{\frac{4}{10^2}}_{a_3} + \underbrace{\frac{1}{10^3}}_{a_4} + \underbrace{\frac{5}{10^4}}_{a_5} + \underbrace{\frac{9}{10^5}}_{a_6} + \underbrace{\frac{2}{10^6}}_{a_7} + \underbrace{\frac{6}{10^7}}_{a_8} + \cdots \\ &= a_1 + a_2 + a_3 + \cdots \end{aligned}$$

Heuristically, for a given sequence $\{a_n\}$, we want to consider whether the sum of all terms makes sense. But, in mathematics, adding infinite numbers is not doable. Hence, the sum

$$a_1 + a_2 + a_3 + \cdots + a_n + \cdots$$

does not make sense.

Question: How to define the sum of infinite numbers (terms)?

Consider the "*partial sum*" of $\{a_n\}$

$$\begin{aligned} s_1 &= a_1 && \text{(first partial sum)} \\ s_2 &= a_1 + a_2 && \text{(second partial sum)} \\ s_3 &= a_1 + a_2 + a_3 && \text{(third partial sum)} \\ &\vdots \\ s_n &= a_1 + a_2 + \cdots + a_n = \sum_{k=1}^{n} a_k && \text{(nth partial sum).} \end{aligned}$$

Then, for every $n \in \mathbb{N}$, s_n is well-defined and $\{s_n\}_{n=1}^{\infty}$ forms a new sequence. Suppose that sum of the infinite terms of $\{a_n\}$ is well-defined. It is supposed to be the limit of $\{s_n\}$.

Definition 11.2.1. We say that a sequence $\{a_n\}$ is "*summable*" if the sequence $\{s_n\}$ converges. The symbol $\sum_{n=1}^{\infty} a_n$ denotes the limit $\lim_{n\to\infty} s_n$ and we call it the sum of the sequence $\{a_n\}$. If $\lim_{n\to\infty} s_n = s$, we write $\sum_{n=1}^{\infty} a_n = s$.

Remark.

(i) $\displaystyle\sum_{n=1}^{\infty} a_n$ is usually called an "*infinite series*".

(ii) The statement "$\{a_n\}$ is summable" is conventionally replaced by the statement "$\displaystyle\sum_{n=1}^{\infty} a_n$ converges". That is,

$$\text{"the series } \sum_{n=1}^{\infty} a_n \text{ converges"} \quad \text{if and only if} \quad \text{"the sequence } \{s_n\}_{n=1}^{\infty} \text{ converges".}$$

(iii) If the sequence $\{s_n\}$ is divergent, we say that the series $\displaystyle\sum_{n=1}^{\infty} a_n$ is divergent.

Example 11.2.2.

(1) Let $a_n = \dfrac{1}{2^n}$. Then $s_n = \dfrac{1}{2} + \dfrac{1}{2^2} + \cdots + \dfrac{1}{2^n} = 1 - \dfrac{1}{2^n}$.

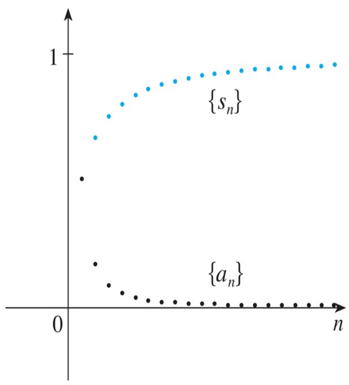
n	Sum of first n terms
1	0.50000000
2	0.75000000
3	0.87500000
4	0.93750000
5	0.96875000
6	0.98437500
7	0.99218750
10	0.99902344
15	0.99996948
20	0.99999905
25	0.99999997

Hence,

$$\sum_{n=1}^{\infty} a_n = \lim_{n\to\infty} \sum_{k=1}^{n} a_k = \lim_{n\to\infty} s_n = \lim_{n\to\infty} 1 - \frac{1}{2^n} = 1.$$

(2) (Telescoping series) Let $a_n = \dfrac{1}{n(n+1)}$. Then

$$\begin{aligned} s_n &= \sum_{k=1}^{n} \frac{1}{k(k+1)} = \frac{1}{1\cdot 2} + \frac{1}{2\cdot 3} + \cdots + \frac{1}{n(n+1)} \\ &= (1 - \cancel{\tfrac{1}{2}}) + (\cancel{\tfrac{1}{2}} - \cancel{\tfrac{1}{3}}) + \cdots + (\cancel{\tfrac{1}{n}} - \frac{1}{n+1}) \\ &= 1 - \frac{1}{n+1} \end{aligned}$$



Since $\displaystyle\lim_{n\to\infty} s_n = \lim_{n\to\infty} 1 - \frac{1}{n+1} = 1$, the series $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = 1$. The series $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n(n+1)}$ is convergent and the sequence $\left\{\dfrac{1}{n(n+1)}\right\}$ is summable.

(3) Let $a_n = (-1)^n$. Then

$$\begin{aligned} s_{2n} &= (-1) + 1 + (-1) + 1 + \cdots + 1 = 0 \\ s_{2n+1} &= (-1) + 1 + (-1) + 1 + \cdots + 1 + (-1) = -1 \end{aligned}$$

Hence, the limit $\lim_{n\to\infty} s_n$ does not exist. (That is, $\{s_n\}$ is divergent, $\{(-1)^n\}$ is not summable or $\sum_{n=1}^{\infty}(-1)^n$ is divergent.)

■ Geometric Series

A geometric series with ratio r is a series of the form

$$\sum_{n=0}^{\infty} ar^n = a + ar + ar^2 + \cdots + ar^n + \cdots, \qquad a \neq 0$$

Note: The series starts with the 0th term rather than 1st term.

(1) For $r = 1$, $s_n = \underbrace{a + a + \cdots + a}_{n} = na \to \pm\infty$ as $n \to \infty$. Hence $\lim_{n\to\infty} s_n$ is divergent.

(2) For $r \neq 1$,

$$\begin{aligned} s_n &= a + ar + \cdots + ar^n \\ rs_n &= \phantom{a + {}} ar + \cdots + ar^n + ar^{n+1} \end{aligned}$$

We have $(r-1)s_n = a(r^{n+1} - 1)$ and hence

$$s_n = \frac{a(r^{n+1}-1)}{r-1}.$$

Consider the limit $\lim_{n\to\infty} s_n = \lim_{n\to\infty} \frac{a(r^{n+1}-1)}{r-1}$ provided $r \neq 1$.

(i) If $|r| < 1$, then $\lim_{n\to\infty} r^{n+1} = 0$. Hence, $\sum_{n=0}^{\infty} ar^n = \lim_{n\to\infty} s_n = \frac{a}{1-r}$.

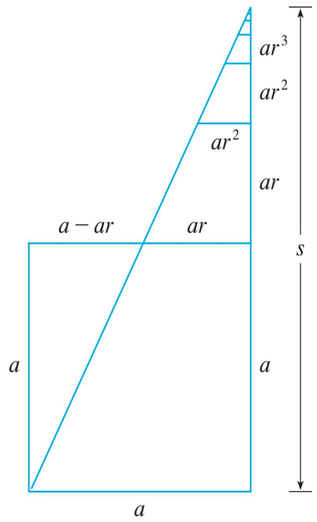
(ii) If $|r| > 1$, then $\lim_{n\to\infty} r^{n+1}$ diverges. Hence, $\sum_{n=0}^{\infty} ar^n = \lim_{n\to\infty} s_n$ diverges.

(iii) If $r = -1$, $s_n = a - a + a - a + \cdots + (-1)^{n-1}a = \begin{cases} 0 & n \text{ is even} \\ a & n \text{ is odd.} \end{cases}$ Hence, $\sum_{n=0}^{\infty} ar^n = \lim_{n\to\infty} s_n$ diverges.

Conclusion: The geometric series $\sum_{n=0}^{\infty} ar^n$, $a \neq 0$

(1) converges if $|r| < 1$ and $\sum_{n=0}^{\infty} ar^n = \frac{1}{1-r}$.

(2) diverges if $|r| \geq 1$.



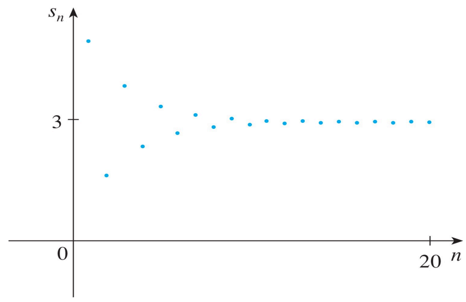
In the figure, $\frac{s}{a} = \frac{a}{a-ar}$. Then $s = \frac{a}{1-r}$.

Example 11.2.3.

(1) Evaluate $5 - \frac{10}{3} + \frac{20}{9} - \frac{40}{27} + \cdots$.

Proof. For the series, the first term $a = 5$ and the ratio $r = -\frac{2}{3}$. Since $|r| = |-\frac{2}{3}| = \frac{2}{3} < 1$, the series is convergent and

$$\sum_{n=0}^{\infty} 5\left(-\frac{2}{3}\right)^n = \frac{5}{1-(-\frac{2}{3})} = 3.$$



□

(2) Evaluate $\sum_{n=0}^{\infty} 2 \cdot \left(\frac{5}{3}\right)^n$.

Proof. Since the ratio of the geometric series is $r = \frac{5}{3} > 1$. The series is divergent. □

(3) Write $0.1232323\cdots = 0.1\overline{23}$ as a ratio of integer.

Proof.

$$\begin{aligned}
0.1\overline{23} &= 0.1 + 0.023 + 0.00023 + 0.0000023 + \cdots \\
&= \frac{1}{10} + \frac{23}{10^3} + \frac{23}{10^5} + \frac{23}{10^7} + \cdots \\
&= \frac{1}{10} \cdot \frac{23}{10^3}\Big(\underbrace{1}_{a} + \underbrace{\frac{1}{10^2}}_{r} + \frac{1}{10^4} + \cdots\Big) \\
&= \frac{1}{10} + \frac{23}{10^3} \cdot \frac{1}{1 - \frac{1}{10^2}} \\
&= \frac{122}{99}.
\end{aligned}$$

□

(4) Find the sum of the series $\sum_{n=0}^{\infty} x^n$, where $|x| < 1$.

Proof.

$$\sum_{n=0}^{\infty} x^n + 1 + x + x^2 + x^3 + \cdots .$$

The first term of the series is $a = 1$ and the ratio $r = x$ with $|r| = |x| < 1$. Hence, the series is convergent and $\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$. □

■ Harmonic Series

A harmonic series has the form

$$\sum_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots .$$

We claim that $\sum_{n=1}^{\infty} \frac{1}{n} = \infty$. It sufficies to show that for any number $M > 0$, $\sum_{n=1}^{\infty} \frac{1}{n} > M$. Consider

$$\begin{aligned}
\sum_{n=1}^{2^k} \frac{1}{n} &= 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots + \frac{1}{8} + \cdots + \frac{1}{16} + \cdots + \frac{1}{2^k} \\
&> 1 + \frac{1}{2} + \left(\frac{1}{4} + \frac{1}{4}\right) + \left(\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}\right) + \Big(\underbrace{\frac{1}{16} + \cdots + \frac{1}{16}}_{8 \text{ times}}\Big) + \cdots \\
&\quad + \Big(\underbrace{\frac{1}{2^k} + \cdots + \frac{1}{2^k}}_{2^{k-1} \text{ times}}\Big) \\
&> 1 + \underbrace{\frac{1}{2} + \frac{1}{2} + \cdots + \frac{1}{2}}_{k \text{ times}} \\
&= 1 + \frac{k}{2}
\end{aligned}$$

Choose $k > 2M$. Then $\sum_{n=1}^{2^k} \frac{1}{n} > M$. Hence, $\sum_{n=1}^{\infty} \frac{1}{n} > \sum_{n=1}^{2^k} \frac{1}{n} > M$. Since M is an arbitrary positive number,

$$\sum_{n=1}^{\infty} \frac{1}{n} = \infty.$$

❏ Laws of Series

Theorem 11.2.4. *If $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ are convergent series and c is a constant. Then*

(1) $\sum_{n=1}^{\infty} (a_n \pm b_n)$ converges and $\sum_{n=1}^{\infty} (a_n \pm b_n) = \sum_{n=1}^{\infty} a_n \pm \sum_{n=1}^{\infty} b_n$.

(2) $\sum_{n=1}^{\infty} (ca_n)$ converges and $\sum_{n=1}^{\infty} (ca_n) = c \sum_{n=1}^{\infty} a_n$.

Remark. The result of Theorem 11.2.4 is false if one of the series $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ is divergent.

Example 11.2.5. Evaluate $\sum_{n=1}^{\infty} \left[\frac{3}{n(n+1)} + \frac{1}{2^n}\right]$.

Proof. Since $\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = 1$ (converges), we have $\sum_{n=1}^{\infty} \frac{3}{n(n+1)} = 3\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = 3$. For the series $\sum_{n=1}^{\infty} \frac{1}{2^n}$, it is a geometric series with the first term $a = \frac{1}{2}$ and the ratio $r = \frac{1}{2}$. Then it converges and $\sum_{n=1}^{\infty} \frac{1}{2^n} = \frac{\frac{1}{2}}{1-\frac{1}{2}} = 1$. Hence,

$$\sum_{n=1}^{\infty} \left[\frac{3}{n(n+1)} + \frac{1}{2^n}\right] = \sum_{n=1}^{\infty} \frac{3}{n(n+1)} + \sum_{n=1}^{\infty} \frac{1}{2^n} = 3 + 1 = 4.$$

□

11.3 Test for Divergence

For most series, it is difficult to find their limit even if they have nice patterns. Therefore, we usually don't expect to compute the exact limit of a convergent series. Instead of this, we want to study some tests for convergence or divergence of a series and estimate their limits if they converge in the present and next sections.

Theorem 11.3.1. *(Cauchy Criterion) Let $\{a_n\}$ be a sequence and $\{s_n\}$ be the sequence of partial sums of $\{a_n\}$. Then the following statements are equivalent.*

(a) $\{a_n\}$ is summable. ($\sum_{n=1}^{\infty} a_n$ is convergent.)

(b) $\{s_n\}$ converges.

(c) $\lim_{m,n\to\infty} s_n - s_m = 0$.

(d) $\lim_{m,n\to\infty} a_{n+1} + a_{n+2} + \cdots + a_m = 0$.

Proof. (Exercise) □

Theorem 11.3.2. *If the series $\sum_{n=1}^{\infty} a_n$ is convergent, then $\lim_{n\to\infty} a_n = 0$.*

Proof. Since $\sum_{n=1}^{\infty} a_n$ is convergent, the sequence $\{s_n\}$ is convergent. Consider $a_n = s_n - s_{n-1}$. Then

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty}(s_n - s_{n-1}) = \lim_{n\to\infty} s_n - \lim_{n\to\infty} s_{n-1} = 0.$$

□

Remark.

(i) The converse of Theorem 11.3.2 is false. That is, even if $\lim_{n\to\infty} a_n = 0$, it cannot imply that the series $\sum_{n=1}^{\infty} a_n$ converges. For example, $a_n = \frac{1}{n}$. Then $a_n \to 0$ as $n \to \infty$ but $\sum_{n=1}^{\infty} a_n = \infty$.

(ii) That $\sum_{n=1}^{\infty} a_n$ diverges cannot imply $\lim_{n\to\infty} a_n \neq 0$. (For example, $a_n = \frac{1}{n}$.)

■ **Test for Divergence**

Theorem 11.3.3. *(Test for Divergence) If* $\lim_{n\to\infty} a_n$ *does not converge to* 0 *(either* $\lim_{n\to\infty} a_n$ *DNE or* $\lim_{n\to\infty} a_n = L \neq 0$*), then the series* $\sum_{n=1}^{\infty} a_n$ *is divergent.*

Example 11.3.4. Determine whether the series $\sum_{n=1}^{\infty} \frac{n^2}{5n^2+4}$ is convergent or divergent.

Proof. Consider the limit

$$\lim_{n\to\infty} \frac{n^2}{5n^2+4} = \lim_{n\to\infty} \frac{1}{5+\frac{4}{n^2}} = \frac{1}{5} \neq 0.$$

By the test for divergence, the series $\sum_{n=1}^{\infty} \frac{n^2}{5n^2+4}$ is divergent. □

Remark. In Chapter 2, we understand that, for a sequence $\{a_n\}$, a finite number of terms of $\{a_n\}$ doesn't affect the convergence or divergence of the sequence. A series has similar results. If we only concern whether a series $\sum a_n$ is convergent or divergent (but not the exact value of the sereis), the sum of a finite number terms does not change its convergence or divergence. That is, for any number $n_0 \in \mathbb{N}$, the series $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=n_0}^{\infty} a_n$ both converg or both diverge.

11.4 Tests for Convergence

So far, we can compute the sum of some special series (for example, the geometric series with ratio $|r| < 1$, $\sum \frac{1}{n(n+1)}$ etc). But even for a simple series, like $\sum \frac{1}{n^2}$, it is not easy to find its sum since the formula of its partial sum is difficult to be obtained.

In the present section, we will introduce some methods to determine whethere a series is convergent. First of all, we consider the sequence $\{a_n\}$ whose all terms have the same sign. Because of this, its partial sum $\{s_n\}$ is a monotonic sequence and we can use the bounded criterion to determine whether the sequence of partial sum is convergent or not.

Definition 11.4.1. We say that $\{a_n\}$ is a "*nonnegative sequence*" ("*nonpositive sequence*") if $a_n \geq 0$ ($a_n \leq 0$) for all $n \in \mathbb{N}$.

Remark. If $\{a_n\}$ is a nonnegative (nonpositive) sequence, then the sequence of the partial sum $\{s_n\}$ is a nondecreasing (nonincreasing) sequence.

Theorem 11.4.2. *(Bounded Criterion) A nonnegative sequence* $\{a_n\}$ *is summable if and only if the sequence of partial sum* $\{s_n\}$ *is bounded (above).*

Proof. (Exercise) □

11.4.1 The Integral Test

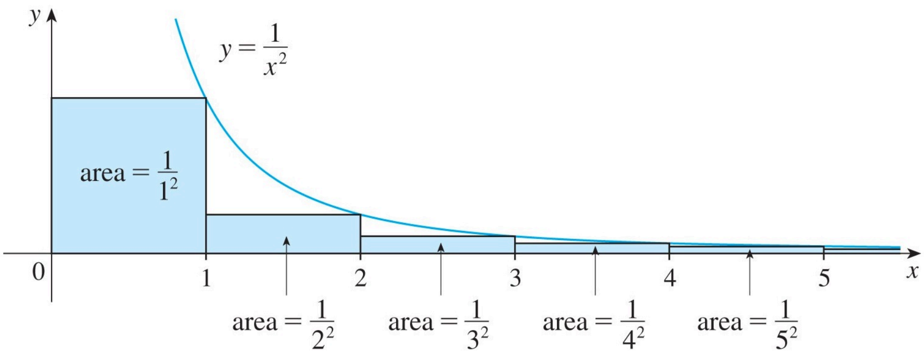
Observe the series $\sum_{n=1}^{\infty} \frac{1}{n^2}$. The nth partial sum

$$s_n = \sum_{k=1}^{n} \frac{1}{k^2} = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2}$$

is an increasing sequence. To determine whether the sequence $\{s_n\}_{n=1}^{\infty}$ converges, it sufficies to show that the seqnence is bounded above since it is increasing. Let's consider the function $f(x) = \frac{1}{x^2}$ on $[1, \infty)$. We have

$$s_n = \sum_{k=1}^{n} \frac{1}{k^2} = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2} < 1 + \int_1^n \frac{1}{x^2}\, dx < 1 + \int_1^{\infty} \frac{1}{x^2}\, dx = 2.$$

Hence, $\{s_n\}$ is bounded above (by 2). By the bounded criterion, the series $\sum_{n=1}^{\infty} \frac{1}{n^2}$ is convergent.



n	$s_n = \sum_{i=1}^{n} \frac{1}{i^2}$
5	1.4636
10	1.5498
50	1.6251
100	1.6350
500	1.6429
1000	1.6439
5000	1.6447

Also, observe another series $\sum_{n=1}^{\infty} \frac{1}{\sqrt{n}}$. The nth partial sum

$$s_n = \sum_{k=1}^{n} \frac{1}{\sqrt{k}} = 1 + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \cdots + \frac{1}{\sqrt{n}}$$

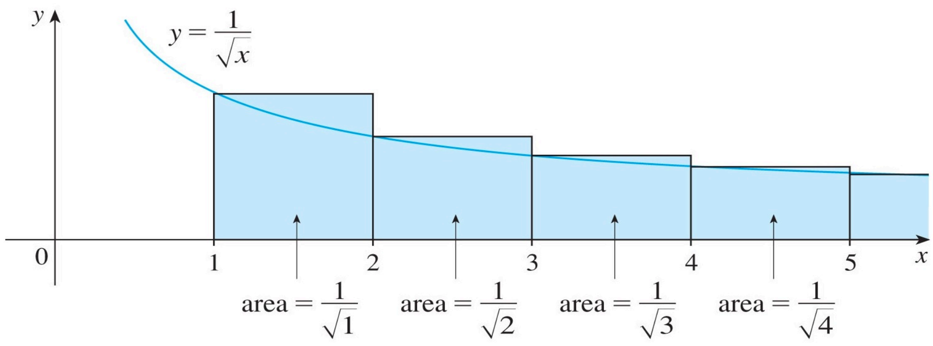
is an increasing sequence. Consider the function $f(x) = \frac{1}{\sqrt{x}}$ on $[1, \infty)$. We have

$$s_n = \sum_{k=1}^{n} \frac{1}{\sqrt{k}} = 1 + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \cdots + \frac{1}{\sqrt{n}} > \int_1^{n-1} \frac{1}{\sqrt{x}}\, dx = 2\sqrt{n-1} - 1.$$

Then

$$\lim_{n\to\infty} s_n \geq \lim_{n\to\infty} (2\sqrt{n-1} - 1) = \infty.$$

and the series $\sum_{n=1}^{\infty} \frac{1}{\sqrt{n}}$ is divergent.



n	$s_n = \sum_{i=1}^{n} \frac{1}{\sqrt{i}}$
5	3.2317
10	5.0210
50	12.7524
100	18.5896
500	43.2834
1000	61.8010
5000	139.9681

Theorem 11.4.3. *(Integral Test) Suppose that f is a positive and decreasing function on $[1, \infty)$ and $f(n) = a_n$. Then*

$$\sum_{n=1}^{\infty} a_n \quad \textit{converges} \quad \textit{if and only if} \quad \int_1^{\infty} f(x)\,dx \quad \textit{converges}.$$

That is, the series $\sum_{n=1}^{\infty} a_n$ and the improper integral $\int_1^{\infty} f(x)\,dx$ either both converge or both diverge.

Proof. Since f is decreasing, for every $k \in \mathbb{N}$,

$$f(k+1) \cdot 1 \leq \int_k^{k+1} f(x)\,dx \leq f(k) \cdot 1.$$

Since f is positive, for every $n \in \mathbb{N}$,

$$0 \leq \underbrace{\sum_{k=1}^{n-1} a_{k+1}}_{s_n - a_1} = \sum_{k=1}^{n-1} f(k+1) \leq \underbrace{\sum_{k=1}^{n-1} \int_k^{k+1} f(x)\,dx}_{\int_1^n f(x)\,dx} \leq \sum_{k=1}^{n-1} f(k) = \underbrace{\sum_{k=1}^{n-1} a_k}_{s_{n-1}}.$$

Hence,

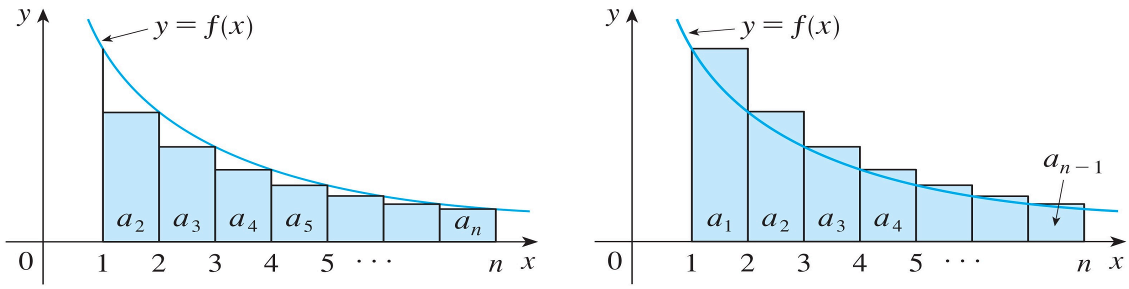
$$\sum_{n=2}^{\infty} a_n \leq \int_1^{\infty} f(x)\,dx \leq \sum_{n=1}^{\infty} a_n.$$

This inequality implies that $\sum_{n=1}^{\infty} a_n$ and $\int_1^{\infty} f(x)\,dx$ either both converge or both diverge.

□

Remark.

(i) To determine whether a series is convergent or divergent, it is not necessary to start with the first term. That is, the series $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=n_0}^{\infty} a_n$ either both converge or both diverge.



Hence, to use the integral test, it sufficies to compute the integral with lower limit at $x = n_0$ instead of $x = 1$. That is,

$$\int_{n_0}^{\infty} f(x)\,dx \quad \text{converges (diverges)} \iff \sum_{n=n_0}^{\infty} a_n \quad \text{converges (diverges)}$$
$$\iff \sum_{n=1}^{\infty} a_n \quad \text{converges (diverges).}$$

(ii) It is not necessary that f is "always" decreasing. We can use the integral test as long as the function f is positive and decreasing on (n_0, ∞) and $f(n) = a_n$ for some large number n_0 and $n \geq n_0$.

Example 11.4.4. Determine whether the series $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^2+1}$ is convergent or divergent.

Proof. The function $f(x) = \dfrac{1}{x^2+1}$ is positive and decreasing on $[1, \infty)$. Also, $f(n) = \dfrac{1}{n^2+1}$ for all $n \in \mathbb{N}$. Since the improper integral

$$\int_1^{\infty} \frac{1}{x^2+1}\,dx = \lim_{t\to\infty} \int_1^t \frac{1}{x^2+1}\,dx = \lim_{t\to\infty} \tan^{-1} x\Big|_1^t = \lim_{t\to\infty} \left(\tan^{-1} t - \tan^{-1} 1\right) = \frac{\pi}{2} - \frac{\pi}{4} = \frac{\pi}{4},$$

by the integral test, the series $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^2+1}$ converges. □

Example 11.4.5. (p-series) For what values of p is the series $\dfrac{1}{n^p}$ convergent?

Proof. If $p \leq 0$, $\dfrac{1}{n^p} = n^{-p} \geq 1$ for all $n \in \mathbb{N}$. Hence $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^p}$ diverges.

Consider the cases $0 < p < \infty$. The function $f(x) = \dfrac{1}{x^p}$ is positive and decreasing on $[1, \infty)$, and $f(n) = \dfrac{1}{n^p}$. Since

$$\int_1^{\infty} \frac{1}{x^p}\,dx = \begin{cases} \infty & \text{when } 0 < p < 1 \quad \text{(divergent)} \\ \frac{1}{p-1} & \text{when } p > 1 \quad \text{(convergent).} \end{cases}$$

By the integral test, the series $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^p}$ converges when $p > 1$ and diverges when $p \leq 1$. □

Example 11.4.6.

(i) $\sum_{n=1}^{\infty} \frac{1}{n^3}$ converges (p-series with $p = 3 > 1$)

(ii) $\sum_{n=1}^{\infty} \frac{1}{n^{1/3}}$ diverges (p-series with $p = \frac{1}{3} < 1$)

Note. The integral test can only determine whether a series is convergent (or divergent). But it cannot give the sum of the series.

Example 11.4.7. Determine whether the series $\sum_{n=1}^{\infty} \frac{\ln n}{n}$ converges or diverges.

Proof. Let $f(x) = \frac{\ln x}{x}$. Then $f'(x) = \frac{1 - \ln x}{x^2} < 0$ when $x > e$. Hence, $f(x)$ is positive and decreasing on (e, ∞). Since the integral

$$\int_e^{\infty} \frac{\ln x}{x}\, dx = \lim_{t\to\infty} \int_e^t \frac{\ln x}{x}\, dx = \lim_{t\to\infty} \frac{(\ln x)^2}{2}\Big|_e^t = \lim_{t\to\infty} \frac{(\ln t)^2 - 1}{2} = \infty,$$

by the integarl test, the series $\sum_{n=1}^{\infty} \frac{\ln n}{n}$ diverges. □

■ Estimating the Sum of a Series

Although it is difficult to use the integral test to find the limit of a series $\sum a_n$, it can still help us to approximate the sum of the series. Recall that "$s = \sum_{n=1}^{\infty} a_n$ converges" means that the partial sum $s_n = \sum_{k=1}^{n} a_k \to s$ as $n \to \infty$. Hence, in order to evaluate the sum s, we want to estimate the differenece s_n and s. Define

$$R_n = s - s_n = a_{n+1} + a_{n+2} + \cdots = \sum_{k=n+1}^{\infty} a_k \qquad \text{as the "remainder".}$$

Theorem 11.4.8. *(Remainder Estimate for the Integral Test) Let f be a positive and decreasing function for every $x \geq n_0$, and $f(n) = a_n$ for every $n \in \mathbb{N}$ and $n \geq n_0$. Then*

$$\int_{n+1}^{\infty} f(x)\, dx \leq \sum_{k=n+1}^{\infty} a_k = R_n \leq \int_n^{\infty} f(x)\, dx$$

Note.

$$s_n + \int_{n+1}^{\infty} f(x)\, dx \leq s \leq s_n + \int_n^{\infty} f(x)\, dx.$$

Example 11.4.9.

(a) Approximate the sum of the series $\sum_{n=1}^{\infty}\frac{1}{n^3}$ by using the sum of the first 10 terms. Estimate the error involved in the approximation.

Proof. Let $f(x) = \frac{1}{x^3}$. Then $\int_n^{\infty} \frac{1}{x^3}\,dx = \frac{1}{2n^2}$ and

$$R_{10} \le \int_{10}^{\infty} \frac{1}{x^3}\,dx = \frac{1}{200}.$$

□

(b) How many terms are required to ensure that the sum is accurate to within 0.0005?

Proof. Consider

$$R_n \le \int_n^{\infty} \frac{1}{x^3}\,dx = \frac{1}{2n^2} \le 0.0005.$$

Then $n^2 \ge 1000$ and hence $n \ge 31.6$. We need 32 terms to ensure accuracy to within 0.0005. □

(c) Use $n = 10$ to estimate the sum of the series $\sum_{n=1}^{\infty}\frac{1}{n^3}$.

Proof.

$$s_{10} + \frac{1}{2(11)^2} = \sum_{n=1}^{10}\frac{1}{n^3} + \int_{11}^{\infty}\frac{1}{x^3}\,dx \le s \le \sum_{n=1}^{10}\frac{1}{n^3} + \int_{10}^{\infty}\frac{1}{x^3}\,dx = s_{10} + \frac{1}{2(10)^2}.$$

Since $s_{10} \approx 1.197532$, we have $1.201664 \le s \le 1.202532$. □

Note. In fact, to make the error smaller than 0.0005, it only needs 10 terms by part(c) instead of 32 terms by part(b).

11.4.2 The Comparison Test

In Section 11.3, we know that the geometric series $\sum_{n=1}^{\infty}\frac{1}{2^n}$ is convergent.

Question: Does it say the convergence or divergence of the series $\sum_{n=1}^{\infty}\frac{1}{2^n+1}$?

Observe that the sequence of the partial sum $s_n = \sum_{k=1}^{n}\frac{1}{2^k+1}$ is an increasing sequence. Since $0 < \frac{1}{2^k+1} < \frac{1}{2^k}$ for every $k \in \mathbb{N}$, we have

$$s_n = \sum_{k=1}^{n}\frac{1}{2^k+1} \le \sum_{k=1}^{n}\frac{1}{2^k} \le \sum_{k=1}^{\infty}\frac{1}{2^k} = 1.$$

Hence, $\{s_n\}$ is bounded above. By the bounded criterion, the series $\sum_{n=1}^{\infty} \frac{1}{2^k+1}$ converges. Moreover, $\sum_{n=1}^{\infty} \frac{1}{2^k+1} < 1$.

Heuristically, we may have the insight of two nonnegative series.

(i) If every term of one series is smaller than the corresponding term of another convergent series, then the former series is also convergent.

(ii) If every term of one series is larger than the corresponding term of another divergent series, then the former series is also divergent.

■ The Comparison Test

Theorem 11.4.10. *(The Comparision Test) Suppose that* $\sum_{n=1}^{\infty} a_n$ *and* $\sum_{n=1}^{\infty} b_n$ *are series with nonnegative terms and* $0 \le b_n \le a_n$ *for all* $n \in \mathbb{N}$.

(1) If $\sum_{n=1}^{\infty} a_n$ *is convergent, then* $\sum_{n=1}^{\infty} b_n$ *is convergent.*

(2) If $\sum_{n=1}^{\infty} b_n$ *is divergent, then* $\sum_{n=1}^{\infty} a_n$ *is divergent.*

Proof. Let $s_n = a_1 + a_2 + \cdots + a_n$ and $t_n = b_1 + b_2 + \cdots + b_n$. Then the sequences $\{s_n\}$ and $\{t_n\}$ are increasing and $0 \le t_n \le s_n$ for every $n \in \mathbb{N}$.

(1) If $\sum_{n=1}^{\infty} a_n$ is convergent, $\{s_n\}$ is convergent. Since $\{t_n\}$ is increasing and bounded above, it is convergent and thus $\sum_{n=1}^{\infty} b_n$ is convergent.

(2) If $\sum_{n=1}^{\infty} b_n$ is divergent, then $\lim_{n\to\infty} t_n = \infty$. Therefore, $\lim_{n\to\infty} s_n = \infty$ and thus $\sum_{n=1}^{\infty} a_n$ is divergent.

□

Remark.

(i) In order to use the Comparison Test, the "nonnegative" condition is necessary. For example, $b_n = -1$ and $a_n = \frac{1}{n^2}$ for all $n \in \mathbb{N}$. Then $b_n < a_n$. But the series $\sum_{n=1}^{\infty} b_n = \sum_{n=1}^{\infty}(-1) = -\infty$ is divergent and the series $\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} \frac{1}{n^2}$ is convergent.

(ii) In the use of the Comparsion Test, we need to know some convergent or divergent series. Some important series are:

- p-series $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^p}$ $\begin{cases} \text{converges when } p > 1 \\ \text{diverges when } p \leq 1 \end{cases}$
- geometric series $\displaystyle\sum_{n=1}^{\infty} ar^n$ $\begin{cases} \text{converges when } |r| < 1 \\ \text{diverges when } |r| \geq 1 \end{cases}$

Example 11.4.11. Determine whether the series $\displaystyle\sum_{n=1}^{\infty} \frac{5}{2n^2+4n+3}$ is convergent or divergent.

Proof. That the series $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^2}$ is convergent (p-series, $p = 2$) implies the series $\displaystyle\sum_{n=1}^{\infty} \frac{5}{2n^2}$ is also convergent. Since $\dfrac{5}{2n^2+4n+3} < \dfrac{5}{2} \cdot \dfrac{1}{n^2}$ for every $n \in \mathbb{N}$, by the Comparison Test, the series $\displaystyle\sum_{n=1}^{\infty} \frac{5}{2n^2+4n+3}$ is convergent.

□

Remark. To determine whether a series is convergent, it sufficies to consider the convergence of the "tail " ($\displaystyle\sum_{n=n_0}^{\infty} a_n$) of the series. Therefore, in the use of the Comparison Test, we can replace the condition $0 \leq b_n \leq a_n$ "$\underline{\text{for every } n \geq 1}$" by "$\underline{\text{ for every } n \geq n_0}$" and for some integer n_0, and the test still holds.

Example 11.4.12. Determine whether the series $\displaystyle\sum_{n=1}^{\infty} \frac{\ln n}{n}$ is convergent or divergent.

Proof. Since $\ln n > 1$ for $n > e$, we have $\dfrac{\ln n}{n} > \dfrac{1}{n}$ when $n \geq 3$. Also, the series $\displaystyle\sum_{n=1} \frac{1}{n}$ diverges (p-series, $p = 1$). By the Comparison Test, the series $\displaystyle\sum_{n=3}^{\infty} \frac{\ln n}{n}$ diverges and thus the series $\displaystyle\sum_{n=1}^{\infty} \frac{\ln n}{n}$ also diverges. □

Example 11.4.13. Determine whether the series $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^3-5n-2}$ is convergent or divergent.

Proof.

Observe that

(i) Not all terms are positive

(ii) We guess the series is convergent and hope $\dfrac{1}{n^3-5n-2} < \dfrac{2}{n^3}$ for all $n \geq n_0$. To find n_0, consider

$$2n^3 - 10n - 4 > n^3 \iff n^3 > 10n + 4 \quad \Rightarrow n \geq 4.$$

When $n \geq 4$, the term $\dfrac{1}{n^3 - 5n - 2} > 0$ and $\dfrac{1}{n^3 - 5n - 2} < \dfrac{2}{n^3}$. Also, $\displaystyle\sum_{n=4}^{\infty} \frac{2}{n^3}$ converges (p-series, $p = 3 > 1$). By the Comparison Test, the series $\displaystyle\sum_{n=4}^{\infty} \frac{1}{n^3 - 5n - 2}$ converges. Therefore, the series $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^3 - 5n - 2}$ converges. □

Note. Recall that for $\sum a_n$ and $\sum b_n$ with $0 \leq b_n \leq a_n$ for all $n \in \mathbb{N}$, the Comparison Test says that

(1) $\sum a_n$ converges $\implies$ $\sum b_n$ converges;

(2) $\sum b_n$ diverges $\implies$ $\sum a_n$ diverges.

But the converse is false. That is,

(1) $\sum b_n$ converges $\not\Longrightarrow$ $\sum a_n$ converges;

(2) $\sum a_n$ diverges $\not\Longrightarrow$ $\sum b_n$ diverges.

Example 11.4.14. Consider the series $\displaystyle\sum_{n=1}^{\infty} \frac{1}{2^n - 1}$. In order to use the Comparison Test to show $\sum \dfrac{1}{2^n - 1}$ converges, we cannot choose the known convergent series $\sum \dfrac{1}{2^n}$ because $\dfrac{1}{2^n - 1} > \dfrac{1}{2^n}$. However, $\dfrac{1}{2^n - 1}$ looks very close to $\dfrac{1}{2^n}$. It is reasonable to guess that the series $\sum \dfrac{1}{2^n - 1}$ also converges.

■ **<u>The Limit Comparison Test</u>**

Theorem 11.4.15. *(Limit Comparison Test) Let $\{a_n\}$ and $\{b_n\}$ be two nonnegative sequences. If*

$$\lim_{n\to\infty} \frac{a_n}{b_n} = L$$

for some $0 < L < \infty$, then $\displaystyle\sum_{n=1}^{\infty} a_n$ converges if and only if $\displaystyle\sum_{n=1}^{\infty} b_n$ converges. That is, either both series $\displaystyle\sum_{n=1}^{\infty} a_n$ and $\displaystyle\sum_{n=1}^{\infty} b_n$ converge or both diverge.

Proof. (Exercise) □

Example 11.4.16. Determine whether the series $\displaystyle\sum_{n=1}^{\infty} \frac{3}{2^n - 1}$ is convergent or divergent.

Proof. Consider the geometric series $\sum_{n=1}^{\infty} \frac{1}{2^n}$. Since $\sum_{n=1}^{\infty} \frac{1}{2^n}$ converges (geometric series with $r = \frac{1}{2} < 1$) and

$$\lim_{n\to\infty} \frac{\frac{3}{2^n-1}}{\frac{1}{2^n}} = \lim_{n\to\infty} \frac{3}{1-\frac{1}{2^n}} = 3,$$

by the Limit Comparison Test, the series $\sum_{n=1}^{\infty} \frac{3}{2^n-1}$ is convergent. □

Example 11.4.17. Determine whether the series $\sum_{n=1}^{\infty} \frac{2n^2+3n}{\sqrt{5+n^5}}$ is convergent or divergent.

Proof. Consider the series $\sum_{n=1}^{\infty} \frac{1}{n^{1/2}}$. Since $\sum_{n=1}^{\infty} \frac{1}{n^{1/2}}$ diverges (p-series, $p = \frac{1}{2} < 1$) and

$$\lim_{n\to\infty} \frac{\frac{2n^2+3n}{\sqrt{5+n^5}}}{\frac{1}{n^{1/2}}} = \lim_{n\to\infty} \frac{2+\frac{3}{n}}{\sqrt{\frac{5}{n^5}+1}} = 2,$$

by the Limit Comparison Test, the series $\sum_{n=1}^{\infty} \frac{2n^2+3n}{\sqrt{5+n^5}}$ diverges. □

■ Estimating Sums

Suppose that $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ are two convergent series with nonnegative terms and $0 \le b_n \le a_n$ for all $n \in \mathbb{N}$. Let

$$s = \sum_{n=1}^{\infty} a_n, \quad s_n = \sum_{k=1}^{n} a_k \quad \text{and} \quad R_n = s - s_n = a_{n+1} + a_{n+2} + \cdots$$
$$t = \sum_{n=1}^{\infty} b_n, \quad t_n = \sum_{k=1}^{n} b_k \quad \text{and} \quad T_n = t - t_n = b_{n+1} + b_{n+2} + \cdots$$

then $0 \le T_n \le R_n$ for all $n \in \mathbb{N}$. Hence, if we can estimate R_n, then we have an upper bound of T_n.

Example 11.4.18. Use the sum of the first 100 terms to approximate the sum of the series $\sum \frac{1}{n^3+1}$. Estimate the error involved in this approximation.

Proof. Since $\frac{1}{n^3+1} < \frac{1}{n^3}$ for all $n \in \mathbb{N}$, we have

$$T_{100} = \sum_{n=101}^{\infty} \frac{1}{n^3+1} \le \sum_{n=101}^{\infty} \frac{1}{n^3} < \int_{100}^{\infty} \frac{1}{x^3}\, dx = \frac{1}{2(100)^2}.$$

The error is less than $\frac{1}{2(100)^2}$ and $\sum_{n=1}^{\infty} \frac{1}{n^3+1} \approx \sum_{n=1}^{100} \frac{1}{n^3+1} \approx 0.6864538.$

□

11.5 Alternating Series

In the previous section, we consider the convergence tests for the nonnegative series (because of the bounded criterion). In the present section, we want to relax the condition and discuss the convergence for some special series which includes positive and negative terms alternatively.

❑ Alternating Series

Definition 11.5.1. An alternating series $\sum_{n=1}^{\infty} a_n$ is a series whose terms are alternatively positive and negative.

Let $b_n = |a_n|$. The general form of an alternating series is

$$\sum_{n=1}^{\infty} a_n = \begin{cases} \sum_{n=1}^{\infty}(-1)^n b_n & \text{if } a_1 < 0 \\ \sum_{n=1}^{\infty}(-1)^{n-1} b_n & \text{if } a_1 \geq 0. \end{cases}$$

Example 11.5.2. The series $\sum_{n=1}^{\infty}(-1)^n$ is an alternating series.

■ Alternating Series Test

Theorem 11.5.3. *If the alternating series*

$$\sum_{n=1}^{\infty}(-1)^{n-1} b_n = b_1 - b_2 + b_3 - b_4 + \cdots \quad \textit{where } b_n > 0$$

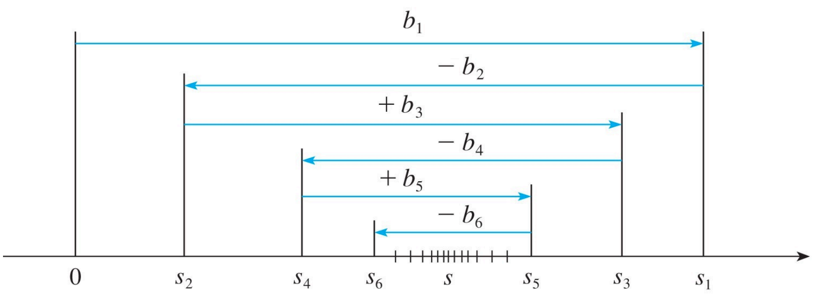
satisfies

(i) $b_{n+1} \leq b_n$ *for all* $n \in \mathbb{N}$

(ii) $\lim_{n\to\infty} b_n = 0.$

then the series is convergent.

Proof.



Let $\{s_n\}$ be the sequence of the partial sums of the alternating series. The condition (i) implies that, for every $n \in \mathbb{N}$,

$$s_{2n+2} = s_{2n} + \underbrace{(b_{2n+1} - b_{2n+2})}_{\geq 0} \geq s_{2n}$$

and

$$s_{2n} = b_1 - \underbrace{(b_2 - b_3)}_{\geq 0} - \cdots - \underbrace{(b_{2n-1} - b_{2n})}_{\geq 0} \leq b_1.$$

We have

$$0 \leq s_2 \leq s_4 \leq s_6 \leq \cdots \leq s_{2n} \leq \cdots \leq b_1$$

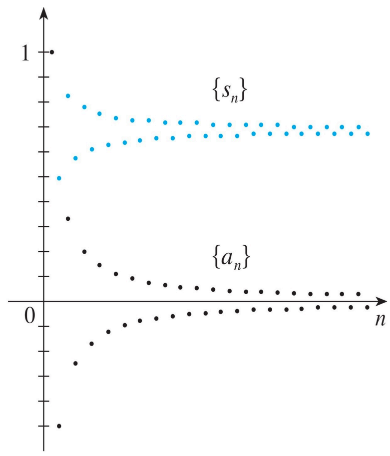
which is increasing and bounded above by b_1. By the bounded criterion, $\lim\limits_{n\to\infty} s_{2n} = s$ is convergent. Since $s_{2n+1} = s_{2n} + b_{2n+1}$, by condition (ii),

$$\lim_{n\to\infty} s_{2n+1} = \lim_{n\to\infty} s_{2n} + \lim_{n\to\infty} b_{2n+1} = s + 0 = s.$$

Hence $\lim\limits_{n\to\infty} s_n = s$ and the alternating series is convergent. □

Example 11.5.4. (alternating harmonic series) Determine whether the series $\sum\limits_{n=1}^{\infty} \dfrac{(-1)^{n-1}}{n}$ is convergent or divergent.

Proof. Let $b_n = \dfrac{1}{n}$. Then $\sum\limits_{n=1}^{\infty} \dfrac{(-1)^{n-1}}{n} = \sum\limits_{n=1}^{\infty} b_n$. Since $b_{n+1} = \dfrac{1}{n+1} < \dfrac{1}{n} = b_n$ for all $n \in \mathbb{N}$ and $\lim\limits_{n\to\infty} b_n = \lim\limits_{n\to\infty} \dfrac{1}{n} = 0$, by the alternating series test, the series $\sum\limits_{n=1}^{\infty} \dfrac{(-1)^{n-1}}{n}$ is convergent. □



Example 11.5.5. Determine whether the series $\sum\limits_{n=1}^{\infty} \dfrac{(-1)^n 3n}{4n-1}$ is convergent or divergent.

Proof. Let $b_n = \dfrac{3n}{4n-1}$ and $a_n = \dfrac{(-1)^n 3n}{4n-1} = (-1)^n b_n$. Then $|a_n| = b_n$ for every $n \in \mathbb{N}$.

Since $\lim\limits_{n\to\infty} b_n = \lim\limits_{n\to\infty} \dfrac{3n}{4n-1} = \dfrac{3}{4} \neq 0$, the limit $\lim\limits_{n\to\infty} a_n$ is not equal to 0 (in fact, the limit does not exist). By the test for divergent, the series $\sum\limits_{n=1}^{\infty} \dfrac{(-1)^n 3n}{4n-1} = \lim\limits_{n\to\infty} a_n$ is divergent. □

Example 11.5.6. Determine whether the series $\sum\limits_{n=1}^{\infty} \dfrac{(-1)^{n+1} n^2}{n^3+1}$ is convergent or divergent.

Proof. Let $b_n = \dfrac{n^2}{n^3+1}$ Then $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^{n+1}n^2}{n^3+1} = \sum_{n=1}^{\infty} (-1)^{n+1} b_n$. Since

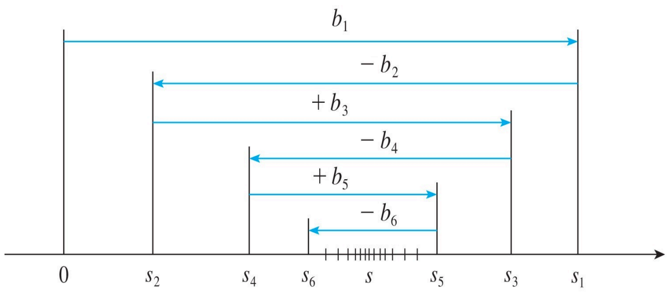
$$b_{n+1} - b_n = \frac{(n+1)^2}{(n+1)^3+1} - \frac{n^2}{n^3+1} = \frac{-n^4 - 2n^3 - n^2 + 2n + 1}{[(n+1)^3+1](n^3+1)} < 0 \quad \text{for all } n \in \mathbb{N},$$

we have $b_{n+1} \leq b_n$ for all $n \in \mathbb{N}$. Also, $\displaystyle\lim_{n\to\infty} b_n = \lim_{n\to\infty} \frac{n^2}{n^3+1} = 0$. By the alternating series test, the series $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^{n+1}n^2}{n^3+1}$ is convergent. □

Note. In this example, we can compute $\dfrac{d}{dx}\left(\dfrac{x^2}{x^3+1}\right) = \dfrac{x(2-x^3)}{(x^3+1)^2} < 0$ for $x \geq 2$ to obtain $b_{n+1} \leq b_n$ for all $n \in \mathbb{N}$.

Remark. As the similar discussion as before, in the use of the alternating series test, it only needs that the series satisfies conditions (i) in Theorem 11.5.3 for every $n \geq n_0$ for some fixed integer n_0.

■ Estimating Sums



Observe the structure of an alternating series satisfying the two conditions (i) and (ii) in Theorem 11.5.3. Let $R_n = s - s_n$ be the remainder of the series, then

$$|R_n| = |s - s_n| \leq b_{n+1}.$$

Theorem 11.5.7. *(Alternating Series Estimation Theorem) If* $s = \displaystyle\sum_{n=1}^{\infty} (-1)^{n-1} b_n$ *is the sum of an alternating series that satisfies*

$$(i)\ 0 \leq b_{n+1} \leq b_n \quad \textit{for every } n \in \mathbb{N} \qquad \textit{and} \qquad (ii)\ \lim_{n\to\infty} b_n = 0$$

then

$$|R_n| = |s - s_n| \leq b_{n+1}$$

Example 11.5.8. Find the sum of the series $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^n}{n!}$ correct to three decimal places.

Proof. The series $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^n}{n!}$ is an alternating series. Let $b_n = \dfrac{1}{n!}$. Then $b_{n+1} = \dfrac{1}{(n+1)!} < \dfrac{1}{n!} = b_n$ and $\displaystyle\lim_{n\to\infty} b_n = \lim_{n\to\infty} \frac{1}{n!} = 0$. To find n such that $b_n = \dfrac{1}{n!} < 0.001$, we have $n \geq 7$. Hence, by the alternating series estimation,

$$|R_6| = |s - s_6| \leq b_7 < 0.001 \qquad (\text{in fact, } b_7 < 0.0002).$$

Then $s_6 = 1 - 1 + \dfrac{1}{2} - \dfrac{1}{6} + \dfrac{1}{24} - \dfrac{1}{120} + \dfrac{1}{720} \approx 0.368056$. In fact $s = \dfrac{1}{e} \approx 0.36787944$. □

11.6 Absolute Convergence

In the present section, we will continue to discuss the convergence of general series (without alternating patterns). Intuitively, it is difficult to give a nice test for every series because they may have too many varieties. Therefore, we hope to use some known results (discussed in the previous sections) to deal with the convergence of certain general series.

❑ Absolute Convergence

Definition 11.6.1. (a) A series $\sum_{n=1}^{\infty} a_n$ is called "*absolutely convergent*" if the series of absolute values $\sum_{n=1}^{\infty} |a_n|$ is convergent.

(b) A series $\sum_{n=1}^{\infty} a_n$ is called "*conditionally convergent*" if it is convergent but not absolutely convergent.

Example 11.6.2.

(1) The series $\sum_{n=1}^{\infty} \frac{(-1)^n}{n}$ is convergent by the alternating series test. But $\sum_{n=1}^{\infty} \left|\frac{(-1)^n}{n}\right| = \sum_{n=1}^{\infty} \frac{1}{n} = \infty$ is divergent (harmonic series, p-series with $p = 1$). Therefore, $\sum_{n=1}^{\infty} \frac{(-1)^n}{n}$ is a conditionally convergent series.

(2) The series $\sum_{n=1}^{\infty} \frac{(-1)^n}{n^2}$ is convergent by the alternating series test and $\sum_{n=1}^{\infty} \left|\frac{(-1)^n}{n^2}\right| = \sum_{n=1}^{\infty} \frac{1}{n^2}$ is also convergent (p-series with $p = 2$). Therefore, $\sum_{n=1}^{\infty} \frac{(-1)^n}{n^2}$ is absolutely convergent.

Question: For the two series $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} |a_n|$, can the convergence of one series imply the convergence of the other one?

Theorem 11.6.3. *If a series $\sum_{n=1}^{\infty} a_n$ is absolutely convergent, then it is convergent. That is, if $\sum_{n=1}^{\infty} |a_n|$ converges then $\sum_{n=1}^{\infty} a_n$ converges.*

Proof. Observe that $0 \le a_n + |a_n| \le 2|a_n|$. If $\sum_{n=1}^{\infty} a_n$ is absolutely convergent, then $\sum_{n=1}^{\infty} 2|a_n|$

converges. By the Comparison Test, the series $\sum_{n=1}^{\infty}(a_n + |a_n|)$ converges. Hence, the series

$$\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty}(a_n + |a_n| - |a_n|) = \sum_{n=1}^{\infty}(a_n + |a_n|) - \sum_{n=1}^{\infty}|a_n|$$

converges.

$\square$

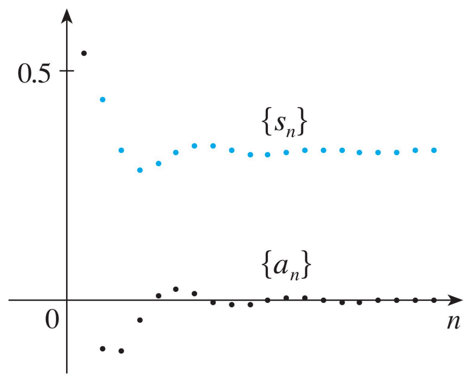
Note.

(1) The converse of Theorem 11.6.3 is false. That is, the convergence of $\sum_{n=1}^{\infty} a_n$ cannot imply the convergence of $\sum_{n=1}^{\infty}|a_n|$. For example, $\sum_{n=1}^{\infty}\frac{(-1)^n}{n}$ is convergent but $\sum_{n=1}^{\infty}\left|\frac{(-1)^n}{n}\right|$ is divergent.

(2) If $\sum_{n=1}^{\infty} a_n$ is divergent, then $\sum_{n=1}^{\infty}|a_n|$ must be divergent.

Example 11.6.4. Determine whether the series $\sum_{n=1}^{\infty}\frac{\cos n}{n^2}$ is convergent or divergent.

Proof. The series $\sum_{n=1}^{\infty}\frac{\cos n}{n^2}$ is not an alternating series. Consider $\left|\frac{\cos n}{n^2}\right| \le \frac{1}{n^2}$ for every $n \in \mathbb{N}$.
Since $\sum_{n=1}^{\infty}\frac{1}{n^2}$ converges (p-series, $p = 2$), by the Comparison Test, the series $\sum_{n=1}^{\infty}\left|\frac{\cos n}{n}\right|$ converges. Hence, the series $\sum_{n=1}^{\infty}\frac{\cos n}{n^2}$ is absolutely convergent and this implies that it is convergent. $\square$



Exercise. Let $\{a_n\}$ be a sequence and define

$$a_n^+ = \begin{cases} a_n, & \text{if } a_n \ge 0 \\ 0, & \text{if } a_n < 0 \end{cases} \quad \text{and} \quad a_n^- = \begin{cases} 0, & \text{if } a_n \ge 0 \\ a_n, & \text{if } a_n < 0 \end{cases}$$

Prove that the series $\sum_{n=1}^{\infty}|a_n|$ converges if and only if both of the series $\sum_{n=1}^{\infty} a_n^+$ and $\sum_{n=1}^{\infty} a_n^-$ converge and moreover,

$$\sum_{n=1}^{\infty}|a_n| = \sum_{n=1}^{\infty} a_n^+ - \sum_{n=1}^{\infty} a_n^-.$$

Hint: ($\Longrightarrow$) Using the Comarison Test with the fact $0 \le |a_n^{\pm}| \le |a_n|$ for every $n \in \mathbb{N}$ and moreover, the equality holds from the laws for series.
($\Longleftarrow$) Using the laws for series with the fact $|a_n| = a_n^+ - a_n^-$ for every $n \in \mathbb{N}$.

❑ Rearrangement

Consider an example of a paradox. Let

$$\begin{aligned}
x &= 1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\cdots+\frac{(-1)^{n+1}}{n}+\cdots \\
&\stackrel{?}{=} (1-\frac{1}{2})+\frac{1}{4}+(\frac{1}{3}-\frac{1}{6})-\frac{1}{8}+(\frac{1}{5}-\frac{1}{10})-\frac{1}{12}+(\frac{1}{7}-\frac{1}{14})-\frac{1}{16}+(\frac{1}{9}-\frac{1}{18})-\frac{1}{20}+\cdots \\
&= \frac{1}{2}-\frac{1}{4}+\frac{1}{6}-\frac{1}{8}+\cdots \\
&= \frac{1}{2}(1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\cdots) \\
&= \frac{1}{2}x.
\end{aligned}$$

Hence $x=\frac{1}{2}x$ and we obtain a contradiction that $x=0$.

Question: What's wrong with this?

For a sum of finitely many numbers, we obtain the same value if arbitrarily rearraneging the order of those numbers.

Question: Can we get the same value of the sum of infinitely many numbers if we arbitrarily rearrange the order of these numbers?

Definition 11.6.5. Let $\{a_n\}$ and $\{b_n\}$ be two sequences. We say that $\{b_n\}$ is a "*rearrangement*" of $\{a_n\}$ if there exists a one-to-one and onto function f on $\mathbb{N}$ such that $b_n = a_{f(n)}$ for every $n \in \mathbb{N}$.

Note. In general, $\displaystyle\sum_{n=1}^{\infty} a_n \neq \sum_{n=1}^{\infty} b_n$ if $\{b_n\}$ is a rearrangement of $\{a_n\}$.

Theorem 11.6.6. *If $\displaystyle\sum_{n=1}^{\infty} a_n$ is conditionally convergent then, for any number $L \in \mathbb{R}$, there exists a rearrangement $\{b_n\}$ of $\{a_n\}$ such that $\displaystyle\sum_{n=1}^{\infty} a_n = L$.*

Proof. We only sketch the proof by the following steps.

(I) Let $\{p_n\}$ be the nonnegative subsequence of $\{a_n\}$ and $\{q_n\}$ be the negative subsequence of $\{a_n\}$. Since $\displaystyle\sum_{n=1}^{\infty} a_n$ is conditionally convergent, we have $\displaystyle\sum_{n=1}^{\infty} |a_n|$ diverges. Hence, at least one of the series $\displaystyle\sum_{n=1}^{\infty} p_n$ and $\displaystyle\sum_{n=1}^{\infty} q_n$ is divergent. Moreover, the fact that $\displaystyle\sum_{n=1}^{\infty} a_n$ converges implies both series $\displaystyle\sum_{n=1}^{\infty} p_n$ and $\displaystyle\sum_{n=1}^{\infty} q_n$ are divergent. We have that

$$\sum_{n=1}^{\infty} p_n = \infty \quad \text{and} \quad \sum_{n=1}^{\infty} q_n = -\infty.$$

(II) W.L.O.G, say $L > 0$. We construct a sequence $\{b_n\}$ from $\{p_n\}$ and $\{q_n\}$ by the following process. Since $\sum_{n=1}^{\infty} p_n = \infty$, there exists $n_1 \in \mathbb{N}$ such that

$$\sum_{n=1}^{n_1-1} p_n < L \le \sum_{n=1}^{n_1} p_n.$$

Let $S_1 = \sum_{n=1}^{n_1} p_n$. Then $S_1 \ge L$ and $S_1 - p_{n_1} < L$. Hence, $|S_1 - L| < p_{n_1}$.

Since $\sum_{n=1}^{\infty} q_n = -\infty$, there exists $m_1 \in \mathbb{N}$ such that

$$\sum_{n=1}^{n_1} p_n + \sum_{n=1}^{m_1-1} q_n > L \ge \sum_{n=1}^{n_1} p_n + \sum_{n=1}^{m_1} q_n.$$

Let $T_1 = \sum_{n=1}^{n_1} p_n + \sum_{n=1}^{m_1} q_n = S_1 + \sum_{n=1}^{m_1} q_n$. Then $T_1 \le L$ and $T_1 - q_{m_1} > L$. Hence, $|T_1 - L| < q_{m_1}$.

Continue this process, we have $1 \le n_1 < n_2 < \cdots$ and $1 \le m_1 < m_2 < \cdots$ and $\{S_k\}$ and $\{T_k\}$ such that for every $k \in \mathbb{N}$,

$$S_k = T_{k-1} + \sum_{n=n_{k-1}+1}^{n_k} p_n, \quad S_k \ge L, \quad S_k - p_{n_k} < L \quad \Longrightarrow |S_k - L| < p_{n_k}$$

and

$$T_k = S_k + \sum_{n=m_{k-1}+1}^{m_k} q_n, \quad T_k \le L, \quad T_k - q_{m_k} \ge L \quad \Longrightarrow |T_k - L| < q_{m_k}.$$

Define $\{b_n\} = \left\{p_1, p_2, \cdots, p_{n_1}, q_1, q_2, \cdots, q_{m_1}, p_{n_1+1}, \cdots, p_{n_2}, q_{m_1+1}, \cdots q_{m_2}, \cdots\right\}$

(III) To check that $\{b_n\}$ is a rearrangement of $\{a_n\}$, we have to show that

(i) To show that each a_n appears at most once in $\{b_n\}$. Since each a_n is either in $\{p_n\}$ or in $\{q_n\}$, and each p_n or each q_n appears in $\{b_n\}$ at most once by the construction of $\{b_n\}$, we have each a_n appears in $\{b_n\}$ at most once.

(ii) To show that each a_n appears at least once in $\{b_n\}$. For $K \in \mathbb{N}$, a_K must appear in $\{p_n\}_{n=1}^{K}$ or in $\{q_n\}_{n=1}^{K}$. Hence, a_K appears in $\{b_n\}$ at least once.

(IV) Check that $S_k \to L$ and $T_k \to L$ as $k \to \infty$. Since the series $\sum_{n=1}^{\infty} a_n$ converges, $a_n \to 0$ as $n \to \infty$. Then $p_n \to 0$ and $q_n \to 0$ as $n \to \infty$. Hence, by part (II), $S_k \to L$ and $T_k \to L$ as $k \to \infty$.

By the above argument $\{b_n\}$ is a rearrangement of $\{a_n\}$ and $\sum_{n=1}^{\infty} b_n = L$.

□

Theorem 11.6.7. *If* $\sum_{n=1}^{\infty} a_n$ *is absolutely convergent and* $\{b_n\}$ *is a rearrangement of* $\{a_n\}$*, then*

(a) $\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} b_n$ *and*

(b) $\sum_{n=1}^{\infty} b_n$ *is absolutely convergent.*

Proof. Let $s_n = \sum_{k=1}^{n} a_k$ and $t_m = \sum_{k=1}^{m} b_k$.

(a) Since $\sum_{n=1}^{\infty} a_n$ is absolutely convergent and hence it is convergent, the series $\sum_{n=1}^{\infty} a_n$ is a finite number. Given $\varepsilon > 0$, we want to prove $\left| t_m - \sum_{n=1}^{\infty} a_n \right| < \varepsilon$ as m is sufficiently large.

Since $\sum_{n=1}^{\infty} |a_n|$ converges, there exists $N \in \mathbb{N}$ such that

$$|a_{N+1}| + |a_{N+2}| + \cdots < \frac{\varepsilon}{2}.$$

Since $\{b_n\}$ is a rearrangement of $\{a_n\}$, there exists $M \in \mathbb{N}$ such that $\{a_1, a_2, \cdots, a_N\} \subseteq \{b_1, b_2, \cdots, b_M\}$. For $m > M$

$$|t_m - s_N| \leq |a_{N+1}| + |a_{N+2}| + \cdots < \frac{\varepsilon}{2}.$$

Then

$$|t_m - \sum_{n=1}^{\infty} a_n| \leq |t_m - s_N| + |s_N - \sum_{n=1}^{\infty} a_n| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Hence, $\{t_m\}$ converges to $\sum_{n=1}^{\infty} a_n$ and we have $\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} b_n$.

(b) Consider the sequence $\{|a_n|\}$. Since $\sum_{n=1}^{\infty} a_n$ is absolutely convergent, $\sum_{n=1}^{\infty} |a_n|$ is also absolutely convergent. On the other hand, since $\{b_n\}$ is a rearrangement of $\{a_n\}$, $\{|b_n|\}$ is a rearrangement of $\{|a_n|\}$. By part(a),

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

Hence, $\sum_{n=1}^{\infty} |b_n|$ converges; that is, $\sum_{n=1}^{\infty} b_n$ is absutely convergent.

□

■ **Product of two sequences**

Suppose that $\{a_n\}$ and $\{b_n\}$ are summable sequences. We recall that

$$\sum_{n=1}^{\infty}(a_n \pm b_n) = \sum_{n=1}^{\infty} a_n \pm \sum_{n=1}^{\infty} b_n$$
$$\sum_{n=1}^{\infty}(ca_n) = c\sum_{n=1}^{\infty} a_n \quad \text{where } c \text{ is a constant.}$$

Question: Can we express $(\sum_{n=1}^{\infty} a_n)(\sum_{n=1}^{\infty} b_n)$ as a form of series? If yes, what is the expression?

Heuristically, we observe the product of two finite series.

$$\Big(\sum_{n=1}^{N} a_n\Big)\Big(\sum_{m=1}^{M} b_m\Big) = \sum_{k=1}^{L} c_k.$$

where $\{c_k\}$ contains all products of $a_n b_m$.

Question: Is the formula still true for the product of two arbitrary infinite series?
Anserer: In general, it is not true for two summable sequences.

Exercise. Find two summable sequences $\{a_n\}$ and $\{b_n\}$ such that there is no summable sequence $\{c_n\}$ satisfying

$$\Big(\sum_{n=1}^{\infty} a_n\Big)\Big(\sum_{n=1}^{\infty} b_n\Big) = \sum_{n=1}^{\infty} c_n.$$

Theorem 11.6.8. *If $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ converge absolutely and $\{c_n\}$ is any sequence containing all products $a_i b_j$ for each pair (i, j), then*

$$\sum_{n=1}^{\infty} c_n = \Big(\sum_{n=1}^{\infty} a_n\Big)\Big(\sum_{n=1}^{\infty} b_n\Big).$$

Proof. (Exercise) □

11.7 The Ratio and Root Tests

In the previous section, we study that an absolutely convergent series is also convergent. However, it is not easy to check whether a general series is absolutely convergent. In the present section, we will introduce two methods which can determine whether certain series are convergent or divergent. The spirit of these two methods is from the comparison with geometric series.

❑ **The Ratio Test**

Theorem 11.7.1. *(Ratio Test) For the series $\sum_{n=1}^{\infty} a_n$, suppose that $\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| = L$.*

(a) If $L < 1$, then the series $\sum_{n=1}^{\infty} a_n$ is absolutely convergent (and therefore it is convergent).

(b) If $L > 1$ (or $L = \infty$), then the series $\sum_{n=1}^{\infty} a_n$ is divergent.

(c) If $L = 1$ the Ratio Test is inconclusive. (For example, $\sum \frac{1}{n}$ diverges and $\sum \frac{1}{n^2}$ converges).

Proof. (Postponed) □

Example 11.7.2. Determine whether the following series are convergent or divergent.

(1) $\sum_{n=1}^{\infty} \frac{1}{n!}$

Proof. Let $a_n = \frac{1}{n!}$. Then

$$\lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| = \lim_{n\to\infty} \frac{\frac{1}{(n+1)!}}{\frac{1}{n!}} = \lim_{n\to\infty} \frac{1}{n+1} = 0 < 1.$$

By the ratio test, the series $\sum_{n=1}^{\infty} \frac{1}{n!}$ is convergent. □

(2) $\sum_{n=1}^{\infty} \frac{1}{n!}$

Proof. Let $a_n = \frac{1}{n!}$. Then

$$\lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| = \lim_{n\to\infty} \frac{\frac{1}{(n+1)!}}{\frac{1}{n!}} = \lim_{n\to\infty} \frac{1}{n+1} = 0 < 1.$$

By the ratio test, the series $\sum_{n=1}^{\infty} \frac{1}{n!}$ is convergent. □

(3) $\sum_{n=1}^{\infty} \frac{r^n}{(n+1)!}$ for some $r \in \mathbb{R}$.

Proof. Let $a_n = \frac{r^n}{(n+1)!}$. Then

$$\lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| = \lim_{n\to\infty} \frac{\frac{r^{n+1}}{(n+2)!}}{\frac{r^n}{(n+1)!}} = \lim_{n\to\infty} \frac{r}{n+2} = 0 < 1.$$

By the ratio test, the series $\sum_{n=1}^{\infty} \frac{r^n}{(n+1)!}$ is convergent. □

(4) $\displaystyle\sum_{n=1}^{\infty}(-1)^n\frac{n^3}{3^n}$

Proof. Let $a_n = (-1)^n\dfrac{n^3}{3^n}$. Then

$$\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| = \lim_{n\to\infty}\left|\frac{(-1)^{n+1}\frac{(n+1)^3}{3^{n+1}}}{(-1)^n\frac{n^3}{3^n}}\right| = \lim_{n\to\infty}\frac{1}{3}\Big(\frac{n+1}{n}\Big)^3 = \frac{1}{3} < 1.$$

By the ratio test, the series $\displaystyle\sum_{n=1}^{\infty}(-1)^n\frac{n^3}{3^n}$ is convergent. □

(5) $\displaystyle\sum_{n=1}^{\infty}\frac{n^n}{n!}$

Proof. Let $a_n = \dfrac{n^n}{n!}$. Then

$$\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| = \lim_{n\to\infty}\left|\frac{\frac{(n+1)^{n+1}}{(n+1)!}}{\frac{n^n}{n!}}\right| = \lim_{n\to\infty}\Big(\frac{n+1}{n}\Big)^n = \lim_{n\to\infty}(1+\frac{1}{n})^n = e > 1.$$

By the ratio test, the series $\displaystyle\sum_{n=1}^{\infty}\frac{n^n}{n!}$ is divergent. □

Note. Consider $\dfrac{n^n}{n!} = \dfrac{n\cdot n\cdots n}{1\cdot 2\cdots n} \geq n \to \infty$ as $n \to$. By the Test for Divergence, the series $\displaystyle\sum_{n=1}^{\infty}\frac{n^n}{n!}$ is divergent.

<u>*Proof of Ratio Test*</u>

(a) Since $\displaystyle\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| = L < 1$, choosing a number s such that $L < s < 1$, there exists $N \in \mathbb{N}$ such that for every $n \geq N$

$$\frac{|a_{n+1}|}{|a_n|} < s < 1.$$

Hence, $|a_{n+1}| < |a_n|s$ for every $n > N$. We have

$$\begin{array}{rcl}
|a_{N+2}| & < & |a_{N+1}|s \\
|a_{N+3}| & < & |a_{N+2}|s < |a_{N+1}|s^2 \\
& \vdots & \\
|a_{N+k}| & < & |a_{N+k-1}|s < \cdots < |a_{N+1}|s^{k-1} \qquad \text{for } k = 1,2,3,\cdots
\end{array}$$

For every $n > N$, the partial sum s_n of $\sum_{n=1}^{\infty} |a_n|$ satisfies

$$\begin{aligned} s_n &= |a_1| + |a_2| + \cdots + |a_N| + |a_{N+1}| + \cdots + |a_n| \\ &= \sum_{k=1}^{N} |a_k| + |a_{N+1}| + \cdots + |a_n| \\ &< \sum_{k=1}^{N} |a_k| + |a_{N+1}| + |a_{N+1}|s + |a_{N+1}|s^2 + \cdots + |a_{N+1}|s^{n-(N+1)} \\ &= \sum_{k=1}^{N} |a_k| + \frac{|a_{N+1}|(1 - s^{n-N})}{1 - s} \\ &< \sum_{k=1}^{N} |a_k| + \frac{|a_{N+1}|}{1 - s} \qquad \text{since } 0 < s < 1. \end{aligned}$$

Since $\{s_n\}$ is an increasing sequence and bounded above, by the bounded criterion, $\{s_n\}$ converges and hence $\sum_{n=1}^{\infty} a_n$ is absolutely convegent.

(b) Since $\lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| = L > 1$, choosing a number s such that $1 < s < L$, there exists $N \in \mathbb{N}$ such that for every $n \geq N$

$$\frac{|a_{n+1}|}{|a_n|} > s > 1.$$

Hence, $|a_{n+1}| > |a_n|s$ for every $n > N$. We have

$$\begin{aligned} |a_{N+2}| &> |a_{N+1}|s \\ |a_{N+3}| &> |a_{N+2}|s < |a_{N+1}|s^2 \\ &\vdots \\ |a_{N+k}| &> |a_{N+k-1}|s > \cdots < |a_{N+1}|s^{k-1} \qquad \text{for } k = 1, 2, 3, \cdots \end{aligned}$$

W.L.O.G, we may assume that $|a_{N+1}| > 0$. Then

$$\lim_{n\to\infty} |a_n| \geq \lim_{n\to\infty} |a_{N+1}|s^{n-(N+1)} = \infty \qquad (\text{since } s > 1).$$

Hence, $\lim_{n\to\infty} a_n \neq 0$. By the Test for Divergence, the series $\sum_{n=1}^{\infty} a_n$ is divergent.

❑ The Root Test

Theorem 11.7.3. *(Root Test) For the series $\sum_{n=1}^{\infty} a_n$, suppose that $\lim_{n\to\infty} \sqrt[n]{|a_n|} = L$.*

(a) If $L < 1$, then the series $\sum_{n=1}^{\infty} a_n$ is absolutely convergent (and therefore it is convergent).

(b) If $L > 1$ (or $L = \infty$), then the series $\sum_{n=1}^{\infty} a_n$ is divergent.

(c) If $L = 1$ the Ratio Test is inconclusive. (For example, $\sum \frac{1}{n}$ diverges and $\sum \frac{1}{n^2}$ converges).

Proof. (Postponed) □

Example 11.7.4. Determine whether the following series are convergent or divergent.

(1) $\sum_{n=1}^{\infty} \frac{1}{(\ln n)^n}$.

Proof. Let $a_n = \frac{1}{(\ln n)^n}$. Then

$$\lim_{n\to\infty} \sqrt[n]{|a_n|} = \lim_{n\to\infty} \sqrt[n]{\left|\frac{1}{(\ln n)}\right|^n} = \lim_{n\to\infty} \frac{1}{\ln n} = 0 < 1.$$

By the root test, the series $\sum_{n=1}^{\infty} \frac{1}{(\ln n)^n}$ is convergent. □

(2) $\sum_{n=1}^{\infty} \frac{2^n}{n^3}$.

Proof. Let $a_n = \frac{2^n}{n^3}$. Then

$$\lim_{n\to\infty} \sqrt[n]{|a_n|} = \lim_{n\to\infty} \sqrt[n]{\left|\frac{2^n}{n^3}\right|} = \lim_{n\to\infty} \frac{2}{\sqrt[n]{n^3}} = 2 > 1.$$

By the root test, the series $\sum_{n=1}^{\infty} \frac{2^n}{n^3}$ is divergent. □

(3) $\sum_{n=1}^{\infty} \left(\frac{2n+3}{3n+2}\right)^n$.

Proof. Let $a_n = \left(\frac{2n+3}{3n+2}\right)^n$. Then

$$\lim_{n\to\infty} \sqrt[n]{|a_n|} = \lim_{n\to\infty} \sqrt[n]{\left|\frac{2n+3}{3n+2}\right|^n} = \lim_{n\to\infty} \frac{2n+3}{3n+2} = \frac{2}{3} < 1.$$

By the root test, the series $\sum_{n=1}^{\infty} \left(\frac{2n+3}{3n+2}\right)^n$ is convergent. □

Proof of Ratio Test

(a) Since $\lim_{n\to\infty} \sqrt[n]{|a_n|} = L < 1$, choosing a number s such that $L < s < 1$, there exists $N \in \mathbb{N}$ such that for every $n \geq N$

$$\sqrt[n]{|a_n|} < s < 1.$$

Hence, $|a_n| < s^n$ for every $n \geq N$. The partial sum s_n of $\sum_{n=1}^{\infty} |a_n|$ satisfies

$$\begin{aligned} s_n &= |a_1| + |a_2| + \cdots + |a_N| + |a_{N+1}| + \cdots + |a_n| \\ &< \sum_{k=1}^{N} |a_k| + s^{N+1} + s^{N+2} + \cdots + s^n \\ &= \sum_{k=1}^{N} |a_k| + \frac{s^{N+1}(1 - s^{n-N})}{1 - s} \\ &< \sum_{k=1}^{N} |a_k| + \frac{s^{N+1}}{1 - s} \qquad \text{since } 0 < s < 1. \end{aligned}$$

Since $\{s_n\}$ is an increasing sequence and bounded above, by the bounded criterion, $\{s_n\}$ converges and hence $\sum_{n=1}^{\infty} a_n$ is absolutely convegent.

(b) Since $\lim_{n\to\infty} \sqrt[n]{|a_n|} = L > 1$, choosing a number s such that $1 < s < L$, there exists $N \in \mathbb{N}$ such that for every $n \geq N$

$$\sqrt[n]{|a_n|} > s > 1.$$

Hence, $|a_n| > s^n$ for every $n > N$. We have

$$\lim_{n\to\infty} |a_n| \geq \lim_{n\to\infty} s^n = \infty \qquad \text{(since } s > 1\text{)}.$$

Hence, $\lim_{n\to\infty} a_n \neq 0$. By the Test for Divergence, the series $\sum_{n=1}^{\infty} a_n$ is divergent.

11.8 Strategy for Testing Series

In the present section, we will organize all tests introduced in previous sections. The following steps are some strategies for convergence or divergence for series.

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

1. ***p*-series:**

$$\sum_{n=1}^{\infty} \frac{1}{n^p} \text{ is } \begin{cases} \text{convergent} & \text{when } p > 1 \\ \text{divergent} & \text{when } p \leq 1. \end{cases}$$

2. **geometric series:**

$$\sum_{n=1}^{\infty} ar^n \ (a \neq 0) \text{ is } \begin{cases} \text{convergent} & \text{when } |r| < 1 \\ \text{divergent} & \text{when } |r| \geq 1. \end{cases}$$

3. When the form of the series is similar to a p-series or a geometric series (for example, $\sum \frac{2}{n^2+3n+1}$ or $\sum \frac{2^{n+1}-5}{3^n+2}$), we could determine the convergence or divergence by using the comparison test (or limit comparison test).

4. **Test for Divergence:**

$$\lim_{n\to\infty} a_n \neq 0 \qquad \Longrightarrow \qquad \sum_{n=1}^{\infty} a_n \quad \text{is divergent.}$$

5. **Alternating Series Test:** If the series has the form $\sum_{n=1}^{\infty}(-1)^n b_n$ for $b_n > 0$ satisfying

$$\text{(i) } b_{n+1} \leq b_n \text{ for all } n \in \mathbb{N} \qquad \text{and} \qquad \text{(ii) } \lim_{n\to\infty} b_n = 0$$

then the series $\sum_{n=1}^{\infty}(-1)^n b_n$ is convergent.

6. **Ratio Test:** Suppose that $\lim_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right| = L.$

$$\sum_{n=1}^{\infty} a_n \text{ is } \begin{cases} \text{absolutely convergent} & \text{if } L < 1 \\ \text{divergent} & \text{if } L > 1 \\ \text{inconclusive} & \text{if } L = 1 \end{cases}$$

7. **Root Test:** Suppose that $\lim_{n\to\infty} \sqrt[n]{|a_n|} = L.$

$$\sum_{n=1}^{\infty} a_n \text{ is } \begin{cases} \text{absolutely convergent} & \text{if } L < 1 \\ \text{divergent} & \text{if } L > 1 \\ \text{inconclusive} & \text{if } L = 1 \end{cases}$$

8. **Integral Test:** Suppose that f is positive and nonincreasing on $[1, \infty)$, and $a_n = f(n)$. Then

$$\sum_{n=1}^{\infty} a_n \quad \text{is convergent (divergent)} \qquad \Longleftrightarrow \qquad \int_1^{\infty} f(x)\,dx \quad \text{is convergent (divergent).}$$

Chapter 12

Power Series and Taylor Series

12.1 Approximation by polynomial functions

Although most elementary functions have nice differentiable and integrable properties, it is not easy to compute their exact values like $\sin 1$, e^2, $\ln 3$ etc. Polynomial functions are a family of best functions. For a polynomial function $P(x)$, we can easily find its value $P(a)$ by basic algebraic algorithms. Naturally, we want to study whether a function can be approximated by polynomial functions. In Section 5.6, we knew that a differentiable function can be approximated by a 1-degree polynomial, at least near a certain point. We expect to (locally) approximate elementary functions by higher degree polynomials

■ Coefficients and Derivatives of $P(x)$

Lemma 12.1.1. *(a) Let* $P(x) = c_0 + c_1 x + c_2 x^2 + c_n x^n$. *Then*

$$c_k = \frac{P^{(k)}(0)}{k!} \qquad \text{for } k = 0, 1, 2, \cdots, n$$

(b) Let $P(x) = c_0 + c_1(x-a) + c_2(x-a)^2 + \cdots + c_n(x-a)^n$ *be "polynomial in* $(x-a)$*". Then*

$$c_k = \frac{P^{(k)}(a)}{k!} \qquad \text{for } k = 0, 1, 2, \cdots, n.$$

Proof. Compute them directly. □

■ Taylor Polynomial

Question: For a given function $f(x)$ (for example, e^x), can we find a polynomial function $P(x)$ such that $P(x)$ is close to $f(x)$ for every $x \in \mathbb{R}$?

In general, it is impossible to obtain such a nice approximation "for every x". Therefore, we will usually consider an approximation by polynomial functions "near a given point".

Question: Which polynomial function will give nice approximations (near a given point a)? How to find such a polynomial function?

Heuristically, for a given function $f(x)$ with sufficiently many times derivatives at a, we expect that $P(x)$ is an appropriate polynomial to approximate f (near a) if both $f(x)$ and $P(x)$ have the same first n times derivatives at a. That is, $f^{(k)}(a) = P^{(k)}(a)$ for $k = 0, 1, 2, \cdots, n$. Therefore, if we write $P(x) = c_0 + c_1(x-a) + c_2(x-a)^2 + \cdots + c_n(x-a)^n$, then

$$c_k = \frac{P^{(k)}(a)}{k!} = \frac{f^{(k)}(a)}{k!}$$

for $k = 0, 1, 2, \cdots, n$.

Definition 12.1.2. Suppose that f is a function such that $f'(a), f''(a), \cdots, f^{(n)}(a)$ exist. Define

$$P_{n,a,f}(x) = c_0 + c_1(x-a) + c_2(x-a)^2 + \cdots + c_n(x-a)^n = \sum_{k=1}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k$$

where $c_k = \dfrac{f^{(k)}(a)}{k!}$ for $k = 0, 1, \cdots, n$. The polynomial $P_{n,a,f}(x)$ is called the "*Taylor polynomial* of degree n for f at a".

Remark.

(i) If there is no confusion, we may replace $P_{n,a,f}(x)$ by $P_{n,a}(x)$.

(ii) $P^{(k)}(a) = f^{(k)}(a)$ for every $k = 0, 1, 2, \cdots, n$.

Example 12.1.3. Find the Taylor polynomial of degree n for f at the given point.

(1) $f(x) = \sin x$, at $a = 0$.

Proof.

$$f^{(4n)}(0) = 0, \quad f^{(4n+1)}(0) = 1, \quad f^{(4n+2)}(0) = 0, \quad f^{(4n+3)}(0) = -1.$$

Then

$$\begin{aligned} P_{2n+1,0}(x) &= x - \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots + (-1)^n \frac{x^{2n+1}}{(2n+1)!} \\ P_{2n+2,0}(x) &= P_{2n+1,0}(x) \end{aligned}$$

□

(2) $f(x) = \tan^{-1} x$, at $a = 0$. Find $P_{3,0}(x)$

Proof.

$$f'(x) = \frac{1}{1+x^2}, \quad f''(x) = \frac{-2x}{(1+x^2)^2} \quad \text{and} \quad f'''(x) = \frac{-2(1+x^2)^2 + 2x \cdot 2(1+x^2)^2 \cdot 2x}{(1+x^2)^4}.$$

Then $f(0) = 0,\ f'(0) = 1,\ f''(0) = 0$ and $f'''(0) = -2$. Hence,

$$P_{3,0}(x) = x - \frac{2x^3}{3!} = x - \frac{x^3}{3}.$$

□

(3) $f(x) = e^x$, at $a = 0$. Find $P_{n,0}(x)$.

Proof. For every $k = 0, 1, 2, \cdots$, $f^{(k)}(x) = e^x$. Hence, $f^{(k)}(0) = 1$. We have

$$P_{n,0}(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!}.$$

□

Exercise. If $f(x)$ is a n-degree polynomial in $x-a$, say $f(x) = c_0 + c_1(x-a) + c_2(x-a)^2 + \cdots + c_n(x-a)^n$, then

(1) $P_{k,a}(x) = c_0 + c_1(x-a) + c_2(x-a)^2 + \cdots + c_n(x-a)^k$ for every $0 \le k \le n$.

(2) $f(x) = P_{k,a}(x)$ for every $k \ge n$.

(3) $f(x) = P_{k,b}(x)$ for every $k \neq n$ and every $b \in \mathbb{R}$.

❑ Approximation of $f(x)$ by $P_{n,a}(x)$

So far, we only know that the Taylor polynomial $P_{n,a}(x)$ is defined by the first n times derivatives of f at a. But we don't figure out the connection between $f(x)$ and $P_{n,a}(x)$.

Observe that $P_{1,a}(x) = f(a) + f'(a)(x-a)$ is the linear approximation (introduced in Section 5.6). Then

$$\frac{f(x) - P_{1,a}(x)}{x-a} = \frac{f(x) - f(a) - f'(a)(x-a)}{x-a} = \frac{f(x) - f(a)}{x-a} - f'(a) \longrightarrow 0 \quad \text{as} \quad x \to a.$$

Consider $P_{2,a}(x) = f(a) + f'(a)(x-a) + \dfrac{f''(a)}{2}(x-a)^2$. Then

$$\begin{aligned}\frac{f(x) - P_{2,a}(x)}{(x-a)^2} &= \frac{f(x) - f(a) - f'(a)(x-a) - \frac{f''(a)(x-a^2)}{2}}{(x-a)^2} \\ &= \frac{f(x) - f(a) - f'(a)(x-a)}{(x-a)^2} - \frac{f''(a)}{2}\end{aligned}$$

Therefore,

$$\lim_{x \to a} \frac{f(x) - P_{2,a}(x)}{(x-a)^2} \overset{L.H}{=} \lim_{x \to a} \frac{f'(x) - f'(a)}{2(x-a)} - \frac{f''(a)}{2} = 0$$

provided $f'(x)$ exists as x near a.

Question: Is there similar result for every $n \in \mathbb{N}$?

Theorem 12.1.4. *Suppose that f is a function such that $f'(a),\ f''(a), \cdots, f^{(n)}(a)$ exist. Then*

$$\lim_{x\to a}\frac{f(x)-P_{n,a}(x)}{(x-a)^n}=0$$

Proof. Consider

$$\frac{f(x)-P_{n,a}(x)}{(x-a)^n}=\frac{f(x)-\sum_{k=0}^{n-1}\frac{f^{(k)}(a)}{k!}(x-a)^k}{(x-a)^n}-\frac{f^{(n)}(a)}{n!}.$$

Let $Q(x)=f(x)-\displaystyle\sum_{k=0}^{n-1}\frac{f^{(k)}(a)}{k!}(x-a)^k$ and $g(x)=(x-a)^n$. Then, for $1\le i\le n-1$,

$$Q^{(i)}(x)=f^{(i)}(x)-f^{(i)}(a)-f^{(i+1)}(a)(x-a)-\cdots-\frac{f^{(n-1)}(a)(x-a)^{n-i-1}}{1\cdot 2\cdots(n-i-1)}.$$

Hence, $\displaystyle\lim_{x\to a}Q^{(i)}(x)=0$ for $i=0,1,2,\cdots,n-1$. On the other hand,

$$g^{(i)}(x)=n(n-1)\cdots(n-i+1)(x-a)^{n-i}$$

and hence $\displaystyle\lim_{x\to a}g^{(i)}(x)=0$ for $i=0,1,2,\cdots,n-1$. By applying L'Hôpital's Rule $n-1$ times,

$$\begin{aligned}
\lim_{x\to a}\frac{f(x)-P_{n,a}(x)}{(x-a)^n} &= \lim_{x\to a}\frac{Q(x)}{g(x)}-\frac{f^{(n)}(a)}{n!}\\
&\overset{L.H.}{=} \lim_{x\to a}\frac{Q'(x)}{g'(x)}-\frac{f^{(n)}(a)}{n!}\\
&\overset{L.H.}{=} \quad\vdots\\
&\overset{L.H.}{=} \lim_{x\to a}\frac{Q^{(n-1)}(x)}{g^{(n-1)}(x)}-\frac{{}^{(n)}(a)}{n!}\\
&= \lim_{x\to a}\frac{f^{(n-1)}(x)-f^{(n-1)}(a)}{n!(x-a)}-\frac{f^{(n)}(a)}{n!}\\
&= 0.
\end{aligned}$$

□

Note. Theorem 12.1.4 says that the more differentiabilities of f at a has, the better approximation of f by $P_{n,a}(x)$ is when x is near a.

❑ Local behaviors of functions

We recall the Second Derivative Test: $f'(x)$ is continuous near a and $f'(a)=0$.

(i) If $f''(a)>0$ then f has a minimum at a.

(ii) If $f''(a)<0$ then f has a maximum at a.

(iii) If $f''(a)=0$ then the test is inconclusive.

Heuristically, we follow the same idea that if $f'(a) = f''(a) = \cdots = f^{(n-1)}(a) = 0$ and $f^{(n)}(a) \neq 0$, then the sign of $f^{(n)}(a)$ might give some information about the local behavior of f near a. For example

$$P_{5,a}(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \frac{f^{(4)}(a)}{4!}(x-a)^4 + \frac{f^{(5)}(a)}{5!}(x-a)^5.$$

Then

$$\frac{f(x) - P_{5,a}(x)}{(x-a)^5} \overset{x\to a}{\longrightarrow} 0 \quad \Longrightarrow \quad \left|f(x) - P_{5,a}(x)\right| \ll |x-a|^5 \quad \text{as } x \text{ is close to } a. \tag{12.1}$$

Suppose that $f'(a) = f''(a) = 0$ and $f^{(3)}(a) \neq 0$. Then

$$P_{5,a}(x) = f(a) + (x-3)^3\Big[\frac{f'''(a)}{3!} + \frac{f^{(4)}(a)}{4!}(x-a) + \frac{f^{(5)}(a)}{5!}(x-a^2)\Big]. \tag{12.2}$$

For the bracket in (12.2), the term $\frac{f'''(a)}{3!}$ is dominated as x is near a. Hence, $\frac{f'''(a)}{3!}(x-a)^3$ determines the behavior of $P_{5,a}(x)$ when x is near a. Also, we may obtain that the behavior of $f(x)$, as x is near a, is like $\frac{f'''(a)}{3!}(x-a)^3$ by (12.1).

Roughly speaking, if a function f has sufficiently many derivatives at a, the first nonzero derivative $f^{(k)}(a)$ will dominate the behavior of f when x is near a.

Theorem 12.1.5. *Suppose that a function $f(x)$ satisfies*

$$f'(a) = f''(a) = \cdots = f^{(n-1)}(a) = 0 \quad \textit{and} \quad f^{(n)}(a) \neq 0.$$

(1) If n is even and $f^{(n)}(a) > 0$, then f has a local minimum at a.

(2) If n is even and $f^{(n)}(a) < 0$, then f has a local maximum at a.

(3) If n is odd, then f has neither a local maximum nor a local minimum at a.

Proof. W.L.O.G, we may assume that $f(a) = 0$. Otherwise we may take $f(x) - f(a)$.

$$\begin{aligned} P_{n,a}(x) &= f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots \frac{f^{(n-1)}(a)}{(n-1)!}(x-a)^{n-1} + \frac{f^{(n)}(a)}{n!}(x-a)^n \\ &= \frac{f^{(n)}(a)}{n!}(x-a)^n. \end{aligned}$$

By Theorem 12.1.4, $0 = \lim\limits_{x\to a}\frac{f(x) - P_{n,a}(x)}{(x-a)^n} = \lim\limits_{x\to a}\Big[\frac{f(x)}{(x-a)^n} - f^{(n)}(a)\Big]$. Then $\frac{f(x)}{(x-a)^n}$ has the same sign as $f^{(n)}(a)$ when x is sufficiently close to a.

(I) If n is even, then $(x-a)^n > 0$ for $x \neq a$ and therefore $f(x)$ and $f^{(n)}(a)(x-a)^n$ have the same sign when x is close to a.

  (i) For $f^{(n)}(a) > 0$, $f(x) > 0 = f(a)$ when x is close to a. Hence, f has a local minimum at a.

(ii) For $f^{(n)}(a) < 0$, $f(x) < 0 = f(a)$ when x is close to a. Hence, f has a local maximum at a.

(II) If n is odd, $(x-a)^n \begin{cases} > 0, & \text{when } x > a \\ < 0, & \text{when } x < a \end{cases}$, then $f(x)$ and $f^{(n)}(a)(x-a)^n$ have the same sign for $x > a$ and different sign for $x < a$. That is,

(i) For $f^{(n)}(a) > 0$, $f(x) \begin{cases} > 0, & \text{when } x > a, \\ < 0, & \text{when } x < a. \end{cases}$

(ii) For $f^{(n)}(a) < 0$, $f(x) \begin{cases} < 0, & \text{when } x > a, \\ > 0, & \text{when } x < a. \end{cases}$

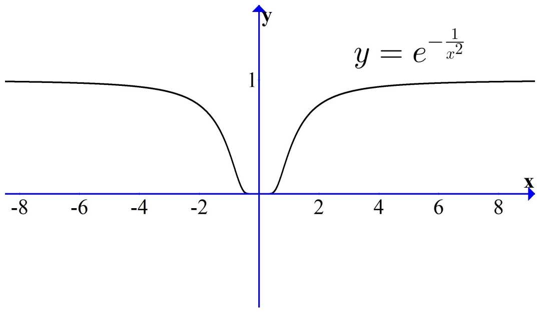
Hence, f has neither maximum nor minimum at a.

□

Note. Theorem 12.1.5 is inconclusive if $f^{(k)}(a) = 0$ for every $k \in \mathbb{N}$. For example,

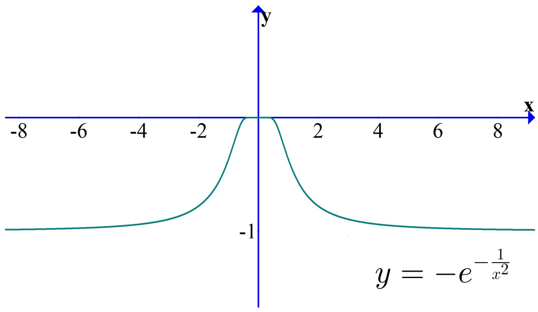
$$f(x) = \begin{cases} e^{-\frac{1}{x^2}} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

$f^{(k)}(0) = 0 \quad$ for all $k \in \mathbb{N}$.



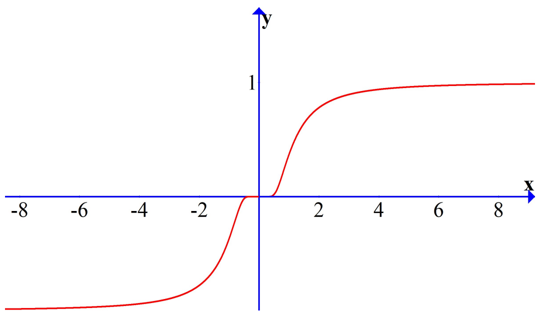
$$f(x) = \begin{cases} -e^{-\frac{1}{x^2}} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

$f^{(k)}(0) = 0 \quad$ for all $k \in \mathbb{N}$.



$$f(x) = \begin{cases} e^{-\frac{1}{x^2}} & \text{if } x > 0 \\ -e^{-\frac{1}{x^2}} & \text{if } x < 0 \\ 0 & \text{if } x = 0 \end{cases}$$

$f^{(k)}(0) = 0 \quad$ for all $k \in \mathbb{N}$.



❑ Uniqueness of $P_{n,a,f}$

Definition 12.1.6. Let f and g be two functions. We say that f and g are "*equal up to order n at a*" if

$$\lim_{x\to a} \frac{f(x) - g(x)}{(x-a)^n} = 0.$$

Remark. If $f(x)$ has nth derivative at a, then $f(x)$ and $P_{n,a,f}(x)$ are equal up to order n at a.

Question: Is there any polynomial $Q(x)$, different from $P_{n,a,f}(x)$, of degree less than or equal to n such that $f(x)$ and $Q(x)$ are equal up to order n at a?
Answer: No, by the following theorem.

Theorem 12.1.7. *Let P and Q be two polynomials in $(x-a)$, of degree less than or equal to n. Suppose that P and Q are equal up to order n at a. Then $P = Q$.*

Proof. We claim that if $R(x)$ is a polynomial of degree less than or equal to n and $\lim\limits_{x\to a} \dfrac{R(x)}{(x-a)^n} = 0$, then $R(x) \equiv 0$.

Proof of claim: Expressing $R(x)$ as a polynomial in $(x-a)$

$$R(x) = b_0 + b_1(x-a) + b_2(x-a)^2 + \cdots + b_n(x-a)^n,$$

we want to show that $b_i = 0$ for $i = 0, 1, 2, \cdots, n$ by induction.

Since $\lim\limits_{x\to a} \dfrac{R(x)}{(x-a)^n} = 0$, we have

$$0 \le \lim_{x\to a} |R(x)| \le \lim_{x\to a} |(x-a)|^n = 0.$$

Then $R(a) = \lim\limits_{x\to a} R(x) = 0$. Thus, for $i = 0$, $b_0 = 0$ and $R(x) = b_1(x-a) + \cdots + b_n(x-a)^n$.

If $b_0 = b_1 = \cdots = b_i = 0$ for $1 \le i < n$, then $R(x) = b_{i+1}(x-a)^{i+1} + \cdots + b_n(x-a)^n$. By using the similar argument as above, since $\lim\limits_{x\to a} \dfrac{R(x)}{(x-a)^n} = 0$, we have

$$\lim_{x\to a} \left| \frac{R(x)}{(x-a)^{i+1}} \right| \le \lim_{x\to a} |x-a|^{n-(i+1)} = 0.$$

Hence,

$$0 = \lim_{x\to a} \frac{R(x)}{(x-a)^{i+1}} = \lim_{x\to a} b_{i+1} + b_{i+2}(x-a) + \cdots + b_n(x-a)^{n-(i+1)} = b_{i+1}.$$

By the induction, we have $b_0 = b_1 = \cdots = b_n = 0$ and the claim is proved.

Now, define $R(x) = P(x) - Q(x)$. Since P and Q are equal up to order n at a, $R(x)$ is a polynomial of degree less than or equal to n and

$$\lim_{x\to a} \frac{R(x)}{(x-a)^n} = \lim_{x\to a} \frac{P(x)-Q(x)}{(x-a)^n} = 0.$$

By the claim, $R(x) \equiv 0$ and hence $P(x) \equiv Q(x)$. □

Corollary 12.1.8. *Suppose that f has nth derivative at a and P is a polynomial in $(x-a)$ of degree less than or equal to n which equals f up to order n at a. Then $P(x) = P_{n,a,f}(x)$.*

Proof. Since

$$\lim_{x\to a} \frac{P(x) - P_{n,a,f}(x)}{(x-a)^n} = \lim_{x\to a} \frac{P(x) - f(x)}{(x-a)^n} + \lim_{x\to a} \frac{f(x) - P_{n,a,f}(x)}{(x-a)^n} = 0,$$

$P(x)$ and $P_{n,a,f}(x)$ are equal up to order n at a. Also, P and $P_{n,a,f}(x)$ are polynomials of degree less than or equal to n. By Theorem 12.1.7, $P(x) = P_{n,a,f}(x)$. □

Remark. In Corollary 12.1.8, the hypothesis that "f has nth derivative at a" is necessary. There exists some function f such that some polynomials are equal to f of order n at a, but f does not have n times derivatives at a. Hence, $P_{n,a,f}(x)$ does not exists. For example,

$$f(x) = \begin{cases} x^{n+1} & \text{if } x \text{ is irrational} \\ 0 & \text{if } x \text{ is rational.} \end{cases}$$

Then $P(x) = 0$ and f are equal up to order n at 0. On the other hand, $f'(0)$ exists but $f'(x)$ does not exist for every $x \neq 0$. Hence, $f''(0)$ does not exist.

Remark. Corollary 12.1.8 gives another method to find the nth degree Taylor polynomial of f. That is, to find a polynomial $P(x)$ of degree n such that

$$\lim_{x\to a} \frac{f(x) - P(x)}{(x-a)^n} = 0.$$

Then $P(x) = P_{n,a,f}(x)$. For example,

$$\begin{aligned} \tan^{-1} x = \int_0^x \frac{1}{1+t^2}\,dt &= \int_0^x 1 - t^2 + t^4 - t^6 + \cdots + (-1)^n t^{2n} + \frac{(-1)^{n+1} t^{2n+2}}{1+t^2}\,dt \\ &= \underbrace{x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots + \frac{(-1)^n x^{2n+1}}{2n+1}}_{P(x)\,:\,2n+1 \text{ degree polynomial}} + (-1)^{n+1} \int_0^x \frac{t^{2n+2}}{1+t^2}\,dt \end{aligned}$$

Consider

$$\left|\frac{\tan^{-1}(x) - P(x)}{x^{2n+1}}\right| = \left|\frac{\int_0^x \frac{t^{2n+2}}{1+t^2}\,dt}{x^{2n+1}}\right| \le \left|\frac{\int_0^x t^{2n+2}\,dt}{x^{2n+1}}\right| = \frac{1}{2n+3}\left|\frac{x^{2n+3}}{x^{2n+1}}\right| \longrightarrow 0 \quad \text{as} \quad x \to 0.$$

Since $P(x)$ is equal to $\tan^{-1} x$ of order $2n+1$ at 0,

$$P_{2n+1,0}(x) = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots + \frac{(-1)^n x^{2n+1}}{2n+1}.$$

Observe that

$$\begin{aligned} \tan^{-1} x &= x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots + \frac{(-1)^n x^{2n+1}}{2n+1} + (-1)^{n+1} \int_0^x \frac{t^{2n+2}}{1+t^2}\,dt \\ &= P_{2n+1,0}(x) + (-1)^{n+1} \int_0^x \frac{t^{2n+2}}{1+t^2}\,dt. \end{aligned}$$

Then

$$\left|\tan^{-1} x - P_{2n+1,0}(x)\right| = \left|\int_0^x \frac{t^{2n+2}}{1+t^2}\,dt\right| \le \frac{|x|^{2n+3}}{2n+3}.$$

Hence, for some $|x_0| \le 1$,

$$|\tan^{-1} x_0 - P_{2n+1,0}(x_0)| \le \frac{|x_0|^{2n+3}}{2n+3} < \frac{1}{2n+3}.$$

We can estimate $\tan^{-1} x_0$ by computing $P_{2n+3,0}(x_0)$ with error less than $\dfrac{1}{2n+3}$.

12.2 Estimating Error and Taylor Theorem

In the previous section, we have learned that if f has n times derivatives at a, then

$$|f(x) - P_{n,a}(x)| \ll |x-a|^n \quad \text{as } x \text{ is sufficiently close to } a.$$

Question: Can we estimate the difference between $f(x)$ and $P_{n,a}(x)$ when x is in some interval of a?

Definition 12.2.1. We define the remainder term $R_{n,a}(x)$ by

$$R_{n,a}(x) = f(x) - P_{n,a}(x)$$

By the definition of the remainder,

$$f(x) = P_{n,a}(x) + R_{n,a}(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + R_{n,a}(x).$$

Observe that

$$\begin{aligned}
f(x) &\overset{F.T.C}{=} f(a) + \underbrace{\int_a^x f'(t)\,dt}_{R_{0,a}(x)} \\
&\overset{I.B.P}{=} f(a) + f'(t)t\Big|_a^x - \int_a^x f''(t)t\,dt \\
&= f(a) + f'(x)x - f'(a)a - \int_a^x f''(t)t\,dt \\
&= f(a) + f'(a)(x-a) - f'(a)x + f'(x)x - \int_a^x f''(t)t\,dt \\
&= f(a) + f'(a)(x-a) + \big(f'(x) - f'(a)\big)x - \int_a^x f''(t)t\,dt \\
&\overset{I.B.P}{=} f(a) + f'(a)(x-a) + \Big(\int_a^x f''(t)\,dt\Big)x - \int_a^x f''(t)t\,dt \\
&= f(a) + f'(a)(x-a) + \underbrace{\int_a^x f''(t)(x-t)\,dt}_{R_{1,a}(x)} \\
&\overset{I.B.P}{=} f(a) + f'(a)(x-a) - f''(t)\cdot\frac{(x-t)^2}{2}\Big|_a^x + \int_a^x \frac{f'''(t)}{2}(x-t)^2\,dt \\
&= f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + \underbrace{\int_a^x \frac{f'''(a)}{2}(x-t)^2\,dt}_{R_{2,a}(x)}
\end{aligned}$$

By induction, if $f^{(n+1)}$ is continuous on $[a, x]$, then

$$R_{n,a}(x) = \int_a^x \frac{f^{(n+1)}(t)}{n!}(x-t)^n\,dt \qquad \text{(integral form)}$$

❑ <u>Taylor Theorem</u>

Theorem 12.2.2. *(Taylor Theorem) Let $f(t)$ be a $n+1$ times differentiable function on $[a, x]$ and $R_{n,a}(x)$ be defined by*

$$f(x) = f(a) + f'(a)(x-a) + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + R_{n,a}(x).$$

Then

(a) (Cauchy form)

$$R_{n,a}(x) = \frac{f^{(n+1)}(\xi)}{n!}(x-\xi)^n(x-a) \qquad \text{for some } \xi \in (a,x).$$

(b) (Lagrange form)

$$R_{n,a}(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1} \qquad \text{for some } \xi \in (a,x).$$

(c) (Integral form)

$$R_{n,a}(x) = \int_a^x \frac{f^{(n+1)}(t)}{n!}(x-t)^n \, dt$$

Proof.

> Recall the Cauchy Mean Value Theorem: If F and G are continous on $[a, x]$ and differentiable on (a, x), there exists $\xi \in (a, x)$ such that
>
> $$\frac{F(x)-F(a)}{G(x)-G(a)} = \frac{F'(\xi)}{G'(\xi)}.$$

Define F on $[a, x]$ by

$$F(t) = f(t) + f'(t)(x-t) + \cdots + \frac{f^{(n)}(t)}{n!}(x-t)^n.$$

Let G be a differentiable function on $[a, x]$ such that $G'(t) \neq 0$ on (a, x). By the Cauchy Mean Value Theorem, there exists a number $\xi \in (a, x)$ such that

$$\frac{F(x)-F(a)}{G(x)-G(a)} = \frac{F'(\xi)}{G'(\xi)}. \tag{12.3}$$

Also,

$$F(x) - F(a) = f(x) - \left[f(a) + f'(a)(x-a) + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n\right] = R_{n,a}(x)$$

and

$$F'(\xi) = \cancel{f'(\xi)} - \cancel{f'(\xi)} + \cancel{f''(\xi)(x-\xi)} - \cancel{f''(\xi)(x-\xi)} + \cancel{\cdots} + \frac{f^{(n+1)}(\xi)}{n!}(x-\xi)^n = \frac{f^{(n+1)}(\xi)}{n!}(x-\xi)^n.$$

By (12.3),

$$R_{n,a}(x) = \frac{f^{(n+1)}(\xi)}{n!}(x-\xi)^n \cdot \frac{G(x)-G(a)}{G'(\xi)}.$$

(a) Let $G(t) = t - a$. Then $G(x) - G(a) = x - a$ and $G'(\xi) = 1$. Hence,

$$R_{n,a}(x) = \frac{f^{(n+1)}(\xi)}{n!}(x-\xi)^n(x-a).$$

(b) Let $G(t) = (x-t)^{n+1}$. Then $G(x) - G(a) = -(x-a)^{n+1}$ and $G'(\xi) = -(n+1)(x-\xi)^n$. Hence,

$$R_{n,a}(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1}.$$

The part(c) is proved by using integration by parts □

Remark. In Theorem 12.2.2,

(i) the ξ in part(a) and part(b) are usually different.

(ii) the ξ in part(a) and part(b) depend on a and x.

(iii) by part(b), if $\left|f^{(n+1)}(t)\right| < M$ for all $t \in [a, x]$, then

$$\left|R_{n,a}(x)\right| < M \cdot \frac{|x-a|^{n+1}}{(n+1)!}.$$

(iv) by part(c), if $\left|f^{(n+1)}(t)\right| < M$, then

$$\left|R_{n,a}(x)\right| \leq \frac{M}{n!}\left|\int_a^x (x-t)^n\, dt\right| = \frac{M}{(n+1)!}\left| -(x-t)^{n+1}\Big|_a^x \right| = \frac{M}{(n+1)!}|x-a|^{n+1}.$$

Example 12.2.3. Estimate $\sin 2$ with error less than 0.0001.

Proof. Let $f(x) = \sin x$. Then $|f^{(n)}(x)| \leq 1$ $(= M)$ for every $x \in \mathbb{R}$ and $n = 0, 1, 2, \cdots$. The Taylor polynomial for f at 0 is

$$\begin{aligned} P_{2n+1,0}(x) &= \sum_{k=0}^{2n+1} \frac{f^{(k)}(0)}{k!}x^k \\ &= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots + (-1)^n \frac{x^{2n+1}}{(2n+1)!1} \\ &= \sum_{k=1}^{n} \frac{(-1)^k}{(2k+1)!}x^{2k+1}. \end{aligned}$$

Let $M = 1$. The remainder

$$|R_{2n+1,0}(x)| = \left|\sin x - P_{2n+1,0}(x)\right| \leq \left|\frac{f^{(2n+2)}(\xi)}{(2n+2)!}\right||x|^{2n+2} \leq \frac{|x|^{2n+2}}{(2n+2)!}.$$

Consider $\dfrac{2^{2n+2}}{(2n+2)!} < 0.0001$. Then $n \geq 5$ and

$$P_{11,0}(2) = 2 - \frac{2^3}{3!} + \frac{2^5}{5!} - \frac{2^7}{7!} + \frac{2^9}{9!} - \frac{2^{11}}{11!} \approx 0.909296136$$

and $\sin 2 \approx 0.90929743$. □

Remark. For a given number $x_0 \in \mathbb{R}$,

$$\left| \sin x_0 - P_{2n+1,0}(x_0) \right| \le \frac{1}{(2n+2)!} \underbrace{|x_0|^{2n+2}}_{\text{fixed number}} \longrightarrow 0 \qquad \text{as} \quad n \to \infty.$$

Hence, for any $x \in \mathbb{R}$, we can use n-degree polynomial to approximate $\sin x$ with error arbitrarily small by choosing n sufficiently large.

Notice that the choice of n depends on the error ε and the value of $|x|$. When the point x is far from the center "0", we should choose larger number n in order to keep the error still less than ε.

Example 12.2.4. Let $f(x) = e^x$. Then $f^{(k)}(x) = e^x$ for every $k \in \mathbb{N}$ and

$$P_{n,0}(x) = \sum_{k=0}^{n} \frac{f^k(0)}{k!} x^k = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!}$$

and for $x > 0$,

$$e^x - P_{n,0}(x) = \int_0^x \frac{e^t}{n!}(x-t)^n \, dt \le \frac{e^x}{n!} \int_0^x (x-t)^n \, dt = \frac{e^x}{(n+1)!} x^{n+1}.$$

To estimate the value of e with error less than 0.0001. Since $e < 3$, we have

$$\left| e - P_{n,0}(1) \right| \le \frac{e^1}{(n+1)!} < \frac{3}{(n+1)!} < 0.0001$$

Choose $n = 7$ and $R_8 \le \frac{1}{13440}$. We have $P_{7,0}(1) = 2.7182$.

Example 12.2.5. For $f(x) = \ln(1+x)$,

$$\begin{aligned} f'(x) &= \frac{1}{1+x} = (1+x)^{-1}, \quad f''(x) = -(1+x)^{-2}, \quad f'''(x) = 2(1+x)^{-3}, \cdots \\ f^{(k)}(x) &= (-1)^{k+1}(k-1)!(1+x)^{-k} \qquad \text{for } k = 1, 2, \cdots . \end{aligned}$$

Then

$$P_{n,0}(x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots + (-1)^{n-1} \frac{x^n}{n}$$

and

$$\begin{aligned} R_{n,0}(x) &= \frac{1}{(n+1)!} f^{(n+1)}(\xi)(x-0)^{n+1} \qquad \text{for some } \xi \in (0, x) \\ &= \frac{(-1)^{n+1} x^{n+1}}{n+1} \cdot \frac{1}{(1+\xi)^{n+1}}. \end{aligned}$$

If $x > 0$, then $\xi \in (0, x)$ and hence $\dfrac{1}{(1+\xi)^{n+1}} < 1$. We have $\left| R_{n,0}(x) \right| \le \dfrac{|x|^{n+1}}{n+1}$.

Example 12.2.6.

(a) Approximate the function $f(x) = \sqrt[3]{x}$ by a Taylor polynomial of degree 2 at $a = 8$.

(b) How accurate is this approximation when $7 \le x \le 9$.

Proof. (a) Compute

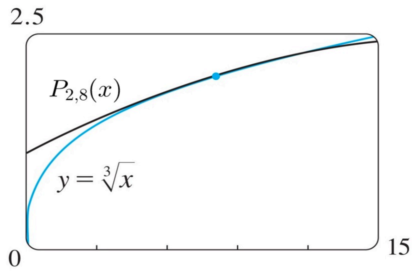
$$f'(x) = -\frac{1}{3}x^{-\frac{2}{3}},\ f''(x) = -\frac{2}{9}x^{-\frac{5}{3}},\ f'''(x) = \frac{10}{27}x^{-\frac{8}{3}}.$$

Then

$$f(8) = 2,\ f'(8) = \frac{1}{12},\ f''(8) = -\frac{1}{144}.$$

Hence, $P_{2,8}(x) = 2 + \frac{1}{12}(x-8) - \frac{1}{288}(x-8)^2$ and the approximation is

$$\sqrt[3]{x} \approx 2 + \frac{1}{12}(x-8) - \frac{1}{288}(x-8)^2.$$



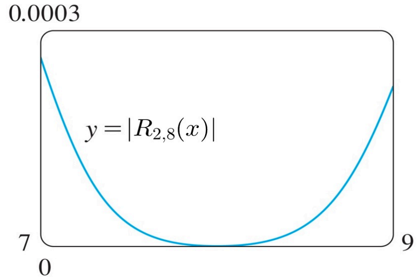
(b) To find a bound M such that $|f'''(x)| \le M$ for $7 \le x \le 9$, consider

$$|f'''(x)| = \frac{10}{27}|x|^{-\frac{8}{3}} \le \frac{10}{27} \cdot 7^{-\frac{8}{3}} \quad \text{for } 7 \le x \le 9.$$

Hence, for $7 \le x \le 9$,

$$|R_{2,8}(x)| \le \frac{1}{3!} \cdot \frac{10}{27} \cdot 7^{-\frac{8}{3}}|x-8|^3 \le \frac{0.0021}{3!} \cdot 1 < 0.0004.$$

Note. In fact, $|R_{2,8}(x)| < 0.0003$ for $7 \le x \le 9$.



□

Example 12.2.7. (a) What is the maximum error possible in using the approximation

$$\sin x \approx x - \frac{x^3}{3!} + \frac{x^5}{5!}$$

when $-0.3 \le x \le 0.3$? Use the approximation to find 12° correct to six decimal places.

(b) For what values of x is this approximation accurate to within 0.00005?

Proof. (a) (**Method 1: Alternating Series**) When $-0.3 \le x \le 0.3$, the series is an alternating series and

$$\frac{|x|^{2k+1}}{(2k+1)!} \le \frac{|x|^{2k-1}}{(2k-1)!} \quad \text{and} \quad \frac{|x|^{2k+1}}{(2k+1)!} \longrightarrow 0 \quad \text{as } k \to \infty.$$

By the alternating series estimation,

$$\left| \sin x - (x - \frac{x^3}{3!} + \frac{x^5}{5!}) \right| \le \frac{|x|^7}{7!} \le \frac{(0.3)^7}{7!} \approx 4.3 \times 10^{-8}.$$

Then

$$\sin 12^\circ = \sin(\frac{\pi}{15}) \approx \frac{\pi}{15} - \frac{1}{3!}(\frac{\pi}{15})^3 + \frac{1}{5!}(\frac{\pi}{15})^5 \approx 0.20791169$$

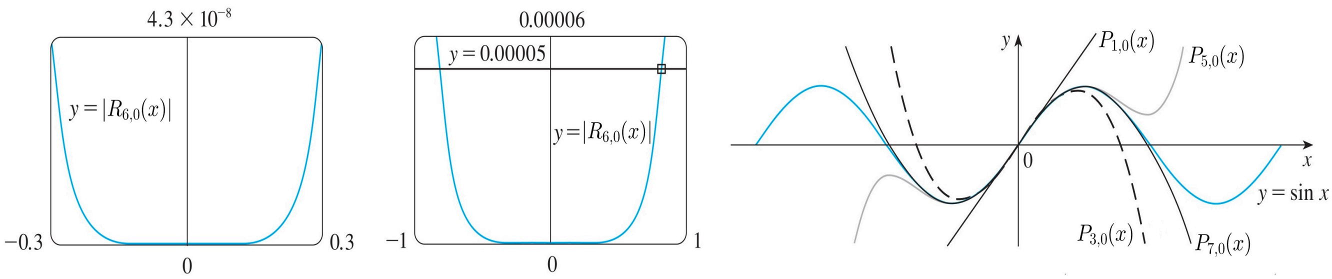
(**Method 2: Taylor's Inequality**) For $(x) = \sin x$, the polynomial $x - \frac{x^3}{3!} + \frac{x^5}{5!}$ is the 6th degree Taylor's polynomial for f at 0. The error

$$|R_{6,0}(x)| \le \frac{M}{7!}|x|^7$$

where M is a number such that $|f^{(7)}(z)| \le M$ for $-0.3 \le z \le 0.3$. To find M, consider $f^{(7)}(x) = -\cos x$. When $-0.3 \le z \le 0.3$, $|-\cos z| \le |\cos 0| = 1 = M$. Then

$$|R_{6,0}(x)| \le \frac{1}{7!} \cdot (0.3)^7 < 4.3 \times 10^{-8}.$$

(b) Consider $|R_{6,0}(x)| \le \frac{|x|^7}{7!} < 0.00005$. Then $|x| \le (0.252)^{1/7} \approx 0.821$.



□

12.3 Power Series

In the previous section, we know that a function f with sufficiently many times derivatives at a could be approximated by its Taylor polynomial $P_{n,a}(x)$. Some examples reveal that the approximation become better (at least near a) if we choose larger degree Taylor polynomials. In fact, this observation is not exactly true (and we will discuss in the later sections). We want to ask whether a smooth function can be expressed as a "power series"

$$f(x) = c_0 + c_1(x-a) + c_2(x-a)^2 + \cdots + c_n(x-a)^n + \cdots .$$

Moreover, if f has the power series expression, what is it?

Definition 12.3.1. (a) A power series is a series of the form

$$\sum_{n=0}^{\infty} c_n x^n = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + \cdots$$

where x is a variable and the c_n are constants called the "*coefficients*" of the series.

(b) In general, a series of the form

$$\sum_{n=0}^{\infty} c_n(x-a)^n = c_0 + c_1(x-a) + c_2(x-a)^2 + \cdots$$

is called a "*power series in* $(x-a)$" or a "*power series centered at* a" or a "*power series about* a".

Example 12.3.2. $\sum_{n=0}^{\infty} x^n = 1 + x + x^2 + x^3 + \cdots$ is a power series.

For given $x = x_0$, we should determine whether the series $\sum_{n=0}^{\infty} c_n(x_0-a)^n$ converges or diverges.

Definition 12.3.3. (a) We say that a power series $\sum_{n=0}^{\infty} c_n(x-a)^n$ converges

(i) at x_0 if $\sum_{n=0}^{\infty} c_n(x_0-a)^n$ converges;

(ii) on the set S if $\sum_{n=0}^{\infty} c_n(x-a)^n$ converges at each $x \in S$.

(b) If we regard a series $f(x) = \sum_{n=0}^{\infty} c_n(x-a)^n$ as a function, then the domain of $f(x)$ is the set of all x for which the series converges.

Remark. A poswer series $\sum_{n=0}^{\infty} c_n(x-a)^n$ always converges at its center a. In fact, it converges to the constant term c_0.

Example 12.3.4. Consider the series $\sum_{n=0}^{\infty} x^n = 1 + x + x^2 + x^3 + \cdots$ as a geometric series with ratio x. Then the series converges when $|x| < 1$ and diverges when $|x| \geq 1$. Therefore, the domain of $\sum_{n=0}^{\infty} x^n$ is $(-1, 1)$.

Example 12.3.5. For what values of x is the series $\sum_{n=0}^{\infty} n!x^n$ convergent?

Proof. (Idea: using the ratio test or root test)

Let $a_n = n!x^n$. Then $\left|\frac{a_{n+1}}{a_n}\right| = \left|\frac{(n+1)!x^{n+1}}{n!x^n}\right| = (n+1)|x|$. If $x = 0$, $\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| = 0 < 1$ and if $x \neq 0$, $\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| = \infty$.

By the Ratio Test, the series converges when $x = 0$. □

Example 12.3.6. For what values of x does the series $\sum_{n=1}^{\infty} \frac{(x-3)^n}{n}$ converge?

Proof. Let $a_n = \dfrac{(x-3)^n}{n}$. Then

$$\left|\frac{a_{n+1}}{a_n}\right| = \left|\frac{\frac{(x-3)^{n+1}}{n+1}}{\frac{(x-3)^n}{n}}\right| = \frac{n}{n+1}|x-3| \longrightarrow |x-3| \quad \text{as } n \to \infty.$$

By the Ratio Test, if $|x-3| < 1$ (i.e. $2 < x < 4$), the series $\displaystyle\sum_{n=1}^{\infty} \frac{(x-3)^n}{n}$ converges and if $|x-3| > 1$ (i.e. $x < 2$ or $x > 4$) the series $\displaystyle\sum_{n=1}^{\infty} \frac{(x-3)^n}{n}$ diverges.

For $|x-3| = 1$,

(i) When $x-3 = 1$ (i.e. $x = 4$), $\displaystyle\sum_{n=1}^{\infty} \frac{(x-3)^n}{n} = \sum_{n=1}^{\infty} \frac{1}{n}$ diverges (p-series, $p = 1$).

(ii) When $x-3 = -1$ (i.e. $x = 2$), $\displaystyle\sum_{n=1}^{\infty} \frac{(x-3)^n}{n} = \sum_{n=1}^{\infty} \frac{(-1)^n}{n}$ converges by the alternating series test.

Hence, the power series $\displaystyle\sum_{n=1}^{\infty} \frac{(x-3)^n}{n}$ converges on $[2, 4)$ and diverges on $(-\infty, 2) \cup [4, \infty)$. □

Example 12.3.7. (Bessel function of order 0) Find the domain of the Bessel function

$$J_0(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{2^{2n}(n!)^2}.$$

Proof. Let $a_n = \dfrac{(-1)x^{2n}}{2^{2n}(n!)^2}$. Then

$$\left|\frac{a_{n+1}}{a_n}\right| = \left|\frac{\frac{(-1)x^{2(n+1)}}{2^{2(n+1)}[(n+1)!]^2}}{\frac{(-1)x^{2n}}{2^{2n}(n!)^2}}\right| = \frac{1}{2^2(n+1)^2}|x|^2.$$

For every $x \in \mathbb{R}$,

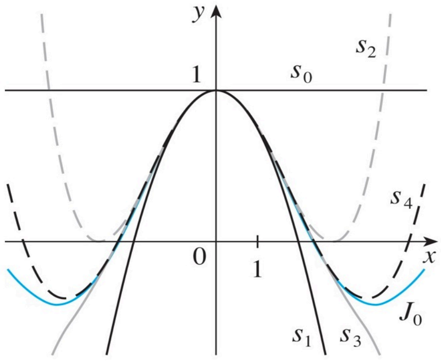
$$\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| = \lim_{n\to\infty} \frac{1}{2^2(n+1)^2}|x|^2 = 0 < 1.$$

By the Ratio Test, the series converges for every x and the domain of $J_0(x)$ is $\mathbb{R}$. □

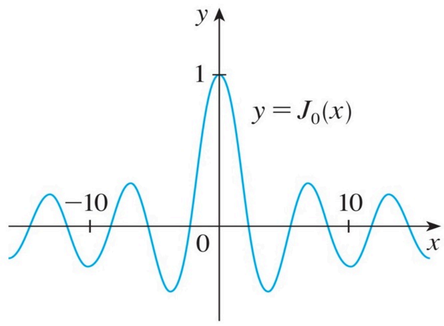
From the above examples, we observe that the region where the power series $\displaystyle\sum_{n=0}^{\infty} c_n(x-a)^n$ is convergent has always turned out to be an interval (e.g. $\{a\}$, finite interval, $(-\infty, \infty)$ etc).

Question: Is the set where a power series converges an interval (including the case that converges at a single point)?

Theorem 12.3.8. *For a given power series* $\displaystyle\sum_{n=0}^{\infty} c_n(x-a)^n$*,*

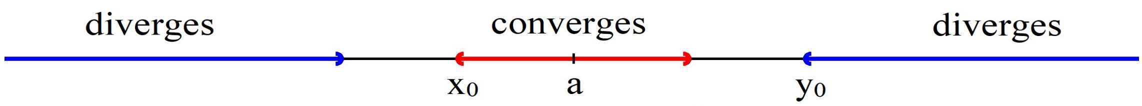


Partial sums of the Bessel function J_0



(a) if the series converges at $x_0 \neq a$, then it converges absolutely at every x with $|x - a| < |x_0 - a|$.

(b) if the series diverges at y_0, then it diverges at every x with $|x - a| > |y_0 - a|$.



Proof.

(a) Since $\displaystyle\sum_{n=0}^{\infty} c_n(x_0 - a)^n$ converges, we have $\displaystyle\lim_{n\to\infty} |c_n(x_0 - a)^n| = 0$. Thus, there exists $N \in \mathbb{N}$ such that for every $n > N$ such that $|c_n(x_0 - a)^n| < 1$.

Let x satisfy $|x-a| < |x_0-a|$. Since $\left|\dfrac{x-a}{x_0-a}\right| < 1$, the series $\displaystyle\sum_{n=N+1}^{\infty} \left|\frac{x-a}{x_0-a}\right|^n$ converges. Also,

$$|c_n(x-a)^n| = |c_n(x_0-a)^n| \left|\frac{x-a}{x_0-a}\right|^n < \left|\frac{x-a}{x_0-a}\right|^n \qquad \text{for } n > N.$$

By the comparison test, the series $\displaystyle\sum_{n=N+1}^{\infty} |c_n(x-a)^n|$ conveges and hence $\displaystyle\sum_{n=1}^{\infty} |c_n(x-a)^n|$ also converges.

(b) Let z_0 be a number such that $|y_0 - a| < |z_0 - a|$. Assume that the series $\sum_{n=0}^{\infty} c_n(z_0 - a)^n$ converges. By part(a), for every x with $|x - a| < |z_0 - a|$, the series $\sum_{n=0}^{\infty} c_n(x - a)^n$ converges. Hence the series $\sum_{n=0}^{\infty} c_n(y_0 - a)^n$ converges. It contradicts the hypothesis that the series $\sum_{n=0}^{\infty} c_n(y_0 - a)^n$ diverges. Therefore, $\sum_{n=0}^{\infty} c_n(z_0 - a)^n$ must diverges.

Since z_0 is an arbitrary number with $|y_0 - a| < |z_0 - a|$, part(b) is proved.

□

Theorem 12.3.9. *For a given power series* $\sum_{n=0}^{\infty} c_n(x - a)^n$*, there are only three possibilities:*

(i) The series converges only when $x = a$*.*

(ii) The series converges for all x*.*

(iii) There is a positive number R *such that the series conveges if* $|x - a| < R$ *and diverges if* $|x - a| > R$*.*

Proof. (Exercise) □

Note.

(a) The number R in part(c) of Theorem 12.3.9 is called the "*radius of convergence*".

(b) By convention, we define the radius of convergence as $R = 0$ in part(a), and as $R = \infty$ in part(b).

(c) The interval which consists of all values of x for which the series converges is called the "*interval of convergence*" of the power series.

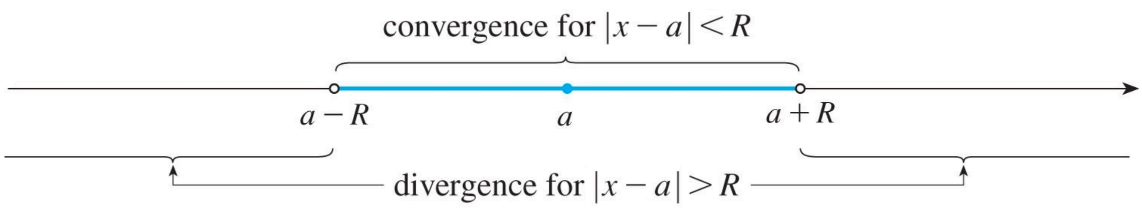
(d) In order to find the interval of convergence in part(c) if the radius of convergence is obtained, we still need to consider the endpoints of the interval. That is, to consider whether the series converges at the endpoints $x = a - R$ and $x = a + R$. All situations would occur. Hence, the interval of convergence could be $(a - R, a + R)$, $[a - R, a + R)$, $(a - R, a + R]$ or $[a - R, a + R]$.

Example 12.3.10.

Question: How to find the radius of convergece for a given power series? What is the connection between the coefficients and the radius of convergence?

Suppose that $\lim_{n\to\infty} \left|\frac{c_{n+1}}{c_n}\right| = L$. Let $a_n = c_n(x - a)^n$. Then

$$\left|\frac{a_{n+1}}{a_n}\right| = \left|\frac{c_{n+1}}{c_n}\right| |x - a| \longrightarrow L|x - a| \qquad \text{as } n \to \infty.$$



	Series	Radius of convergence	Interval of convergence
Geometric series	$\sum_{n=0}^{\infty} x^n$	$R=1$	$(-1, 1)$
Example 1	$\sum_{n=0}^{\infty} n!\, x^n$	$R=0$	$\{0\}$
Example 2	$\sum_{n=1}^{\infty} \frac{(x-3)^n}{n}$	$R=1$	$[2, 4)$
Example 3	$\sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{2^{2n}(n!)^2}$	$R=\infty$	$(-\infty, \infty)$

By the ratio test,

$$\text{if } L|x-a|<1 \iff |x-a|<\frac{1}{L}, \text{ then the series } \sum_{n=0}^{\infty} c_n(x-a)^n \text{ is convergent;}$$

$$\text{if } L|x-a|>1 \iff |x-a|>\frac{1}{L}, \text{ then the series } \sum_{n=0}^{\infty} c_n(x-a)^n \text{ is divergent.}$$

Hence, the radius of convergence of the series is $R=\frac{1}{L}$ where $\lim_{n\to\infty}\left|\frac{c_{n+1}}{c_n}\right|=L$.

Example 12.3.11. Find the radius and interval of convergence of the series $\sum_{n=0}^{\infty}\frac{x^n}{n!}$.

Proof. Let $a_n=\frac{x^n}{n!}$. Then

$$\left|\frac{a_{n+1}}{a_n}\right| = \left|\frac{\frac{x^{n+1}}{(n+1)!}}{\frac{x^n}{n!}}\right| = \frac{|x|}{n+1}.$$

Hence, for every $x\in\mathbb{R}$, $\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| = \lim_{n\to\infty}\frac{|x|}{n+1}=0$. The series $\sum_{n=0}^{\infty}\frac{x^n}{n!}$ converges for every $x\in\mathbb{R}$. The radius of convergence is ∞ and the interval of convergence is $\mathbb{R}$. □

Example 12.3.12. Find the radius and interval of convergence of the series $\sum_{n=0}^{\infty} n^n x^n$.

Proof. For every $x \neq 0$, if $n \in \mathbb{N}$ and $n > \frac{2}{|x|}$, then $|nx| > 2$. Hence,

$$\lim_{n\to\infty} |n^n x^n| = \lim_{n\to\infty} |nx|^n \geq \lim_{n\to\infty} 2^n = \infty.$$

By the test for divergence, the series $\sum_{n=0}^{\infty} n^n x^n$ diverges at every $x \in 0$. The radius of convergence is 0 and the interval of convergence is $\{0\}$.

□

Example 12.3.13. Find the radius and interval of convergence of the series $\sum_{n=0}^{\infty} \frac{(-3)^n x^n}{\sqrt{n+1}}$.

Proof. Let $a_n = \frac{(-3)^n x^n}{\sqrt{n+1}}$. Then

$$\left|\frac{a_{n+1}}{a_n}\right| = \left|\frac{\frac{(-3)^{n+1} x^{n+1}}{\sqrt{n+2}}}{\frac{(-3)^n x^n}{\sqrt{n+1}}}\right| = 3\sqrt{\frac{n+1}{n+2}} \longrightarrow 3|x| \quad \text{as } n \to \infty.$$

By the Ratio Test,

(1) When $3|x| < 1 \iff |x| < \frac{1}{3}$, the power series is convergent.

(2) When $3|x| > 1 \iff |x| > \frac{1}{3}$, the power series is divergent.

(3) At the endpoints,

(i) if $x = \frac{1}{3}$, the series is $\sum_{n=0}^{\infty} \frac{(-1)^n}{\sqrt{n+1}}$ is convergnet by the alternating series test.

(ii) if $x = -\frac{1}{3}$, the series is $\sum_{n=0}^{\infty} \frac{1}{\sqrt{n+1}}$ is divergent (p-series, $p = \frac{1}{2} < 1$).

Hence, the radius of convergence is $\frac{1}{3}$ and the interval of convergence is $(-\frac{1}{3}, \frac{1}{3}]$. □

Example 12.3.14. Find the radius and interval of convergence of the series $\sum_{n=0}^{\infty} \frac{n(x+2)^n}{3^{n+1}}$.

Proof. Let $a_n = \frac{n(x+2)^n}{3^{n+1}}$. Then

$$\left|\frac{a_{n+1}}{a_n}\right| = \left|\frac{\frac{(n+1)(x+2)^{n+1}}{3^{n+2}}}{\frac{n(x+2)^n}{3^{n+1}}}\right| = \frac{n}{3(n+1)}|x+2| \longrightarrow \frac{1}{3}|x+2| \quad \text{as } n \to \infty.$$

By the Ratio Test,

(1) When $\frac{1}{3}|x + 2 < 1 \iff |x+2| < 3$, the power series is convergent.

(2) When $\frac{1}{3}|x+2>1 \iff |x+2|>3$, the power series is divergent.

(3) At the endpoints, consider $\frac{1}{3}|x+2|=1 \iff |x+2|=3$.

(i) If $x=1$, the series is $\frac{1}{3}\sum_{n=0}^{\infty} n=\infty$ is divergent.

(ii) If $x=-5$, the series is $\frac{1}{3}\sum_{n=0}^{\infty}(-1)^n n$ is divergent by the test for divergence.

Hence, the radius of convergence is 3 and the interval of convergence is $(-5,1)$. □

Remark.

(i) The Ratio Test (or Root Test) do not apply for the endpoints of the interval of convergence.

(ii) Theorem 12.3.9 is false for general series $\sum_{n=0}^{\infty} f_n(x)$.

❑ Operations for Power Series

When regarding power series as functions, we want to know whether some operations (such as addition, subtraction, multiplication, division, differentiation or integration) for functions also apply for power series and how they work.

Theorem 12.3.15. *Let $f(x)=\sum_{n=0}^{\infty} b_n(x-a)^n$ and $g(x)=\sum_{n=0}^{\infty} c_n(x-a)^n$ with the intervals of convergence $(a-L,a+L)$ and $(a-M,a+M)$ respectively. Let $R=\min(L,M)$. Then*

(a) $(f\pm g)(x)=\sum_{n=0}^{\infty}\big(b_n\pm c_n\big)(x-a)^n$ *on* $(a-R,a+R)$.

(b) $\big(f\cdot g\big)(x)=\sum_{n=0}^{\infty} d_n(x-a)^n$ *on* $(a-R,a+R)$ *where* $d_n=\sum_{k=0}^{n} b_k c_{n-k}$.

(c) $\dfrac{f(x)}{g(x)}=\sum_{n=0}^{\infty} e_n(x-a)^n$ *where* e_n *satisfies* $b_n=\sum_{k=0}^{n} c_k e_{n-k}$ *on* $(a-R,a+R)$.

Proof. (Exercise) □

Remark. Suppose that $f(x)=\sum_{n=0}^{\infty} b_n(x-a)^n$ and $g(x)=\sum_{n=0}^{\infty} c_n(x-a)^n$ both converge at $x_0\neq a$. Theorem 12.3.15 hold for all x with $|x-a|<|x_0-a|$ by using Theorem 12.3.8.

■ Term-by-term Differentiation and Integration

Suppose that $f(x) = \sum_{n=0}^{\infty} c_n(x-a)^n$ on $(a-R, a+R)$. It is natural to ask whether f is continuous, differentiable, integrable or has other properites. Moreover, if f is differentiable or integrable, what are its derivative function or antiderivative function?

Intuitively, we guess that

$$\begin{aligned} \frac{d}{dx}f(x) &= \frac{d}{dx}\Big(\sum_{n=0}^{\infty} c_n(x-a)^n\Big) \overset{??}{=} \sum_{n=0}^{\infty} \frac{d}{dx}\big[c_n(x-a)^n\big] = \sum_{n=0}^{\infty} c_n n(x-a)^{n-1} \\ \int f(x)\,dx &= \int \sum_{n=0}^{\infty} c_n(x-a)^n\,dx \overset{??}{=} \sum_{n=0}^{\infty} \int c_n(x-a)^n\,dx = C + \sum_{n=0}^{\infty} \frac{c_n}{n+1}(x-a)^{n+1} \end{aligned}$$

The interchange of infinite sum and limit ($\lim\limits_{x\to x_0} \sum\limits_{n=0}^{\infty} \overset{??}{=} \sum\limits_{n=0}^{\infty} \lim\limits_{x\to x_0}$), infinite sum and differentiation ($\frac{d}{dx}\sum\limits_{n=0}^{\infty} \overset{??}{=} \sum\limits_{n=0}^{\infty} \frac{d}{dx}$) or infinite sum and integration($\int \sum\limits_{n=0}^{\infty} \overset{??}{=} \sum\limits_{n=0}^{\infty} \int$) involve the concept of the interchange of two limits ($\lim\limits_{n\to\infty} \lim\limits_{m\to\infty} \overset{??}{=} \lim\limits_{m\to\infty} \lim\limits_{n\to\infty}$). It is a very importnat issue in mathematics. The above equalities are true if the summations are just sum of "finite terms". However, when the summations are sum of "infinite terms", the results could be totally different and the equalities are usually false.

In the future, we will discuss general series of functions $\sum\limits_{n=0}^{\infty} f_n(x)$ and the term-by-term differentiation and integration will be important topics.

For a power series $\sum\limits_{n=0}^{\infty} c_n(x-a)^n$, it can be regarded as a series of functions $\sum\limits_{n=0}^{\infty} f_n(x)$ with special forms of power functions. Since it has such a nice structure, some operations are applied for the power series (term-by-term).

Lemma 12.3.16. *Suppose that* $\sum\limits_{n=0}^{\infty} c_n(x-a)^n$ *converges on* $(a-R, a+R)$. *Then*

(a) the series $\sum\limits_{n=0}^{\infty} nc_n(x-a)^n$ *converges on* $(a-R, a+R)$ *and*

(b) the series $\sum\limits_{n=0}^{\infty} \frac{c_n}{n+1}(x-a)^{n+1}$ *converges on* $(a-R, a+R)$.

Proof. We will prove part(a) here and the proof of part(b) is similar and left to the readers.

For $z_0 \in (a-R, a+R)$, fix a number $x_0 \in (a-R, a+R)$ such that $|z_0 - a| < |x_0 - a| < R$. By Theorem 12.3.8, the series $\sum_{n=0}^{\infty} c_n(x_0 - a)^n$ converges aboslutely. That is,

$$\sum_{n=0}^{\infty} |c_n(x_0 - a)^n| = \sum_{n=0}^{\infty} |c_n|\big|x_0 - a\big|^n < \infty. \tag{12.4}$$

On the other hand, since $\left|\dfrac{z_0 - a}{x_0 - a}\right| < 1$, we have $n \cdot \left|\dfrac{z_0 - a}{x_0 - a}\right|^{n-1} \longrightarrow 0$ as $n \to \infty$. Thus, there exists $N \in \mathbb{N}$ such that for $n > N$, $n \cdot \left|\dfrac{z_0 - a}{x_0 - a}\right|^{n-1} < 1$. Consider

$$|nc_n(z_0 - a)^{n-1}| = \frac{1}{|x_0 - a|}|c_n(x_0 - a)^n| \cdot \underbrace{n \cdot \left|\frac{z_0 - a}{x_0 - a}\right|^{n-1}}_{<1 \text{ as } n \geq N+1} \leq \frac{1}{|x_0 - a|}|c_n(x_0 - a)^n|$$

where the last inequality holds when $n \geq N+1$. From the comparison test and (12.4), the series $\sum_{n=N+1}^{\infty} |nc_n(z_0 - a)^{n-1}|$ conveges. Then the series $\sum_{n=0}^{\infty} c_n(z_0 - a)^n$ absolutely conveges and hence it also converges. Since z_0 is an arbitrary number in $(a-R, a+R)$, the series $\sum_{n=0}^{\infty} nc_n(x-a)^n$ converges on $(a-R, a+R)$.

□

Theorem 12.3.17. *Let* $f(x) = \sum_{n=0}^{\infty} c_n(x-a)^n$ *with the radius of convergence* $R > 0$. *Then*

(a) for any $0 < L < R$, f *is integrable on* $[a-L, a+L]$ *and*

$$\int f(x)\,dx = \int \sum_{n=0}^{\infty} c_n(x-a)^n\,dx = \sum_{n=0}^{\infty} \int c_n(x-a)^n\,dx = C + \sum_{n=0}^{\infty} \frac{c_n}{n+1}(x-a)^{n+1}.$$

(b) f *is differentiable (and therefore continuous) on* $(a-R, a+R)$ *and*

$$f'(x) = \frac{d}{dx}\Big(\sum_{n=0}^{\infty} c_n(x-a)^n\Big) = \sum_{n=0}^{\infty} \frac{d}{dx}\big[c_n(x-a)^n\big] = \sum_{n=0}^{\infty} nc_n(x-a)^{n-1}.$$

Proof. We postpone the proof in the end of this section. □

Remark.

(i) The derivative and antiderivative functions of power series are another power series. Although the radius of convergence of derivative and antiderivative function are the same as the one of the origianl power series, their intervals of convergence may be different from the one of the original power series. (See the example $\sum_{n=1}^{\infty} \dfrac{x^n}{n^2}$ or $\tan^{-1} x$).

(ii) The term-by-term differentiation and integration gives a powerful method to solve differential equations.

❑ Proof of Theorem 12.3.17

Recall that "the convergence of a series $\sum_{n=0}^{\infty} c_n(x-a)^n$ on $(a-R, a+R)$" means that the sequence of partial sums $\left\{s_n(x) = \sum_{k=0}^{n} c_k(x-a)^k\right\}_{n=0}^{\infty}$ converges on $(a-R, a+R)$. That is, for every $x_0 \in (a-R, a+R)$, the limit $\lim_{n\to\infty} s_n(x_0)$ converges. Hence, the series $\sum_{n=0}^{\infty} c_n(x-a)^n$ should be expressed as

$$\sum_{n=0}^{\infty} c_n(x-a)^n = \lim_{n\to\infty}\Big[\sum_{k=0}^{n} c_k(x-a)^k\Big] = \lim_{n\to\infty} s_n(x) =: s(x).$$

For $x_0, y_0 \in (a-R, a+R)$, we have $\lim_{n\to\infty} s_n(x_0) = s(x_0)$ and $\lim_{n\to\infty} s_n(y_0) = s(y_0)$. But the "rate of convergence" may be different. That is, for $\varepsilon > 0$, the above two limits say that

$$|s_n(x_0) - s(x_0)| < \varepsilon \quad \text{for } n > N_1 \qquad \text{and} \qquad |s_n(y_0) - s(y_0)| < \varepsilon \quad \text{for } n > N_2.$$

The numbers N_1 and N_2 may be different. These numbers N depend not only on the error ε but also on the number x.

Since the power series has a good structure, we can obtain a nice result such that "in a certain restricted interval" the corresponding integer N in the definition of limit only depends on ε but is indpendent of x. Hence, the rate of convergence is "*uniform*".

Lemma 12.3.18. *Suppose that $\sum_{n=0}^{\infty} c_n(x-a)^n$ converges to $f(x)$ on $(a-R, a+R)$. Let $0 < L < R$ be a fixed number. For every $\varepsilon > 0$, there exists $N = N(\varepsilon) \in \mathbb{N}$ (depending on ε only), such that for $n \geq N$,*

$$\left| f(x) - \sum_{k=0}^{n} c_k(x-a)^k \right| < \varepsilon \qquad \textit{for every } x \in [a-L, a+L].$$

Proof. Let $x_0 = a + L \in (a-R, a+R)$. Then $\sum_{n=0}^{\infty} c_n(x_0-a)^n$ converges absolutely. That is, $\sum_{n=0}^{\infty} \left|c_n(x_0-a)^n\right| < \infty$. Therefore, for given $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$\sum_{k=N+1}^{\infty} \left|c_k(x_0-a)^k\right| = \sum_{k=N+1}^{\infty} \left|c_k L^k\right| < \varepsilon.$$

For every $x \in [a-L, a+L]$ and $n > N$, since $\sum_{n=0}^{\infty} c_n(x-a)^n$ converges to $f(x)$, we have

$$\begin{aligned}\left|f(x) - \sum_{k=0}^{n} c_k(x-a)^k\right| &= \left|\sum_{k=n+1}^{\infty} c_k(x-a)^k\right| \le \sum_{k=n+1}^{\infty} \left|c_k(x-a)^k\right| \\ &\le \sum_{k=N+1}^{\infty} \left|c_k(x-a)^k\right| \le \sum_{k=N+1}^{\infty} \left|c_k L^k\right| < \varepsilon.\end{aligned}$$

Note that the number N only depends on L but is independent of $x \in [a-L, a+L]$. □

Note. When the sequence of partial sum $s_k(x) = \sum_{n=0}^{k} c_n(x-a)^n$ satisfies Lemma 12.3.18 on $[a-L, a+L]$, we call $\{s_k(x)\}$ "*converges uniformly on* $[a-L, a+L]$".

Proof of Theorem 12.3.17

(a) Denote $s_n(x) = \sum_{k=0}^{n} c_k(x-a)^k$ as the partial sum of the series $\sum_{k=0}^{\infty} c_k(x-a)^k$. For every $n \in \mathbb{N}$, $s_n(x)$ is a polynomial function and thus it is integrable on $(a-R, a+R)$.

Let $0 < L < R$. By Lemma 12.3.18, for given $\varepsilon > 0$, there exists $N = N(\varepsilon) \in \mathbb{N}$ such that for every $x \in [a-L, a+L]$

$$\left|f(x) - s_n(x)\right| < \varepsilon \qquad \text{whenever } n \ge N. \tag{12.5}$$

Since $s_{N+1}(x)$ is integrable on $[a-L, a+L]$, there exists a partition $P = \{x_0, x_1, \cdots, x_k\}$ of $[a-L, a+L]$ such that

$$U(P, s_{N+1}) - L(P, s_{N+1}) < \varepsilon.$$

By (12.5),

$$\begin{aligned}\left|U(P,f) - U(P, s_{N+1})\right| &= \left|\sum_{i=1}^{k}(M_i - M_i^{(N+1)})(x_i - x_{i-1})\right| \\ &\le \sum_{i=1}^{k} \left|M_i - M_i^{(N+1)}\right|\left|x_i - x_{i-1}\right| \\ &\le \underbrace{\max_{1\le i\le k}\left|M_i - M_i^{(N+1)}\right|}_{<\varepsilon \text{ (check it)}} \underbrace{\sum_{i=1}^{k}\left|x_i - x_{i-1}\right|}_{=2L} \\ &< 2L\varepsilon\end{aligned}$$

where $M_i = \max_{x\in[x_{i-1},x_i]} f(x)$ and $M_i^{(N+1)} = \max_{x\in[x_{i-1},x_i]} s_{N+1}(x)$.

Similarly, $\Big|L(P, f) - L(P, s_{N+1})\Big| < 2L\varepsilon$. Then

$$\begin{aligned}\Big|U(P, f) - L(P, f)\Big| &\leq \Big|U(P, f) - U(P, s_{N+1})\Big| + \Big|U(P, s_{N+1}) - L(P, s_{N+1})\Big| \\ &\quad + \Big|L(P, s_{N+1}) - L(P, f)\Big| \\ &< 4L\varepsilon + \varepsilon = (4L+1)\varepsilon.\end{aligned}$$

Hence, f is integrable on $[a - L, a + L]$.

On the other hand, for $x \in (a - R, a + R)$, we choose $0 < L < R$ such that $x \in [a - L, a + L]$. For given $\varepsilon > 0$, let N be the integer such that the inequality (12.5) hold when $n > N$. Then

$$\begin{aligned}\Big|\int_a^x f(t)\,dt - \sum_{k=0}^{n} \frac{c_k}{k+1}(x-a)^{k+1}\Big| &= \Big|\int_a^x f(t)\,dt - \int_a^x s_n(t)\,dt\Big| = \Big|\int_a^x f(t) - s_n(t)\,dt\Big| \\ &\leq \int_a^x |f(t) - s_n(t)|\,dt \leq \int_a^x \varepsilon\,dt \leq L\varepsilon.\end{aligned}$$

Let $n \to \infty$, we have $\displaystyle\int_a^x f(t)\,dt = \sum_{n=0}^{\infty} \frac{c_n}{n+1}(x-a)^{n+1}$.

(b) By Lemma 12.3.16, the derivative of partial sum $s_k'(x) = \displaystyle\sum_{n=0}^{k} nc_n(x-a)^{n-1}$ converges on $(a - R, a + R)$. Fix $0 < L < R$, by Lemma 12.3.18, $s_k'(x)$ converges to $\displaystyle\sum_{n=0}^{\infty} nc_n(x-a)^{n-1} =: g(x)$ uniformly on $[a-L, a+L]$. It sufficies to show that $f'(x) = g(x)$ for every $x \in [a-L, a+L]$. By part(a),

$$\int_a^x g(t)\,dt = \lim_{n\to\infty} \int_a^x s_n'(t)\,dt \overset{F.T.C}{=} \lim_{n\to\infty} \Big(s_n(x) - s_n(a)\Big) = f(x) - f(a).$$

Then, by the Fundamental Theorem of Calculus,

$$f'(x) = g(x) = \sum_{n=0}^{\infty} nc_n(x-a)^{n-1} \qquad \text{for every } x \in [a - L, a + L].$$

Since L is an arbitrary number with $0 < L < R$, we obtain $f'(x) = \displaystyle\sum_{n=0}^{\infty} nc_n(x-a)^{n-1}$ on $(a - R, a + R)$.

Remark.

(i) The series $\displaystyle\sum_{n=0}^{\infty} c_n(x-a)^n$ and $\displaystyle\sum_{n=1}^{\infty} nc_n(x-a)^{n-1}$ have the same radius of convergence. But they may have different interval of convergence. For example,

$$\sum_{n=1}^{\infty} \frac{1}{n^2}x^n \qquad \text{converges on } [-1, 1]$$

$$\sum_{n=1}^{\infty} \frac{1}{n}x^{n-1} \qquad \text{converges on } [-1, 1)$$

(ii) For every $k \in \mathbb{N}$,

$$\frac{d^k}{dx^k}\Big(\sum_{n=0}^{\infty} c_n(x-a)^n\Big) = \sum_{n=0}^{\infty} \frac{d^k}{dx^k}\Big(c_n(x-a)^n\Big) = \sum_{n=k}^{\infty} n(n-1)(n-2)\cdots(n-k+1)(x-a)^{n-k}$$

All the above series have the same radii of convergence.

12.4 Power Series Representation

In Section 12.2, we know that if a function f has sufficiently many derivatives at a point a, it can be approximated by polynomials $P_{n,a}(x)$ (at least near a). As n becomes large, the approximation becomes better. This suggests us that if n tends to infinity, f might be expressed as a power series.

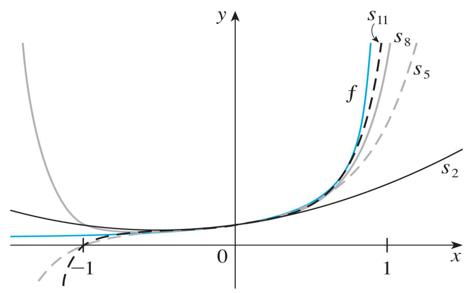
Some reasons also motivate us to find power series representation for a function. Many functions have no elementary antiderivatives or it is difficult to solve differential equations, or the approximation of them are difficult to find. We hope to express those functions as sums of power series and do the differentiation or integration on the power series rather than dealing with the original functions.

Example 12.4.1.

Consider the power series $\sum_{n=0}^{\infty} x^n = 1 + x + x^2 + \cdots$. If we regard the series as a geometric series with ratio x, then the series diverges when $|x| > 1$ and converges when $|x| < 1$. Moreover,

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x} \qquad \text{for } |x| < 1. \tag{12.6}$$

Hence, the power series is regarded as expressing the function $f(x) = \dfrac{1}{1-x}$.



Note. Observe that the domain of $f(x) = \dfrac{1}{1-x}$ is $\mathbb{R}\backslash\{1\}$ but the domain of the series $\sum_{n=0}^{\infty} x^n$ is $(-1, 1)$. This says that a power series representation of a function may equal this function only on a proper subset of its domain rather than the whole domain.

Question: For a given function, does it have a power series representation? If yes, for what values of x does $f(x)$ equal $\sum_{n=0}^{\infty} c_n x^n$? If $f(x) = \sum_{n=0}^{\infty} c_n x^n$, can we take differentation or integration on the power series term-by-term?

Example 12.4.2. Express $\dfrac{1}{1+x^2}$ as the sum of a power series and find the interval of convergence.

Proof. Consider $\frac{1}{1+x^2} = \frac{1}{1-(-x^2)}$. Replacing x by $-x^2$ in Equation (12.6), we have

$$\frac{1}{1+x^2} = \frac{1}{1-(-x^2)} = \sum_{n=0}^{\infty}(-x^2)^n = \sum_{n=0}^{\infty}(-1)^n x^{2n}.$$

The geometric series converes when $|-x^2| < 1$. Thus, the interval of convergence is $(-1, 1)$. □

Example 12.4.3. Find the power series representation of $\frac{1}{x+2}$.

Proof. Consider $\frac{1}{x+2} = \frac{1}{2} \cdot \frac{1}{1-(-\frac{x}{2})}$. Replacing x by $-\frac{x}{2}$ in Equation (12.6), we have

$$\frac{1}{x+2} = \frac{1}{2} \cdot \frac{1}{1-(-\frac{x}{2})} = \frac{1}{2}\sum_{n=0}^{\infty}(-\frac{x}{2})^n = \sum_{n=0}^{\infty}\frac{(-1)^n}{2^{n+1}}x^n.$$

The power series converges when $|-\frac{x}{2}| < 1$. The interval of convergence is $(-2, 2)$. □

Example 12.4.4. Find a power series representation of $\frac{x^3}{x+2}$.

Proof. The power series representation is

$$\frac{x^3}{x+2} = x^3 \cdot \frac{1}{x+2} = x^3\sum_{n=0}^{\infty}\frac{(-1)^n x^n}{2^{n+1}} = \sum_{n=0}^{\infty}\frac{(-1)^n x^{n+3}}{2^{n+1}}.$$

The interval of convergence is $(-2, 2)$.

□

Example 12.4.5. (Bessel function) The function

$$J_0(x) = \sum_{n=0}^{\infty}\frac{(-1)^n x^{2n}}{2^{2n}(n!)^2} \quad \text{is defined for all } x \in \mathbb{R}$$

Then

$$J_0'(x) = \frac{d}{dx}\Big[\sum_{n=0}^{\infty}\frac{(-1)^n x^{2n}}{2^{2n}(n!)^2}\Big] = \sum_{n=0}^{\infty}\frac{d}{dx}\Big[\frac{(-1)^n x^{2n}}{2^{2n}(n!)^2}\Big] = \sum_{n=1}^{\infty}\frac{(-1)^n 2n x^{2n-1}}{2^n(n!)^2} \quad \text{on } \mathbb{R}.$$

Example 12.4.6. Express $\frac{1}{(1-x)^2}$ as a power series by differentiating $\frac{1}{1-x}$. What is the radius of convergence?

Proof. Since $\frac{1}{1-x} = 1 + x + x^2 + \cdots = \sum_{n=0}^{\infty} x^n$ for $|x| < 1$,

$$\begin{aligned}\frac{1}{(1-x)^2} &= \frac{d}{dx}\Big[\frac{1}{1-x}\Big] = \frac{d}{dx}\Big[\sum_{n=0}^{\infty}x^n\Big] = \sum_{n=1}^{\infty}\frac{d}{dx}(x^n) = \sum_{n=1}^{\infty}nx^{n-1}\ \Big(= \sum_{n=0}^{\infty}(n+1)x^n\Big)\\ &= 1 + 2x + 3x^2 + \cdots.\end{aligned}$$

The radius of convergence of the power series of $\frac{1}{(1-x)^2}$ is 1 which is the same as the radius of convergence of the power series of $\frac{1}{1-x}$. □

Example 12.4.7. Find a power series representation for $\ln(1+x)$ and its radius of convergence.

Proof. Since $\ln(1+x) = \displaystyle\int \frac{1}{1+x}\,dx$ and $\displaystyle\frac{1}{1-(-x)} = \sum_{n=0}^{\infty}(-x)^n = \sum_{n=0}^{\infty}(-1)^n x^n$ for $|x| < 1$,

$$\ln(1+x) = \int \frac{1}{1+x}\,dx = \int \sum_{n=0}^{\infty}(-1)^n x^n\,dx = C + \sum_{n=0}^{\infty}(-1)^n \frac{x^{n+1}}{n+1}.$$

To determine C, taking $x = 0 \in (-1, 1)$, we have $0 = \ln(1+0) = C$ and hence

$$\ln(1+x) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}x^n}{n}.$$

Since the radius of convergence of the series for $\displaystyle\frac{1}{1+x}$ is 1, the radius of convergence of the series for $\ln(1+x)$ is also 1. □

Example 12.4.8. Find a power series representation for $f(x) = \tan^{-1} x$.

Proof. Since $f'(x) = \displaystyle\frac{1}{1+x^2} = \frac{1}{1-(-x^2)} = \sum_{n=0}^{\infty}(-1)^n x^{2n}$ on $|x| < 1$, we have

$$f(x) = \tan^{-1} x = \int \sum_{n=0}^{\infty}(-1)^n x^{2n}\,dx = C + \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{2n+1}.$$

To determine C, taking $x = 0$, we have $0 = \tan^{-1} 0 = C$ and hence

$$\tan^{-1} x = \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} x^{2n+1}.$$

Since the radius of convergence of the series for $\displaystyle\frac{1}{1+x^2}$ is 1, the radius of convergence of the series for $\tan^{-1} x$ is also 1. □

Note. In fact, the power series representation is also true when $x = \pm 1$. But this result is not given by the above theorem.

Example 12.4.9. Express $\dfrac{\pi}{4}$ as a series.

Proof. From Example 12.4.8,

$$\frac{\pi}{4} = \tan^{-1} 1 = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots + \frac{(-1)^n}{2n+1} + \cdots$$

In fact, $\dfrac{\pi}{4}$ has several different series representations. For example,

$$\begin{aligned}
\frac{\pi}{4} &= \tan^{-1}\frac{1}{2} + \tan^{-1}\frac{1}{3} \\
&= \left[\frac{1}{2} - \frac{1}{3}\left(\frac{1}{2}\right)^3 + \frac{1}{5}\left(\frac{1}{2}\right)^5 - \frac{1}{7}\left(\frac{1}{2}\right)^7 + \cdots\right] + \left[\frac{1}{3} - \frac{1}{3}\left(\frac{1}{3}\right)^3 + \frac{1}{5}\left(\frac{1}{3}\right)^5 - \frac{1}{7}\left(\frac{1}{3}\right)^7 + \cdots\right]
\end{aligned}$$

□

Note. If we use the idnetity $\pi = 48\tan^{-1}\frac{1}{18} + 32\tan^{-1}\frac{1}{57} - 20\tan^{-1}\frac{1}{239}$ to approximate π, it will give more rapid rate of convergence than the above series representation since $\frac{1}{18}$, $\frac{1}{57}$ and $\frac{1}{239}$ are much smaller than $\frac{1}{2}$ and $\frac{1}{3}$. This implies that the reminder of the former decays to zero much more rapidly than the one of latter.

Example 12.4.10. (a) Evaluate $\displaystyle\int \frac{1}{1+x^7}\,dx$ as a power series

(b) Approximate $\displaystyle\int_0^{0.5} \frac{1}{1+x^7}\,dx$ correct to within 10^{-7}.

Proof. (a) Since $\displaystyle\frac{1}{1+x^7} = \frac{1}{1-(-x^7)} = \sum_{n=0}^{\infty}(-x^7)^n = \sum_{n=0}^{\infty}(-1)^n x^{7n}$ for $|x < 1$, we have

$$\int \frac{1}{1+x^7}\,dx = \int \sum_{n=0}^{\infty}(-1)^n x^{7n}\,dx = C + \sum_{n=0}^{\infty}(-1)^n \frac{x^{7n+1}}{7n+1} \quad \text{for } |x| < 1.$$

(b)

$$\int_0^{0.5} \frac{1}{1+x^7}\,dx = \sum_{n=0}^{\infty} \frac{(-1)^n}{7n+1} x^{7n+1}\Big|_0^{0.5} = \sum_{n=0}^{\infty}(-1)^n \frac{(0.5)^{7n+1}}{7n+1}.$$

By the alternating series estimation, for $\displaystyle\sum_{n=0}^{\infty}(-1)b_n$ with $b_n > 0$, the estimate of remainder $|R_n| < b_{n+1}$. Hence, for $b_n = \dfrac{(0.5)^{7n+1}}{7n+1} < 10^{-7}$, we have $n \geq 4$.

Therefore,

$$\int_0^{0.5} \frac{1}{1+x^7}\,dx \approx \frac{1}{2} - \frac{1}{8\cdot 2^8} + \frac{1}{15\cdot 2^{15}} - \frac{1}{22\cdot 2^{22}} \approx 0.49951374.$$

□

Remark. Suppose that $f(x) = \displaystyle\sum_{n=0}^{\infty} c_n(x-a)^n$ converges for $|x-a| < R$. Then $f'(x) = \displaystyle\sum_{n=1}^{\infty} nc_n(x-a)^n$ also converges for $|x-a| < R$. Hence $f'(x)$ has a power series representation on $(x-R, x+R)$. We can also take term-by-term differentiation and obtain

$$f''(x) = \sum_{n=2}^{\infty} n(n-1)(x-a)^{n-2} \qquad \text{converges on } (a-R, a+R)$$

$$\vdots$$

$$f^{(k)}(x) = \sum_{n=k}^{\infty} n(n-1)(n-2)\cdots(n-k+1)(x-a)^{n-k} \qquad \text{converges on } (a-R, a+R).$$

12.5 Taylor and Maclaurin Series

So far, we can find power series representations for a centain restricted class of functions.

Question: Which functions do have power series representations?

Suppose that f has a power series representation

$$f(x) = c_0 + c_1(x-a) + c_2(x-a)^2 + \cdots + c_n(x-a)^n + \cdots = \sum_{n=0}^{\infty} c_n(x-a)^n \qquad \text{for } |x-a| < R$$

Question: what are the coefficients c_n?

By the term-by-term differentiation, we can take $\dfrac{d^k}{dx^k}$ on f and obtain

$$f^{(k)}(x) = \sum_{n=k}^{\infty} n(n-1)\cdots(n-k+1)c_n(x-a)^{n-k}.$$

Plugging $x = a$ into the equation, we have

$$c_k = \frac{f^{(k)}(a)}{k!} \quad \text{for } k = 0, 1, 2, \cdots .$$

Remark. We have seen this coefficient formula in Taylor polynomials.

Definition 12.5.1. (a) Let f be a function with infinitely many times derivatives at a, that is, $f'(a), f''(a), \cdots, f^{(k)}(a), \cdots$ exist for $k = 1, 2, \cdots$. Then the series

$$f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(k)}(a)}{k!}(x-a)^k + \cdots = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$$

is called the "*Taylor series for f at a*" (or "*Taylor series for f about a*" or "*Taylor series for f centered at a*").

(b) For the special case $a = 0$, the Taylor series at 0, $\displaystyle\sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!}x^n$ is also called the "*Maclaurin series for f*".

Note. If f can be represented as a power series about a with radius of convergence $R > 0$, then f is equal to the sum of its Taylor series about a.

Example 12.5.2. Find the Taylor series for the following functions at the given points.

(1) $f(x) = e^x$ at $x = 0$.

Proof. Since $f^{(k)}(x) = e^x$, we have $f^{(k)}(0) = 1$ for $k = 0, 1, 2, \cdots$. Hence, the Taylor series for f at 0 (Maclaurin series) is

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!}x^n = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} + \cdots .$$

Moreover, let $a_n = \dfrac{x^n}{n!}$. Then $\left|\dfrac{a_{n+1}}{a_n}\right| = \dfrac{|x|}{n+1} \longrightarrow 0 < 1$ as $n \to \infty$ for every x. By the Ratio Test, the Taylor series converges for all x. □

(2) $f(x) = \sin x$ at $x = 0$.

Proof. For $k \in \mathbb{N}$,

$$f^{(4n)}(x) = \sin x,\ f^{(4n+1)}(x) = \cos x,\ f^{(4n+2)}(x) = -\sin x,\ f^{(4n+3)}(x) = -\cos x$$
$$f^{(4n)}(0) = \sin x,\ f^{(4n+1)}(0) = 1,\ f^{(4n+2)}(0) = 0,\ f^{(4n+3)}(0) = -1$$

The Taylor series for f at 0 (Maclaurin series) is

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1}.$$

Let $a_n = \dfrac{(-1)^n}{(2n+1)!} x^{2n+1}$. Then

$$\left|\frac{a^{n+1}}{a_n}\right| = \left|\frac{\frac{(-1)^{n+1}}{(2n+3)!} x^{2n+3}}{\frac{(-1)^n}{(2n+1)!} x^{2n+1}}\right| = \left|\frac{x^2}{(2n+1)(2n+2)}\right| \longrightarrow 0 \quad \text{for all } x.$$

Therefore, the Taylor series $\displaystyle\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1}$ converges for all $x \in \mathbb{R}$. □

By the definition of Taylor series, as long as a function f has infinitely many derivatives at a, the Taylor series for f about a is defined. It is natural to ask the following questions:

Question:

(i) What values of x for which the Taylor sereis is convergent or divergent?

(ii) If the Taylor series converges at x, does it converge to $f(x)$? That is, $f(x) \overset{??}{=} \displaystyle\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$

We usually determine whether and where a Taylor series converges by using the Ratio test or Root test. Even if the Taylor series for f about a converges at some number $x \neq a$, it may not converge to $f(x)$. For example,

$$f(x) = \begin{cases} e^{-\frac{1}{x^2}} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

We can evaluate that $f(0) = f'(0) = f''(0) = \cdots = f^{(k)}(0) = \cdots = 0$. Hence, the Taylor series for f at 0 is the zero function which does not converge to f except at the center 0.

The Taylor series for f at a is defined by the limit of its partial sum

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n = \lim_{N\to\infty} \sum_{n=0}^{N} \frac{f^{(n)}(a)}{n!}(x-a)^n = \lim_{N\to\infty} P_{N,a}(x)$$

To check whether the Taylor series converges to f, we should show that

$$f(x) = \lim_{N\to\infty} P_{N,a}(x) \quad \text{if and only if} \quad \lim_{N\to\infty} R_{N,a}(x) = \lim_{N\to\infty} f(x) - P_{N,a}(x) = 0.$$

We recall the Taylor Theorem here (see Theorem 12.2.2).

Let $f(t)$ be a $n+1$ times differentiable function on $[a, x]$ and $R_{n,a}(x)$ be defined by

$$f(x) = P_{n,a}(x) + R_{n,a}(x).$$

Then

(a) (Cauchy form)

$$R_{n,a}(x) = \frac{f^{(n+1)}(\xi)}{n!}(x-\xi)^n(x-a) \qquad \text{for some } \xi \in (a, x).$$

(b) (Lagrange form)

$$R_{n,a}(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1} \qquad \text{for some } \xi \in (a, x).$$

(c) (Integral form)

$$R_{n,a}(x) = \int_a^x \frac{f^{(n+1)}(t)}{n!}(x-t)^n\, dt$$

By using the part(b) of Taylor Theorem, we can derive the Taylor inequality

Lemma 12.5.3. *Let $f(t)$ be a $n+1$ times differentiable function on $[a, x]$ and $|f^{(n+1)}(z)| \le M$ for all $z \in [a, x]$. Then the remainder $R_{n,a}(x)$ of the Taylor series satisfies the inequaltiy*

$$|R_{n,a}(x)| \le \frac{M}{(n+1)!}|x-a|^{n+1}$$

Corollary 12.5.4. *Let $f(t)$ be a $n+1$ times differentiable function on $(a-R, a+R)$ and $|f^{(n+1)}(z)| \le M$ for all $z \in (a-R, a+R)$. Then for every $x \in (a-R, a+R)$,*

$$|R_{n,a}(x)| = \underbrace{\frac{|f^{(n+1)}(z_0)|}{(n+1)!}}_{\text{for some } z_0\in[a,x]} |x-a|^{n+1} \le \frac{M}{(n+1)!}|x-a|^{n+1}.$$

Example 12.5.5. Determine whether the equality $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ holds. If yes, for what values of x does the equality hold?

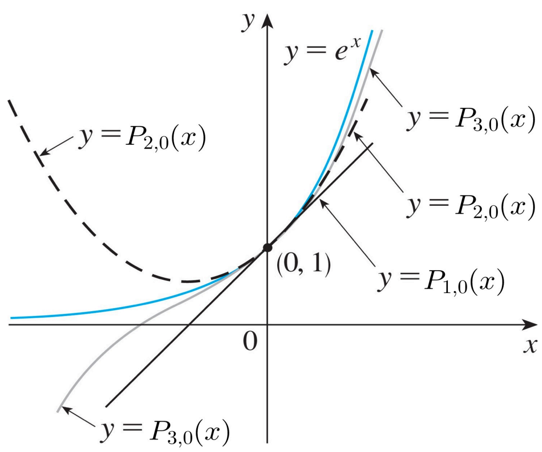
Proof. Let $f(x) = e^x$. Then $f^{(n)}(x) = e^x$ for all $n \in \mathbb{N}$ and

$$f(x) = e^x = \sum_{k=0}^{n} \frac{x^k}{k!} + R_{n,0}(x).$$

Fix a number x_0 and choose a number $d \geq |x_0|$. Then $|f^{(n+1)}(z)| \leq e^{|z|} \leq e^d$ for all $0 \leq |z| \leq |x_0| \leq d$. By the Taylor inequality,

$$0 \leq |R_{n,0}(x_0)| \leq \frac{e^d}{(n+1)!}|x_0 - 0|^{n+1} \leq e^d \frac{d^{n+1}}{(n+1)!}.$$

By the Squeeze Theorem, $\lim_{n\to\infty} |R_{n,0}(x_0)| = 0$. Hence, the Taylor series $\sum_{n=0}^{\infty} \frac{x^n}{n!}$ converges to e^x at x_0. Since x_0 is an arbitrary number in $\mathbb{R}$, the Taylor series $\sum_{n=0}^{\infty} \frac{x^n}{n!}$ converges to e^x for every number in $\mathbb{R}$. □



Example 12.5.6. Find the Taylor series for $f(x) = e^x$ at $a = 2$, and determine whether and for what values of x, $f(x)$ equals its Taylor series about $a = 2$.

Proof. Since $f^{(n)}(x) = e^x$, $f^{(n)}(2) = e^2$. The Taylor series for f at $a = 2$ is

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(2)}{n!}(x-2)^n = \sum_{n=0}^{\infty} \frac{e^2}{n!}(x-2)^n$$

- To determine for which values of x the Taylor series conveges.
  Let $a_n = \frac{e^2}{n!}(x-2)^n$. Then for every $x \in \mathbb{R}$,

$$\left|\frac{a_{n+1}}{a_n}\right| = \left|\frac{\frac{e^2}{(n+1)!}(x-2)^{n+1}}{\frac{e^2}{n!}(x-2)^n}\right| = \frac{1}{n+1}|x-2| \longrightarrow 0 \quad \text{as } n \longrightarrow \infty.$$

- To determine whether $e^x = \sum_{n=0}^{\infty} \frac{e^2}{n!}(x-2)^n$.

  Fix a number $d > 0$. By the Taylor theorem, for x with $|x-2| < d$, there exists z_x between 2 and x such that
  $$R_{n,2}(x) = \frac{f^{(n+1)}(z_x)}{(n+1)!}|x-2|^{n+1} = \frac{e^{z_x}}{(n+1)!}|x-2|^{n+1}.$$
  Hence, for $|x-2| < d$,
  $$0 \le |R_{n,2}(x)| \le \frac{e^{2+d}}{(n+1)!}|x-2|^{n+1} \le e^{2+d}\frac{d^{n+1}}{(n+1)!}.$$
  By the Squeeze Theorem, $\lim_{n\to\infty} R_{n,2}(x) = 0$ for every $|x-2| < d$ and this imiplies that $e^x = \sum_{n=0}^{\infty} \frac{e^2}{n!}(x-2)^n$ for every $|x-2| < d$. Sicne d is arbitrary number, we have
  $$e^x = \sum_{n=0}^{\infty} \frac{e^2}{n!}(x-2)^n \qquad \text{for every } x \in \mathbb{R}.$$

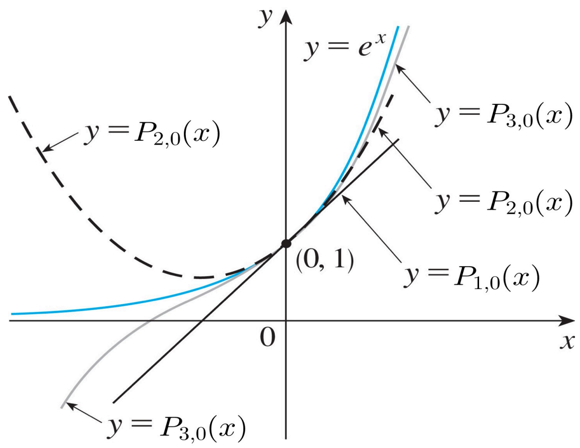
□

Example 12.5.7. Find the Maclaurin series for $f(x) = \sin x$ and prove that it represents $\sin x$ for all x.

Proof. The derivatives of f are
$$f^{(4k)}(x) = \sin x,\ f^{(4k+1)}(x) = \cos x,\ f^{(4k+2)}(x) = -\sin x,\ f^{(4k+3)}(x) = -\cos x.$$
Then
$$f^{(4k)}(0) = 0,\ f^{(4k+1)}(0) = 1,\ f^{(4k+2)}(0) = 0,\ f^{(4k+3)}(0) = -1.$$
The Maclaurin series for $\sin x$ is
$$\sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!}x^n = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!}x^{2n+1}.$$
Since $|f^{(n+1)(x)}| = |\pm \sin x|$ or $|\pm \cos x| \le 1$ for all $x \in \mathbb{R}$ and $n \in \mathbb{N}$, we have
$$|R_{n,0}(x) \le \frac{1}{(n+1)!}|x|^{n+1}.$$
Hence, for every fixed x, $R_{n,0}(x) \to 0$ as $n \to \infty$. That is,
$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!}x^{2n+1} \quad \text{for all } x \in \mathbb{R}.$$

□

Example 12.5.8. Prove that $\cos x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!}$.



Proof.

$$\begin{aligned}\cos x &= \frac{d}{dx}(\sin x) = \frac{d}{dx}\Big(\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1}\Big)\\ &= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!}\Big(\frac{d}{dx} x^{2n+1}\Big)\\ &= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n} \quad \text{for all } x \in \mathbb{R}.\end{aligned}$$

□

Example 12.5.9. Find the Maclaurin series for the function $f(x) = x \cos x$

Proof. Since $\cos x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n}$ for all x, we have

$$x \cos x = x \cdot \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n} = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n+1} \quad \text{for all } x.$$

□

Exercise. Find the Taylor series for $f(x) = \ln(1 + x)$ and for what values of x the Taylor series converges to $f(x)$.

Answer: $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} x^n \quad$ for $-1 < x \leq 1$.

■ Binomial Series

Example 12.5.10. (Binomial Series) Use the Maclaurin series for $f(x) = (1 + x)^k$ to deduce the formula of the binomial series where k is any real number.

Proof. The derivatives of f is

$$f^{(n)}(x) = k(k-1)(k-2)\cdots(k-n+1)(1+x)^{k-n} \quad \text{for } n = 1, 2, \cdots.$$

Then

$$f^{(n)}(0) = k(k-1)(k-2)\cdots(k-n+1) \quad \text{for } n = 1, 2, \cdots.$$

The Maclaurin series for $f(x) = (1+x)^k$ is

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n = \sum_{n=0}^{\infty} \frac{k(k-1)(k-2)\cdots(k-n+1)}{n!} x^n \qquad \text{(binomial series)}$$

□

Note. (1) (Convergence)

(i) For $k \in \mathbb{N}$, $k - n + 1 = 0$ when $n = k + 1$. Then the binomial series is a finite sum and a k degree polynomial. Therefore, the series converges for all x.

(ii) For $k \in \mathbb{R}\backslash\mathbb{N}$, let $a_n = \dfrac{k(k-1)(k-2)\cdots(k-n+1)}{n!} x^n$. Consider

$$\left|\frac{a_{n+1}}{a_n}\right| = \frac{|k-n|}{n+1}|x| = \frac{|1-\frac{k}{n}|}{1+\frac{1}{n}}|x| \longrightarrow |x| \quad \text{as } n \to \infty.$$

By the Ratio Test, the binomial series converges if $|x| < 1$ and diverges if $|x| > 1$.
Question: How about $x = \pm 1$?
Answer: depending on k.

- If $-1 < k \le 0$, the series converges at 1.
- If $k \ge 0$, the series converges at ± 1.

(2) Denote the coefficients in the binomial series

$$\binom{k}{n} = \frac{k(k-1)(k-2)\cdots(k-n+1)}{n!} \qquad \text{(binomial coefficients)}$$

If $k \in \mathbb{N}$ and $k \ge n$, then $\dbinom{k}{n} = \dfrac{k!}{n!(k-n)!}$.

(3) The binomial series: if $k \in \mathbb{R}$ and $|x| < 1$, then

$$\begin{aligned}(1+x)^k &= \sum_{n=0}^{\infty} \binom{k}{n} x^n \\ &= 1 + kx + \frac{k(k-1)}{2!}x^2 + \frac{k(k-1)(k-2)}{3!}x^3 + \cdots + \frac{k(k-1)(k-2)\cdots(k-n+1)}{n!}x^n.\end{aligned}$$

Example 12.5.11. Find the Maclaurin series for the function $f(x) = \dfrac{1}{\sqrt{4-x}}$ and its radius of convergence.

Proof. The function $f(x) = \dfrac{1}{\sqrt{4-x}} = (4-x)^{-\frac{1}{2}}$. By the binomial series with $k = -\frac{1}{2}$ and replacing x by $-\dfrac{x}{4}$, we have

$$\begin{aligned} f(x) &= \frac{1}{2}\sum_{n=0}^{\infty}\binom{-\frac{1}{2}}{n}\Big(-\frac{x}{4}\Big)^n \\ &= = \frac{1}{2}\Big[1 + \frac{1}{8}x + \frac{1\cdot 3}{2!\,8^2}x^2 + \frac{1\cdot 3\cdot 5}{3!\,8^3}x^3 + \cdots + \frac{1\cdot 3\cdot 5\cdots(2n-1)}{n!\,8^n}x^n + \cdots\Big] \end{aligned}$$

The series converges when $|-\dfrac{x}{4}| < 1$, that is, on $(-4, 4)$. □

Example 12.5.12. Find the sum of the series

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

Proof. Consider

$$\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n\cdot 2^n} = \sum_{n=1}^{\infty}(-1)^{n-1}\frac{\left(\frac{1}{2}\right)^n}{n}.$$

Using the Maclarin series for $\ln(1+x)$ by taking $x = \frac{1}{2}$, we have

$$\sum_{n=1}^{\infty}(-1)^{n-1}\frac{\left(\frac{1}{2}\right)^n}{n} = \ln(1+\frac{1}{2}) = \ln\frac{3}{2}.$$

□

Exercise. Evaluate the sum of the series $\displaystyle\sum_{n=0}^{\infty}(-1)^n\frac{2n+2}{(2n+1)!}$.

Answer: $\displaystyle\sum_{n=0}^{\infty}(-1)^n\frac{2n+2}{(2n+1)!} = \sin 1 + \cos 1$.

■ **Multiplication and Divison of Power Series**

Recall that if $f(x) = \displaystyle\sum_{n=0}^{\infty} b_n(x-a)^n$ and $g(x) = \displaystyle\sum_{n=0}^{\infty} c_n(x-a)^n$, then

$$\begin{aligned} f(x)g(x) &= \sum_{n=0}^{\infty} d_n(x-a)^n \qquad \text{where } d_n = \sum_{k=0}^{n} b_k c_{n-k} \\ \frac{f(x)}{g(x)} &= \sum_{n=0}^{\infty} e_n(x-a)^n \quad \text{where } e_n \text{ satisfying } b_n = \sum_{k=0}^{n} c_k e_{n-k}. \end{aligned}$$

Example 12.5.13.

(1) Find the first three nonzero terms in the Maclaurin series for $e^x \sin x$.

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n = 1 + x + x^2 + x^3 + \cdots \qquad R = 1$$

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \qquad R = \infty$$

$$\sin x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots \qquad R = \infty$$

$$\cos x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots \qquad R = \infty$$

$$\tan^{-1} x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{2n+1} = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots \qquad R = 1$$

$$\ln(1+x) = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{x^n}{n} = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots \qquad R = 1$$

$$(1+x)^k = \sum_{n=0}^{\infty} \binom{k}{n} x^n = 1 + kx + \frac{k(k-1)}{2!}x^2 + \frac{k(k-1)(k-2)}{3!}x^3 + \cdots \qquad R = 1$$

Proof. Since

$$\begin{aligned} e^x &= 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} + \cdots \qquad \text{and} \\ \sin x &= x - \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots + \frac{(-1)^n}{(2n+1)^n} x^{2n+1} + \cdots, \end{aligned}$$

we have

$$\begin{aligned} e^x \sin x &= \left(1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots\right)\left(x - \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots\right) \\ &= x + x^2 + \frac{x^3}{3} + \cdots . \end{aligned}$$

□

(2) Find the first three nonzero terms in the Maclaurin series for $\tan x$.

Proof. Since

$$\begin{aligned} \sin x &= x - \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots + \frac{(-1)^n}{(2n+1)!} x^{2n+1} + \cdots \qquad \text{and} \\ \cos x &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots + \frac{(-1)^n}{(2n)!} + \cdots, \end{aligned}$$

we have

$$\begin{aligned}\tan x &= \frac{\sin x}{\cos x} = \frac{x - \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots}{1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots} \\ &= x + \frac{1}{3}x^3 + \frac{2}{15}x^5 + \cdots .\end{aligned}$$

□

Note. One reason that Taylor series are important is that they enable us to integrate functions which we cannot find and express their antiderivatives as elementary functions.

Example 12.5.14.

(1) Evaluate $\int e^{-x^2}\,dx$ as an infinite series.

Proof. Since $e^{-x^2} = \sum_{n=0}^{\infty} \frac{(-x^2)^n}{n!}$ for any x, we obtain

$$\begin{aligned}\int e^{-x^2}\,dx &= \int \sum_{n=0}^{\infty} \frac{(-x^2)^n}{n!} = \sum_{n=0}^{\infty} \int \frac{(-1)^n}{n!} x^{2n}\,dx \\ &= C + \sum_{n=0}^{\infty} \frac{(-1)^n}{n!\,(2n+1)} x^{2n+1}.\end{aligned}$$

□

(2) Evaluate $\int_0^1 e^{-x^2}\,dx$ correct to within an error of 0.001.

Proof. Consider

$$\begin{aligned}\int_0^1 e^{-x^2}\,dx &= \sum_{n=0}^{\infty} \frac{(-1)^n}{n!\,(2n+1)} x^{2n+1}\Big|_0^1 \\ &= 1 - \frac{1}{3} + \frac{1}{10} - \frac{1}{42} + \frac{1}{216} \qquad \text{(alternating series)}\end{aligned}$$

By the alternating series estimation, $|s - \sum_{k=0}^{n} b_n| \le b_{n+1}$. Consider

$$\left|\frac{(-1)^n}{n!\,(2n+1)} \cdot 1^{2n+1}\right| < 0.001.$$

Then $n \ge 5$ and $\int_0^1 e^{-x^2}\,dx \approx 0.7475$. □

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

Proof.

$$\begin{aligned}\lim_{x\to 0}\frac{e^x-1-x}{x^2} &= \lim_{x\to 0}\frac{(1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\cdots)-1-x}{x^2}\\ &= \lim_{x\to 0}\Big(\frac{1}{2!}+\frac{x}{3!}+\cdots+\frac{x^{n-2}}{n!}+\cdots\Big)\\ &= \frac{1}{2}.\end{aligned}$$

Note: we can also obtain the above limit by the L'Hôpital Rule. □

Exercise.

(1) Find the Taylor series for the function $f(x) = \sin^{-1} x$ and find its interval of convergence. (Hint: $\sin^{-1}(x) = \int \frac{1}{\sqrt{1-t^2}\, dt}$ and using the binomial series.)

(2) Express the following functions as their Taylor series and find the limits

(i) $\lim_{x\to 1}\frac{\ln x}{x-1}$.

(ii) $\lim_{x\to 0}\frac{\sin x - \tan x}{x^3}$.

(iii) $\lim_{x\to 0}\frac{(e^{2x}-1)\ln(1+x^3)}{(1-\cos 3x)^2}$.

Chapter 13

Vector Functions

In this chapter, we will use the vector-valued functions to describe curves, surfaces and the motion of objects through space.

13.1 Preliminary

❏ n-dimensional Spaces

Definition 13.1.1. Let A and B be two sets. We define $A \times B$ by

$$A \times B = \{(a, b) \mid a \in A,\ b \in B\}.$$

We call $A \times B$ the "*product (set) of A and B*".

Example 13.1.2.

$$\begin{aligned}
\mathbb{R}^2 &= \mathbb{R} \times \mathbb{R} = \{(a, b) \mid a \in \mathbb{R},\ b \in \mathbb{R}\} \\
\mathbb{R}^3 &= \mathbb{R} \times \mathbb{R} \times \mathbb{R} = \{(a, b, c) \mid a \in \mathbb{R},\ b \in \mathbb{R},\ c \in \mathbb{R}\} \\
&\ \ \vdots \\
\mathbb{R}^n &= \overbrace{\mathbb{R} \times \cdots \times \mathbb{R}}^{n} = \{(a_1, a_2, \cdots, a_n) \mid a_i \in \mathbb{R},\ i = 1, 2, \cdots, n\}
\end{aligned}$$

Example 13.1.3.

(1) Let $f : \mathbb{R} \to \mathbb{R}^n$. For example,

$$\begin{aligned}
t &\xrightarrow{f} \big(x(t), y(t), z(t)\big) \\
t &\xrightarrow{f} \big(x_1(t), x_2(t), \cdots, x_n(t)\big)
\end{aligned}$$

There exist $f_1, f_2, \cdots f_n : \mathbb{R} \to \mathbb{R}$ such that

$$f(t) = \big(f_1(t), f_2(t), \cdots, f_n(t)\big)$$

(2) Let $f : \mathbb{R}^n \to \mathbb{R}$. For example, $f(x, y) = x^2 + y^2$.

(3) Let $f : \mathbb{R}^n \to \mathbb{R}^m$. There exist $f_1, f_2, \cdots f_m : \mathbb{R}^n \to \mathbb{R}$ such that

$$f(x_1, \cdots, x_n) = \Big(f_1(x_1, \cdots, x_n), \cdots, f_m(x_1, \cdots x_n)\Big).$$

■ Graph

The graph of a function f is the set consisting of the ordered pairs $\big(x, f(x)\big)$ where $x \in Dom(f)$.

$$Graph(f) = \Big\{\big(x, f(x)\big) \;\Big|\; x \in Dom(f)\Big\}.$$

❑ Vectors

The n-dimensional space $\mathbb{R}^n$ consists of the points with coordinates $\{(x_1, \cdots, x_n) \,|\, x_i \in \mathbb{R},\ i = 1, 2, \cdots, n\}$. A point P in $\mathbb{R}^n$ has coordinate

$$P = (a_1, a_2, \cdots, a_n) \qquad \text{(usually written as } P(a_1, \cdots, a_n)\text{).}$$

The distance between $P(a_1, \cdots, a_n)$ and $Q(b_1, \cdots, b_n)$ is

$$|\overline{PQ}| = \sqrt{(a_1 - b_1)^2 + \cdots (a_n - b_n)^2}.$$

A vector $\mathbf{v}$ in a n dimensional vector space can be written as $\mathbf{v} =< a_1, \cdots, a_n >$. The "*length*" (or "*magnitude*") of a vector is

$$||\mathbf{v}|| = \sqrt{a_1^2 + \cdots a_n^2}.$$

In this chapter, we will take more attention on the vectors in 3-dimensional vector spaces and most of the following results also hold for vectors in n-dimensional vector spaces.

■ Laws and Operations of Vectors

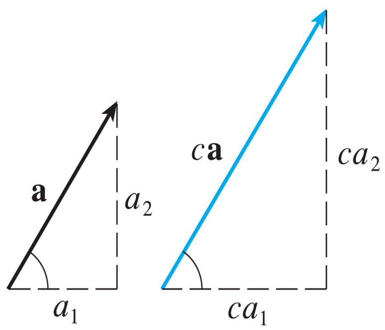
Let $\mathbf{a} =< a_1, a_2, a_3 >$ and $\mathbf{b} =< b_1, b_2, b_3 >$ be two vectors and c be a real number. Then

(a) $\mathbf{a} \pm \mathbf{b} =< a_1 \pm b_1, a_2 \pm b_2, a_3 \pm b_3 >$

(b) $c\mathbf{a} =< ca_1, ca_2, ca_3 >$.

- if $c = 0$, then $c\mathbf{a} = \mathbf{0} =< 0, 0, 0 >$.
- if $c > 0$, then $c\mathbf{a}$ and $\mathbf{a}$ have the same direction.
- if $c < 0$, then $c\mathbf{a}$ and $\mathbf{a}$ have the opposite directions.

Note that $c\mathbf{a} = \mathbf{0}$ if and only if $c = 0$ or $\mathbf{a} = \mathbf{0}$.



(c) The *dot product* (*inner product*) of $\mathbf{a}$ and $\mathbf{b}$ is defined by

$$\mathbf{a} \cdot \mathbf{b} = a_1b_1 + a_2b_2 + a_3b_3.$$

Remark. Let θ be the angle between $\mathbf{a}$ and $\mathbf{b}$.

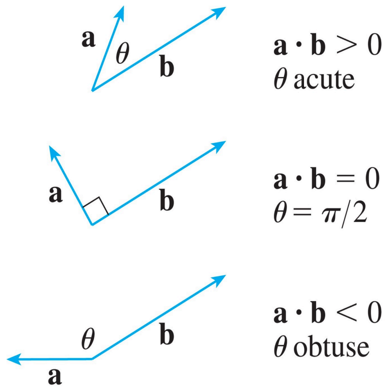
(i) $\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\|\|\mathbf{b}\| \cos\theta$

(ii) $\mathbf{a} \cdot \mathbf{b} = \mathbf{0}$ if and only if $\theta = \dfrac{\pi}{2}$.

In this case, we say that $\mathbf{a}$ and $\mathbf{b}$ are "*perpendicular*" (or "*orthogonal*") (usually denoted $\mathbf{a} \perp \mathbf{b}$).

Hence, $\mathbf{0}$ is perpendicular to any vector.

(iii) $\mathbf{a} \cdot \mathbf{a} = \|\mathbf{a}\|^2$.

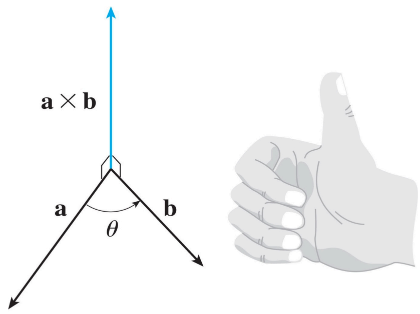


(d) (*cross product of* $\mathbf{a}$ *and* $\mathbf{b}$) satisfies

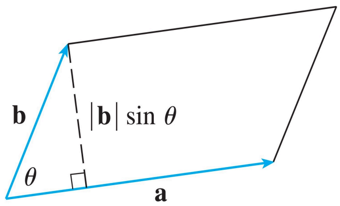
(i) $\mathbf{a} \times \mathbf{b}$ is a vector perpendicular to both $\mathbf{a}$ and $\mathbf{b}$;

(ii) $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{a} \times \mathbf{b}$ satisfy right hand rule

(iii) the magnitude of $\mathbf{a} \times \mathbf{b}$ is equal to the area of the paralellogram with sides $\mathbf{a}$ and $\mathbf{b}$



The right-hand rule gives the direction of $\mathbf{a} \times \mathbf{b}$.

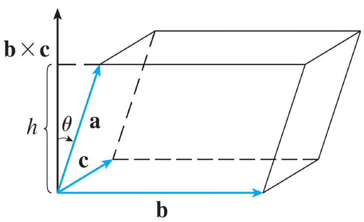


From the above conditions, we have

$$\|\mathbf{a} \times \mathbf{b}\| = \|\mathbf{a}\|\|\mathbf{b}\| \sin\theta$$

and

$$\begin{aligned} \mathbf{a} \times \mathbf{b} &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} \\ &= (a_2 b_3 - a_3 b_2)\mathbf{i} + (a_3 b_1 - a_1 b_3)\mathbf{j} + (a_1 b_2 - a_2 b_1)\mathbf{k}. \end{aligned}$$



The volume of the parallelepiped whose adjacent sides are the vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ is $\left|\mathbf{c} \cdot (\mathbf{a} \times \mathbf{b})\right|$.

Remark.

(i) $(\mathbf{a} \times \mathbf{b}) \times \mathbf{c} \neq \mathbf{a} \times (\mathbf{b} \times \mathbf{c})$;

(ii) $(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} = 0$ if and only if the vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ are in the same plane.

Definition 13.1.4.

(a) The "*standard basis vectors*" in $\mathbb{R}^3$ denote

$$\mathbf{i} =< 1, 0, 0 >, \quad \mathbf{j} =< 0, 1, 0 >, \quad \mathbf{k} =< 0, 0, 1 > .$$

Then $\mathbf{a} =< a_1, a_2, a_3 >= a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$.

(b) A unit vector is a vector with magnitude 1.

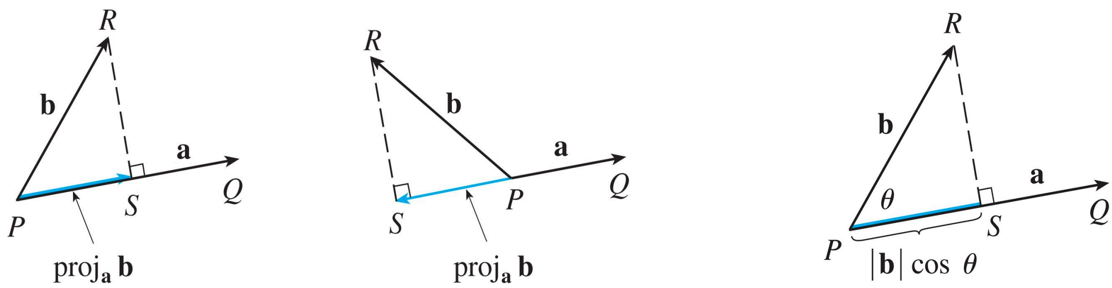
(c) We say that the two vectors $\mathbf{a}$ and $\mathbf{b}$ are "*parallel*" (usually denoted $\mathbf{a} /\!/ \mathbf{b}$) if there exists a number c such that $\mathbf{a} = c\mathbf{b}$.

(d) A vector (denoted by $\text{proj}_{\mathbf{a}}\mathbf{b}$) is called the "*projection of* $\mathbf{b}$ *onto* $\mathbf{a}$" if

$$\text{proj}_{\mathbf{a}}\mathbf{b} /\!/ \mathbf{a} \qquad \text{and} \qquad (\mathbf{b} - \text{proj}_{\mathbf{a}}\mathbf{b}) \perp \mathbf{a}.$$

Note. We can compute that

$$\text{proj}_{\mathbf{a}}\mathbf{b} = \Big(\mathbf{b} \cdot \frac{\mathbf{a}}{||\mathbf{a}||}\Big) \frac{\mathbf{a}}{||\mathbf{a}||} = \frac{\mathbf{a} \cdot \mathbf{b}}{||\mathbf{a}||^2}\,\mathbf{a} \quad \text{and} \quad ||\text{proj}_{\mathbf{a}}\mathbf{b}|| = ||\mathbf{b}|| \cos\theta.$$



Vector projections

Scalar projection

13.2 Vector Functions and Space Curves

As we know, we can regard $\mathbb{R}^n$ as a n-dimensional vector space. Every element in $\mathbb{R}^n$ can be expressed as a vector $\mathbf{a} =< a_1, \cdots, a_n >$. In this chapter, we consider the function whose range consisting of vectors in 3-dimensional vector space $\mathbb{R}^3$.

Definition 13.2.1. A vector-valued function (vector function) is a function whose domain is a set of real numbers and whose range is a set of vectors

$$\mathbf{r}(t) : \{\text{subset in } \mathbb{R}\} \longrightarrow \{\text{set of vectors}\}.$$

Note. In the present chapter, we will focus the vector function $\mathbf{r}(t)$ whose values are three-dimensional vectors.

We recall the expressions of vectors and vector-valued functions.

$$\begin{aligned}\mathbf{a} &= < a_1, a_2, a_3 >= a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k} \\ \mathbf{r}(t) &= < f(t), g(t), h(t) >= f(t)\mathbf{i} + g(t)\mathbf{j} + h(t)\mathbf{k}\end{aligned}$$

where $f, g, h : \mathbb{R} \to \mathbb{R}$ are component functions

Example 13.2.2. Let $\mathbf{r}(t) =< 2t^2, 3t - 4, \sqrt{t} >$ be a vector-valued function. The domain of $\mathbf{r}(t)$ is $[0, \infty)$.

❑ Limits of Vector-valued Functons

To study the calculus of vector-valued functions, motivated by the concepts of real-valued functions, we will discuss the limits and continuity of vector-valued functions. We heuristically consider that

(i) a limit of a vector valued function is supposed to be a vector; and

(ii) if $\mathbf{L}$ is the limit of a vector valued function $\mathbf{r}(t)$ as $t \to a$, we expect that $\mathbf{r}(t)$ arbitrarily approaches to $\mathbf{L}$ by taking t arbitrarily close to a.

Definition 13.2.3. Let $\mathbf{r}(t)$ be a vector valued function defined on an open interval I and $a \in I$. We say that the limit of $\mathbf{r}(t)$ exists, as t approaches a if there exists a vector $\mathbf{L}$ such that

$$\lim_{t\to a} \|\mathbf{r}(t) - \mathbf{L}\| = 0.$$

The vector $\mathbf{L}$ is called the "*limit of* $\mathbf{r}(t)$ *as t arpproaches a*" and we write

$$\lim_{t\to a} \mathbf{r}(t) = \mathbf{L}.$$

Remark. Suppose that $\mathbf{r}(t) =< f(t), g(t), h(t) >$ and $\mathbf{L} =< L_1, L_2, L_3 >$. Then $\lim\limits_{t\to a} \mathbf{r}(t)$ exists if and only if $\lim\limits_{t\to a} f(t)$, $\lim\limits_{t\to a} g(t)$ and $\lim\limits_{t\to a} h(t)$ exist. Moreover,

$$\lim_{t\to a} \mathbf{r}(t) = \mathbf{L} \quad \text{if and only if} \quad \lim_{t\to a} f(t) = L_1,\ \lim_{t\to a} g(t) = L_2 \text{ and } \lim_{t\to a} h(t) = L_3.$$

This implies that

$$\lim_{t\to a} \mathbf{r}(t) =< \lim_{t\to a} f(t),\ \lim_{t\to a} g(t),\ \lim_{t\to a} h(t) > .$$

Example 13.2.4. Suppose that $\mathbf{r}(t) = (1 + t^3)\mathbf{i} + te^{-t}\mathbf{j} + \dfrac{\sin t}{t}\mathbf{k}$. Then

$$\lim_{t\to 0} \mathbf{r}(t) = [\lim_{t\to 0}(1 + t^3)]\ \mathbf{i} + [\lim_{t\to 0} te^{-t}]\ \mathbf{j} + [\lim_{t\to 0} \frac{\sin t}{t}]\ \mathbf{k} = \mathbf{i} + \mathbf{k}.$$

■ Laws of limts

Theorem 13.2.5. *Let* $\mathbf{r}(t)$ *and* $\mathbf{s}(t)$ *be vector valued functions defined on I, u be a real-valued function defined on I and* α *be a number. Suppose that*

$$\lim_{t\to a} \mathbf{r}(t) = \mathbf{L},\ \lim_{t\to a} \mathbf{s}(t) = \mathbf{M} \quad \textit{and} \quad \lim_{t\to a} u(t) = c.$$

Then

(a) $\lim_{t\to a} \left(\mathbf{r} \pm \mathbf{s}\right)(t) = \mathbf{L} \pm \mathbf{M}$.

(b) $\lim_{t\to a} \alpha\mathbf{r}(t) = \alpha\mathbf{L}$.

(c) $\lim_{t\to a} \mathbf{r}(t)\cdot\mathbf{s}(t) = \mathbf{L}\cdot\mathbf{M}$.

(d) $\lim_{t\to a} u(t)\mathbf{r}(t) = c\mathbf{L}$.

(e) $\lim_{t\to a} \mathbf{r}(t)\times\mathbf{s}(t) = \mathbf{L}\times\mathbf{M}$.

Proof. (Exercise) □

❑ Continuity of Vector-valued Functons

Definition 13.2.6. Let $\mathbf{r}(t)$ be a vector valued function defined on $I \subseteq \mathbb{R}$ and $a \in I$. We say that

(a) $\mathbf{r}$ is continuous at a if

$$\lim_{t\to a} \mathbf{r}(t) = \mathbf{r}(a).$$

(b) $\mathbf{r}$ is continuous on I if $\mathbf{r}$ is continuous at every point of I.

Note. If $\mathbf{r}(t) =< f(t), g(t), h(t) >$ is continuous at a, then

$$< \lim_{t\to a} f(t),\ \lim_{t\to a} g(t),\ \lim_{t\to a} h(t) >= \lim_{t\to a} \mathbf{r}(t) = \mathbf{r}(a) =< f(a),\ g(a),\ h(a) > .$$

We have

$$\lim_{t\to a} f(t) = f(a), \quad \lim_{t\to a} g(t) = g(a) \quad \text{and} \quad \lim_{t\to a} h(t) = h(a)$$

Thus, $\mathbf{r}(t)$ is continuous at a if and only if all its component functions f, g and h are continuous at a.

Theorem 13.2.7. *Let $\mathbf{r}(t)$ and $\mathbf{s}(t)$ be vector valued functions defined on I, u be a real-valued function defined on I and α be a number. Suppose that $\mathbf{r}$, $\mathbf{s}$ and u are continuous at a. Then $\mathbf{r}\pm\mathbf{s}$, $\alpha\mathbf{r}$, $u\mathbf{r}$, $\mathbf{r}\cdot\mathbf{s}$ and $\mathbf{r}\times\mathbf{s}$ are continuous at a.*

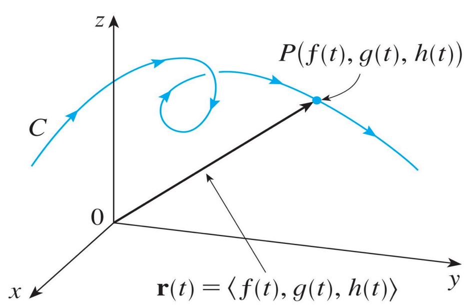
Proof. Exercise. □

❑ Space Curves

Consider the vector valued function $\mathbf{r}(t) =< f(t), g(t), h(t) >$. The tip of $\mathbf{r}(t)$ is the point $P\big(f(t), g(t), h(t)\big)$.

As t ranges over an interval I, the point P traces out some path C in $\mathbb{R}^3$. That is,

$$C = Range\big(\mathbf{r}(t)\big), \quad t \in I.$$



C is traced out by the tip of a moving position vector $\mathbf{r}(t)$.

Definition 13.2.8. Let $f(t)$, $g(t)$ and $h(t)$ be three functions defined on an interval I. The set C of all points (x, y, z) in space, where

$$x = f(t), \quad y = g(t), \quad z = h(t) \quad \text{for } t \in I \tag{13.1}$$

is called a "*space curve*".

Note.

(1) The equation (13.1) is called "*parametric equation of C*" and t is called a "*parameter*".

(2) The space curve C is "*oriented*" in the direction as t increases.

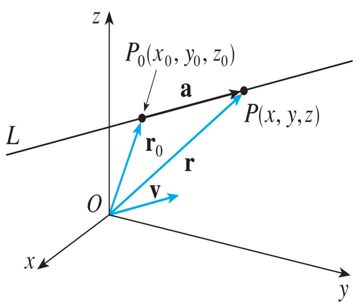
Example 13.2.9. Describe the curve defined by the vector function

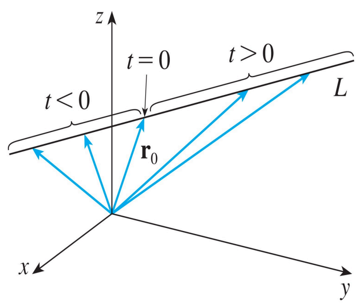
$$\mathbf{r}(t) =< 5 + t, 1 + 4t, 3 - 2t >$$

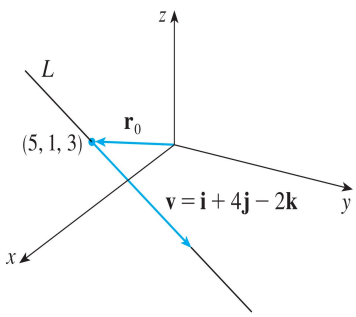
Proof. From the parametric equation, the coordinates are

$$x = 5 + t, \quad y = 1 + 4t, \quad z = 3 - 2t.$$

The curve represents a line passing through $(5, 1, 3)$ and parallel to the vector $< 1, 4, -2 >$. Let $\mathbf{r_0} =< 5, 1, -3 >$ and $\mathbf{v} =< 1, 4, -2 >$. Then $\mathbf{r}(t) = \mathbf{r_0} + t\mathbf{v}$. □







Example 13.2.10. Sketch the curve whose vector equation is

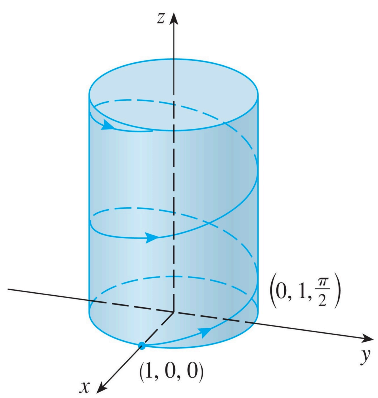
$$\mathbf{r}(t) = \cos t\ \mathbf{i} + \sin t\ \mathbf{j} + t\ \mathbf{k}$$

Proof.

The parametric equation represents the curve with coordinates

$$x = \cos t, \quad y = \sin t, \quad z = t.$$

The curve is called a "*helix*".

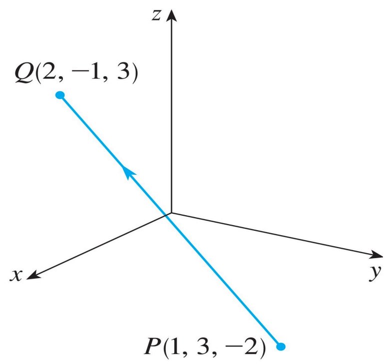


□

Example 13.2.11. Find a vector equation and parametric equations for the line segment that joins the point $P(1, 3, -2)$ to the point $Q(2, -1, 3)$.

Proof.
The line segment joining the tip of $\mathbf{r_0} =< 1, 3, -2 >$ to the tip of $\mathbf{r_1} =< 2, -1, 3 >$ is

$$\mathbf{r}(t) = (1 - t)\mathbf{r_0} + t\mathbf{r_1}, \quad 0 \le t \le 1.$$



The vector equation of the line segment is

$$\begin{aligned}\mathbf{r}(t) &= (1 - t) < 1, 3, -2 > +t < 2, -1, 3 > \\ &= < 1 + t, 3 - 4t, -2 + 5t >, \quad 0 \le t \le 1.\end{aligned}$$

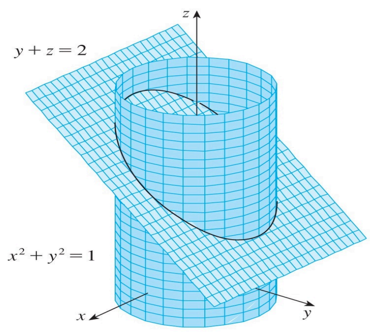
The parametric equation of the line segment is

$$x = 1 + t, \quad y = 3 - 4t, \quad z = -2 + 5t \qquad 0 \le t \le 1.$$

□

Example 13.2.12. Find a vector function that represents the curve of intersection of the cylinder $x^2 + y^2 = 1$ and the plane $y + z = 2$.

Proof.

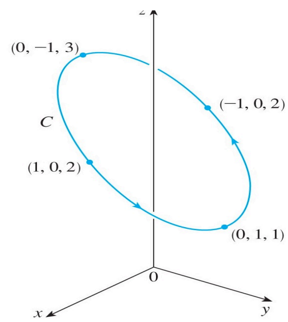


For (x, y, z) on the cylinder $x^2 + y^2 = 1$,

$$x = \cos t, \quad y = \sin t \quad 0 \le t \le 2\pi.$$

Also, for (x, y, z) on the plane $y + z = 2$, $z = 2 - y$. Then for (x, y, z) on the intersection of $x^2 + y^2 = 1$ and $y + z = 2$,

$$z = 2 - y = 2 - \sin t, \quad 0 \le t \le 2\pi.$$



Hnece, the parametric equation for the curve is

$$x = \cos t, \quad y = \sin t, \quad z = 2 - \sin t \quad 0 \le t \le 2\pi.$$

The parametrization of the curve is

$$\mathbf{r}(t) = \cos t\,\mathbf{i} + \sin t\,\mathbf{j} + (2 - \sin t)\,\mathbf{k} \quad 0 \le t \le 2\pi.$$

□

13.3 Derivatives and Integrals of Vector Functions

❑ Derivatives

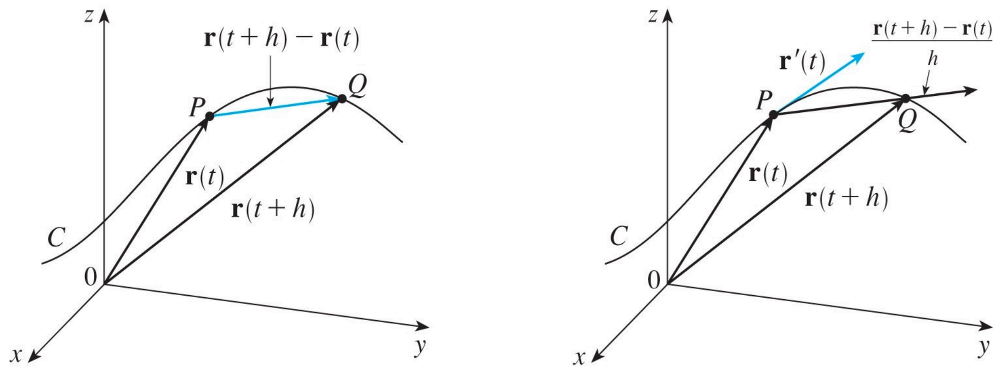
Recall that the derivative of a real-valued function f is defined by

$$\frac{df}{dx} = f'(x) = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h}$$

Let $\mathbf{r}(t)$ be a vector-valued function. Consider

$$\frac{d\mathbf{r}}{dt} = \mathbf{r}'(t) = \lim_{h \to 0} \frac{\mathbf{r}(t+h) - \mathbf{r}(t)}{h}$$

if the limit exists.



(a) The secant vector (b) The tangent vector

Note. (1) The numernator $\mathbf{r}(t+h) - \mathbf{r}(t) = \overrightarrow{PQ}$ means a secant vector.

(2) The term $\dfrac{\mathbf{r}(t+h) - \mathbf{r}(t)}{h}$ represents the vector $\dfrac{1}{h}\big(\mathbf{r}(t+h) - \mathbf{r}(t)\big)$ which has the same direction as $\mathbf{r}(t+h) - \mathbf{r}(t)$.

(3) As $h \to 0$, the vector $\dfrac{1}{h}\big(\mathbf{r}(t+h) - \mathbf{r}(t)\big)$ approaches a vector which lies on the tangent line.

Definition 13.3.1. Let $\mathbf{r}(t)$ be a vector function defined on $I \subseteq \mathbb{R}$, C be the curve consisting of the graph of $\mathbf{r}(t)$ and $P = \mathbf{r}(a)$ be a point on C.

(a) We say that $\mathbf{r}(t)$ is differentiable at a if the limit $\displaystyle\lim_{h \to 0} \frac{\mathbf{r}(a+h) - \mathbf{r}(a)}{h}$ exists. The limit is called the "*derivative of* $\mathbf{r}$ *at* a" and denoted by $\mathbf{r}'(a)$. Moreover, we say $\mathbf{r}(t)$ is differentiable on I if it is differentiable at every point in I.

(b) If the derivative $\mathbf{r}'(a)$ exists, it is the "*tangent vector*" to the curve C at the point P provided $\mathbf{r}'(a) \neq \mathbf{0}$.

(c) The "tangent line" to C at P is defined to be the line through P parallel to the tangnet vector $\mathbf{r}'(a)$.

(d) The unit tangent vector is

$$\mathbf{T}(t) = \frac{\mathbf{r}'(t)}{\|\mathbf{r}'(t)\|}.$$

Note. From the definition of part(c), the parametric equation of the tangent line to C at P is

$$\mathbf{r}(a) + t\mathbf{r}'(a), \quad t \in \mathbb{R}.$$

Theorem 13.3.2. *If* $\mathbf{r}(t) = < f(t), g(t), h(t) > = f(t)\,\mathbf{i} + g(t)\,\mathbf{j} + h(t)\,\mathbf{k}$*, where* f, g *and* h *are differentiable functions, then*

$$\mathbf{r}'(t) = < f'(t), g'(t), h'(t) > = f'(t)\,\mathbf{i} + g'(t)\,\mathbf{j} + h'(t)\,\mathbf{k}.$$

Proof. (Exercise) □

Example 13.3.3. Suppose that $\mathbf{r}(t) = (1+t^3)\,\mathbf{i} + te^{-t}\,\mathbf{j} + \sin 2t\,\mathbf{k}$.

(a) The tangent vector function is $\mathbf{r}'(t) = 3t^2\,\mathbf{i} + (1-t)e^{-t}\,\mathbf{j} + 2\cos 2t\,\mathbf{k}$.

(b) To find the unit tangent vector at the point where $t = 0$, consider the position vector $\mathbf{r}(0) = \mathbf{i}$ and the tangnet vector $\mathbf{r}'(0) = \mathbf{j} + 2\mathbf{k}$. Therefore, the unit tangent vector at the point $(1, 0, 0)$ is

$$\mathbf{T}(0) = \frac{\mathbf{r}'(0)}{\|\mathbf{r}'(0)\|} = \frac{1}{\sqrt{5}}(\mathbf{j} + 2\mathbf{k}) = \frac{1}{\sqrt{5}}\mathbf{j} + \frac{2}{\sqrt{5}}\mathbf{k}.$$

Example 13.3.4. For the curve $\mathbf{r}(t) = \sqrt{t}\,\mathbf{i} + (2-t)\,\mathbf{j}$, find $\mathbf{r}'(t)$ and sketch the position vector $\mathbf{r}(1)$ and the tangent vector $\mathbf{r}'(1)$.

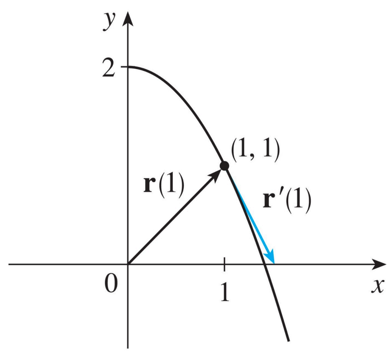
Proof. The tangent vector is $\mathbf{r}'(t) = \dfrac{1}{2\sqrt{t}}\,\mathbf{i} - \mathbf{j}$. Then $\mathbf{r}(1) = \mathbf{i} + \mathbf{j}$ and $\mathbf{r}'(1) = \dfrac{1}{2}\,\mathbf{i} - \mathbf{j}$.

To sketch the position vector and the tangent vector, consider the parametric equation

$$x = \sqrt{t}, \quad y = 2 - t \qquad \Rightarrow \qquad y = 2 - x^2,\ x \geq 0.$$

Then parametric equation of the tangent line to the plane curve at $(1, 1)$ is

$$\ell(t) = \mathbf{r}(1) + t\mathbf{r}'(1) = (\mathbf{i}+\mathbf{j}) + t(\frac{1}{2}\mathbf{i} - \mathbf{j}) = (1 + \frac{1}{2}t)\,\mathbf{i} + (1-t)\,\mathbf{j}$$



□

Example 13.3.5. Find parametric equations for the tangent line to the helix with parametric equation

$$x = 2\cos t, \quad y = \sin t, \quad z = t.$$

at the point $(0, 1, \frac{\pi}{2})$.

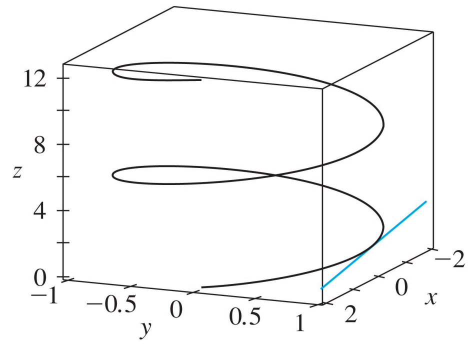
Proof.

The vector function is $\mathbf{r}(t) =< 2\cos t, \sin t, t >$. Then the tangent vector function is

$$\mathbf{r}'(t) =< -2\sin t, \cos t, 1 > .$$

At the point $(0, 1, \frac{\pi}{2})$, $\mathbf{r}(t) =< 2\cos t, \sin t, t >=< 0, 1, \frac{\pi}{2} >$. Thus, $t = \frac{\pi}{2}$. The tangent vector is

$$\mathbf{r}'(\frac{\pi}{2}) =< -2, 0, 1 >.$$



Hence, the parametric equation of the tangent line through $(0, 1, \frac{\pi}{2})$ is

$$x = 0 + (-2)t = -2t, \quad y = 1 + 0t = 1, \quad z = \frac{\pi}{2} + t.$$

□

Theorem 13.3.6. *Suppose that $\mathbf{r}(t)$ is differentiable at a. Then it is continuous at a.*

Proof. Let $\mathbf{r}(t) =< f(t), g(t), h(t) >$. Since $\mathbf{r}(t)$ is differentiable at a, f, g and h are also differentiable at a and hence they are continuous at a. This implies that $\mathbf{r}(t)$ is continuous at a. □

■ Second Derivatives

$$\mathbf{r}(t) \quad \xrightarrow{\frac{d}{dt}} \quad \mathbf{r}'(t) \quad \xrightarrow{\frac{d}{dt}} \quad \left(\mathbf{r}'\right)' = \mathbf{r}''(t)$$

$$\begin{aligned} \mathbf{r}(t) =< f(t), g(t), h(t) > \quad &\Rightarrow \quad \mathbf{r}'(t) =< f'(t), g'(t), h'(t) > \\ &\Rightarrow \quad \mathbf{r}''(t) =< f''(t), g''(t), h''(t) > . \end{aligned}$$

Similarly, if $\mathbf{r}^{(k)}(t)$ exists, then

$$\mathbf{r}^{(k)}(t) =< f^{(k)}(t), g^{(k)}(t), h^{(k)}(t) > .$$

■ Differentiation Rules

Theorem 13.3.7. *Let $\mathbf{r}$ and $\mathbf{s}$ be two differentiable vector functions, c be a number and u be a real-valued function. Then*

(a) $\dfrac{d}{dt}[\mathbf{r}(t) + \mathbf{s}(t)] = \mathbf{r}'(t) + \mathbf{s}'(t)$.

(b) $\dfrac{d}{dt}[c\mathbf{r}(t)] = c\mathbf{r}'(t)$.

(c) $\dfrac{d}{dt}[u(t)\mathbf{r}(t)] = u'(t)\mathbf{r}(t) + u(t)\mathbf{r}'(t)$.

(d) $\dfrac{d}{dt}[\mathbf{r}(t) \cdot \mathbf{s}(t)] = \mathbf{r}'(t) \cdot \mathbf{s}(t) + \mathbf{r}(t) \cdot \mathbf{s}'(t)$. *(real-valued function)*

(e) $\dfrac{d}{dt}[\mathbf{r}(t) \times \mathbf{s}(t)] = \mathbf{r}'(t) \times \mathbf{s}(t) + \mathbf{r}(t) \times \mathbf{s}'(t).$

Exercise. Let $\mathbf{r}(t) =< e^{3t},\ \sin(t^2),\ 2t^2 - t >$, $\mathbf{s}(t) =< \dfrac{t^2}{t+1},\ \sec(2t),\ \ln(t^2+1) >$ and $\mathbf{u}(t) =< 1,\ t,\ t^2 >$. Find $\dfrac{d}{dt}\Big((\mathbf{r} \times \mathbf{s}) \cdot \mathbf{u}\Big)$.

Proposition 13.3.8. *Let $\mathbf{r}(t)$ be a differentiable vector function on I and $\mathbf{r}'(t) \neq \mathbf{0}$ for every $t \in I$. Then*

(a) $\dfrac{d}{dt}\|\mathbf{r}(t)\| = \dfrac{\mathbf{r}(t) \cdot \mathbf{r}'(t)}{\|\mathbf{r}(t)\|}.$

(b) $\dfrac{d}{dt}\left(\dfrac{\mathbf{r}(t)}{\|\mathbf{r}(t)\|}\right) = \dfrac{-\|\mathbf{r}\|'}{\|\mathbf{r}\|^2}\mathbf{r} + \dfrac{1}{\|\mathbf{r}\|}\mathbf{r}' \overset{(n=3)}{=} \dfrac{1}{\|\mathbf{r}\|^2}\big[(\mathbf{r} \times \mathbf{r}') \times \mathbf{r}\big].$

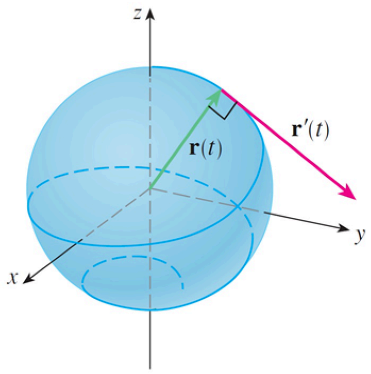
Proof. (Direct computation! We left the proof to the readers as exercise)

□

Remark. The results of Proposition 13.3.8 are true for all n-dimensional vector valued functions except for the last equality of part(b) which is true for 3-dimensional vector valued functions.

Example 13.3.9. Show that if $\|\mathbf{r}(t)\| = C$, then $\mathbf{r}'(t)$ is orthogonal to $\mathbf{r}(t)$ for all t.

Proof.



Since $\mathbf{r}(t) \cdot \mathbf{r}(t) = \|\mathbf{r}(t)\|^2 = C^2$ (constant), we have

$$2\mathbf{r}(t) \cdot \mathbf{r}'(t) = \frac{d}{dt}[\mathbf{r}(t) \cdot \mathbf{r}(t)] = \frac{d}{dt}(C^2) = 0$$

Hence, $\mathbf{r}(t)$ is orthogonal to $\mathbf{r}'(t)$ for all t.

For example, $\mathbf{r}(t) =< \cos t, \sin t >$. □

■ Chain Rules

Theorem 13.3.10. *(Chain Rule) Let $u(t)$ be a real valued function defined on I and $\mathbf{r}(t)$ be a vector valued function whose domain containing the range of u. Suppose that u is differentiable at a and $\mathbf{r}$ is differentiable at $u(a)$, then $(\mathbf{r} \circ u)(t) = \mathbf{r}\big(u(t)\big)$ is differentiable at a and*

$$(\mathbf{r} \circ u)'(a) = u'(a)\mathbf{r}'\big(u(a)\big).$$

Proof. Exercise □

❑ Integrals

Recall that the integral of a real-valued function $f(t)$ over $[a, b]$ is defined by the limit of Riemann sums.

$$\int_a^b f(t)\,dt = \lim_{\|P\|\to 0} \sum_{i=1}^n f(t_i^*)\triangle t_i$$

We try to use the same strategy to define the definite integral of vector-valued functions. Let $\mathbf{r}(t)$ be a continuous vector-valued function defined on $[a, b]$. Let $P = \{t_0, t_1, \cdots, t_n\}$ be a partition of $[a, b]$ and $\triangle t_i = |t_i - t_{i-1}|$. Define

$$\begin{aligned}
\int_a^b \mathbf{r}(t)\,dt &= \lim_{\|P\|\to 0} \sum_{i=1}^n \mathbf{r}(t_i^*)\triangle t_i \\
&= \lim_{\|P\|\to 0} \Big[\sum_{i=1}^n < f(t_i^*), g(t_i^*), h(t_i^*) > \triangle t_i\Big] \\
&= \lim_{\|P\|\to 0} < \sum_{i=1}^n f(t_i^*)\triangle t_i, \sum_{i=1}^n g(t_i^*)\triangle t_i, \sum_{i=1}^n h(t_i^*)\triangle t_i > \\
&= < \int_a^b f(t)\,dt, \int_a^b g(t)\,dt, \int_a^b h(t)\,dt > \\
&= \Big(\int_a^b f(t)\,dt\Big)\,\mathbf{i} + \Big(\int_a^b g(t)\,dt\Big)\,\mathbf{j} + \Big(\int_a^b h(t)\,dt\Big)\,\mathbf{k}
\end{aligned}$$

Definition 13.3.11. Let $\mathbf{r}(t)$ be a vector valued function defined on $[a, b]$ where $\mathbf{r}(t) =< f(t),\ g(t),\ h(t) >$. We say that $\mathbf{r}$ is integrable on $[a, b]$ if f, g,and h are integrable on $[a, b]$ and

$$\begin{aligned}
\int_a^b \mathbf{r}(t)\,dt &= < \int_a^b f(t)\,dt, \int_a^b g(t)\,dt, \int_a^b h(t)\,dt > \\
&= \Big(\int_a^b f(t)\,dt\Big)\,\mathbf{i} + \Big(\int_a^b g(t)\,dt\Big)\,\mathbf{j} + \Big(\int_a^b h(t)\,dt\Big)\,\mathbf{k}.
\end{aligned}$$

Remark. (Integral Rule) If $\mathbf{r}(t)$ is continuous on $[a, b]$, then $\mathbf{r}(t)$ is integrable on $[a, b]$.

Theorem 13.3.12. *Let $\mathbf{r}(t)$ and $\mathbf{s}(t)$ be integrable vector valued functions on $[a, b]$, $\mathbf{c}$ be a vector, and α and β be two numbers. Then*

(a) The vector valued function $(\alpha\mathbf{r} + \beta\mathbf{s})(t)$ is also integrable on $[a, b]$ and

$$\int_a^b (\alpha\mathbf{r} + \beta\mathbf{s})(t)\,dt = \alpha\int_a^b \mathbf{r}(t)\,dt + \beta\int_a^b \mathbf{s}(t)\,dt.$$

(b) $\displaystyle\int_a^b \mathbf{c}\cdot\mathbf{r}(t)\,dt = \mathbf{c}\cdot\int_a^b \mathbf{r}(t)\,dt.$

(c) $\displaystyle\Big\|\int_a^b \mathbf{r}(t)\,dt\Big\| \le \int_a^b \|\mathbf{r}(t)\|\,dt.$

Proof. The proofs of part(a) and (b) are easy and left to the readers. We will prove part(c) here.

Let $\mathbf{R} = \displaystyle\int_a^b \mathbf{r}(t)\, dt$. Then

$$\begin{aligned} \|\mathbf{R}\| \left\| \int_a^b \mathbf{r}(t)\, dt \right\| &= \|\mathbf{R}\|^2 = \mathbf{R} \cdot \mathbf{R} \\ &= \mathbf{R} \cdot \int_a^b \mathbf{r}(t)\, dt = \int_a^b \mathbf{R} \cdot \mathbf{r}(t)\, dt \\ &\leq \int_a^b \left\| \mathbf{R} \cdot \mathbf{r}(t) \right\| dt \leq \int_a^b \|\mathbf{R}\| \|\mathbf{r}(t)\|\, dt \\ &= \left\| \mathbf{R} \right\| \int_a^b \left\| \mathbf{r}(t) \right\| dt. \end{aligned}$$

Hence,

$$\left\| \int_a^b \mathbf{r}(t)\, dt \right\| \leq \int_a^b \|\mathbf{r}(t)\|\, dt.$$

□

■ **Fundamental Theorem of Caluclus**

$$\int_a^b \mathbf{r}(t)\, dt = \mathbf{R}(t)\Big|_a^b = \mathbf{R}(b) - \mathbf{R}(a)$$

where $\mathbf{R}$ is an antiderivative of $\mathbf{r}$, that is $\mathbf{R}'(t) = \mathbf{r}(t)$. Denote

$$\mathbf{R}(t) = \int \mathbf{r}(t)\, dt.$$

Example 13.3.13. Let $\mathbf{r}(t) = 2\cos t\, \mathbf{i} + \sin t\, \mathbf{j} + 2t\, \mathbf{k}$. Then

$$\int \mathbf{r}(t)\, dt = 2\sin t\, \mathbf{i} - \cos t\, \mathbf{j} + t^2\, \mathbf{k} + C$$

and

$$\int_0^{\frac{\pi}{2}} \mathbf{r}(t)\, dt = 2\sin t\Big|_0^{\frac{\pi}{2}} \mathbf{i} - \cos t\Big|_0^{\frac{\pi}{2}} \mathbf{j} + t^2\Big|_0^{\frac{\pi}{2}} \mathbf{k} = 2\,\mathbf{i} + \mathbf{j} + \frac{\pi^2}{4}\,\mathbf{k}.$$

13.4 Arc Length and Curvature

❏ Length of a Curve

In Section 10.1, we have learned how to evaluate the arc length of a parametric curve. Let

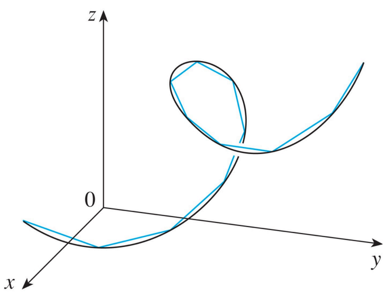
$$x = f(t), \quad y = g(t), \quad a \leq t \leq b.$$

The arc length of the curve is

$$L = \int_a^b \sqrt{[f'(t)]^2 + [g'(t)]^2}\, dt = \int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt.$$

Consider the space curve with the vector equations

$$\mathbf{r}(t) =< f(t), g(t), h(t) >, \quad a \leq t \leq b$$



The length of a space curve is the limit of lengths of inscribed polygons.

If the curve is traversed exactly once as t increases from a to b, the arc length is

$$L = \int_a^b \sqrt{[f'(t)]^2 + [g'(t)]^2 + [h'(t)]^2}\, dt = \int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}\, dt.$$

Note. (1) If $\mathbf{r}(t)$ is the position vector of an object at time t, then $\mathbf{r}'(t)$ is the velocity vector and $\|\mathbf{r}'(t)\|$ is the speed.

(2) Since $\mathbf{r}'(t) =< f'(t), g'(t), h'(t) >$, we have $\|\mathbf{r}'(t)\| = \sqrt{[f'(t)]^2 + [g'(t)]^2 + [h'(t)]^2}$. The arc length is

$$L = \int_a^b \|\mathbf{r}'(t)\|\, dt.$$

We give a precise proof of formula of arc length here.

Theorem 13.4.1. *Let $\mathbf{r}(t)$ be a continuously differentiable vector function on $[a, b]$. Let C be the curve parametrized by $\mathbf{r}$. The arc length of C is*

$$L(C) = \int_a^b \|\mathbf{r}'(t)\|\, dt.$$

Proof. Let $P = \{t_0, t_1, \cdots, t_n\}$ be a partitition of $[a, b]$. By the Fundamental Theorem of Calculus,

$$\|\mathbf{r}(t_i) - \mathbf{r}(t_{i-1})\| = \left\| \int_{t_{i-1}}^{t_i} \mathbf{r}'(t)\, dt \right\| \le \int_{t_{i-1}}^{t_i} \|\mathbf{r}'(t)\|\, dt.$$

Then

$$\sum_{i=1}^n \|\mathbf{r}(t_i) - \mathbf{r}(t_{i-1})\| \le \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \|\mathbf{r}'(t)\|\, dt = \int_a^b \|\mathbf{r}'(t)\|\, dt.$$

Since P is an arbitrary partition of $[a, b]$, we have

$$L(C) = \sup_P \sum_{i=1}^n \|\mathbf{r}(t_i) - \mathbf{r}(t_{i-1})\| \le \int_a^b \|\mathbf{r}'(t)\|\, dt. \tag{13.2}$$

On the other hand, define $s(t)$ as arc length of the curve from $\mathbf{r}(a)$ to $\mathbf{r}(t)$. Then $s(t+h) - s(t)$ is the arc length from $\mathbf{r}(t)$ to $\mathbf{r}(t + h)$. □

By (13.2),

$$\|\mathbf{r}(t + h) - \mathbf{r}(t)\| \le s(t + h) - s(t) \le \int_t^{t+h} \|\mathbf{r}'(u)\|\, du.$$

Then, for $h > 0$,

$$\left\| \frac{\mathbf{r}(t + h) - \mathbf{r}(t)}{h} \right\| = \frac{\|\mathbf{r}(t + h) - \mathbf{r}(t)\|}{h} \le \frac{s(t + h) - s(t)}{h} \le \frac{1}{h} \int_t^{t+h} \|\mathbf{r}'(u)\|\, du.$$

By the Fundamental Theorem of Calculus, as $h \to 0$,

$$\|\mathbf{r}'(t)\| \le \underbrace{\lim_{h\to 0} \frac{s(t + h) - s(t)}{h}}_{=s'(t)} \le \|\mathbf{r}'(t)\|.$$

Therefore, the arc length of C is

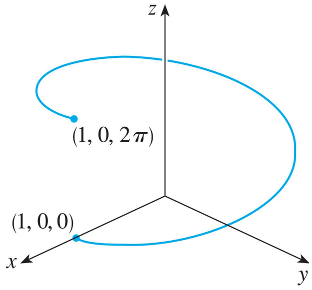
$$s(b) = \int_a^b s'(t)\, dt = \int_a^b \|\mathbf{r}'(t)\|\, dt.$$

Example 13.4.2. Find the length of the arc of the circular helix with vector equation $\mathbf{r}(t) = \cos t\,\mathbf{i} + \sin t\,\mathbf{j} + t\,\mathbf{k}$, from the point $(1, 0, 0)$ to the point $(1, 0, 2\pi)$.

Proof.

Compute $\mathbf{r}'(t) = -\sin t\,\mathbf{i} + \cos t\,\mathbf{j} + \mathbf{k}$ and then $\|\mathbf{r}'(t)\| = \sqrt{(-\sin t)^2 + (\cos t)^2 + 1^2} = \sqrt{2}$. The length of the arc is

$$L = \int_0^{2\pi} \|\mathbf{r}'(t)\|\,dt = \int_0^{2\pi} \sqrt{2}\,dt = 2\sqrt{2}\pi.$$

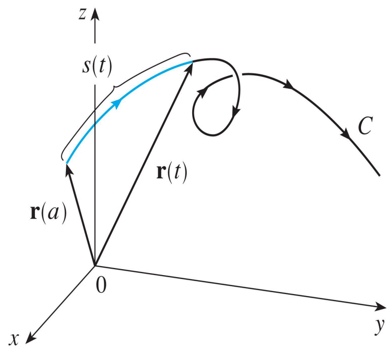


□

■ The Arc Length Function

Let C be a curve with vector function $\mathbf{r}(t) = f(t)\,\mathbf{i} + g(t)\,\mathbf{j} + h(t)\,\mathbf{k}$, $a \le t \le b$. Suppose that $\mathbf{r}'(t)$ is continuous and C is traversed exactly once as t increases from a to b. The arc length function is

$$s(t) = \int_a^t \|\mathbf{r}'(u)\|\,du = \int_a^t \sqrt{\left(\frac{dx}{du}\right)^2 + \left(\frac{dy}{du}\right)^2 + \left(\frac{dz}{du}\right)^2}\,du$$



Note. The value of $s(t)$ is the arc length of the part of C between $\mathbf{r}(a)$ and $\mathbf{r}(t)$. By the Fundamental Theorem of Calculus,

$$\frac{ds}{dt} = \|\mathbf{r}'(t)\|.$$

Observe that the arc length function $s(t)$ is one-to-one. Hence, we may also regard t as a function of s, say $t = t(s)$. Then we can "parametrize a curve with respect to are length.

$$\mathbf{r} = \mathbf{r}\big(t(s)\big).$$

For example, when $s = 3$, $\mathbf{r}\big(t(3)\big)$ is the position vector of the point 3 unit of length along the curve from its starting point.

Example 13.4.3. Reparametrize the helix $\mathbf{r}(t) = \cos t\,\mathbf{i} + \sin t\mathbf{j} + t\,\mathbf{k}$ with respect to arc length measured from $(1, 0, 0)$ in the direction of increasing t

Proof. Find the arc length function from the starting time $t = 0$.

$$s(t) = \int_0^t \|\mathbf{r}'(u)\|\,du = \int_0^t \sqrt{2}\,du = \sqrt{2}t.$$

Hence, $t = t(s) = \frac{1}{\sqrt{2}}s$. We have

$$\mathbf{r}\big(t(s)\big) = \cos(\frac{1}{\sqrt{2}}s)\,\mathbf{i} + \sin(\frac{1}{\sqrt{2}}s)\,\mathbf{j} + \frac{1}{\sqrt{2}}s\,\mathbf{k}.$$

□

❏ Curvature

Question: How do we feel the "curvature" of a curve?

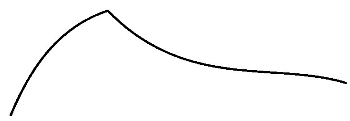
From our expericence, when we ride a bike at a constant speed, it is more difficult to turn the direction along a path with "larger curvature" than the one with a smaller curvature.[*][†]

Small curvature   Large curvature

To discuss the curvature of a curve, we should discard some cases:

(i) Discontinuous curve

(ii) The curve has sharp corners or cusps



(iii) Imagine a particle moves along a curve, we don't expect that it "stays" at a point for a period since it cannot decide whether the direction changes there. Thus, we assume $\|\mathbf{r}'(t)\| \neq 0$. We parametrize the curve with respect to arc length parameter "s" rather than time parameter "t".

Definition 13.4.4.

(a) A parametrization $\mathbf{r}(t)$ is called "*smooth* " on an interval I if $\mathbf{r}'$ is continuous and $\mathbf{r}'(t) \neq \mathbf{0}$ on I.

(b) A curve is called "smooth" if it has a smooth parametrization.

---

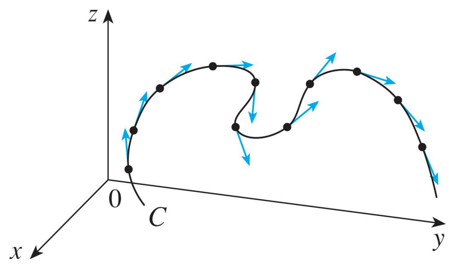
*Heuristically speaking, along the larger curvature path, we need to change directions more at the same time. The constant speed says that the same period is corresponding to the same travelling distance. Thus, we can also explain the larger curvature path as, when travelling the same distance, the direction changes more.

†The "curvature" is a geometric word. It is supposed to only depend on distance and direction but not "time". Hence, to define "curvature", we usually parametrize in s.

Suppose that C is a smooth curve defined by the vector function $\mathbf{r}$. The unit tangnet vector

$$\mathbf{T}(t) = \frac{\mathbf{r}'(t)}{\|\mathbf{r}'(t)\|}$$

indicates the direction of the curve.



Unit tangent vectors at equally spaced points on C

Heuristically, the curvature of C at a given point is a measure of how quickly the curve changes direction at that point.

Specifically, we define it to be the magnitude of the rate of change of the unit tangent vector with respect to arc length.

Definition 13.4.5. The curvature of a curve is

$$\kappa = \left\| \frac{d\mathbf{T}}{ds} \right\|$$

where $\mathbf{T}$ is the unit tangent vector.

Note.

(1) The unit tangent vector $\mathbf{T}$ is usually expressed as a vector function in "t". By the chain rule

$$\frac{d\mathbf{T}}{dt} = \frac{d\mathbf{T}}{ds}\frac{ds}{dt}.$$

Then

$$\kappa = \left\| \frac{d\mathbf{T}}{ds} \right\| = \left\| \frac{d\mathbf{T}/dt}{ds/dt} \right\|.$$

(2) Since the arc length function $s(t) = \int_0^t \|\mathbf{r}'(u)\|\, du$, by the Fundamental Theorem of Calculus, $\frac{ds}{dt} = \|\mathbf{r}'(t)\|$. Hence,

$$\kappa = \frac{\|\mathbf{T}'(t)\|}{\|\mathbf{r}'(t)\|}.$$

Example 13.4.6. Show that the curvature of a circle of radius a is $\frac{1}{a}$.

Proof. A parametrization of a circle of radius a is $\mathbf{r}(t) = a\cos t\ \mathbf{i} + a\cos t\ \mathbf{j}$. Then $\mathbf{r}'(t) = -a\sin t\ \mathbf{i} + a\cos t\ \mathbf{j}$ and $\|\mathbf{r}'(t)\| = a$. The unit tangent vector function is

$$\mathbf{T}(t) = \frac{\mathbf{r}'(t)}{\|\mathbf{r}'(t)\|} = -\sin t\ \mathbf{i} + \cos t\ \mathbf{j}.$$

Then

$$\mathbf{T}'(t) = -\cos t\ \mathbf{i} - \sin t\ \mathbf{j} \quad \text{and} \quad \left\| \frac{d\mathbf{T}}{dt} \right\| = 1.$$

The curvature is

$$\kappa = \frac{\|\mathbf{T}'(t)\|}{\|\mathbf{r}'(t)\|} = \frac{1}{a}.$$

□

Note. Small circles have large curvature and large circles have small curvature.

Theorem 13.4.7. *The curvature of the curve given by the vector function* $\mathbf{r}$ *is*

$$\kappa(t) = \frac{||\mathbf{r}'(t) \times \mathbf{r}''(t)||}{||\mathbf{r}'(t)||^3}.$$

Proof. Since $\mathbf{T} = \dfrac{\mathbf{r}'}{||\mathbf{r}'||}$ and $||\mathbf{r}'|| = \dfrac{ds}{dt}$, we have

$$\mathbf{r}' = ||\mathbf{r}'||\mathbf{T} = \frac{ds}{dt}\mathbf{T}.$$

By the product rule,

$$\mathbf{r}'' = \frac{d^2s}{dt^2}\mathbf{T} + \frac{ds}{dt}\mathbf{T}'.$$

Consider

$$\mathbf{r}' \times \mathbf{r}'' = \frac{ds}{dt}\frac{d^2s}{dt^2}\underbrace{\mathbf{T} \times \mathbf{T}}_{=\mathbf{0}} + \left(\frac{ds}{dt}\right)^2 \mathbf{T} \times \mathbf{T}'.$$

Since $||\mathbf{T}|| = 1$, we have $\mathbf{T}(t) \perp \mathbf{T}'(t)$. Then $||\mathbf{T} \times \mathbf{T}'|| = \underbrace{||\mathbf{T}||}_{=1}\,||\mathbf{T}'|| = ||\mathbf{T}'||$. Also,

$$||\mathbf{r} \times \mathbf{r}''|| = \left(\frac{ds}{dt}\right)^2 ||\mathbf{T} \times \mathbf{T}'|| = \left(\frac{ds}{dt}\right)^2 \underbrace{||\mathbf{T}||}_{=1}\,||\mathbf{T}'|| = \left(\frac{ds}{dt}\right)^2 ||\mathbf{T}'||.$$

Hence,

$$||\mathbf{T}'|| = \frac{||\mathbf{r}' \times \mathbf{r}''||}{\left(\frac{ds}{dt}\right)^2} = \frac{||\mathbf{r}' \times \mathbf{r}''||}{||\mathbf{r}'||^2}.$$

The curvature is

$$\kappa = \frac{||\mathbf{T}'||}{||\mathbf{r}'||} = \frac{||\mathbf{r}' \times \mathbf{r}''||}{||\mathbf{r}'||^3}.$$

□

Example 13.4.8. Find the curvature of the twisted cubic $\mathbf{r}(t) =< t, t^2, t^3 >$ at general point and at $(0, 0, 0)$.

Proof. Since $\mathbf{r}'(t) =< 1, 2t, 3t^2 >$ and $\mathbf{r}''(t) =< 0, 2, 6t >$, we have

$$\mathbf{r}'(t) \times \mathbf{r}''(t) = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 2t & 3t^2 \\ 0 & 2 & 6t \end{vmatrix} =< 6t^2, -6t, 2 > .$$

Then $||\mathbf{r}' \times \mathbf{r}''|| = \sqrt{36t^4 + 36t^2 + 4} = 2\sqrt{9t^4 + 9t^2 + 1}$ and $||\mathbf{r}'|| = \sqrt{1 + 4t^2 + 9t^4}$. The curvature is

$$\kappa = \frac{2\sqrt{9t^4 + 9t^2 + 1}}{(1 + 4t^2 + 9t^4)^{3/2}}.$$

At $t = 0$, $\kappa(0) = 2$. □

• **Special Case** $y = f(x)$

Suppose that the curve C is the graph of $f(x)$. We can express it as vector-valued function.

$$\mathbf{r}(x) = x\,\mathbf{i} + f(x)\,\mathbf{j}\,\Big(+0\,\mathbf{k}\Big).$$

Then

$$\mathbf{r}'(x) = \mathbf{i} + f'(x)\,\mathbf{j} \quad \text{and} \quad \mathbf{r}''(x) = f''(x)\,\mathbf{j}.$$

The cross product is

$$\mathbf{r}'(x) \times \mathbf{r}''(x) = f''(x)\,\mathbf{k}.$$

We have

$$\|\mathbf{r}' \times \mathbf{r}''\| = |f''(x)| \quad \text{and} \quad \|\mathbf{r}'\| = \sqrt{1 + [f'(x)]^2}.$$

Hence, the curvature is

$$\kappa = \frac{\|\mathbf{r}' \times \mathbf{r}''\|}{\|\mathbf{r}'\|^3} = \frac{|f''(x)|}{(1 + [f'(x)]^2)^{3/2}}.$$

Example 13.4.9. Find the curvature of the parabola $y = x^2$ at the point $(0,0)$, $(1,1)$ and $(2,4)$.

Proof.

Compute that $y' = 2x$ and $y'' = 2$. The curvature of the curve is

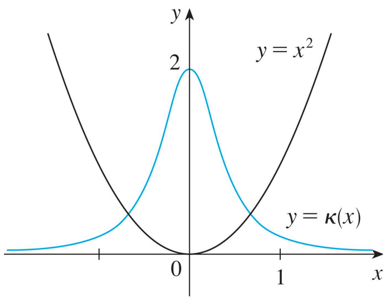
$$\kappa(x) = \frac{|y''|}{[1 + (y')^2]^{3/2}} = \frac{2}{(1 + 4x^2)^{3/2}}$$

At $(0,0)$, $\kappa(0) = 2$.

At $(1,1)$, $\kappa(1) = \dfrac{2}{5^{3/2}} \approx 0.18$.

At $(2,4)$, $\kappa(2) = \dfrac{2}{17^{3/2}} \approx 0.03$.

We can observe that $\kappa(x) \to 0$ as $x \to \pm\infty$.



The parabola $y = x^2$ and its curvature function $y = \kappa(x)$

□

Chapter 14

Partial Derivatives and Differentiability

14.1 Functions of Several Variables

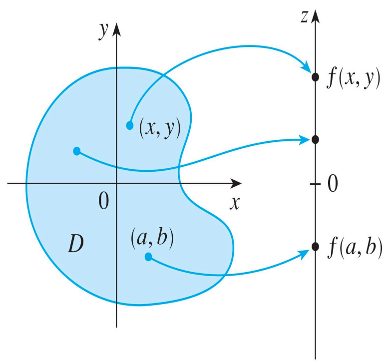
❏ Functions of Two Variables

Example 14.1.1.

(1) Let $T = f(x, y)$ represent the temperature at the position (x, y) where x and y indicate the longitude and latitude respectively.

(2) Let $V = V(r, h)$ represent the volume of a circular cylinder where r and h indicate the raidus and the height of the cylinder respectively.

Definition 14.1.2. A function f of two variables is a rule that assigns to each ordered pair of real numbers (x, y) in a set D a unique real number denoted by $f(x, y)$. The set D is the "*domain*" of f and its "*range*" is the set of values that f takes on. That is, $Range(f) = \{f(x, y) \mid (x, y) \in D\}$.



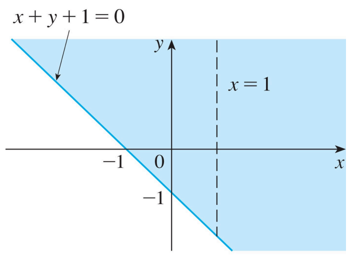
Sometimes, we express $z = f(x, y)$ where x and y are independent variables and z is a dependent variable.

Remark. If a function is given by a formula and no domain is specified, then the domain of f is understood to be the set of all pair(x, y) for which the given expression is a well-defined real number.

Example 14.1.3.

(1) Let $f(x, y) = \dfrac{\sqrt{x+y+1}}{x-1}$. The domain of f is

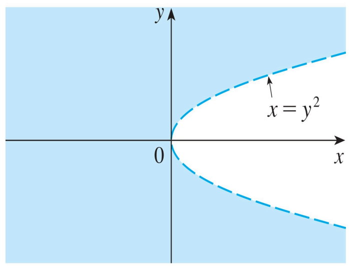
$$\begin{aligned} Dom(f) &= \{(x, y) \mid x+y+1 \geq 0,\ x-1 \neq 0\} \\ &= \{(x, y) \mid y \geq -x-1,\ x \neq 1\}. \end{aligned}$$



Domain of $f(x, y) = \dfrac{\sqrt{x+y+1}}{x-1}$

(2) Let $f(x, y) = x\ln(y^2 - x)$. The domain of f is

$$\begin{aligned} Dom(f) &= \{(x, y) \mid y^2 - x > 0\} \\ &= \{(x, y) \mid x < y^2\}. \end{aligned}$$



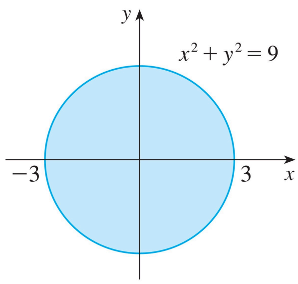
Domain of $f(x, y) = x\ln(y^2 - x)$

(3) Let $g(x, y) = \sqrt{9 - x^2 - y^2}$. The domain of g is

$$\begin{aligned} Dom(f) &= \{(x, y) \mid 9 - x^2 - y^2 \geq 0\} \\ &= \{(x, y) \mid x^2 + y^2 \leq 9\}. \end{aligned}$$

The range of g is

$$\begin{aligned} Range(g) &= \{z \mid z = \sqrt{9 - x^2 - y^2},\ (x, y) \in Dom(g)\} \\ &= \{z \mid 0 \leq z \leq 3\}. \end{aligned}$$



Domain of $g(x, y) = \sqrt{9 - x^2 - y^2}$

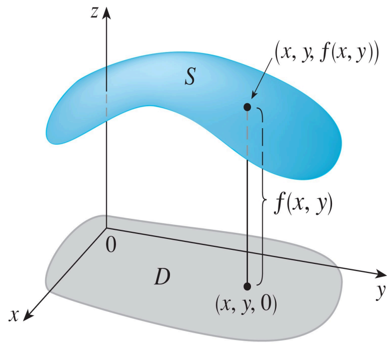
■ Some ways to figure out two variables functions

We introduce some visual methods to understand functions of two variables.

• Graph of a function

Definition 14.1.4. If f is a function of two variables with domain D, then the "*graph*" of f is the set of all points $(x, y, z) \in \mathbb{R}^3$ such that $z = f(x, y)$ and (x, y) is in D. That is,

$$Graph(f) = \{(x, y, z) \mid z = f(x, y),\ (x, y) \in D\}.$$

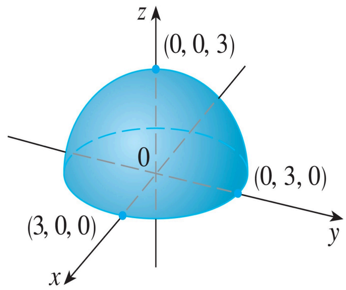


Example 14.1.5. Sketch the graph of $g(x, y) = \sqrt{9 - x^2 - y^2}$.

Proof. Let $z = \sqrt{9 - x^2 - y^2}$. Then the graph of g is

$$\begin{aligned} Graph(g) &= \{(x, y, z) \mid z^2 = 9 - x^2 - y^2,\ z \geq 0\} \\ &= \{(x, y, z) \mid x^2 + y^2 + z^2 = 9,\ z \geq 0\} \end{aligned}$$

□



Graph of $g(x, y) = \sqrt{9 - x^2 - y^2}$

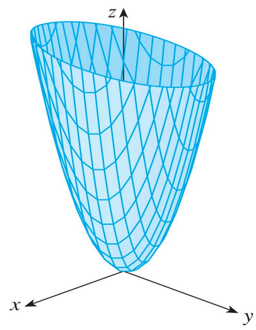
Note. An entire sphere cannot be represented by a single function of x and y. The lower hemisphere is represented by the function $h(x, y) = -\sqrt{9 - x^2 - y^2}$.

Example 14.1.6. Find the domain and range and sketch the graph of $h(x, y) = 4x^2 + y^2$.

Proof. $Dom(h) = \mathbb{R}^2$ and $Range(f) = [0, \infty)$. The graph of h

$$Graph(h) = \{(x, y, z) \mid z = 4x^2 + y^2,\ (x, y) \in \mathbb{R}^2\}$$

is an elliptic paraboliod. □



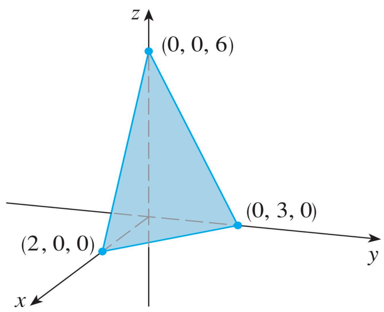
Graph of $h(x, y) = 4x^2 + y^2$

Example 14.1.7. Sketch the graph of the function

$$f(x, y) = 6 - 3x - 2y.$$

Proof. Let $z = 6 - 3x - 2y$ or $3x + 2y + z = 6$. The intercepts of the function are $(2, 0, 0)$, $(0, 3, 0)$ and $(0, 0, 6)$.

□

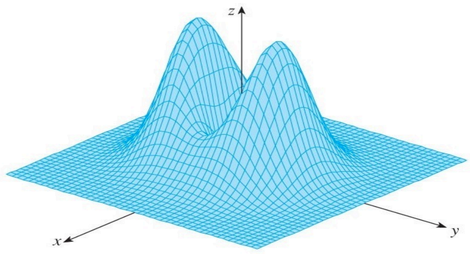


Graph of $f(x, y) = 6 - 3x - 2y$

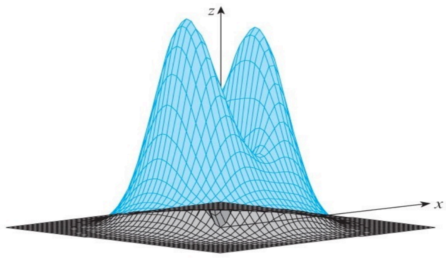
Note. The function $f(x, y) = ax + by + c$ is called a "*linear function*". The graph of such a function is a plane and has the equation $z = ax + by + c$ or $ax + by - z + c = 0$.

- **Computer-generated graphs**

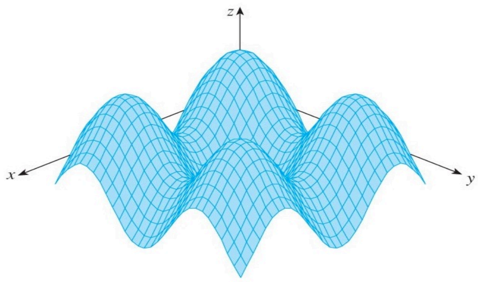
In general, it is difficult to sketch the graph of a two-variables function. A nice method to sketch the traces in the vertical plne $x = k$ and $y = h$. For example, fix $x = k$ and sketch the graph of a single variable function $z = f(k, y)$. It is a curve on the plane $x = k$. Draw all such curve as x ranges over all possible values in the x direction.



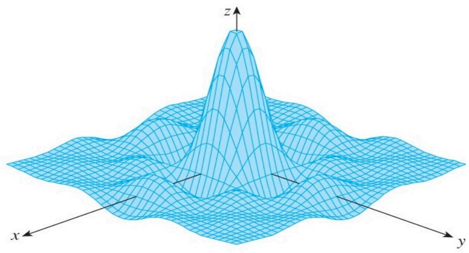
(a) $f(x, y) = (x^2 + 3y^2)e^{-x^2-y^2}$



(b) $f(x, y) = (x^2 + 3y^2)e^{-x^2-y^2}$



(c) $f(x, y) = \sin x + \sin y$



(d) $f(x, y) = \dfrac{\sin x \sin y}{xy}$

• **Level Curves**

So far, we have two methods for visualizing functions: arrow diagrams and graphs. A third method is to consider a contour map on which points of constant elevation are joined to form "*contour curves*", or "*level curves*".

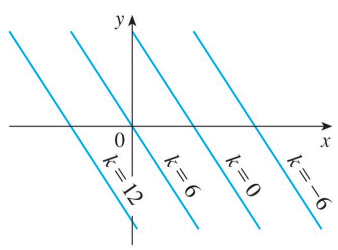
Definition 14.1.8. The "*level curves*" of a function f of two variables are the curves with equation $f(x, y) = k$, where k is a constant (in the range of f). The level curve is the set $\{(x, y) \in D \mid f(x, y) = k\}$.

Note. (1) A level curve $f(x, y) = k$ is the set of all points in the domain of f at which f takes on a given value k. (It shows where the graph of f has height k).

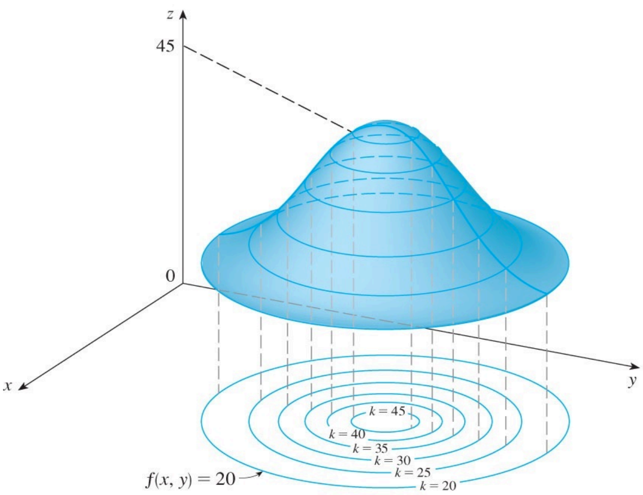
(2) Level curves are useful in the reality. For example, isothermals, contour map, contour line.

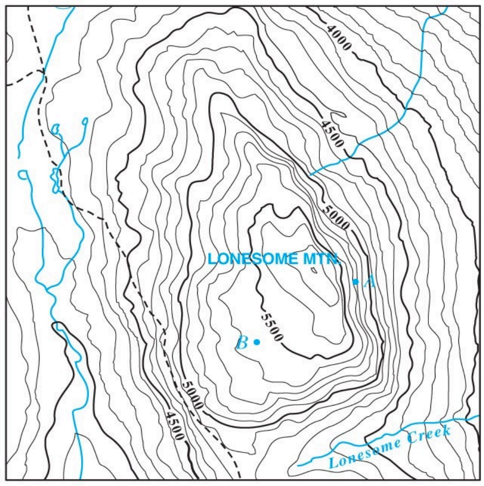
Example 14.1.9. Sketch the level curves of the function $f(x, y) = 6 - 3x - 2y$ for the values $k = -6, 0, 6, 12$.

Proof. Consider the curves $6 - 3x - 2y = k$ in the domain. For $k = -6, 0, 6, 12$, the corresponding level curves are $3x+2y-12 = 0$, $3x + 2y - 6 = 0$, $3x + 2y = 0$ and $3x + 2y + 6 = 0$. □



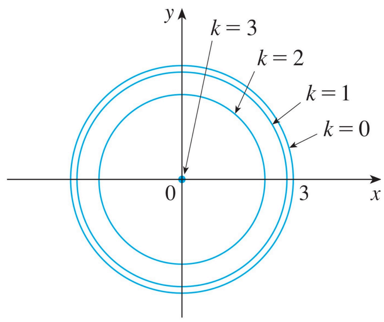
Contour map of $f(x, y) = 6 - 3x - 2y$





Example 14.1.10. Sketch the level curves of the function $g(x, y) = \sqrt{9 - x^2 - y^2}$ for the values $k = 0, 1, 2, 3$.

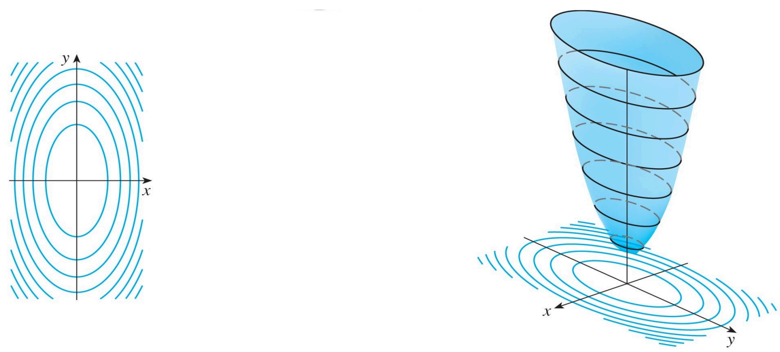
Proof. Consider the curves $\sqrt{9 - x^2 - y^2} = k$ in the domain. For $k = 0, 1, 2, 3$, the corresponding level curves are $x^2 + y^2 = 9$, $x^2 + y^2 = 8$, $x^2 + y^2 = 5$ and $x^2 + y^2 = 0$. □



Contour map of $g(x, y) = \sqrt{9 - x^2 - y^2}$

Example 14.1.11. Sketch the level curves of the function $h(x, y) = 4x^2 + y^2 + 1$.

Proof. Consider the curves $4x^2 + y^2 + 1 = k$ in the domain. We can rewrite the equation as $\dfrac{x^2}{\frac{1}{4}(k-1)} + \dfrac{y^2}{k-1} = 1$. For $k > 1$, the level curves are a family of ellipses with semiaxes $\frac{1}{2}\sqrt{k-1}$ and $\sqrt{k-1}$.



(a) Contour map

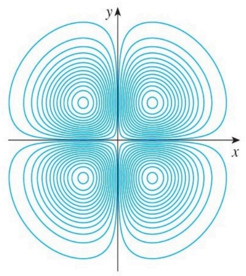
(b) Horizontal traces are raised level curves

The graph of $h(x, y) = 4x^2 + y^2 + 1$
is formed by lifting the level curves.

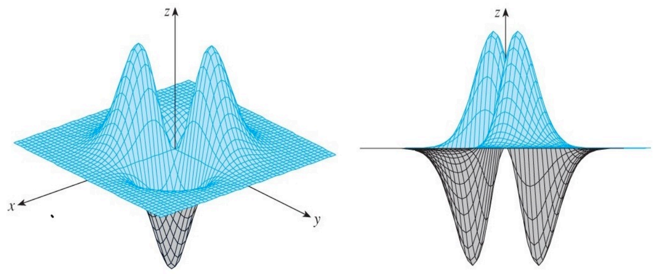
□

Note. The following two figures show different visualized concepts to figure out a two variables functions $f(x, y)$.

(1) $f(x, y) = -xye^{-x^2-y^2}$.

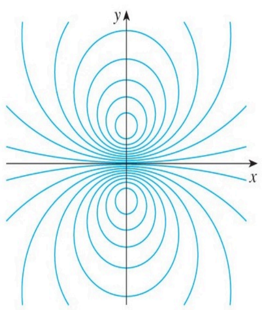


Level curves of $f(x, y) = -xye^{-x^2-y^2}$

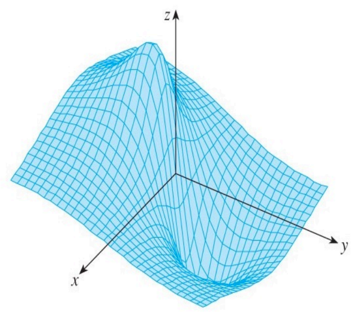


Two views of $f(x, y) = -xye^{-x^2-y^2}$

(2) $f(x, y) = \dfrac{-3y}{x^2 + y^2 + 1}$.



Level curves of $f(x, y) = \dfrac{-3y}{x^2 + y^2 + 1}$



$f(x, y) = \dfrac{-3y}{x^2 + y^2 + 1}$

❏ Functions of Three or More Variables

■ Three variables functions

A function of three variables, f, is a rule that assigns to each ordered triple (x, y, z) in a domain $D \subseteq \mathbb{R}^3$ a unique real number denoted by $f(x, y, z)$.

Example 14.1.12. The function $f(x, y, z) = \ln(z - y) + xy \sin z$ has the domain

$$Dom(f) = \{(x, y, z) \mid z - y > 0\} = \{(x, y, z) \mid z > y\}.$$

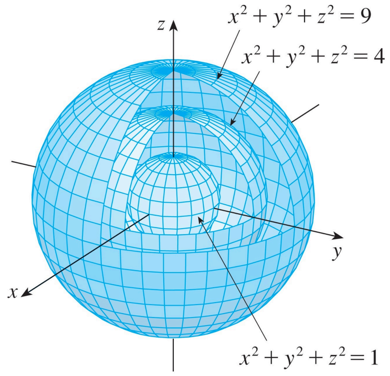
Note. It is difficult to visualize a function f of three variables by its graph since that would lie in four-dimensional space.

We obtain some insight into f by examining its "level surfaces", which are surfaces with equation $f(x, y, z) = k$, where k is a constant in the range of f.

Example 14.1.13. Find the level surfaces of the function $f(x, y, z) = x^2 + y^2 + z^2$.

Proof. Consider the surface with equation $x^2 + y^2 + z^2 = k$, $k \geq 0$. The corresponding level surfaces form a family of concentric spheres with radius $\sqrt{k}$.

□



■ ***n* variables functions**

A function of n variables is a rule that assigns a number $z = f(x_1, x_2, \cdots, x_n)$ to an n-tuple $(x_1, x_2, \cdots, x_n)$ of real numbers.

Example 14.1.14. (Cost function) Let C_i be the cost per unit of the ith ingredient and x_i be the units of the ith ingredient are used. The total cost is

$$C = f(x_1, x_2, \cdots, x_n) = C_1 x_1 + C_2 x_2 + \cdots + C_n x_n.$$

which is a n-variable function.

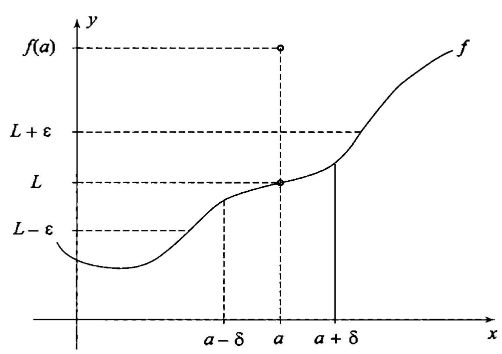
Remark. Since the point $(x_1, x_2, \cdots, x_n)$ and the vector $\mathbf{x} =< x_1, x_2, \cdots, x_n >$ are one-to-one correspondence, we have three ways of looking at a function f defined on a subset of $\mathbb{R}^n$.

1. As a function of n real variables $x_1, x_2, \cdots, x_n$, denote $f(x_1, x_2, \cdots, x_n)$.
2. As a function of a single point variable $(x_1, x_2, \cdots, x_n)$, denote $f\big((x_1, x_2, \cdots, x_n)\big)$.
3. As a function of a single vector variable $\mathbf{x} =< x_1, x_2, \cdots, x_n >$, denote $f(\mathbf{x}) = f(< x_1, x_2, \cdots, x_n >)$.

14.2 Limits and Continuity

❑ Limits

Recall that the limit of a single variable function $f(x)$ as x approaches a is followed by the concept that the value of $f(x)$ approaches L as x tends to a. The precise $\varepsilon - \delta$ definition is given in Chapter 3.



Question: How about the limit of a two variables function f as (x, y) approaches a point (a, b)?

Definition 14.2.1. (Heuristic definition) Let f be a function of two variables whose domain D containing a neighborhood of (a, b) (possibly except (a, b) itself). We say that the limit of

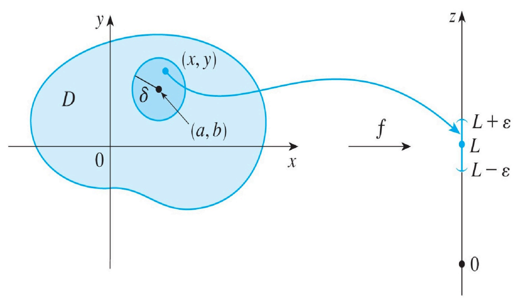
$f(x, y)$ as (x, y) approaches (a, b) exists if there is a number L such that we can make $f(x, y)$ as close to L as we like by taking (x, y) sufficiently close to (a, b).

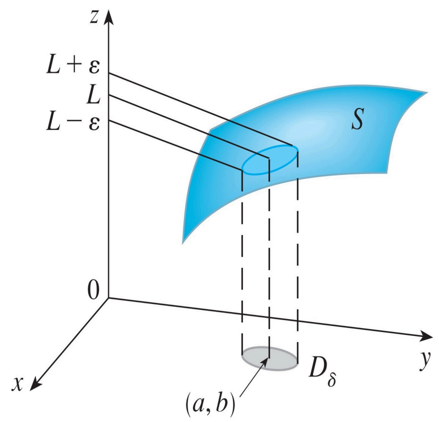
Definition 14.2.2. (Precise definition) Let f be a function of two variables whose domain D containing a neighborhood of (a, b) (possibly except (a, b) itself). We say that the limit of $f(x, y)$, as (x, y) approaches (a, b), exists if there is a number L such that for every number $\varepsilon > 0$ there exists a corresponding number $\delta > 0$ such that

$$|f(x, y) - L| < \varepsilon$$

whenever $(x, y) \in D$ and $0 < \sqrt{(x-a)^2 + (y-a)^2} < \delta$. Denote

$$\lim_{(x,y)\to(a,b)} f(x, y) = L \qquad \text{or} \qquad f(x, y) \to L \quad \text{as } (x, y) \to (a, b).$$



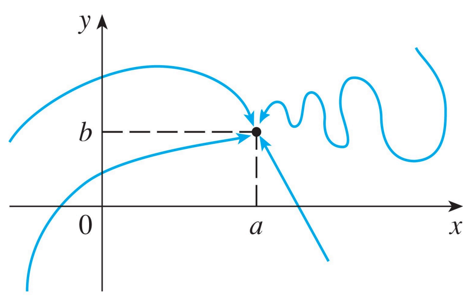


Remark. For functions of a single variable, we only need to consider two possible direction when x approaches a (from the left and from the right).

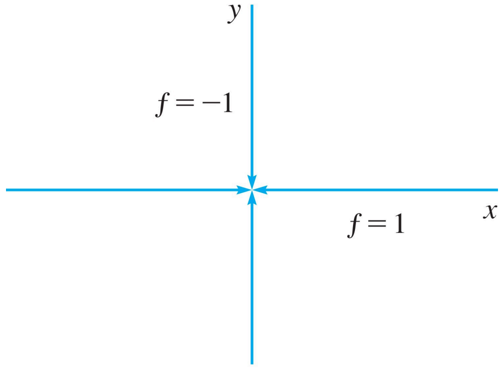
For functoins of two variables, we have to consider an infinite numbers of directions in any manner whatsover as long as (x, y) stays within the domain of f.

Hence, if the limit $\lim_{(x,y)\to(a,b)} f(x, y)$ exists, then $f(x, y)$ must approach the same limit no matter which direction and how (x, y) approaches (a, b).

Note. From the above remark, if $f(x, y) \to L_1$ and (x, y) approaches (a, b) along a path C_1 and $f(x, y) \to L_2$ when (x, y) approaches (a, b) along another path C_2 where $L_1 \neq L_2$, then the limit $\lim_{(x,y)\to(a,b)} f(x, y)$ does not exist.



Example 14.2.3. Let $f(x,y) = \dfrac{x^2 - y^2}{x^2 + y^2}$. Consider the limit of $f(x,y)$ as (x,y) approaches $(0,0)$.



Proof. Along the x-axis ($y = 0$),

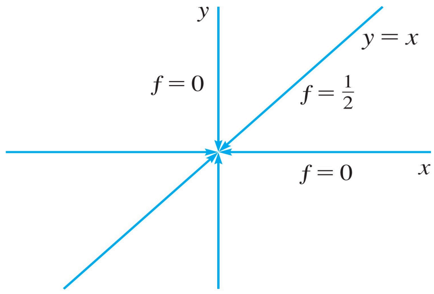
$$\lim_{\substack{(x,y)\to(a,b)\\ y=0}} \frac{x^2 - y^2}{x^2 + y^2} = \lim_{x\to 0} \frac{x^2}{x^2} = 1.$$

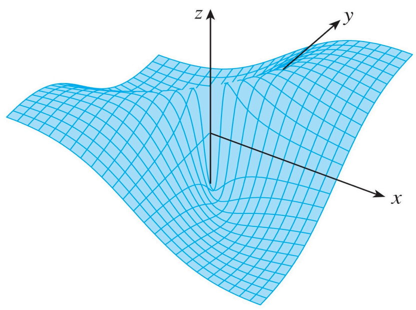
Along the y-axis ($x = 0$),

$$\lim_{\substack{(x,y)\to(a,b)\\ x=0}} \frac{x^2 - y^2}{x^2 + y^2} = \lim_{y\to 0} \frac{-y^2}{y^2} = -1.$$

Hence, the limit $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x^2 - y^2}{x^2 + y^2}$ does not exist. □

Example 14.2.4. If $f(x,y) = \dfrac{xy}{x^2 + y^2}$, does $\displaystyle\lim_{(x,y)\to(0,0)} f(x,y)$ exist?





$f(x,y) = \dfrac{xy}{x^2 + y^2}$

Proof. Along the x-axis ($y = 0$),

$$\lim_{\substack{(x,y)\to(a,b)\\ y=0}} \frac{xy}{x^2 + y^2} = \lim_{x\to 0} \frac{0}{x^2} = 0.$$

Along the y-axis ($x = 0$),

$$\lim_{\substack{(x,y)\to(a,b)\\ x=0}} \frac{xy}{x^2 + y^2} = \lim_{y\to 0} \frac{0}{y^2} = 0.$$

But, along the line $y = x$,

$$\lim_{\substack{(x,y)\to(a,b)\\ x=y}} \frac{xy}{x^2 + y^2} = \lim_{x\to 0} \frac{x^2}{2x^2} = \frac{1}{2}.$$

Hence, the limit $\displaystyle\lim_{(x,y)\to(0,0)} \frac{xy}{x^2 + y^2}$ does not exist. □

Example 14.2.5. If $f(x,y) = \dfrac{xy^2}{x^2 + y^4}$, does $\displaystyle\lim_{(x,y)\to(0,0)} f(x,y)$ exist?

Proof. Along the the line $y = mx$ (not y-axis),

$$\lim_{\substack{(x,y)\to(a,b)\\ y=mx}} \frac{xy^2}{x^2 + y^4} = \lim_{x\to 0} \frac{x(mx)^2}{x^2 + (mx)^4} = \lim_{x\to 0} \frac{x^3(1 + m^2)}{x^2(1 + m^4x^2)} = 0.$$

Along the curve $x = y^2$,

$$\lim_{\substack{(x,y)\to(a,b)\\ x=y^2}} \frac{xy^2}{x^2+y^4} = \lim_{y\to 0} \frac{y^2\cdot y^2}{(y^2)^2+y^4} = \frac{1}{2}.$$

Hence, the limit $\displaystyle\lim_{(x,y)\to(0,0)} \frac{xy^2}{x^2+y^4}$ does not exist.

□

■ **Laws of Limits and Squeeze Theorem**

Theorem 14.2.6. *(Laws of Limits) Let f and g be two variables functions defined on D containing a neighborhood of (a,b) (possibly except (a,b) itself) and c be a constant number. Suppose that the limits $\displaystyle\lim_{(x,y)\to(a,b)} f(x,y)$ and $\displaystyle\lim_{(x,y)\to(a,b)} g(x,y)$ exist. Then*

(a) $\displaystyle\lim_{(x,y)\to(a,b)} [f\pm g](x,y)$ *exists and* $\displaystyle\lim_{(x,y)\to(a,b)} [f\pm g](x,y) = \lim_{(x,y)\to(a,b)} f(x,y) \pm \lim_{(x,y)\to(a,b)} g(x,y)$.

(b) $\displaystyle\lim_{(x,y)\to(a,b)} [cf](x,y)$ *exists and* $\displaystyle\lim_{(x,y)\to(a,b)} [cf](x,y) = c\lim_{(x,y)\to(a,b)} f(x,y)$.

(c) $\displaystyle\lim_{(x,y)\to(a,b)} [fg](x,y)$ *exists and* $\displaystyle\lim_{(x,y)\to(a,b)} [fg](x,y) = \Big(\lim_{(x,y)\to(a,b)} f(x,y)\Big)\Big(\lim_{(x,y)\to(a,b)} g(x,y)\Big)$.

(d) $\displaystyle\lim_{(x,y)\to(a,b)} \left[\frac{f}{g}\right](x,y)$ *exists if* $\displaystyle\lim_{(x,y)\to(a,b)} g(x,y) \neq 0$ *and*

$$\lim_{(x,y)\to(a,b)} \left[\frac{f}{g}\right](x,y) = \frac{\lim_{(x,y)\to(a,b)} f(x,y)}{\lim_{(x,y)\to(a,b)} g(x,y)}$$

provided $\displaystyle\lim_{(x,y)\to(a,b)} g(x,y) \neq 0$.

(e) In particular,

$$\lim_{(x,y)\to(a,b)} x = a, \qquad \lim_{(x,y)\to(a,b)} y = b, \qquad \lim_{(x,y)\to(a,b)} c = c$$

Theorem 14.2.7. *(Squeeze Theorem) Let $f(x,y)$, $g(x,y)$ and $h(x,y)$ be three functions defined near (a,b). Suppose that $f(x,y)\le g(x,y)\le h(x,y)$ for every (x,y) near (a,b). If*

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

then the limit $\displaystyle\lim_{(x,y)\to(a,b)} g(x,y)$ exists and

$$\lim_{(x,y)\to(a,b)} g(x,y) = L.$$

Example 14.2.8. Find $\displaystyle\lim_{(x,y)\to(0,0)} \frac{3x^2y}{x^2+y^2}$ if it exists.

Proof. First of all, we may try the limits when (x, y) approaches $(0, 0)$ along several paths. We observe that all the limits are 0. Therefore, we guess that the limit could exist and equal 0.

Let $\varepsilon > 0$. We want to find $\delta > 0$ such that if $0 < \sqrt{(x-0)^2 + (y-0)^2} < \delta$, then $\left|\dfrac{3x^2y}{x^2+y^2} - 0\right| < \varepsilon$. Consider

$$\left|\frac{3x^2y}{x^2+y^2}\right| = \underbrace{\left|\frac{x^2}{x^2+y^2}\right|}_{<1} \cdot 3|y| < 3|y|.$$

Choose $\delta = \frac{1}{3}\varepsilon$. If $0 < \sqrt{x^2+y^2} < \delta = \frac{1}{3}\varepsilon$, then $|y| \leq \sqrt{x^2+y^2} < \frac{1}{3}\varepsilon$. Therefore,

$$|f(x,y) - 0| = \left|\frac{3x^2y}{x^2+y^2}\right| < 3|y| < 3 \cdot \frac{1}{3}\varepsilon = \varepsilon$$

whenever $0 < \sqrt{x^2+y^2} < \delta$ and this implies that $\displaystyle\lim_{(x,y)\to(0,0)} \frac{3x^2y}{x^2+y^2} = 0$. □

■ **Limt at Infinity**

In the previous chapter, we regard R^n as a vector space and every point $(x_1, \cdots, x_n)$ is identified as a vector $\mathbf{x} =< x_1, \cdots, x_n >$. The length of a vector is denoted by

$$\|\mathbf{x}\| = \sqrt{x_1^2 + \cdots + x_n^2}.$$

Hence, if we want to describe a point (or a vector) $\mathbf{x} \in \mathbb{R}^n$ tending to infinity, we will use the notation "$\|\mathbf{x}\| \to \infty$" (or $\|(x_1, \cdots, x_n)\| \to \infty$ or $\| < x_1, \cdots, x_n > \| \to \infty$)

Remark. We usually use the words "*as* $\|\mathbf{x}\|$ *is sufficiently large*" which means that there exists a positive number M such that for every point $\mathbf{x}$ with $\|\mathbf{x}\| > M$ then $\cdots$. For example, "$f(x, y) > 1$ when $\|(x, y)\|$ is sufficiently large" means that there exists a number $M > 0$ such that $f(x, y) > 1$ for every $\|(x, y)\| > M$.

Definition 14.2.9. (Limit at infinity) Let f be a function of two variables whose domain D containing all points which are sufficiently large. We say that the limit of $f(x, y)$, as (x, y) approaches infinity, exists if there is a number L such that for every number $\varepsilon > 0$ there exists a corresponding number $M > 0$ such that

$$|f(x,y) - L| < \varepsilon$$

whenever $\sqrt{x^2+y^2} > M$. Denote

$$\lim_{\|(x,y)\|\to\infty} f(x,y) = L \qquad \text{or} \qquad f(x,y) \to L \quad \text{as } \|(x,y)\| \to \infty.$$

Example 14.2.10. Let $f(x, y) = x$. Determine whether the limit $\displaystyle\lim_{\|(x,y)\|\to\infty} f(x, y)$ exists.

Proof. Fix $x = 1$ and let $y \to \infty$, then $\|(x, y)\| \to \infty$ and $\displaystyle\lim_{x=1,y\to\infty} f(x, y) = 1$.

Similarly, fix $x = 2$ and let $y \to \infty$, then $\|(x, y)\| \to \infty$ and $\displaystyle\lim_{x=2,y\to\infty} f(x, y) = 2$. Hence, the limit $\displaystyle\lim_{\|(x,y)\|\to\infty} f(x, y)$ does not exist. □

Example 14.2.11. Let $f(x, y) = \dfrac{1}{x^2 + y^2}$. Determine whether the limit $\lim\limits_{\|(x,y)\|\to\infty} f(x, y)$ exists.

Proof. Given $\varepsilon > 0$, choose $M = \dfrac{1}{\sqrt{\varepsilon}}$ and $L = 0$. For $\|(x, y)\| = \sqrt{x^2 + y^2} > M$,

$$|f(x, y) - L| = \left|\frac{1}{x^2 + y^2} - 0\right| < \frac{1}{M^2} = \varepsilon.$$

Hence, $\lim\limits_{\|(x,y)\|\to\infty} f(x, y) = 0.$ □

❑ Continuity

Recall that the continuity of a single variable function $f(x)$ at a is defined by

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

A slogan is that "the limit of f at a is equal to the value of f at a". We attempt to use the same idea to define the continuity of a multi-variables function.

Definition 14.2.12.

(a) A two variables function f is called "*continuous at* (a, b)" if

$$\lim_{(x,y)\to(a,b)} f(x, y) = f(a, b).$$

(b) f is called continuous on D if f is continuous at every point in D.

Remark.

(i) A surface that is the graph of a continuous function has no hole or break.

(ii) The sums, differeneces, products and quotients of continuous functions are continuous on their domains

(iii) Every polynomial function or every rational function of two variables is continuous. For example, $f(x, y) = 3x^5 + 6y^4 + 10x^7y^6 + 5x - 7y + 6$ is continuous everywhere.

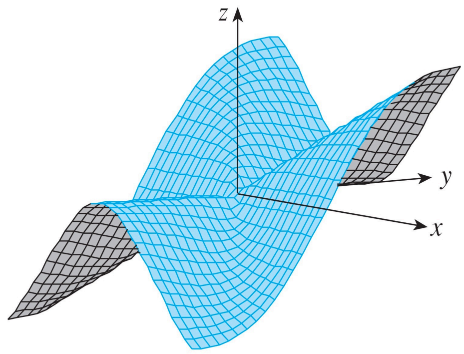
Example 14.2.13. Where is the function $f(x, y) = \dfrac{x^2 - y^2}{x^2 + y^2}$ continuous?

Proof. Since f is a rational function, it is continuous on its domain. That is, f is continuous on $Dom(f) = \big\{(x, y) \mid x^2 + y^2 \neq 0\big\} = \big\{(x, y) \mid (x, y) \neq (0, 0)\big\} = \mathbb{R}\backslash\big\{(0, 0)\big\}$. □

Example 14.2.14. Let $g(x, y) = \begin{cases} \dfrac{x^2 - y^2}{x^2 + y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0). \end{cases}$. Since the limit $\lim\limits_{(x,y)\to(0,0)} g(x, y)$ does not exist, g is not continuous at $(0, 0)$.

Example 14.2.15. Let

$$f(x,y) = \begin{cases} \dfrac{3x^2y}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases}$$

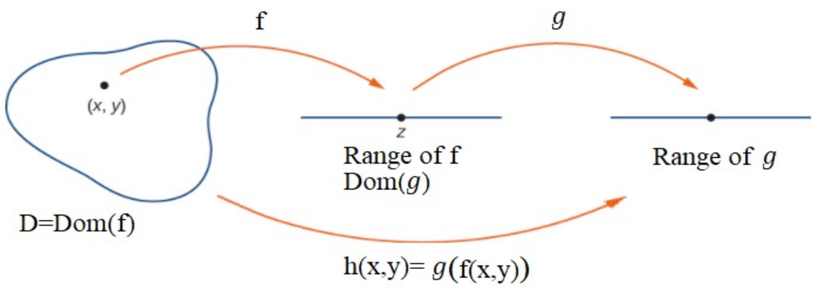


Since f is a rational function for $(x,y) \neq (0,0)$, it is continuous on $\mathbb{R}^2\backslash\{(0,0)\}$. Also, $\displaystyle\lim_{(x,y)\to(0,0)} \frac{3x^2y}{x^2+y^2} = 0 = f(0,0)$. Thus, f is continuous at $(0,0)$ and f is continuous on $\mathbb{R}^2$.

■ Composite Functions

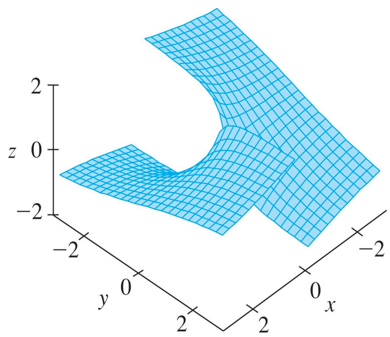
We consider the composition of a two variables function and a single variable function.

Let $f(x,y)$ be a continuous function of two variables and $g(t)$ be a continuous function of a single variable that define on the range of f. Then $h = g \circ f$ defined by $h(x,y) = g\big(f(x,y)\big)$ is also a continuous function.



Example 14.2.16. Where is the function $h(x,y) = \arctan\left(\dfrac{y}{x}\right)$ continuous?

Proof. Let $f(x,y) = \dfrac{y}{x}$ be continuous except on the line $x = 0$. Let $g(t) = \arctan t$ be continuous everywhere. Then the composite function $h(x,y) = \arctan\left(\dfrac{y}{x}\right) = g\big(f(x,y)\big)$ is continuous except the line $x = 0$. □



The function $h(x, y) = \arctan(y/x)$ is discontinuous where $x = 0$.

■ Functions of Three or more Variables

The definitions of limits and continuity of n-variables functions are similar as the ones of two variables functions. We ignore the details of their definitions here.

14.3 Partial Derivatives

Recall that for a single variable function $f(x)$, the derivative of f is defined by

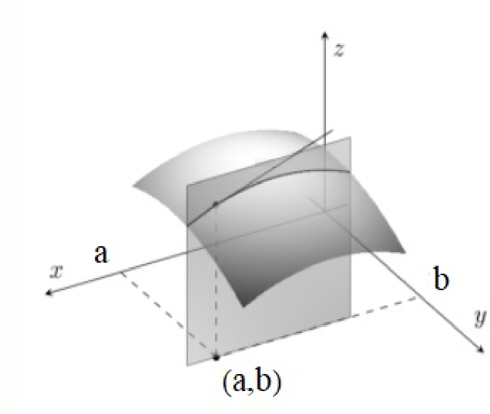
$$f'(a) = \lim_{h\to 0} \frac{f(a+h) - f(a)}{h}$$

which represents the instantaneous rate of change of f with respect to x.

For a two variables function $f(x, y)$, let x vary while keeping y fixed, say $y = b$, where b is a constant. We can regard $f(x, b)$ as a single variable function.
Let $g(x) = f(x, b)$, then $g(a) = f(a, b)$. The derivative of $g(x)$ at $x = a$ is

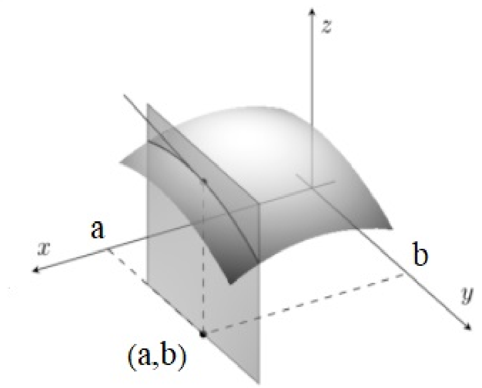
$$g'(a) = \lim_{h\to 0} \frac{g(a+h) - g(a)}{h} = \lim_{h\to 0} \frac{f(a+h, b) - f(a, b)}{h}..$$

We call it the "*partial derivative of f with respect to x at* (a, b)".



Similarly, let y vary while keeping x fixed, say $x = a$. Let $k(y) = f(a, y)$. The partial derivative of f with respect to y at (a, b) is

$$\lim_{h\to 0} \frac{k(b+h) - k(b)}{h} = \lim_{h\to 0} \frac{f(a, b+h) - f(a, b)}{h}$$



Definition 14.3.1. (Partial Derivatives) Let f be a function of two variables. The partial derivatives of f with respect to x and with respect to y are the functions f_x and f_y defined by setting

$$f_x(x, y) = \lim_{h\to 0} \frac{f(x+h, y) - f(x, y)}{h}$$

$$f_y(x, y) = \lim_{h\to 0} \frac{f(x, y+h) - f(x, y)}{h}$$

provided these limits exist.

Notation: Let $z = f(x, y)$. We write

$$\begin{aligned} f_x(x, y) &= f_x = \frac{\partial f}{\partial x} = \frac{\partial}{\partial x} f(x, y) = \frac{\partial z}{\partial x} = D_x f = D_1 f = f_1 \\ f_y(x, y) &= f_x = \frac{\partial f}{\partial y} = \frac{\partial}{\partial y} f(x, y) = \frac{\partial z}{\partial y} = D_y f = D_2 f = f_2 \end{aligned}$$

■ **Find Partial Derivatives of** $z = f(x, y)$

- To find f_x, we regard y as a constant and differentiate $f(x, y)$ with respect to x.
- To find f_y, we regard x as a constant and differentiate $f(x, y)$ with respect to xy.

Example 14.3.2. If $f(x, y) = x^3 + x^2y^3 - 2y^2$, find $f_x(2, 1)$ and $f_y(2, 1)$.

Proof. The partial derivatives of f are

$$f_{(}x, y) = 3x^2 + 2xy^3 \qquad \text{and} \qquad f_y(x, y) = 3x^2y^2 - 4y.$$

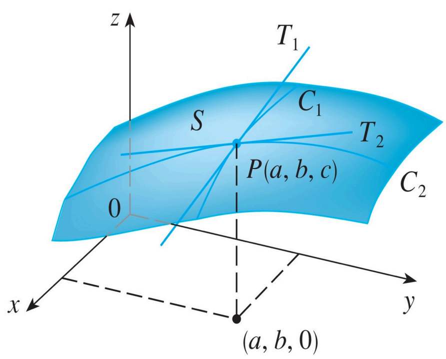
Then $f_x(2, 1) = 12 + 4 = 16$ and $f_y(2, 1) = 12 - 4 = 8$. □

Note. We should consider the single variable function $f(x, 1) = x^3 + x^2 - 4$ and $f(2, y) = 8 + 4y^3 - 2y^2$. Then

$$\begin{aligned} f_x(2, 1) &= \left(\frac{d}{dx} f(x, 1)\right)\Big|_{x=2} = 3x^2 + 2x\Big|_{x=2} = 12 + 4 = 16. \\ f_y(2, 1) &= \left(\frac{d}{dy} f(2, 1)\right)\Big|_{y=1} = 12y^2 - 4y\Big|_{y=1} = 12 - 4 = 8. \end{aligned}$$

■ Interpretation of Partial Derivatives

The equation $z = f(x, y)$ represents a surface S (the graph of f). If $f(a, b) = c$, then the point $P(a, b, c)$ lies on S.
Fix $y = b$, the curve C_1 is the intersection of the vertical plane and S. C_1 is also the graph of the function $g(x) = f(x, b)$, $y = b$. The slope of its tangent line T_1 at P is $g'(a) = f_x(a, b)$. Similar for the curve C_2, the tangnet line T_2 and its slope $f_y(a, b)$.



The partial derivatives of f at (a, b) are the slopes of the tangents to C_1 and C_2.

Example 14.3.3. If $f(x, y) = 4 - x^2 - 2y^2$, find $f_x(1, 1)$ and $f_y(1, 1)$ and interpret these numbers as slopes.

Proof. The partial derivatives of f are

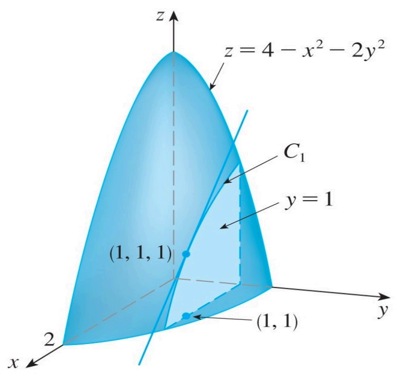
$$f_x(x, y) = -2x \qquad \text{and} \qquad f_y(x, y) - 4y.$$

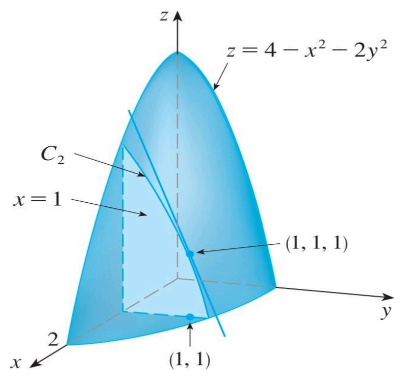
Then $f_x(1, 1) = -2$ and $f_y(1, 1) = -4$.

The equation $z = 4 - x^2 - 2y^2$ represents a paraboloid which is the graph of $f(x, y)$. Fix $y = 1$, $z = 2 - x^2$ is the equation of a parabola which is the intersection of the vertical plane $y = 1$ and the graph of $f(x, y)$. The value $f_x(1, 1) = -2$ is the slope of the tangent line to the parabola $C_1 : z = 2 - x^2$, $y = 1$ at $(1, 1, 1)$.

Similarly, $f_y(1, 1) = -4$ is the slope of the tangnet line to the parabola $C_2 : z = 3 - 2y^2$, $x = 1$ at $(1, 1, 1)$.

□





Note. We can express the curve C_1 as a vector equation $\mathbf{r}(t) =< t, 1, 2 - t^2 >$. Then the tangent vector is $\mathbf{r}'(t) =< 1, 0, -2t >$.

At $(1, 1, 1)$, we have $t = 1$ and then $\mathbf{r}'(1) =< 1, 0, -2 >$. The equation of the tangent line is

$$\mathbf{r}(1) + t\mathbf{r}(1) =< 1 + t, 1, 1 - 2t > .$$

Example 14.3.4. If $f(x, y) = \sin\left(\dfrac{x}{1+y}\right)$, calculate $\dfrac{\partial f}{\partial x}$ and $\dfrac{\partial f}{\partial y}$.

Proof. We can calculate the partial derivatives by the chain rule,

$$\frac{\partial f}{\partial x} = \cos\left(\frac{x}{1+y}\right) \cdot \frac{1}{1+y} \qquad \text{and} \qquad \frac{\partial f}{\partial y} = \cos\left(\frac{x}{1+y}\right) \cdot \frac{-x}{(1+y)^2}.$$

□

■ Implicit Differentiation

Recall that if the two variables x and y satisfy an equation $F(x, y) = 0$, then we can use the implicit differentiation to find the ralated rate of each other ($\dfrac{dy}{dx}$ or $\dfrac{dx}{dy}$).

By following the same idea, if three variables x, y and z satisfy an equation $F(x, y, z) = 0$, we want to find the related rates (partial derivatives) between any two variables.

Example 14.3.5. Find $\dfrac{\partial z}{\partial x}$ and $\dfrac{\partial z}{\partial y}$ if z is defined implicitly as a function of x and y by the equation

$$x^3 + y^3 + z^3 + 6xyz = 1. \tag{14.1}$$

Proof. Differentiating both sides of equation (14.1) with respect to x, we have

$$\frac{\partial}{\partial x}\left[x^3 + y^3 + z^3 + 6xyz\right] = \frac{\partial}{\partial x}(1)$$

Then

$$3x^2 + 3z^2\frac{\partial z}{\partial x} + 6yz + 6xy\frac{\partial z}{\partial x} = 0 \quad \text{and hence}$$
$$\frac{\partial z}{\partial x}\left(3z^3 + 6xy\right) = -\left(3x^2 + 6yz\right).$$

We have

$$\frac{\partial z}{\partial x} = -\frac{x^2 + 2yz}{z^2 + 2xy}.$$

Similarly,

$$\frac{\partial z}{\partial y} = -\frac{y^2 + 2xz}{z^2 + 2xy}.$$

□

■ Functions of Three or More Variables

- For a three variables function $f(x, y, z)$, fix y and z, the partial derivative of f with respect to x is defined by

$$f_x(x, y, z) = \lim_{h \to 0} \frac{f(x + h, y, z) - f(x, y, z)}{h}.$$

  (f_y and f_z have similar definition).

  If $w = f(x, y, z)$, then $\dfrac{\partial w}{\partial x}$ can be interpreted as the rate of change of w with respect to x when y and z are fixed.

- for a n-variables function $f(x_1, x_2, \cdots, x_n)$,

$$f_{x_i}(x_1, x_2, \cdots, x_n) = \lim_{h \to 0} \frac{f(x_1, \cdots, x_i + h, \cdots, x_n) - f(x_1, \cdots, x_i, \cdots x_n)}{h}.$$

  If $u = f(x_1, x_2, \cdots, x_n)$, then $\dfrac{\partial u}{\partial x_i} = \dfrac{\partial f}{\partial x_i} = f_{x_i} = f_i = D_i f$ is the partial deriveative of u with respect to x_i.

Note. Denote $\mathbf{x} = (x_1, \cdots, x_n)$ and $\mathbf{e}_i = (0, \cdots, 0, 1, 0, \cdots 0)$. Then

$$f_{x_i}(\mathbf{x}) = \lim_{h \to 0} \frac{f(\mathbf{x} + h\mathbf{e}_i) - f(\mathbf{x})}{h}.$$

Example 14.3.6. Let $f(x, y, z) = e^{xy} \ln z$, then

$$f_x(x, y, z) = e^{xy} \cdot y \ln z = ye^{xy} \ln z, \quad f_y(x, y, z) = xe^{xy} \ln z, \quad f_z(x, y, z) = e^{xy} \cdot \frac{1}{z}.$$

■ Higher Derivatives

When study a single variable function $f(x)$, we can regard its derivative $f'(x)$ as a new function and consider its second derivative $f''(x)$.

For a two variables function $f(x, y)$, we can also regard its partial derivatives $f_x(x, y)$ and $f_y(x, y)$ as new functions and consider the "*second partial derivatives*". Let $z = f(x, y)$. Then

$$\begin{aligned}
(f_x)_x &= f_{xx} = \frac{\partial}{\partial x}\Big(\frac{\partial f}{\partial x}\Big) = \frac{\partial^2 f}{\partial x^2} = \frac{\partial^2 z}{\partial x^2} = f_{11} \\
(f_x)_y &= f_{xy} = \frac{\partial}{\partial y}\Big(\frac{\partial f}{\partial x}\Big) = \frac{\partial^2 f}{\partial y \partial x} = \frac{\partial^2 z}{\partial y \partial x} = f_{12} \\
(f_y)_x &= f_{yx} = \frac{\partial}{\partial x}\Big(\frac{\partial f}{\partial y}\Big) = \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 z}{\partial x \partial y} = f_{21} \\
(f_y)_y &= f_{yy} = \frac{\partial}{\partial y}\Big(\frac{\partial f}{\partial y}\Big) = \frac{\partial^2 f}{\partial y^2} = \frac{\partial^2 z}{\partial y^2} = f_{22}
\end{aligned}$$

• **third partial derivative**

$$
\begin{aligned}
(f_{xy})_x &= f_{xyx} = \frac{\partial}{\partial x}\Big(\frac{\partial^2 f}{\partial y \partial x}\Big) = \frac{\partial^3 f}{\partial x \partial y \partial x} = \frac{\partial^3 z}{\partial x \partial y \partial x} \\
(f_{xy})_y &= f_{xyy} = \frac{\partial}{\partial y}\Big(\frac{\partial^2 f}{\partial y \partial x}\Big) = \frac{\partial^3 f}{\partial^2 y \partial x} = \frac{\partial^3 z}{\partial^2 y \partial x}
\end{aligned}
$$

Example 14.3.7. Let $f(x, y) = x^3 + x^2y^3 - 2y^2$. Then the first partial derivatives of f are

$$f_x = 3x^2 + 2xy^3, \qquad f_y = 3x^2y^2 - 4y$$

and the second partial derivatives of f are

$$f_{xx} = 6x + 2y^3, \quad f_{xy} = 6xy^2, \quad f_{yx} = 6xy^2, \quad f_{yy} = 6x^2y - 4.$$

■ **Clairaut's Theorem**

Question: For a multi-variables function, does the second partial derivatives keep unchanged when the order of two partial differentiations exchange? For example, if $f(x, y)$ has all second partial derivatives, can we obtain

$$f_{xy} \stackrel{??}{=} f_{yx}.$$

In general, the answer is false.

Exercise. Let

$$f(x, y) = \begin{cases} \dfrac{xy(x^2 - y^2)}{x^2 + y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0) \end{cases}$$

Check that $f_{xy}(0, 0) \neq f_{yx}(0, 0)$.

Question: What conditions of f can guarantee its second partial derivatives are equal when exchanging their order?

Theorem 14.3.8. *(Clairaut's Theorem) Suppose f is defined on a neighborhood D of (a, b). If the functions f_{xy} and f_{yx} are both continuous at (a, b), then*

$$f_{xy}(a, b) = f_{yx}(a, b).$$

Proof. Consider

$$
\begin{aligned}
f_{xy}(a, b) &= \lim_{k\to 0} \frac{f_x(a, b + k) - f_x(a, b)}{k} \\
&= \lim_{k\to 0} \frac{\lim_{h\to 0}\left[\frac{f(a+h,b+k)-f(a,b+k)}{h} - \frac{f(a+h,b)-f(a,b)}{h}\right]}{k} \\
&= \lim_{k\to 0}\lim_{h\to 0} \frac{f(a + h, b + k) - f(a + h, b) - f(a, b + k) + f(a, b)}{kh}.
\end{aligned}
$$

Define $g(y) = f(a + h, y) - f(a, y)$ Then $f_{xy}(a, b) = \lim\limits_{k\to 0}\lim\limits_{h\to 0} \dfrac{g(b + k) - g(b)}{kh}$.

Since f_y is defined on a neighborhood of (a,b), g is differentiable near b and, by the mean value theorem, $g(b+k)-g(b)=kg'(\xi)$ for some $\xi=\xi(k)\in(0,k)$. Then

$$f_{xy}(a,b)=\lim_{k\to 0}\lim_{h\to 0}\frac{g'\big(\xi(k)\big)}{h}=\lim_{k\to 0}\lim_{h\to 0}\frac{1}{h}\Big[f_y\big(a+h,b+\xi(k)\big)-f_y\big(a,b+\xi(k)\big)\Big].$$

Since f_y is differentiable with respect to x and by the mean value theorem again,

$$f_{xy}(a,b)=\lim_{k\to 0}\lim_{h\to 0}f_{yx}\big(a+\eta(h),b+\xi(k)\big)$$

where $\eta(h)\in(0,h)$ and $\xi(k)\in(0,k)$ and hence $\lim_{h\to 0}\eta(h)=0$ and $\lim_{k\to 0}\xi(k)=0$. Also, the continuity of f_{yx} at (a,b) implies that

$$f_{xy}(a,b)=\lim_{k\to 0}\lim_{h\to 0}f_{yx}\big(a+\eta(h),b+\xi(k)\big)=f_{yx}(a,b).$$

□

Remark. The Clairaut's Theorem still holds if the hypothesis is weaken that one of f_{xy} and f_{yx} is continuous at (a,b).

Example 14.3.9. Let $f(x,y)=\sin(3x+yz)$. Then

$$f_x=3\cos(3x+yz),\quad f_{xx}=-9\sin(3x+yz),\quad f_{xy}=-3z\sin(3x+yz)$$

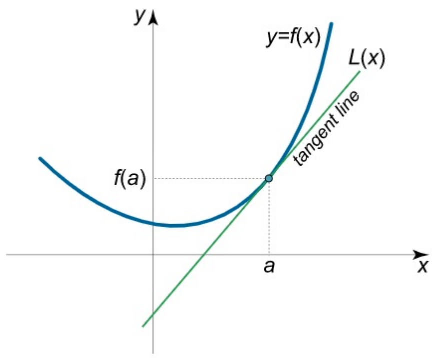
$$f_{xxy}=-9z\cos(3x+yz),\quad f_{xyx}=-9z\cos(3x+yz)=f_{xxy}.$$

14.4 Tangent Planes and Linear Approximations

❑ Tangent Planes

Recall that a single variable function $f(x)$ with derivative $f'(a)$ can be linearly approximated by its "tangent line" *

$$f(x)\approx L(x)=f(a)+f'(a)(x-a)\qquad\text{as } x \text{ is near } a$$



For a two variables function $f(x,y)$, we also expect that it can be linearly approximated by a certain "plane".

Suppose that

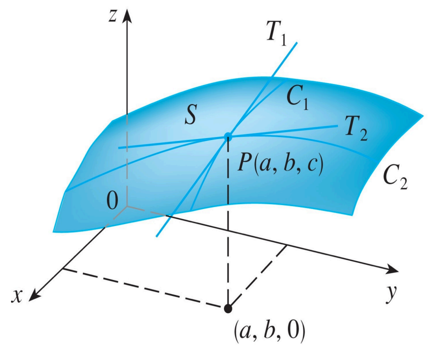
*The figure is download from https://www.math24.net/linear-approximation/

$f(x, y)$ is a two variables function which has continuous first partial derivatives;

S is the surface with equation $z = f(x, y)$ (the graph of f) and $P(a, b, c) \in S$;

C_1 and C_2 are the curves obtained by intersecting the vertical planes $y = b$ and $x = a$ with the sufrace S. Then $P \in C_1 \cap C_2$.

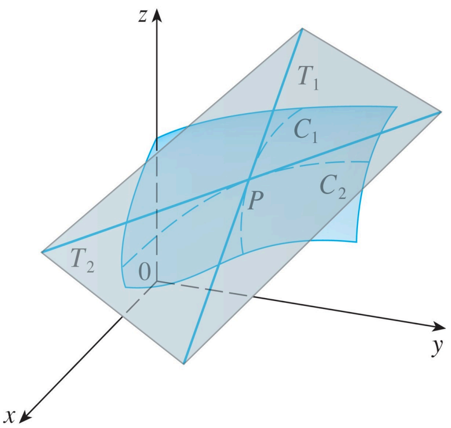
T_1 and T_2 are tangent lines to the curves C_1 and C_2 at the point P.



The partial derivatives of f at (a, b) are the slopes of the tangents to C_1 and C_2.

Definition 14.4.1. The "*tangent plane*" to the surface S at P is defined to be the plane that contains both tangent lines T_1 and T_2.

Note. If C is any curve that lies on S and passes P, then the tangent line to C at P also lies on the tangent plane. Hence, we can think of the tangent plane to S at P as consisting of all possible tangent lines at P to curves that lie on S and pass through P.



The tangent plane contains the tangent lines T_1 and T_2.

■ Equation of the tangent plane

Let the tangent plane to S passing throught $P(a, b, c)$ has equation

$$A(x - a) + B(y - b) + C(z - c) = 0 \tag{14.2}$$

We may assume that it is not a vertical tangent plane and hence $C \neq 0$. Dividing both sides of equation (14.3) by $-C$, the tangent plane has an equivalent equation

$$z - c = \alpha(x - a) + \beta(y - b) \qquad (\alpha = \frac{A}{-C} \text{ and } \beta = \frac{B}{-C}).$$

Since the intersection of the tangent plane and the vertical plane $y = b$ is the tangent line T_1, plugging $y = b$ into equation (14.3),

$$z - c = \alpha(x - a)$$

is the equation of the tangent line T_1. Then α is the slope of T_1 to the curve C_1 at (a, b, c) and hence $\alpha = f_x(a, b)$.

Similarly, $\beta = f_y(a, b)$. Therefore, the equation of the tangent plane to S at P is

$$z - c = f_x(a, b)\big(x - a\big) + f_y(a, b)\big(y - b\big).$$

Example 14.4.2. Find the tangent plane to the elliptic paraboloid $z = 2x^2 + y^2$ at $(1, 1, 3)$.

Proof. Let $f(x,y) = 2x^2 + y^2$. Then $f_x(x,y) = 4x$ and $f_y(x,y) = 2y$. Hence, $f_x(1,1) = 4$ and $f_y(1,1) = 2$. The equation of the tangent plane at $(1,1,3)$ is

$$z - 3 = 4(x-1) + 2(y-1) \qquad \text{or} \qquad z = 4x + 2y - 3.$$

□

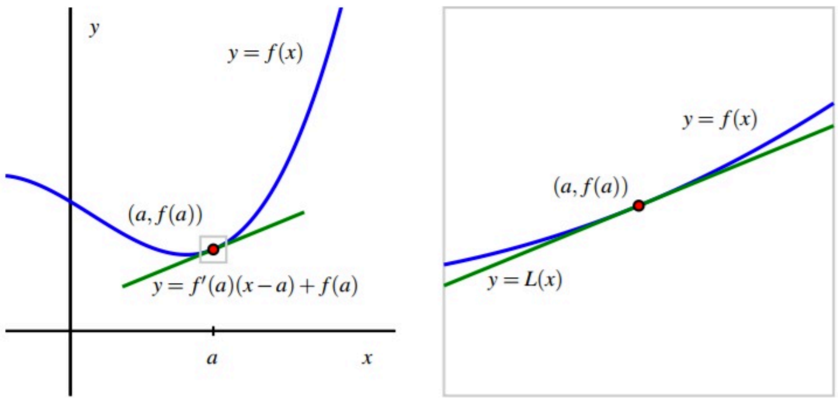
❑ Linear Approximations

We have studied the linear apporximation for a single variable function $f(x)$. We use the tangent line to the graph $y = f(x)$ at a to approxinate the value of f near a and the linearization for f at a is

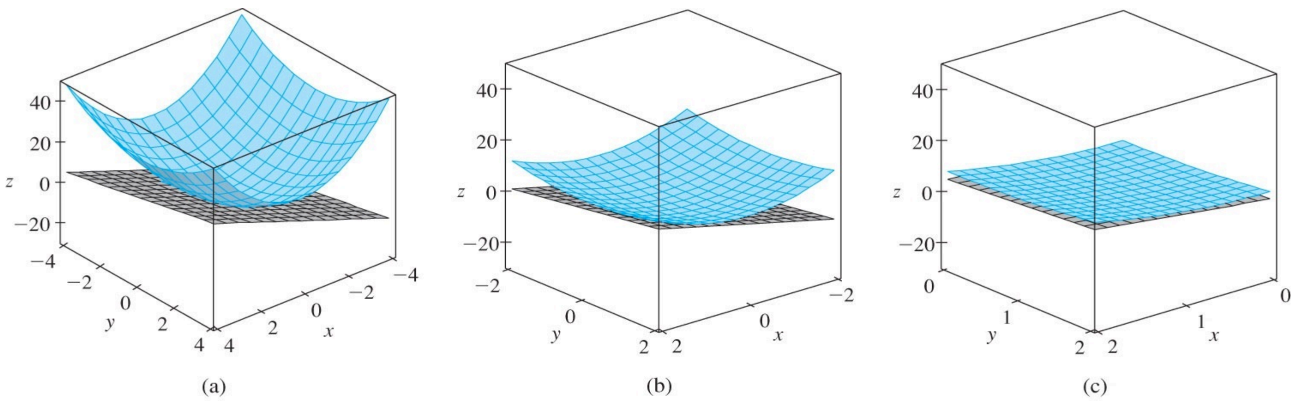
$$L(x) = f(a) + f'(a)(x-a)$$

and

$$f(x) \approx L(x) \qquad \text{as } x \text{ is close to } a.$$



For a two variable function $f(x,y)$, we expect to approximate its values, as (x,y) is near (a,b), by the tangnet plane at (a,b).



The elliptic paraboloid $z = 2x^2 + y^2$ appears to coincide with its tangent plane as we zoom in toward $(1, 1, 3)$.

Suppose that $f(x,y)$ has continuous partial derivative. The tangnet plane to the surface $S : z = f(x,y)$ at $P\big(a, b, f(a,b)\big)$ is

$$z - f(a,b) = f_x(a,b)(x-a) + f_y(a,b)(y-b)$$

or

$$z = f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b).$$

Definition 14.4.3.

(a) We call the function

$$L(x,y) = f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b)$$

the "*linearization of f at* (a,b).

(b) The approximation $f(x,y) \approx L(x,y)$ is called the "*linear approximation*" or "tangent plane approximation" of f at (a,b).

Example 14.4.4. Find the linearization of $f(x,y) = 2x^2+y^2$ at $(1,1,3)$ and use it to approximate the value of $f(1.1, 0.95)$.

Proof. Compute $f_x(x,y) = 4x$ and $f_y(x,y) = 2y$ and hence $f_x(1,1) =$ and $f_y(1,1) = 2$. Then the linearization of f at $(1,1,3)$ is

$$L(x,y) = f(1,1) + f_x(1,1)(x-1) + f_y(1,1)(y-1) = 3 + 4(x-1) + 2(y-1) = 4x + 2y - 3.$$

Also,

$$f(1.1, 0.95) \approx L(1.1, 0.95) = 3 + 4 \cdot 0.1 + 2 \cdot (-0.05) = 3.3.$$

□

We define tangent plane for surface $z = f(x,y)$, where f has continuous partial derivatives. **Question:** What happens if f_x and f_y are not continuous? Consider the following example.

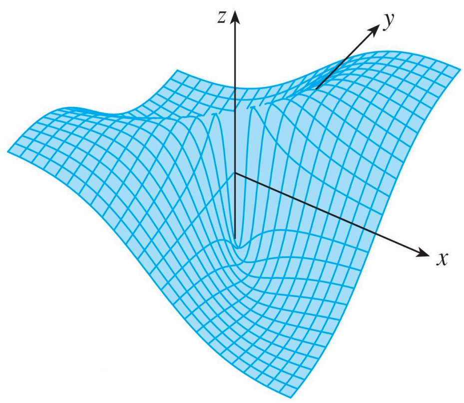
Example 14.4.5.

Let $f(x,y) = \begin{cases} \dfrac{xy}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases}$.

Then $f_x(0,0) = 0 = f_y(0,0)$. For $(x,y) \neq (0,0)$, $f_x(x,y) = \dfrac{y(y^2-x^2)}{(x^2+y^2)^2}$. Along $x = 0$,

$$\lim_{(x,y)\to(0,0),\ x=0} f_x(x,y) = \lim_{y\to 0} \frac{y^3}{y^4} = \infty.$$

Hence, f_x is continuous at $(0,0)$. Also, we can compute that f_y is not continuous at $(0,0)$. Observe that, for (x,y) on the line $x = y$, $f(x,y) = \dfrac{1}{2} \neq 0$. Therefore, f is not continuous at $(0,0)$. This implies that there is linear approximation of f at $(0,0)$.



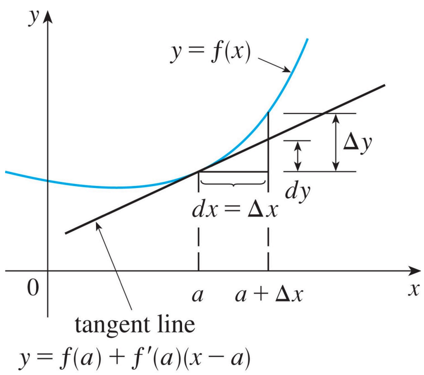
$f(x, y) = \dfrac{xy}{x^2+y^2}$ if $(x, y) \neq (0, 0)$,
$f(0, 0) = 0$

Note. This example says that for the linear approximation, the condition of the continuities of f_x and f_y are necessary.

❑ Differentials

Recall that for a differentiable single variable function $y = f(x)$, dx is the differential of x and $dy = f'(x)\,dx$ is a differential of y.
The symbol $\triangle y$ denotes the change in height of y and dy represents the change in height of the tangent line when x changes $\triangle x = dx$. Hence, as (x, y) is near (a, b),

$$f(x, y) \approx f(a, b) + f'(a, b)\,dx = f(a, b) + dy.$$



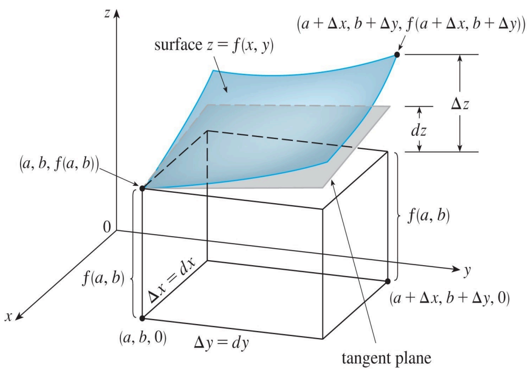
For a differentiable fucntion of two variables $z = f(x, y)$, dx and dy are differentials of x and y respectively, and dz is the differenital of z which is called the "*total differential*". Then

$$dz = f_x(x, y)\,dx + f_y(x, y)\,dy = \frac{\partial z}{\partial x}\,dx + \frac{\partial z}{\partial y}\,dy.$$

Taking $dx = \triangle x = x - a$ and $dy = \triangle y = y - b$, then

$$dz = f_x(x, y)(x - a) + f_y(x, y)(y - b).$$

As (x, y) is near (a, b),



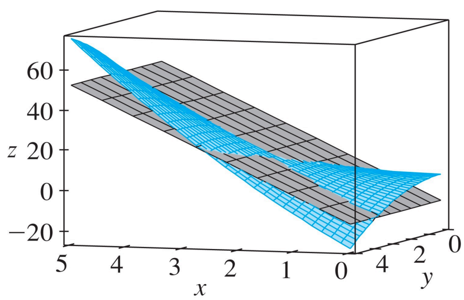
$$f(x, y) \approx f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b) = f(a, b) + dz.$$

Example 14.4.6.
(a) If $z = f(x, y) = x^2 + 3xy - y^2$, find the differential dz.

(b) If x changes from 2 to 2.05 and y changes from 3 to 2.96, compare the values of $\triangle z$ and dz.

Proof.
(a) To find dz, $f_x(x, y) = 2x + 3y$ and $f_y(x, y) = 3x - 2y$. Then

$$dz = \frac{\partial z}{\partial x}dx + \frac{\partial z}{\partial y}dy = (2x + 3y)dx + (3x - 2y)dy.$$



(b) If x changes from x to 2.05 and y changes from 3 to 2.96, compare $\triangle z$ and dz.

$$\begin{aligned}\triangle z &= \; = f(2.05, 2.96) - f(2, 3) = 0.6449\\ dz &= f_x(2, 3)(2.05 - 2) + f_y(2, 3)(2.96 - 3) = 0.65.\end{aligned}$$

□

Example 14.4.7. A cone has the raidus of its base 10 cm and the height 25 cm as the figure.

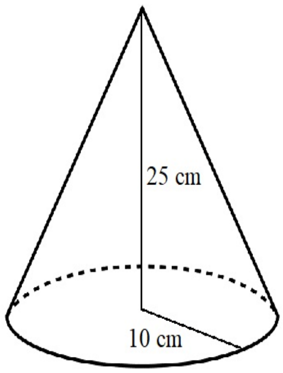
Find a possible error as much as 0.1 cm in radius and height. Use differentials to estimate the maximum error in the caculated volume of the cone.

Proof.

The volume of the cone is $V(r,h) = \frac{1}{3}\pi r^2 h$. Then

$$\frac{\partial V}{\partial r} = \frac{2}{3}\pi rh, \quad \frac{\partial V}{\partial h} = \frac{1}{3}\pi r^2.$$

The differential of V with $dr = 0.1$ and $dh = 0.1$ is



$$\begin{aligned} dV &= \frac{\partial V}{\partial r}(10,25)dr + \frac{\partial V}{\partial h}(10,25)dh \\ &= \frac{500\pi}{3} \cdot 0.1 + \frac{100\pi}{3} \cdot 0.1 = 20\pi \quad (cm^3) \end{aligned}$$

□

❑ Functions of Three or More Variables

■ Linear Approximation

The linearization of f at (a,b,c) is

$$f(x,y,z) \approx L(x,y,z) = f(a,b,c) + f_x(a,b,c)(x-a) + f_y(a,b,c)(y-b) + f_z(a,b,c)(z-c).$$

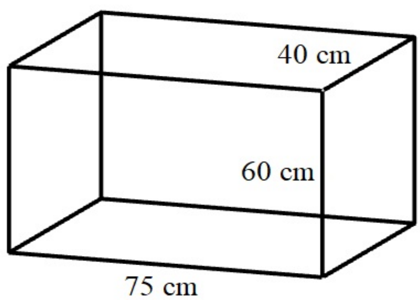
■ Differentials

Let $w = f(x,y,z)$. Then

$$\begin{aligned} \triangle w &= f(x+\triangle x, y+\triangle y, z+\triangle z) - f(x,y,z) \\ dw &= f_x(x,y,z)dx + f_y(x,y,z)dy + f_z(x,y,z)dz = \frac{\partial w}{dx}dx + \frac{\partial w}{\partial y}dy + \frac{\partial w}{\partial z}dz. \end{aligned}$$

Example 14.4.8. A rectangular box has length, width, and height 75cm, 40 cm and 60cm respectively. Use differentials to estimate the largest possible error when the volume of the box is calculatedas each measurement is correct ot within 0.2 cm.

Proof.



Let x, y and z denote the length, width and height of the box. The volume of the box is $V(x,y,z) = xyz$. Then

$$\frac{\partial V}{\partial x} = xy, \ \frac{\partial V}{\partial y} = xz, \ \frac{\partial V}{\partial z} = xy.$$

The differential in V at $(75,40,60)$ with $dx = dy = dz = 0.2$ is

$$\begin{aligned} dV &= \frac{\partial V}{\partial x}dx + \frac{\partial V}{\partial y}dy + \frac{\partial V}{\partial z}dz \\ &= 60 \cdot 40 \cdot 0.2 + 75 \cdot 60 \cdot 0.2 + 75 \cdot 40 \cdot 0.2 = 1980 \ (cm^3). \end{aligned}$$

□

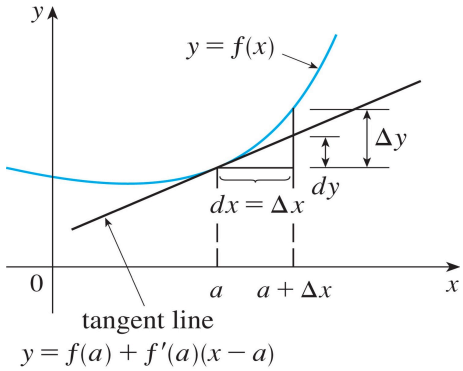
14.5 Differentiability and Gradient

For a two variables function $f(x, y)$, it may have all partial derivatives (f_x and f_y) at (a, b) but f is not continuous there. Hence, f has no linear approximation at (a, b) and it is not "smooth" near (a, b). To understand the differentiability of two variables (n variables) functions, let's observe linear approximation of one variable function and try to give a suitable definition.

We recall the geometric meaning of linear approximation of $y = f(x)$. Let $\triangle y = f(a + \triangle x) - f(a)$. The rate of change of y with respect to x is

$$\frac{\triangle y}{\triangle x} = \frac{f(a + \triangle x) - f(a)}{\triangle x}.$$

If f is differentiable at a, then $\dfrac{\triangle y}{\triangle x} \to f'(a)$ as $\triangle x \to 0$.



Hence,

$$\underbrace{\triangle y}_{\text{increment in } y} = \underbrace{f'(a)\triangle x}_{\text{linear approximation}} + \underbrace{\varepsilon\triangle x}_{\text{error}} \quad \text{where } \varepsilon \to 0 \quad \text{as} \quad \triangle x \to 0.$$

(Note that $\varepsilon = \varepsilon(\triangle x)$ varies as $\triangle x$ varies.)

Formally, we says that a one variable function f is differentiable at x if there exists a number L such that

$$\lim_{h\to 0} \frac{f(x+h) - f(x)}{h} = L \iff \lim_{h\to 0} \frac{f(x+h) - f(x) - Lh}{h} = 0$$
$$\iff \lim_{h\to 0} \frac{|f(x+h) - f(x) - Lh|}{|h|} = 0.$$

Question: How to define the differentiability of two or n variables functions?

If we want to establish an appropriate definition for differentiability, it is natural to expect that the definition should be consistant with the usual derivative when $n = 1$. Also, the definition should reflect the rate of change in any direction.

For the sake to discuss n variables functions conveniently, we will use the following vector symbols to represent the corresponding items in 2, 3 or n dimensional cases.

$\mathbf{x} =< x, y >$	or	$\mathbf{x} =< x, y, z >$	or	$\mathbf{x} =< x_1, \cdots, x_n >$	as variables
$\mathbf{a} =< a_1, a_2 >$	or	$\mathbf{a} =< a_1, a_2, a_3 >$	or	$\mathbf{x} =< a_1, \cdots, a_n >$	as some given point
$\mathbf{h} =< h_1, h_2 >$	or	$\mathbf{h} =< h_1, h_2, h_3 >$	or	$\mathbf{h} =< h_1, \cdots, h_n >$	as small displacement

Then $z = f(\mathbf{x})$ and

$$\triangle z = f(a_1 + h_1, a_2 + h_2) - f(a_1, a_2) = f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}).$$

Inspired by the definition of differentiability of one variable function, we may guess a possible defintion as

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{\triangle z}{\mathbf{h}}.$$

However, the symbol $\frac{\triangle z}{\mathbf{h}}$ is nonsense since the denominator is a vector rather than a scalar. Again, the limit $\lim_{\mathbf{h}\to\mathbf{0}} \frac{\triangle z}{||\mathbf{h}||}$ usually does not exist and it cannot reflect the rate of change of z in the direction $\frac{\mathbf{h}}{||\mathbf{h}||}$. An expectant "derivative" of f at $\mathbf{a}$ is supposed to be an object which sent the direction $\frac{\mathbf{h}}{||\mathbf{h}||}$ to a value. The value will represent the rate of change of z in the direction $\frac{\mathbf{h}}{||\mathbf{h}||}$.

Definition 14.5.1. Let $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$ be a function and $\mathbf{a} \in D$. We say that f is "*differentiable*" at $\mathbf{a}$ if there exists a vector $\mathbf{y} \in \mathbb{R}^2$ such that

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - \mathbf{y}\cdot\mathbf{h}|}{||\mathbf{h}||} = 0$$

The vector $\mathbf{y}$ is denoted by "$\nabla f(\mathbf{a})$" (or "*grad f*") and is called the "*gradient*" of f at $\mathbf{a}$.

Proposition 14.5.2. *The vector* $\mathbf{y}$ [$= \nabla f(\mathbf{a})$ *the gradient of* f *at* $\mathbf{a}$] *in the above definition is unique.*

Proof. If $\mathbf{w}$ is a vector such that

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - \mathbf{w}\cdot\mathbf{h}|}{||\mathbf{h}||} = 0,$$

then

$$\begin{aligned}(*) = \frac{|(\mathbf{y}-\mathbf{w})\cdot\mathbf{h}|}{||\mathbf{h}||} &= \frac{|\mathbf{y}\cdot\mathbf{h} - \mathbf{w}\cdot\mathbf{h}|}{||\mathbf{h}||} \\ &\leq \frac{|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - \mathbf{y}\cdot\mathbf{h}|}{||\mathbf{h}||} + \frac{|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - \mathbf{w}\cdot\mathbf{h}|}{||\mathbf{h}||}.\end{aligned}$$

Choose $\mathbf{h} = \varepsilon(\mathbf{y}-\mathbf{w})$ and let $\varepsilon \to 0$. Then $\mathbf{h} \to \mathbf{0}$ and, by the definition of differentiability, $||\mathbf{y}-\mathbf{w}|| = \frac{\varepsilon||\mathbf{y}-\mathbf{w}||^2}{\varepsilon||\mathbf{y}-\mathbf{w}||} \to 0$. Hence, $||\mathbf{y}-\mathbf{w}|| = 0$ and $\mathbf{y} = \mathbf{w}$. □

Remark.

(i) The gradient, ∇f, of f is a vector-valued function. If $f(\mathbf{x}) : D \subseteq \mathbb{R}^n \to \mathbb{R}$, then the gradient,$\nabla f(\mathbf{x})$, of f is a n component vector-valued function.

(ii) If $f : I \subseteq \mathbb{R} \to \mathbb{R}$ is an one variable function, then $\nabla f(x) = f'(x)$.

(iii) For a fixed $\mathbf{a} \in D$, we can regard $\nabla f(\mathbf{a})$ as an operator which sends every vector $\mathbf{h}$ to a number $\nabla f(\mathbf{a}) \cdot \mathbf{h}$. This number represents the rate of change of f at $\mathbf{a}$ in the direction $\mathbf{h}$.

(iv) $Dom(\nabla f) \subseteq Dom(f)$.

■ **Compute ∇f**

Question: How to compute ∇f?

Example 14.5.3. Let $f(x, y) = x^2 + y^2$. Find $\nabla f(x, y)$.

Proof. Let $\mathbf{h} =< h_1, h_2 >$ and $\mathbf{x} =< x, y >$. Then

$$\begin{aligned} f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) &= [(x + h_1)^2 + (y + h_2)^2] - (x^2 + y^2) \\ &= 2xh_1 + h_1^2 + 2yh_2 + h_2^2 \\ &= < 2x, 2y > \cdot \underbrace{< h_1, h_2 >}_{=\mathbf{h}} + h_1^2 + h_2^2. \end{aligned}$$

Since

$$\begin{aligned} \lim_{<h_1,h_2>\to<0,0>} \frac{|f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - < 2x, 2y > \cdot\mathbf{h}|}{||\mathbf{h}||} &= \lim_{<h_1,h_2>\to<0,0>} \frac{|h_1^2 + h_2^2|}{\sqrt{h_1^2 + h_2^2}} \\ &= \lim_{<h_1,h_2>\to<0,0>} \sqrt{h_1^2 + h_2^2} = 0, \end{aligned}$$

we have $\nabla f(x, y) =< 2x, 2y >$. □

Using the definition to find ∇f is usually complicated. We expect to find a way to compute ∇f more conveniently (at least under certain assumptions).

■ **Sufficient condition for differentiability**

From Example 14.4.5, a two variables function $f(x, y)$ has all partial derivatives at (a, b) cannot guarantee that it is differentiable there. We may need stronger conditions than the existence of all partial derivatives to obtain the differentiability.

Theorem 14.5.4. *Let $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$ be a function. If f has continuous first partial derivatives f_x and f_y at $\mathbf{a}$, then f is differentiable at $\mathbf{a}$ and*

$$\nabla f(\mathbf{a}) = \langle f_x(\mathbf{a}), f_y(\mathbf{a}) \rangle.$$

Proof. Let $\mathbf{a} =< a_1, a_2 >$. For $\mathbf{h} =< h_1, h_2 >$,

$$\begin{aligned} f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) &= f(a_1 + h_1, a_2 + h_2) - f(a_1, a_2) \\ &= \big[f(a_1 + h_1, a_2 + h_2) - f(a_1, a_2 + h_2)\big] + \big[f(a_1, a_2 + h_2) - f(a_1, a_2)\big] \\ &= (I) + (II). \end{aligned}$$

By the Mean Value Theorem, there exists $\theta_1 \in (0, h_1)$ and $\theta_2 \in (0, h_2)$ such that

$$(I) = f_x(a_1 + \theta_1, a_2 + h_2)h_1 \quad \text{and} \quad (II) = f_y(a_1, a_2 + \theta_2)h_2.$$

Note that $\theta_1, \theta_2 \to 0$ as $< h_1, h_2 >\to< 0, 0 >$. Since f_x and f_y are continuous at $< a_1, a_2 >$, as $< h_1, h_2 >\to< 0, 0 >$,

$$f_x(a_1 + \theta_1, a_2 + h_2) - f_x(a_1, a_2) \to 0 \quad \text{and} \quad f_y(a_1, a_2 + \theta_2) - f_y(a_1, a_2) \to 0.$$

Hence,

$$\begin{aligned}
& \lim_{\mathbf{h}\to\mathbf{0}} \frac{|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - < f_x(\mathbf{a}), f_y(\mathbf{a}) > \cdot\mathbf{h}|}{||\mathbf{h}||} \\
= & \lim_{\mathbf{h}\to\mathbf{0}} \frac{|(I) + (II) - f_x(a_1,a_2)h_1 - f_y(a_1,a_2)h_2|}{||\mathbf{h}||} \\
= & \lim_{\mathbf{h}\to\mathbf{0}} \frac{\left|\Big(f_x(a_1+\theta_1, a_2+h_2) - f_x(a_1,a_2)\Big)h_1 + \Big(f_y(a_1, a_2+\theta_2) - f_y(a_1,a_2)\Big)h_2\right|}{\sqrt{h_1^2+h_2^2}} \\
\le & \lim_{\mathbf{h}\to\mathbf{0}} \Big[\Big|f_x(a_1+\theta_1, a_2+h_2) - f_x(a_1,a_2)\Big| + \Big|f_y(a_1, a_2+\theta_2) - f_y(a_1,a_2)\Big| \Big] \\
= & \; 0.
\end{aligned}$$

Therefore, $\nabla f(\mathbf{a}) = \big\langle f_x(\mathbf{a}), f_y(\mathbf{a})\big\rangle$. □

Example 14.5.5. Let $f(x,y) = x^2 + y^2$. Then $f_x(x,y) = 2x$ and $f_y(x,y) = 2y$ are continuous on $\mathbb{R}^2$. Hence, f is differentiable on $\mathbb{R}^2$ and $\nabla f(x,y) = \langle 2x, 2y\rangle$.

Remark. The theorem guarantees that $\nabla f(\mathbf{a}) =< f_x(\mathbf{a}), f_y(\mathbf{a}) >$ if f_x and f_y are continuous at $\mathbf{a}$. Sometimes, the result is still true even if its partial derivatives are not continuous there. For example, $f(x,y) = \begin{cases} x^2\sin\left(\frac{1}{x}\right) & (x,y)\neq(0,0) \\ 0 & (x,y) = (0,0) \end{cases}$. We have seen that f_x and f_y are not continuous at $(0,0)$. On the other hand, $f_x(0,0) = 0 = f_y(0,0)$ and

$$\frac{|f(h_1,h_2) - f(0,0)|}{\sqrt{h_1^2+h_2^2}} = \frac{|h_1^2\sin(\frac{1}{h_1})|}{\sqrt{h_1^2+h_2^2}} \le \frac{|h_1^2\sin(\frac{1}{h_1})|}{|h_1|} \longrightarrow 0$$

as $< h_1, h_2 > \to < 0, 0 >$. Therefore, $\nabla f(0,0) = \langle f_x(0,0), f_y(0,0)\rangle$.

Remark. Let $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ be a function of n variables. Suppose that all first partial derivatives of f are continuous at $\mathbf{a}$. Then f is differentiable at $\mathbf{a}$ and

$$\nabla f(\mathbf{a}) = \langle f_{x_1}(\mathbf{a}), f_{x_2}(\mathbf{a}), \cdots f_{x_n}(\mathbf{a})\rangle.$$

Example 14.5.6. $f(x_1,\cdots,x_n) = \sin(x_1+2x_2+\cdots+nx_n)$ Then $f_{x_k} = k\cos(x_1+2x_2+\cdots+nx_n)$ for $k = 1,2,\cdots,n$. Since $f_{x_k}(x_1,\cdots,x_n)$ is continuous on $\mathbb{R}^n$ for $k = 1,2,\cdots,n$, f is differentiable on $\mathbb{R}^n$ and

$$\nabla f(x_1,\cdots,x_n) = \Big\langle \cos(x_1+\cdots+nx_n), \cdots, n\cos(x_1+\cdots+nx_n) \Big\rangle.$$

Proposition 14.5.7. *Let* $\mathbf{x} =< x_1, x_2, \cdots, x_n >\in\mathbb{R}^n$ *and* $r(\mathbf{x}) = ||\mathbf{x}||$. *Then*

(1) $\nabla r(\mathbf{x}) = \dfrac{\mathbf{x}}{r(\mathbf{x})} = \dfrac{\mathbf{x}}{||\mathbf{x}||}$ *for* $\mathbf{x}\neq\mathbf{0}$.

(2) $\nabla\Big(\dfrac{1}{r(\mathbf{x})}\Big) = -\dfrac{\mathbf{x}}{r^3(\mathbf{x})} = -\dfrac{\mathbf{x}}{||\mathbf{x}||^3}$ *for* $\mathbf{x}\neq\mathbf{0}$.

(3) $\nabla\big(r^m(\mathbf{x})\big) = mr^{m-2}(\mathbf{x})\mathbf{x} = m||\mathbf{x}||^{m-2}\mathbf{x}$ *for* $\mathbf{x}\neq\mathbf{0}$, $m\in\mathbb{N}$.

Proof. (exercise) □

Theorem 14.5.8. *Let $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ be a function. If f is differentiable at $\mathbf{a}$, then it is continuous at $\mathbf{a}$.*

Proof. Since f is differentiable at $\mathbf{a}$, we have

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - \nabla f(\mathbf{a})\cdot\mathbf{h}|}{||\mathbf{h}||} = 0.$$

Then

$$\begin{aligned} 0 \le |f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a})| &= \frac{|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a})|}{||\mathbf{h}||}||\mathbf{h}|| \\ &\le \left[\frac{|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - \nabla f(\mathbf{a})|}{||\mathbf{h}||} + \frac{|\nabla f(\mathbf{a})\cdot\mathbf{h}|}{||\mathbf{h}||}\right]||\mathbf{h}|| \\ &\le \left[\underbrace{\frac{|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - \nabla f(\mathbf{a})|}{||\mathbf{h}||}}_{\to 0 \text{ as } \mathbf{h}\to\mathbf{0}} + \underbrace{||\nabla f(\mathbf{a})||}_{\text{fixed number}}\right]||\mathbf{h}||. \end{aligned}$$

By the squeeze theorem,

$$0 \le \lim_{\mathbf{h}\to\mathbf{0}} |f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a})| \le \lim_{\mathbf{h}\to\mathbf{0}} ||\nabla f(\mathbf{a})||||\mathbf{h}|| = 0.$$

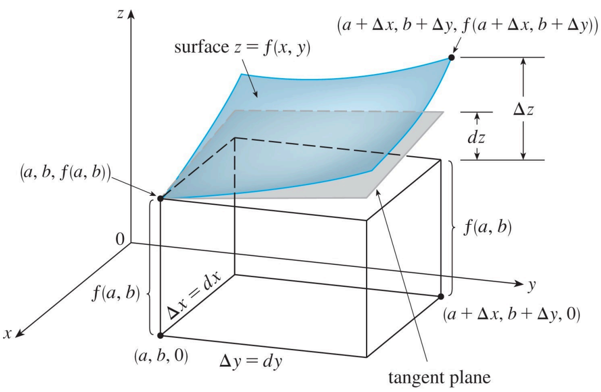
Hence, f is continuous at $\mathbf{a}$. □

■ Geometric viewpoint of defintion of differentiability

For a two variables function $z = f(x, y)$, as x changes from a to $a + \triangle x$ and y changes from b to $b + \triangle y$, the corresponding increment of z is

$$\begin{aligned} \triangle z &= f(a+\triangle x, b+\triangle y) - f(a,b) \\ &= \underbrace{f_x(a,b)\triangle x + f_y(a,b)\triangle y}_{\text{linear approximation}} + \underbrace{\varepsilon_1\triangle x + \varepsilon_2\triangle y}_{\text{error}} \end{aligned}$$

where $\varepsilon_1 = \varepsilon_1(\triangle x, \triangle y)$ and $\varepsilon_2 = \varepsilon_2(\triangle x, \triangle y)$. We expect that $\varepsilon_1, \varepsilon_2 \to 0$ as $(\triangle x, \triangle y) \to (0, 0)$.



$$z - f(a, b) = f_x(a, b)(x - a) + f_y(a, b)(y - b)$$

The following definition is equivalent to Definition 14.5.1.

Definition 14.5.9. Let $z = f(x, y)$. We call that f is "*differentiable*" at (a, b) if $\triangle z$ can be expressed in the form

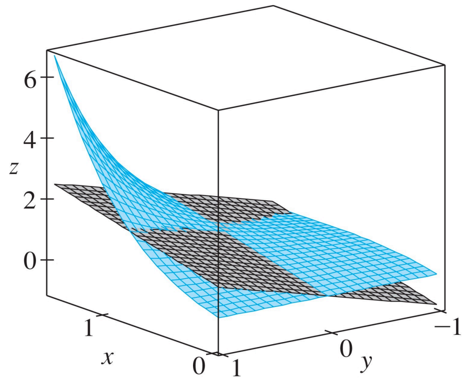
$$\triangle z = f_x(a,b)\triangle x + f_y(a,b)\triangle y + \varepsilon_1\triangle x + \varepsilon_2\triangle y$$

where $\varepsilon_1, \varepsilon_2 \to 0$ as $(\triangle x, \triangle y) \to (0, 0)$.

Example 14.5.10. Show that $f(x, y) = xe^{xy}$ is differentiable at $(1, 0)$ and find its linearization there. Then use it to approximate $f(1.1, -0.1)$.

Proof. Since $f_x(x, y) = e^{xy} + xye^{xy}$ and $f_y(x, y) = x^2e^{xy}$ are continuous functions, $f(x, y)$ is differentiable everywhere. Moreover, $f_x(1, 0) = 1$ and $f_y(1, 0) = 1$. The linearization of f at $(1, 0)$ is

$$\begin{aligned} L(x, y) &= f(1, 0) + f_x(1, 0)(x-1) + f_y(1, 0)(y-0) \\ &= 1 + (x-1) + y \\ &= x + y. \end{aligned}$$



Then

$$f(1.1, -0.1) \approx L(1.1, -0.1) = 1.1 + (-0.1) = 1.$$

In fact, $f(1.1, -0.1) = 1.1e^{-0.1} \approx 0.98542$. □

14.6 The Gradient Vector and Directional Derivatives

❑ Laws of Gradients

Theorem 14.6.1. *Let $f, g : D \subseteq \mathbb{R}^n \to \mathbb{R}$ be differentiable at* $\mathbf{a}$ *and c be a constant number. Then*

(a) $f + g$ is differentiable at $\mathbf{a}$ *and $\nabla(f \pm g)(\mathbf{a}) = \nabla f(\mathbf{a}) \pm \nabla g(\mathbf{a})$.*

(b) cf is differentiable at $\mathbf{a}$ *and $\nabla(cf)(\mathbf{a}) = c\nabla f(\mathbf{a})$.*

(c) fg is differentiable at $\mathbf{a}$ *and $\nabla(fg)(\mathbf{a}) = f(\mathbf{a})\nabla g(\mathbf{a}) + g(\mathbf{a})\nabla f(\mathbf{a})$.*

(d) If $g(\mathbf{a}) \neq 0$, $\dfrac{f}{g}$ is differentiable at $\mathbf{a}$ *and*

$$\nabla\Big(\frac{f}{g}\Big)(\mathbf{a}) = \frac{g(\mathbf{a})\nabla f(\mathbf{a}) - f(\mathbf{a})\nabla g(\mathbf{a})}{g^2(\mathbf{a})}.$$

Proof. We will prove part(c) here and the proofs of part(a)(b)(d) are left to the readers. Consider

$$\begin{aligned} \frac{f(\mathbf{a}+\mathbf{h})g(\mathbf{a}+\mathbf{h}) - f(\mathbf{a})g(\mathbf{a})}{\|\mathbf{h}\|} &= \frac{\big(f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a})\big)g(\mathbf{a}+\mathbf{h}) + f(\mathbf{a})\big(g(\mathbf{a}+\mathbf{h}) - g(\mathbf{a})\big)}{\|\mathbf{h}\|} \\ &= \underbrace{\frac{\big(f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - \nabla f(\mathbf{a})\cdot\mathbf{h}\big)g(\mathbf{a}+\mathbf{h})}{\|\mathbf{h}\|}}_{(I)} + \frac{\big(\nabla f(\mathbf{a})\cdot\mathbf{h}\big)g(\mathbf{a}+\mathbf{h})}{\|\mathbf{h}\|} \\ &+ \underbrace{\frac{f(\mathbf{a})\big(g(\mathbf{a}+\mathbf{h}) - g(\mathbf{a}) - \nabla g(\mathbf{a})\cdot\mathbf{h}\big)}{\|\mathbf{h}\|}}_{(II)} + \frac{f(\mathbf{a})\big(\nabla g(\mathbf{a})\cdot\mathbf{h}\big)}{\|\mathbf{h}\|}. \end{aligned}$$

Hence,

$$\underbrace{\frac{f(\mathbf{a}+\mathbf{h})g(\mathbf{a}+\mathbf{h}) - f(\mathbf{a})g(\mathbf{a}) - \Big(g(\mathbf{a}+\mathbf{h})\nabla f(\mathbf{a}) + f(\mathbf{a})\nabla g(\mathbf{a})\Big)\cdot \mathbf{h}}{\|\mathbf{h}\|}}_{(III)} = (I) + (II).$$

Since f and g are differentiable at $\mathbf{a}$, $\lim\limits_{\mathbf{h}\to\mathbf{0}}(I) = 0$, $\lim\limits_{\mathbf{h}\to\mathbf{0}}(II) = 0$ and $\lim\limits_{\mathbf{h}\to\mathbf{0}} g(\mathbf{a}+\mathbf{h}) = g(\mathbf{a})$. Then $\lim\limits_{\mathbf{h}\to\mathbf{0}}(III) = 0$. Therefore,

$$\nabla(fg)(\mathbf{a}) = f(\mathbf{a})\nabla g(\mathbf{a}) + g(\mathbf{a})\nabla f(\mathbf{a}).$$

□

❑ <u>Directional Derivatives</u>

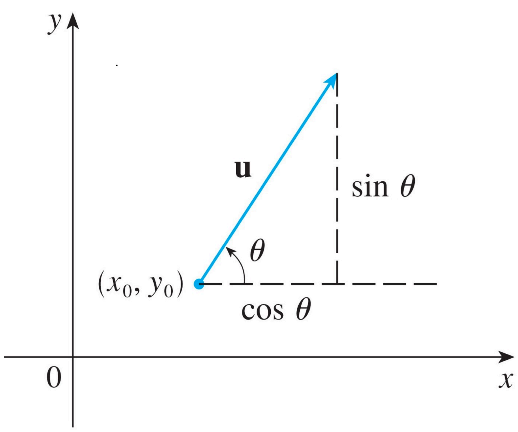
In Section 14.3, we studied the partial derivatives for a two variables function $z = f(x, y)$. The partial derivative

$$f_x(x_0, y_0) = \lim_{h\to 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h}$$

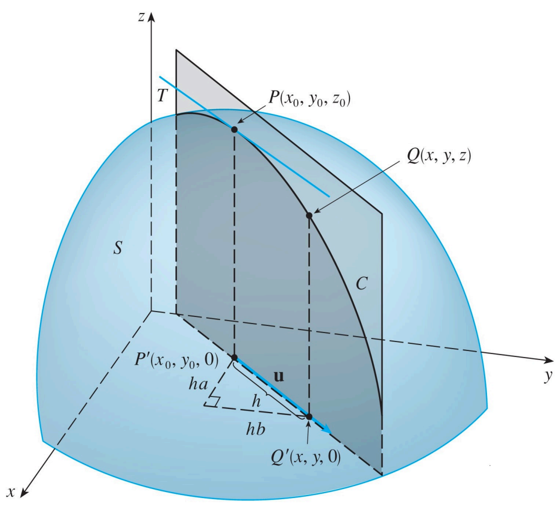
represents the rate of change of z in the x-direction (in the direction of the unit vector $\mathbf{i}$). Similarly,

$$f_y(x_0, y_0) = \lim_{h\to 0} \frac{f(x_0, y_0 + h) - f(x_0, y_0)}{h}$$

represents the rate of change of z in the y-direction (in the direction of the unit vector $\mathbf{j}$).
Question: How about the rate of change of z at (x_0, y_0) in the direction of a unit vector $\mathbf{u} =< a, b >$.



A unit vector $\mathbf{u} = \langle a, b\rangle = \langle \cos\theta, \sin\theta\rangle$



Let $P(x_0, y_0, z_0)$ lie on a surface S. The vertical plane that passes through P in the direction of $\mathbf{u}$ intersects S in a curve C. The slope of the tangent line T to C at the point P is the rate of change of z in the direction $\mathbf{u}$.

Let $\mathbf{u} =< a, b >$ be a unit vector and $z = f(x, y)$. Consider the quotient difference of z in the directional $\mathbf{u}$

$$\frac{\triangle z}{h} = \frac{z - z_0}{h} = \frac{f(x_0 + ha, y_0 + hb) - f(x_0, y_0)}{h}.$$

Taking $h \to 0$, we obtain the rate of change of z in the direction $\mathbf{u}$.

Definition 14.6.2.

(a) Let $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$ be a function and $(x_0, y_0) \in D$. The "directional derivatives" of f at (x_0, y_0) in the direction of a unit vector $\mathbf{u} =< a, b >$ is

$$D_{\mathbf{u}} f(x_0, y_0) = \lim_{h \to 0} \frac{f(x_0 + ha, y_0 + hb) - f(x_0, y_0)}{h}$$

if the limit exists.

(b) In general, let $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ be a function, $\mathbf{a} \in D$ and $\mathbf{u}$ be a unit vector. The directional derivative of f at $\mathbf{a}$ in the direction $\mathbf{u}$ is the limit

$$\lim_{h \to 0} \frac{f(\mathbf{a} + h\mathbf{u}) - f(\mathbf{a})}{h}$$

if it exists and is denoted by $D_{\mathbf{u}} f(\mathbf{a})$.

Remark. (i) In the above definition, the direction $\mathbf{u}$ is a "unit" vector. Hence, if we want to compute the directional derivative of f in the direction $\mathbf{v}$, which is not a unit vector, we should normalize $\mathbf{v}$ by $\mathbf{u} = \dfrac{\mathbf{v}}{||\mathbf{v}||}$.

(ii) If $\mathbf{u} =< 0, \cdots, 0, 1, 0, \cdots, 0 >$, then $D_{\mathbf{u}} f(\mathbf{a}) = f_{x_i}(\mathbf{a})$. The partial derivative of f with respect to x_i is a special directional derivative in the direction x_i.

To compute the directional derivative $D_{\mathbf{u}} f(x_0, y_0)$, there are two common methods:

(i) By the definition

(ii) Under certain assumptions, we can use the following theorem.

Theorem 14.6.3. *If $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\mathbf{a}$, then f has a directional derivative at $\mathbf{a}$ in every direction $\mathbf{u}$ where $\mathbf{u}$ is a unit vector and*

$$D_{\mathbf{u}} f(\mathbf{a}) = \nabla f(\mathbf{a}) \cdot \mathbf{u}.$$

Proof.

Recall that f is differentiable at $\mathbf{a}$. Then

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{|f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) - \nabla f(\mathbf{a}) \cdot \mathbf{h}|}{||\mathbf{h}||} = 0.$$

Let $\mathbf{h} = t\mathbf{u}$ and then $\|\mathbf{h}\| = |t|\|\mathbf{u}\| = |t|$. We have

$$\frac{f(\mathbf{a}+t\mathbf{u})-f(\mathbf{a})}{t} = \frac{f(\mathbf{a}+t\mathbf{u})-f(\mathbf{a})-\nabla f(\mathbf{a})\cdot\mathbf{h}}{t} + \frac{\nabla f(\mathbf{a})\cdot\mathbf{h}}{t}.$$

Hence,

$$\begin{aligned}
\lim_{t\to 0}\left|\frac{f(\mathbf{a}+t\mathbf{u})-f(\mathbf{a})}{t} - \nabla f(\mathbf{a})\cdot\mathbf{u}\right| &= \lim_{t\to 0}\left|\frac{f(\mathbf{a}+t\mathbf{u})-f(\mathbf{a})}{t} - \frac{\nabla f(\mathbf{a})\cdot(t\mathbf{u})}{t}\right| \\
&= \lim_{\mathbf{h}\to\mathbf{0}}\frac{|f(\mathbf{a}+\mathbf{h})-f(\mathbf{a})-\nabla f(\mathbf{a})\cdot\mathbf{h}|}{\|\mathbf{h}\|} \\
&= 0. \quad \text{(since } f \text{ is differentiable at } \mathbf{a}\text{)}
\end{aligned}$$

Therefore,

$$D_{\mathbf{u}}f(\mathbf{a}) = \lim_{t\to 0}\frac{f(\mathbf{a}+t\mathbf{u})-f(\mathbf{a})}{t} = \nabla f(\mathbf{a})\cdot\mathbf{u}.$$

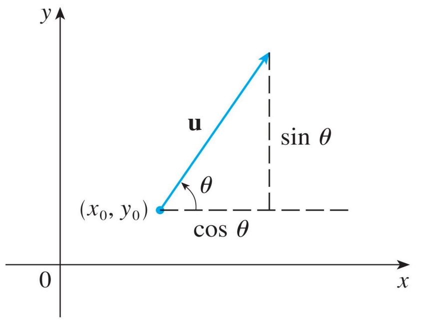
□

Note. In particular, if f is a differentiable function of x and y, then f has a directional derivative in the direction of any unit vector $\mathbf{u} =< a, b >$ and

$$D_{\mathbf{u}}f(x,y) = f_x(x,y)a + f_y(x,y)b.$$

Moreover, if $\mathbf{u} =< \cos\theta, \sin\theta >$, then

$$D_{\mathbf{u}}f(x,y) = f_x(x,y)\cos\theta + f_y(x,y)\sin\theta.$$



A unit vector $\mathbf{u} = \langle a, b\rangle = \langle\cos\theta, \sin\theta\rangle$

Remark. If f is differentiable and $\mathbf{u}$ is a unit vector, then

$$D_{\mathbf{u}}f(\mathbf{a}) = \nabla f(\mathbf{a})\cdot\mathbf{u}.$$

This means that the directional derivative (the rate of change of f) in the direction of a unit vector $\mathbf{u}$ is the scalar projection of the gradient vector $\nabla f(\mathbf{a})$ onto $\mathbf{u}$.

Example 14.6.4. Find the directional derivative $D_{\mathbf{u}}f(x,y)$ if

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

and $\mathbf{u}$ is the unit vector given by angle $\theta = \frac{\pi}{6}$. What is $D_{\mathbf{u}}f(1,2)$?

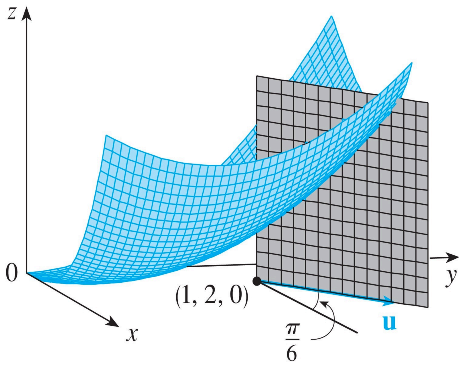
Proof. The gradient of f is

$$\nabla f =< f_x, f_y >=< 3x^2 - 3y, -3x + 8y > .$$

Hence, the directional derivative is

$$\begin{aligned}
D_{\mathbf{u}}f(x,y) &= f_x(x,y)\cos\theta + f_y(x,y)\sin\theta \\
&= (3x^2-3y)\cos\frac{\pi}{6} + (-3x+8y)\sin\frac{\pi}{6} \\
&= \frac{1}{2}[3\sqrt{3}x^2 - 3x + (8-3\sqrt{3})y]
\end{aligned}$$

and $D_{\mathbf{u}}f(1,2) = \dfrac{13-3\sqrt{3}}{2}$.



□

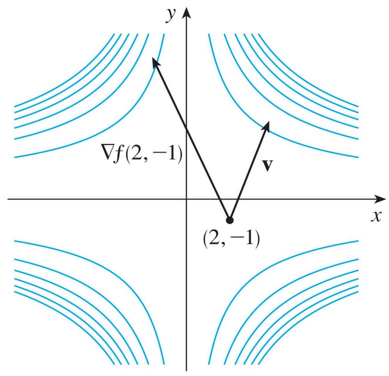
Example 14.6.5. Find the directional derivative of $f(x, y) = x^2y^3 - 4y$ at $(2, -1)$ in the direction $\mathbf{v} = 2\mathbf{i} + 5\mathbf{j}$.

Proof. The gradient of f is

$$\nabla f =< f_x, f_y >=< 2xy^3, 3x^2y^2 - 4 > .$$

Let $\mathbf{u} = \dfrac{\mathbf{v}}{||\mathbf{v}||} = \dfrac{2}{\sqrt{29}}\mathbf{i} + \dfrac{5}{\sqrt{29}}\mathbf{j}$. The directional derivative is

$$D_{\mathbf{u}}f(2, -1) = f_x(2, -1)\cdot\frac{2}{\sqrt{29}} + f_y(2, -1)\cdot\frac{5}{\sqrt{29}} = \frac{32}{\sqrt{29}}.$$



□

❏ Differentiability and Partial Derivatives

From Definition 14.5.1, we can prove that a differentiable function f havs (all) partial derivatives. In fact, it has directional derivatives in every direction. But the converse is false. There indeed exists a function which has all directional derivatives but it is not differentiable.

On the other hand, Theorem 14.5.4 says that continuity of all partial derivatives implies differentiability of f. We hope to understand the connection between the partial derivatives and differentiability.

Theorem 14.6.6. *If $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\mathbf{a}$, then all partial derivatives of f exist at $\mathbf{a}$ and*

$$\nabla f(\mathbf{a}) = \left\langle \frac{\partial f}{\partial x_1}(\mathbf{a}), \frac{\partial f}{\partial x_2}(\mathbf{a}), \cdots, \frac{\partial f}{\partial x_n}(\mathbf{a}) \right\rangle.$$

Proof. Since f is differentiable at $\mathbf{a}$, the gradient vector $\nabla f(\mathbf{a})$ exists and denote

$$\nabla f(\mathbf{a}) =< \alpha_1, \alpha_2, \cdots, \alpha_n > .$$

The partial derivative of f with respect to x_i is

$$\frac{\partial f}{\partial x_i}(\mathbf{a}) = \nabla f(\mathbf{a})\cdot < 0, \cdots, 0, 1, 0, \cdots, 0 >= \alpha_i$$

for $i = 1, 2 \cdots, n$. Hence $\nabla f(\mathbf{a}) = \left\langle \dfrac{\partial f}{\partial x_1}(\mathbf{a}), \dfrac{\partial f}{\partial x_2}(\mathbf{a}), \cdots, \dfrac{\partial f}{\partial x_n}(\mathbf{a}) \right\rangle$.

□

Note. If f is differentiable at $\mathbf{a}$, then we can explicitly write the form of $\nabla f(\mathbf{a})$.

Conclusion: Let $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ be a function. Then

$$\boxed{\text{All partial derivatives of } f \text{ exist and are continuous at } \mathbf{a}}$$

$$\Downarrow$$

$$\boxed{f \text{ is differentiable at } \mathbf{a} \text{ and } \nabla f(\mathbf{a}) \text{ exists and } \nabla f(\mathbf{a}) = \left\langle \frac{\partial f}{\partial x_1}(\mathbf{a}), \frac{\partial f}{\partial x_2}(\mathbf{a}), \cdots, \frac{\partial f}{\partial x_n}(\mathbf{a}) \right\rangle.}$$

$$\Downarrow$$

$$\boxed{\text{All partial derivatives of } f \text{ exist and the directional derivative } D_{\mathbf{u}}f(\mathbf{a}) = \nabla f(\mathbf{a}) \cdot \mathbf{u}}$$

Note. All the converse of the above arrows are false.

❑ Maximizing the Directional Derivatives

Suppose that $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\mathbf{a}$. Then all directional derivatives of f at $\mathbf{a}$ exist and

$$D_{\mathbf{u}}f(\mathbf{a}) = \nabla f(\mathbf{a}) \cdot \mathbf{u}$$

for any unit vector $\mathbf{u}$.

Question: In which direction does f change fastest and what is the maximum rate of change?

Observe that the rate of change of f in the direction $\mathbf{u}$ is

$$D_{\mathbf{u}}f(\mathbf{a}) = \nabla f(\mathbf{a}) \cdot \mathbf{u} = \|\nabla f(\mathbf{a})\| \underbrace{\|\mathbf{u}\|}_{=1} \cos\theta = \|\nabla f(\mathbf{a})\| \cos\theta$$

where θ is the angle between the two vectors $\nabla f(\mathbf{a})$ and $\mathbf{u}$. Hence, the maximum value of $D_{\mathbf{u}}f(\mathbf{a})$ occurs when $\theta = 0$.

Theorem 14.6.7. *Suppose that f is differentiable at* $\mathbf{a}$. *Then*

(a) The maximum value of the directional derivative $D_{\mathbf{u}}f(\mathbf{a})$ is $\|\nabla f(\mathbf{a})\|$ and it occurs when $\mathbf{u}$ *has the same direction as the gradient vector $\nabla f(\mathbf{a})$. That is, the function f at* $\mathbf{a}$ *increases fastest in the same direction of $\nabla f(\mathbf{a})$.*

(b) Similarly, the minimum value of the direction derivative $D_{\mathbf{u}}f(\mathbf{a})$ is $-\|\nabla f(\mathbf{a})\|$ and it occurs when $\mathbf{u}$ *has the opposite direction to the gradient vector $\nabla f(\mathbf{a})$. That is, the function f at* $\mathbf{a}$ *decreases fastest in the opposite direction to $\nabla f(\mathbf{a})$.*

(c) The function does not change in the direction of $\mathbf{u}$ *which is perpendicular to $\nabla f(\mathbf{a})$.*

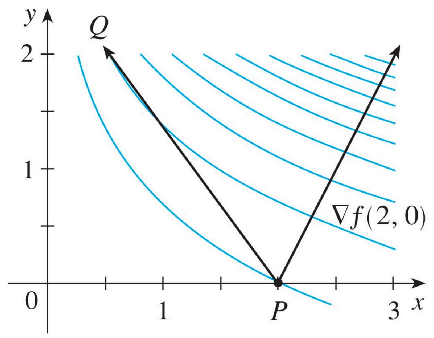
Example 14.6.8. Let $f(x, y) = xe^y$.

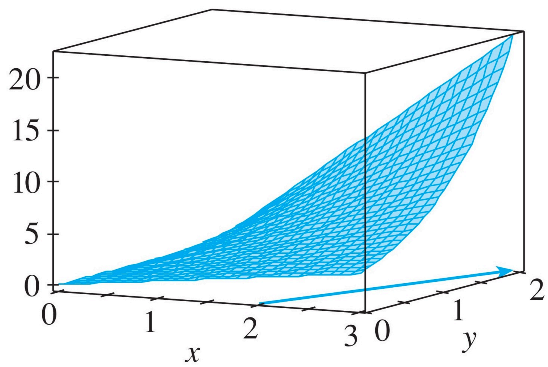
(a) Find the rate of change of f at the point $P(2, 0)$ in the direction from P to $Q(\frac{1}{2}, 2)$.

Proof. The vector $\overrightarrow{PQ} =< -\dfrac{3}{2}, 2 >$ and $\mathbf{u} = \dfrac{\overrightarrow{PQ}}{\| \overrightarrow{PQ} \|} =< -\dfrac{3}{5}, \dfrac{4}{5} >$. The gradient of f is $\nabla f(x, y) =$ $< e^y, xe^y >$ and $\nabla f(2, 0) =< 1, 2 >$. Hence, the rate of change of f in the direction $\overrightarrow{PQ}$ is $D_{\mathbf{u}}f(1, 2) =< 1, 2 > \cdot < -\frac{3}{5}, \frac{4}{5} >= 1$. □

(b) In what direction does f have the maximum rate of change? What is this maximum rate of change?

Proof. f increases fastest in the direction of the gradient vector $\nabla f(2,0) =< 1,2 >$ and the maximum rate of change is $\|\nabla f(2,0)\| = \| < 1,2 > \| = \sqrt{5}$. □





Example 14.6.9. Suppose that the temperature at a point (x, y, z) in space is given by

$$T(x,y,z) = \frac{80}{1 + x^2 + 2y^2 + 3z^2},$$

where T is measured in degree Celsius and x, y, z in meters. In which direction does the temperature increase fastest at the point $(1, 1, -2)$? What is the maximum rate of increase?

Proof. The gradient of T is $\nabla T(x,y,z) = \dfrac{160}{(1 + x^2 + 2y^2 + 3z^2)^2}(-x\mathbf{i} - 2y\mathbf{j} - 3z\mathbf{k})$ and then $\nabla T(1,1,-2) = \frac{5}{8}(-\mathbf{i} - 2\mathbf{j} + 6\mathbf{k})$.

The temperature increases fastest in the direction of the gradient vector $\nabla T(1,1,-2) = \frac{5}{8}(-\mathbf{i} - 2\mathbf{j} + 6\mathbf{k})$ or $-\mathbf{i} - 2\mathbf{j} + 6\mathbf{k}$. The maximum rate of increase is

$$\|\nabla T(1,1,-2)\| = \frac{5}{8}\| - \mathbf{i} - 2\mathbf{j} + 6\mathbf{k}\| = \frac{5\sqrt{41}}{8} \approx 4 \quad ({}^oC/m).$$

□

14.7 The Chain Rule

■ Chain Rule: First Version

Recall the for single variable functions $y = f(x)$, $x = g(t)$, $y = f\big(g(t)\big)$ is a composite function of variable t. Then

$$\frac{dy}{dt} = \frac{dy}{dx}\frac{dx}{dt}.$$

For a two variables function $z = f(x,y)$, if $x = g(t)$ and $y = h(t)$, then $z = f\big(g(t),h(t)\big)$ is indeirectly a function of t, say $z = z(t)$. Suppose that $z = f(x,y)$ is differentiable and, $x = g(t)$ and $y = h(t)$ are differentiable. Then

$$\begin{aligned}\triangle z &= f_x(x,y)\triangle x + f_y(x,y)\triangle y + \varepsilon_1\triangle x + \varepsilon_2\triangle y \\ &= f_x(x,y)\frac{\triangle x}{\triangle t}\triangle t + f_y(x,y)\frac{\triangle y}{\triangle t}\triangle t + \varepsilon_1\frac{\triangle x}{\triangle t}\triangle t + \varepsilon_2\frac{\triangle y}{\triangle t}\triangle t\end{aligned}$$

where $\varepsilon_1, \varepsilon_2 \to 0$ as $(\triangle x, \triangle y) \to (0,0)$. Since $x = g(t)$ and $y = h(t)$ are differentiable in t, we have $\frac{\triangle x}{\triangle t} \to \frac{dx}{dt}$ and $\frac{\triangle y}{\triangle t} \to \frac{dy}{dt}$ as $\triangle t \to 0$. Then, letting $\triangle t \to 0$,

$$\frac{\triangle z}{\triangle t} \to f_x(x,y)\frac{dx}{dt} + f_y(x,y)\frac{dy}{dt} + \underbrace{\lim_{\triangle t\to 0}\varepsilon_1}_{=0}\cdot\frac{dx}{dt} + \underbrace{\lim_{\triangle t\to 0}\varepsilon_2}_{=0}\cdot\frac{dy}{dt}.$$

We obtain

$$\frac{dz}{dt} = \lim_{\triangle t\to 0}\frac{\triangle z}{\triangle t} = f_x(x,y)\frac{dx}{dt} + f_y(x,y)\frac{dy}{dt}.$$

In Chapter 13, we studied the n vector-valued function $\mathbf{r}(t) =< x_1(t), \cdots, x_n(t) >: I \to \mathbb{R}^n$. If $\mathbf{r}(t)$ is differentiable on I, then

$$\mathbf{r}'(t) =< x_1'(t), \cdots, x_n'(t) > .$$

Theorem 14.7.1. *(Chain Rule)*

(a) (Two variables function) Suppose that $z = f(x,y)$ is a differentiable function of x and y where $x = x(t)$ and $y = y(t)$ are both differentiable functions of t. Then z is a differentiable function of t and

$$\frac{dz}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}.$$

(b) (General multiple variables function) Suppose that $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function. If $\mathbf{r} = \mathbf{r}(t)$ is a differentiable curve in D, then $f \circ \mathbf{r}$ is differentiable and

$$\frac{d}{dt}\Big(f\big(\mathbf{r}(t)\big)\Big) = \nabla f\big(\mathbf{r}(t)\big) \cdot \mathbf{r}'(t).$$

Proof. It suffices to prove the case $n = 2$ and the general cases are similar.

Since $x = x(t)$ and $y = y(t)$ are differentiable in t,

$$\triangle x = x(t+\triangle t) - x(t) = \frac{dx}{dt}\triangle t + \varepsilon_1\triangle t \quad \text{and} \quad \triangle y = y(t+\triangle t) - y(t) = \frac{dy}{dt}\triangle t + \varepsilon_2\triangle t$$

where $\varepsilon_1, \varepsilon_2 \to 0$ as $\triangle t \to 0$ as well as

$$\lim_{\triangle t\to 0}\frac{\triangle x}{\triangle t} = \frac{dx}{dt} \quad \text{and} \quad \lim_{\triangle t\to 0}\frac{\triangle y}{\triangle t} = \frac{dy}{dt}.$$

Clearly, $\triangle x, \triangle y \to 0$ as $\triangle t \to 0$.

On the other hand, since f is differentiable,

$$\begin{aligned}\triangle z &= f(x+\triangle x, y+\triangle y) - f(x,y)\\ &= f_x(x,y)\triangle x + f_y(x,y)\triangle y + \varepsilon_3\triangle x + \varepsilon_4\triangle y\end{aligned}$$

where $\varepsilon_3, \varepsilon_4 \to 0$ as $(\triangle x, \triangle y) \to (0,0)$. Then

$$\frac{\triangle z}{\triangle t} = f_x(x,y)\frac{\triangle x}{\triangle t} + f_y(x,y)\frac{\triangle y}{\triangle t} + \varepsilon_3\frac{\triangle x}{\triangle t} + \varepsilon_4\frac{\triangle y}{\triangle t}.$$

Taking limits as $\triangle t \to 0$, we have

$$\begin{aligned}\frac{dz}{dt} = \lim_{\triangle t\to 0}\frac{\triangle z}{\triangle t} &= f_x(x,y)\underbrace{\left(\lim_{\triangle t\to 0}\frac{\triangle x}{\triangle t}\right)}_{=\frac{dx}{dt}} + f_y(x,y)\underbrace{\left(\lim_{\triangle t\to 0}\frac{\triangle y}{\triangle t}\right)}_{=\frac{dy}{dt}}\\ &\quad + \underbrace{\left(\lim_{\triangle t\to 0}\varepsilon_3\right)}_{=0}\left(\lim_{\triangle t\to 0}\frac{\triangle x}{\triangle t}\right) + \underbrace{\left(\lim_{\triangle t\to 0}\varepsilon_4\right)}_{=0}\left(\lim_{\triangle t\to 0}\frac{\triangle y}{\triangle t}\right)\\ &= f_x(x,y)\frac{dx}{dt} + f_y(x,y)\frac{dy}{dt}\\ &= \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}.\end{aligned}$$

□

Example 14.7.2. If $z = x^2y + 3xy^4$, where $x = \sin 2t$ and $y = \cos t$, find $\dfrac{dz}{dt}$ when $t = 0$.

Proof. Compute $\dfrac{\partial z}{\partial x} = 2xy + 3y^4$ and $\dfrac{\partial z}{\partial y} = x^2 + 12xy^3$. Then

$$\begin{aligned}\frac{dz}{dt} &= \frac{\partial z}{\partial x}\frac{dz}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}\\ &= (2xy + xy^4)(2\cos 2t) + (x^2 + 12xy^3)(-\sin t)\\ &= (2\sin 2t\cos t + 3\cos^4 t)(2\cos 2t) + (\sin^2 2t + 12\sin 2t\cos^3 t)(-\sin t).\end{aligned}$$

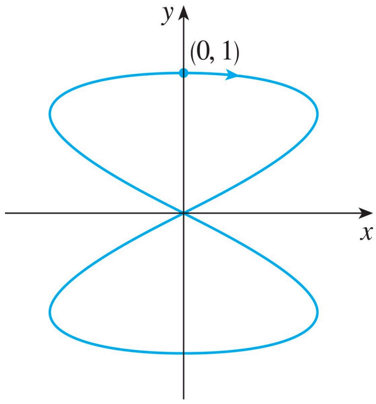
At $t = 0$, $\left.\dfrac{dz}{dt}\right|_{t=0} = 6$.

Note that $\dfrac{dz}{dt}$ represents the rate of change of z with respect to t as the point (x, y) moves along the curve C with parametric equation $r(t) =< \sin 2t, \cos t >$. □

Example 14.7.3. Compute the rate of change of $f(x,y,z) = x^2y + z\cos z$ along the curve $\mathbf{r}(t) =< t, t^2, t^3 >$.

Proof. Compute

$$\nabla f(x,y,z) =< 2xy, x^2, \cos z - z\sin z > \quad \text{and} \quad \mathbf{r}'(t) =< 1, 2t, 3t^2 > .$$



The curve $x = \sin 2t$, $y = \cos t$

Then

$$\begin{aligned} \frac{d}{dt}\big(f\big(\mathbf{r}(t)\big)\big) &= \nabla f\big(\mathbf{r}(t)\big)\cdot \mathbf{r}'(t) \\ &= < 2t^3, t^2, \cos t^3 - t^3 \sin t^3 > \cdot < 1, 2t, 3t^2 > \\ &= 4t^3 + 3t^2 \cos t^3 - 3t^5 \sin t^3. \end{aligned}$$

□

Remark. (1) Suppose that $f(\mathbf{x}) = f(x_1, x_2, \cdots, x_n)$ and $\mathbf{r}(t) =< x_1(t), \cdots, x_n(t) >$. Then

$$\nabla f(\mathbf{x}) =< \frac{\partial f}{\partial x_1}(\mathbf{x}), \frac{\partial f}{\partial x_2}(\mathbf{x}), \cdots, \frac{\partial f}{\partial x_n}(\mathbf{x}) > \quad \text{and} \quad \mathbf{r}'(t) =< x_1'(t), \cdots, x_n'(t) >$$

Hence,

$$\begin{aligned} \frac{d}{dt}\big(f\big(\mathbf{r}(t)\big)\big) &= \nabla f\big(\mathbf{r}(t)\big)\cdot \mathbf{r}'(t) \\ &= < \frac{\partial f}{\partial x_1}(\mathbf{x}), \frac{\partial f}{\partial x_2}(\mathbf{x}), \cdots, \frac{\partial f}{\partial x_n}(\mathbf{x}) > \cdot < x_1'(t), \cdots, x_n'(t) > \\ &= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\big(\mathbf{r}(t)\big) x_i'(t) \\ &= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\big(\mathbf{r}(t)\big) \frac{dx_i}{dt}(t) \end{aligned}$$

(2) Recall that the directional derivative of f at (a, b) in the direction $\mathbf{u}$ (unit vector) is

$$D_{\mathbf{u}} f(a, b) = \nabla f(a, b) \cdot \mathbf{u}.$$

Let the plane curve $\mathbf{r}(t)$ pass $< a, b >$ when $t = t_0$ (that is, $\mathbf{r}(t_0) =< a, b >$). Then

$$\frac{d}{dt}\big(f\big(\mathbf{r}(t)\big)\big)\Big|_{t=t_0} = \nabla f\big(\mathbf{r}(t_0)\big)\cdot \mathbf{r}'(t_0) = \|\mathbf{r}'(t_0)\| D_{\mathbf{u}} f(a, b)$$

where $\mathbf{u} = \dfrac{\mathbf{r}'(t_0)}{\|\mathbf{r}'(t_0)\|}$. This means that the rate of change of the composite function $f\big(\mathbf{r}(t)\big)$ at $t = t_0$ is equal to $\|\mathbf{r}'(t_0)\|$ multiple of the directional derivative of f at $\mathbf{r}(t_0)$ in the direction $\mathbf{r}'(t_0)$.

Corollary 14.7.4. *If $x = x(t)$ and $y = y(t)$ are twice differentiable at t and if $z = f(x, y)$ is twice differentiable at $\big(x(t), y(t)\big)$, then $z = f\big(x(t), y(t)\big)$ is twice differentiable at t and*

$$\frac{d^2z}{dt^2} = \frac{\partial z}{\partial x}\frac{d^2x}{dt^2} + \left(\frac{dx}{dt}\right)^2\frac{\partial^2 z}{\partial x^2} + 2\frac{\partial^2 z}{\partial x \partial y}\frac{dx}{dt}\frac{dy}{dt} + \left(\frac{dy}{dt}\right)^2\frac{\partial^2 z}{\partial y^2} + \frac{\partial z}{\partial y}\frac{d^2y}{dt^2}.$$

Proof. (Exercise) □

■ Chain Rule: Second Version

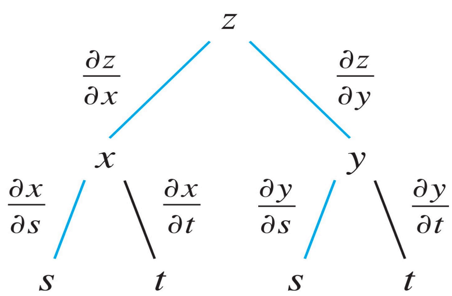
Let $z = f(x, y)$, $x = x(s, t)$ and $y = y(s, t)$ be differentiable functions. Then $z = z(s, t) = f\big(x(s, t), y(s, t)\big)$ is indirectly a function of s and t. Consider the partial derivative of z with respect to t. From the discuss in Section 14.3, fixing s (as a constant w.r.t t) and regarding z as a function of t. We can use the idea of Case1 to find the partial derivative of z with respect to t.

Theorem 14.7.5. *(Chain Rule)*

(a) *Suppose that $z = f(x, y)$ is a differentiable function of x and y, where $x = x(s, t)$ and $y = y(s, t)$ are differentiable functions of s and t. Then*

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s}, \qquad \frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t}.$$

The tree diagram is



(b) *If $x = x(s, t)$ and $y = y(s, t)$ are differentiable at (s, t) and $z = f\big(x_1, \cdots, x_n\big)$ is differentiable at $\big(x_1(t), x_n(t)\big)$ then*

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x_1}\frac{\partial x_1}{\partial s} + \cdots + \frac{\partial z}{\partial x_n}\frac{\partial x_n}{\partial s} = \sum_{i=1}^{n}\frac{\partial z}{\partial x_i}\frac{\partial x_i}{\partial s} \quad \text{and}$$

$$\frac{\partial z}{\partial t} = \frac{\partial z}{\partial x_1}\frac{\partial x_1}{\partial t} + \cdots + \frac{\partial z}{\partial x_n}\frac{\partial x_n}{\partial t} = \sum_{i=1}^{n}\frac{\partial z}{\partial x_i}\frac{\partial x_i}{\partial t}$$

Example 14.7.6. If $z = e^x \sin y$, where $x = st$ and $y = s^2 t$, find $\dfrac{\partial z}{\partial s}$ and $\dfrac{\partial z}{\partial t}$.

Proof. Compute that

$$\frac{\partial z}{\partial x} = e^x \sin y, \quad \frac{\partial z}{\partial y} = e^x \cos y$$

and

$$\frac{\partial x}{\partial s} = t^2, \ \frac{\partial x}{\partial t} = 2st, \ \frac{\partial y}{\partial s} = 2st, \ \frac{\partial y}{\partial t} = s^2.$$

Then

$$\begin{aligned}\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s} &= e^2 \sin y \cdot t^2 + e^2 \cos y \cdot 2st \\ &= t^2 e^{st} \sin(s^2 t) + 2ste^{st} \cos(s^2 t).\end{aligned}$$

and

$$\begin{aligned}\frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t} &= e^2 \sin y \cdot 2st + e^2 \cos y \cdot s^2 \\ &= 2ste^{st} \sin(s^2 t) + s^2 e^{st} \cos(s^2 t).\end{aligned}$$

□

Corollary 14.7.7. *Suppose that $z = f(x, y)$ is a twice differentiable function of x and y, where $x = x(s, t)$ and $y = y(s, t)$ are twice differentiable functions of s and t. Then*

$$\begin{aligned}\frac{\partial^2 z}{\partial s^2} = \frac{\partial}{\partial s}\Big(\frac{\partial z}{\partial s}\Big) &= \frac{\partial}{\partial s}\Big[\frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s}\Big] \\ &= \Big(\frac{\partial^2 z}{\partial x^2}\frac{\partial x}{\partial s} + \frac{\partial^2 z}{\partial y \partial x}\frac{\partial y}{\partial s}\Big)\frac{\partial x}{\partial s} + \frac{\partial z}{\partial x}\frac{\partial^2 x}{\partial s^2} \\ &+ \Big(\frac{\partial^2 z}{\partial x \partial y}\frac{\partial x}{\partial s} + \frac{\partial^2 z}{\partial y^2}\frac{\partial y}{\partial s}\Big)\frac{\partial y}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial^2 y}{\partial s^2}\end{aligned}$$

Example 14.7.8. Let $u = f(s^2 + t^2, st)$ Find $\dfrac{\partial^2 u}{\partial s \partial t}$.

Proof.

$$\frac{\partial u}{\partial t} = \frac{\partial f}{\partial x}(s^2 + t^2, st) \cdot 2t + \frac{\partial f}{\partial y}(s^2 + t^2, st) \cdot s.$$

and

$$\begin{aligned}\frac{\partial^2 u}{\partial s \partial t} = \frac{\partial}{\partial s}\Big(\frac{\partial u}{\partial t}\Big) &= \frac{\partial^2 f}{\partial x^2}(s^2 + t^2, st)(2s)(2t) + \frac{\partial^2 f}{\partial y \partial x}(s^2 + t^2, st)(2t^2) \\ &+ \frac{\partial^2 f}{\partial x \partial y}(s^2 + t^2, st)2s \cdot s + \frac{\partial^2 f}{\partial y^2}(s^2 + t^2, st)t \cdot s + \frac{\partial f}{\partial y}(s^2 + t^2, st) \cdot 1.\end{aligned}$$

□

■ Chain Rule: General Version

Suppose that u is a differentiable function of n variables $x_1, \cdots, x_n$ and each x_i is a differenbitable function of m variables $t_1, \cdots, t_m$. Then u is a differentiable function of $t_1, \cdots, t_m$ and

$$\frac{\partial u}{\partial t_i} = \frac{\partial u}{\partial x_1}\frac{\partial x_1}{\partial t_i} + \frac{\partial u}{\partial x_2}\frac{\partial x_2}{\partial t_i} + \cdots + \frac{\partial u}{\partial x_n}\frac{\partial x_n}{\partial t_i}$$

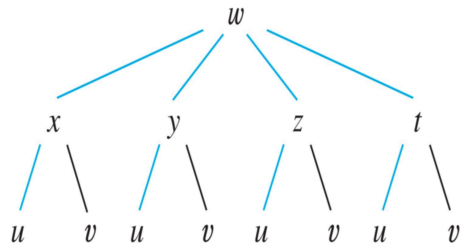
for each $i = 1, 2, \cdots, m$.

Example 14.7.9. Let $w = f(x, y, z, t)$, $x = x(u, v)$, $y = y(u, v)$ and $z = z(u, v)$. Then

$$\frac{\partial w}{\partial u} = \frac{\partial w}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial u} + \frac{\partial w}{\partial z}\frac{\partial z}{\partial u} + \frac{\partial w}{\partial t}\frac{\partial t}{\partial u}$$

and

$$\frac{\partial w}{\partial v} = \frac{\partial w}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial v} + \frac{\partial w}{\partial z}\frac{\partial z}{\partial v} + \frac{\partial w}{\partial t}\frac{\partial t}{\partial v}.$$



Example 14.7.10. If $u = x^4 y + y^2 z^3$, where $x = rse^t$, $y = rs^2 e^{-t}$ and $z = r^2 s \sin t$, find the value of $\frac{\partial u}{\partial s}$ when $r = 2$, $s = 1$ and $t = 0$.

Proof.

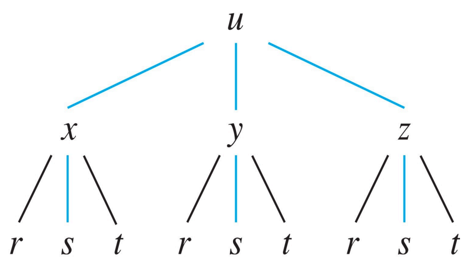
$$\frac{\partial u}{\partial x} = 4x^3 y, \quad \frac{\partial u}{\partial y} = x^4 + 2yz^3, \quad \frac{\partial u}{\partial z} = 3y^2 z^2$$

and

$$\frac{\partial x}{\partial s} = re^t, \quad \frac{\partial y}{\partial s} = 2rse^{-t}, \quad \frac{\partial z}{\partial s} = r^2 \sin t.$$

Then

$$\begin{aligned} \frac{\partial u}{\partial s} &= \frac{\partial u}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial s} + \frac{\partial u}{\partial z}\frac{\partial z}{\partial s} \\ &= 4x^3 y \cdot re^t + (x^4 + 2yz^3) \cdot 2rse^{-t} + 3y^2 z^2 \cdot r^2 \sin t. \end{aligned}$$



When $(r, s, t) = (2, 1, 0)$, $x = 2$, $y = 2$ and $z = 0$. Hence,

$$\left.\frac{\partial u}{\partial s}\right|_{(r,s,t)=(2,1,0)} = 64 \cdot 2 + 16 \cdot 4 + 0 \cdot 0 = 192.$$

□

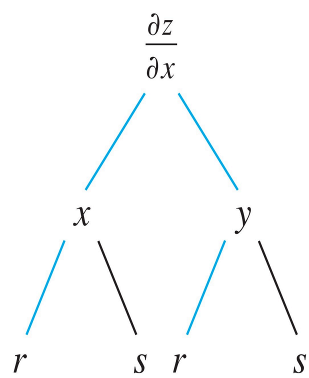
Example 14.7.11. If $z = f(x, y)$ has continuous second-order partial derivatives and $x = r^2 + s^2$ and $y = 2rs$, find $\frac{\partial z}{\partial r}$ and $\frac{\partial^2 z}{\partial r^2}$.

Proof.

$$\frac{\partial z}{\partial r} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial r} = \frac{\partial z}{\partial x}(2r) + \frac{\partial z}{\partial y}(2s).$$

and

$$\begin{aligned} \frac{\partial^2 z}{\partial r^2} = \frac{\partial}{\partial r}\left(\frac{\partial z}{\partial r}\right) &= 2r\left[\frac{\partial^2 z}{\partial x^2}\frac{\partial x}{\partial r} + \frac{\partial^2 z}{\partial y \partial x}\right] + 2\frac{\partial z}{\partial x} \\ &\quad +2s\left[\frac{\partial^2 z}{\partial x \partial y}\frac{\partial x}{\partial r} + \frac{\partial^2 z}{\partial y^2}\frac{\partial y}{\partial r}\right] \\ &= 2\frac{\partial z}{\partial x} + 4r^2\frac{\partial^2 z}{\partial x^2} + 4s^2\frac{\partial^2 z}{\partial y^2} + 8sr\frac{\partial^2 z}{\partial x \partial y}. \end{aligned}$$



Note that $\frac{\partial^2 z}{\partial x \partial y} = \frac{\partial^2 z}{\partial y \partial x}$ since f has continuous second partial derivatives. □

14.8 Mean Value Theorem and Implicit Differentiation

❑ Mean Value Theorem

Theorem 14.8.1. *(Mean Value Theorem) Let $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ be a function and the segment $\overline{\mathbf{ab}} \subset D$. If f is differentiable at every point on $\overline{\mathbf{ab}}$, then there exists $\mathbf{c} \in \overline{\mathbf{ab}}$ such that*

$$f(\mathbf{b}) - f(\mathbf{a}) = \nabla f(\mathbf{c}) \cdot (\mathbf{b} - \mathbf{a}).$$

Proof. Let $g(t) = f(\mathbf{a} + t(\mathbf{b} - \mathbf{a}))$. For $t \in [0, 1]$, $\mathbf{a} + t(\mathbf{b} - \mathbf{a}) \in \overline{\mathbf{ab}}$, $g(0) = f(\mathbf{a})$ and $g(1) = f(\mathbf{b})$. Since f is differentiable at every point on the segment $\overline{\mathbf{ab}}$, $g(t)$ is differentiable on $[0, 1]$ and

$$g'(t) = \nabla f\big(\mathbf{a} + t(\mathbf{b} - \mathbf{a})\big) \cdot (\mathbf{b} - \mathbf{a}).$$

By the mean value theorem for single variable function, there exists $t_0 \in (0, 1)$ such that

$$\begin{aligned} f(\mathbf{b}) - f(\mathbf{a}) &= g(1) - g(0) = g'(t_0)(1 - 0) \\ &= \nabla f\big(\mathbf{a} + t_0(\mathbf{b} - \mathbf{a})\big) \cdot (\mathbf{b} - \mathbf{a}) = \nabla f(\mathbf{c}) \cdot (\mathbf{b} - \mathbf{a}) \end{aligned}$$

where $\mathbf{c} = \mathbf{a} + t_0(\mathbf{b} - \mathbf{a})$. □

Corollary 14.8.2. *Suppose that $f(x, y)$ is differentiable on an open set containing the line segment connecting the point $P(x_0, y_0)$ and $Q(x_0 + h, y + k)$. Then there exists $\theta \in (0, 1)$ such that*

$$f(x_0 + h, y_0 + k) - f(x_0, y_0) = hf_x(x_0 + \theta h, y_0 + \theta k) + kf_y(x_0 + \theta h, y_0 + \theta k).$$

Proof. (Exercise)

□

❑ Implicit Differentiation

Recall that if the two variables x and y have a relation, for example $xy^2 + x \sin y = 1$, we can find $\dfrac{dy}{dx}$. By differentiating of both sides,

$$\frac{d}{dx}\big(xy^2 + x \sin y\big) = \frac{d}{dx}(1)$$

we have

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

In general, for the equation $F(x, y) = 0$ where F is differentiable, we can regard y as a function of x. To find $\dfrac{dy}{dx}$,

$$\frac{\partial}{\partial x}\Big(F(x, y)\Big) = \frac{\partial}{\partial x}(0).$$

We have

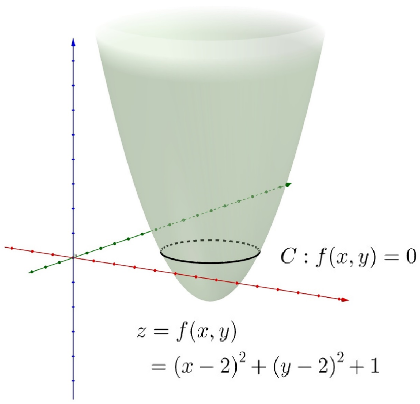
$$\frac{\partial F}{\partial x} \underbrace{\frac{dx}{dx}}_{=1} + \frac{\partial F}{\partial y}\frac{dy}{dx} = 0.$$

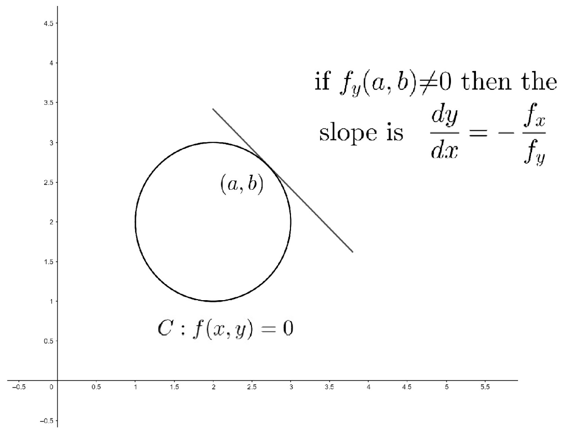
and then

$$\frac{dy}{dx} = -\frac{\frac{\partial F}{\partial x}}{\frac{\partial F}{\partial y}} = -\frac{F_x}{F_y}.$$

Note. The "*Implicit Function Theorem*" give conditions under which this assumption is valid: if F is defined on a disk containing (a, b) where $F(a, b) = 0$, $F_y(a, b) \neq 0$, and F_x and F_y are continuous on the disk, then the equation $F(x, y) = 0$ defines y as a function of x near the point (a, b) and the derivtive of y with respect to x is

$$\frac{dy}{dx} = -\frac{F_x}{F_y}.$$





Example 14.8.3. Find $\frac{dy}{dx}$ if $x^3 + y^3 = 6xy$.

Proof. Let $F(x, y) = x^3 + y^3 - 6xy$. Then $F_x = 3x^2 - 6y$ and $F_y = 3y^2 - 6x$. We have

$$\frac{dy}{dx} = -\frac{F_x}{F_y} = -\frac{3x^2 - 6y}{3y^2 - 6x} = -\frac{x^2 - 2y}{y^2 - 2x}.$$

□

Question: If $z = f(x, y)$ or $F(x, y, z) = 0$, how to find $\frac{\partial z}{\partial x}$ and $\frac{\partial z}{\partial y}$?

For $F(x, y, z) = 0$, we can regard z as a function of x and y, say $z = f(x, y)$. Then $F\big(x, y, f(x, y)\big)$ for all $x, y \in Dom(f)$. Find $\frac{\partial z}{\partial x}$. Consider

$$\frac{\partial}{\partial x}\Big(F(x, y, z)\Big) = \frac{\partial F}{\partial x}\underbrace{\frac{dx}{dx}}_{=1} + \frac{\partial F}{\partial y}\underbrace{\frac{dy}{dx}}_{=0} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial x} = \frac{\partial}{\partial x}(0) = 0.$$

Therefore,

$$\frac{\partial z}{\partial x} = -\frac{F_x}{F_z} \quad \text{provided } F_z \neq 0.$$

Similarly, $\frac{\partial z}{\partial y} = -\frac{F_y}{F_z}$ provided $F_z \neq 0$.

Example 14.8.4. Find $\dfrac{\partial z}{\partial x}$ and $\dfrac{\partial z}{\partial y}$ if $x^3 + y^3 + z^3 + 6xyz = 1$.

Proof. Let $F(x, y, z) = x^3 + y^3 + z^3 + 6xyz - 1$. Then

$$F_x = 3x^2 + 6yz, \; F_y = 3y^2 + 6xz, \; F_z = 3z^2 + 6xy.$$

We have

$$\frac{\partial z}{\partial x} = -\frac{F_x}{F_z} = -\frac{x^2 + 2yz}{z^2 + 2xy} \quad \text{and} \quad \frac{\partial z}{\partial y} = -\frac{F_y}{F_z} = -\frac{y^2 + 2xz}{z^2 + 2xy}.$$

□

We give the Implicit Function Theorem here. It will be discussed in the course of Advanced Calculus.

Theorem 14.8.5. *(Implicit Function Theorem) If F is defined within a sphere containing (a, b, c), where $F(a, b, c) = 0$, $F_z(a, b, c) \neq 0$, and F_x, F_y and F_z aer continuous inside the sphere, then the equation $F(x, y, z) = 0$ define z as a function of x and y near the point (a, b, c) and this function is defferentiable and*

$$\frac{\partial z}{\partial x} = -\frac{F_x}{F_z} \quad \textit{and} \quad \frac{\partial z}{\partial y} = -\frac{F_y}{F_z}.$$

14.9 Tangent Plane to Level Surface

In Section14.4, we have learned that the equation of the tangent plane to the surface $S : z = f(x, y)$ at $P(x_0, y_0, z_0)$ is

$$z - z_0 = f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0). \tag{14.3}$$

Define $F(x, y, z) = z - f(x, y)$. Then

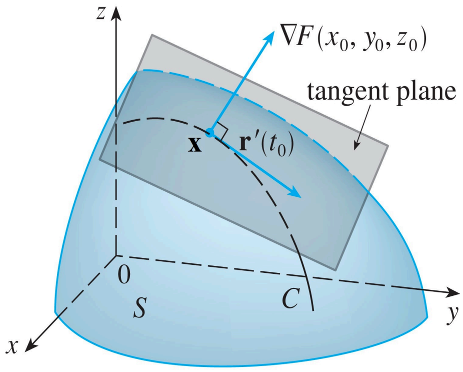
$$S = \left\{(x, y, z) \;\middle|\; z = f(x, y)\right\} = \left\{(x, y, z) \;\middle|\; z - f(x, y) = 0\right\} = \left\{(x, y, z) \;\middle|\; F(x, y, z) = 0\right\}$$

is a level surface of F when the value is equal to 0. Hence, (14.3) also interprets the equation of the tangnet plane to the level surface of F at P.

From the same spirit as above, we consider a differentiable function $F(x, y, z)$ of three variables x, y and z. Let S be a level surface with equation $F(x, y, z) = k$ and $\mathbf{x} =< x_0, y_0, z_0 >\in S$. To find the tangent plane to S at $\mathbf{x}$, it suffices to find the normal vector of S at $\mathbf{x}$.

Theorem 14.9.1. *Let $F : D \subseteq \mathbb{R}^3 \to \mathbb{R}$ be continuously differentiable and $S \subset D$ be a level surface of F. If $\mathbf{x} =< x_0, y_0, z_0 >\in S$ and $\nabla f(\mathbf{x}) \neq \mathbf{0}$, then $\nabla f(\mathbf{x})$ is perpendicular to S at $\mathbf{x}$.*

Proof. In order to prove $\nabla f(\mathbf{x})$ is perpendicular to S at $\mathbf{x}$, it suffices to show that the vector $\nabla f(\mathbf{x})$ is perpendicular to any curve on S passing $\mathbf{x}$ (the tangent vector to the curve at $\mathbf{x}$).



Let $C : \mathbf{r}(t) =< x(t), y(t), z(t) >$ be a differentiable curve that lies on S and passes through $\mathbf{x} =< x_0, y_0, z_0 >$ when $t = t_0$. Let S be the level surface with equation $F(x, y, z) = k$. Then

$$F\big(\mathbf{r}(t)\big) = F\big(x(t), y(t), z(t)\big) = k.$$

Hence,

$$\begin{aligned} 0 = \frac{d}{dt}\Big[F\big(\mathbf{r}(t)\big)\Big] &= \frac{\partial F}{\partial x}\frac{dx}{dt} + \frac{\partial F}{\partial y}\frac{dy}{dt} + \frac{\partial F}{\partial z}\frac{dz}{dt} \\ &= < \frac{\partial F}{\partial x}, \frac{\partial F}{\partial y}, \frac{\partial F}{\partial z} > \cdot < \frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt} > \\ &= \nabla F\big(\mathbf{r}(t)\big) \cdot \mathbf{r}'(t) \end{aligned}$$

Taking $t = t_0$, $\nabla F(\mathbf{x}) \perp \mathbf{r}'(t_0)$.

Note that $\mathbf{r}'(t_0)$ is a tangent vector lying on the tangent plane. Since C is an arbitrary curve on S, any vector on the tangent plane (to S at $\mathbf{x}$) is perpendicular to $\nabla F(\mathbf{x})$. Therefore, $\nabla F(\mathbf{x})$ is the normal vector of the tangent plane to S at $\mathbf{x}$. □

Note. (1) Let S be the level surface with equation $F(x, y, z) = k$ and $\mathbf{x} =< x_0, y_0, z_0 >\in S$. If $\nabla F(\mathbf{x}) \neq \mathbf{0}$, it is natural to define the tangent plane to the level surface S at $\mathbf{x}$ as the plane that passes through $\mathbf{x}$ and has normal vector $\nabla F(\mathbf{x})$. The equation of the tangent plane is

$$\nabla F(x_0, y_0, z_0)\cdot < x - x_0, y - y_0, z - z_0 >= 0.$$

That is,

$$F_x(x_0, y_0, z_0)(x - x_0) + F_y(x_0, y_0, z_0)(y - y_0) + F_z(x_0, y_0, z_0)(z - z_0) = 0.$$

(2) Consider the special case that the surface S with equation $z = f(x, y)$ which is the graph of a function f of two variables. Let $F(x, y, z) = f(x, y) - z$. Then S is with the equation $F(x, y, z) = 0$. Also,

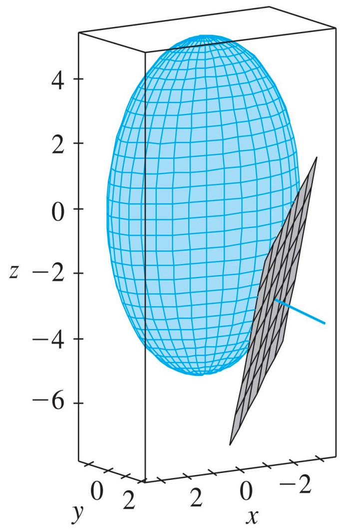
$$F_x(x_0, y_0, z_0) = f_x(x_0, y_0), \quad F_y(x_0, y_0, z_0) = f_y(x_0, y_0), \quad \text{and} \quad F_z(x_0, y_0, z_0) = -1.$$

The equation of the tangent plane to S at (x_0, y_0, z_0) is

$$f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) + (-1)(z - z_0) = 0.$$

Example 14.9.2. Find the equation of the tangne tplane at the point $(-2, 1, -3)$ to the ellipsoid $\frac{x^2}{4} + y^2 + \frac{z^2}{9} = 3$.

Proof.



Let $F(x, y, z) = \dfrac{x^2}{4} + y^2 + \dfrac{z^2}{9}$. Then the ellipsoid is the level surface (with $k = 3$) of $F(x, y, , z)$. Then

$$F_x = \frac{x}{2}, \quad F_y = 2y \quad \text{and} \quad F_z = \frac{2z}{9}.$$

Hence, $F_x(-2, 1, 3) = -1$, $F_y(-2, 1, 3) = 2$ and $F_z(-2, 1, -3) = -\dfrac{2}{3}$. The equation of the tangnet plane is

$$-(x + 2) + 2(y - 1) - \frac{2}{3}(z + 3) = 0$$

or

$$3x - 6y + 2z + 18 = 0.$$

□

❑ Normal Line

The normal line to S at $\mathbf{x}$ is the line passing through $\mathbf{x} =< x_0, y_0, z_0 >$ and perpendicular to the tangent plane. The direction of the normal line is the gradient vector $\nabla F(\mathbf{x})$. The symmetric equation are

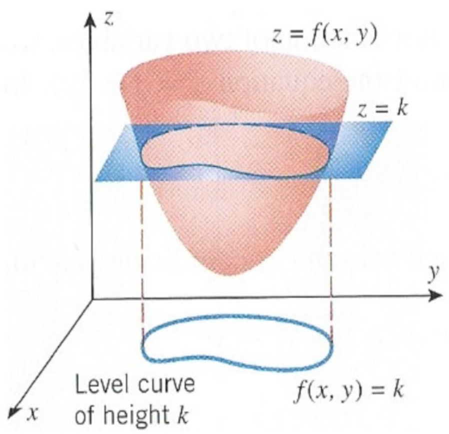
$$\frac{x - x_0}{F_x(x_0, y_0, z_0)} = \frac{y - y_0}{F_y(x_0, y_0, z_0)} = \frac{z - z_0}{F_z(x_0, y_0, z_0)}.$$

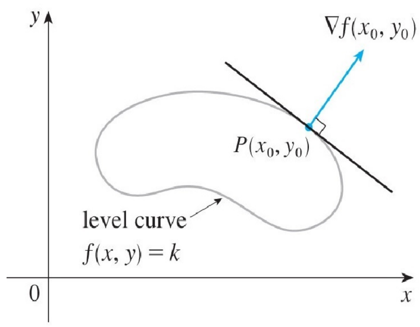
Example 14.9.3. As the above example, the equation of the normal line is

$$\frac{x + 2}{-1} = \frac{y - 1}{2} = \frac{z + 3}{-\frac{2}{3}}.$$

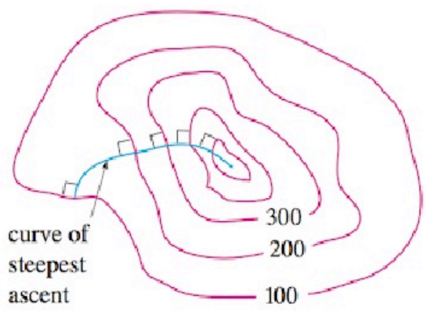
❑ Significance of the Gradient Vector

Consider the function $f(x, y)$ of two variables.

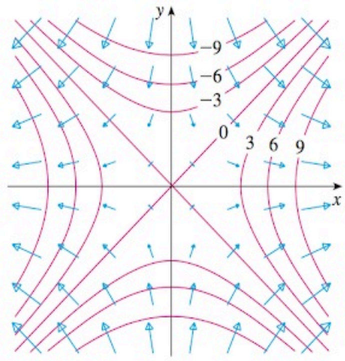




- The gradient vector $\nabla f(x_0, y_0)$ gives the direction of fastest increase of f. Intuitively, it is because the values of f remain constant as we move along the level curve.



a curve of steepest ascent is with direction $\nabla f(x, y)$ It is perpendicular to all of the contour lines.



a gradient vector field for the fuctnion f(x,y)= x^2 - y^2

- $\nabla f(x_0, y_0)$ is perpendicular to the level curve $f(x, y) = k$ that passes throught (x_0, y_0).
- For a plane curve $C : y = f(x)$, define $F(x, y) = y - f(x)$. Then C is a level curve of F. If $(x_0, y_0) \in C$, then $\nabla F(x_0, y_0)$ is the normal vector of C at (x_0, y_0).

  Example 14.9.4. Let C be the curve defined by $C = \{(x, y) \mid x^2 + y^3 = 9\}$. Find the tangent line of C at $(1, 2)$.

  Proof. Let $f(x, y) = x^2 + y^3$. Then C is a level curve of f (with $k = 9$). The gradient vector $\nabla f(1, 2) = < \frac{\partial f}{\partial x}(1, 2), \frac{\partial f}{\partial y}(1, 2) > = < 2, 12 >$ is the normal vector of C at $(1, 2)$. Hence, the tangent vector of C at $(1, 2)$ is $< 12, -2 >$ (perpendicular to $< 2, 12 >$). The equation of the tangent line to C at $(1, 2)$ is

  $$< x - 1, y - 2 > \cdot < 2, 12 > = 0 \quad \text{or} \quad 2(x - 1) + 12(y - 2) = 0.$$

  □

14.10 Maximum and Minimum Values

In the present section, we will study the extreme values of two variables function $f(x, y)$. Recall that, of a single variable funciton $f(x)$, we find the critical points as candinates and determine the extreme values by first derivative test or second derivative test. For a muti-variables functions, we also want to find the critical points by considering the directional derivatives.

Definition 14.10.1. Let f be a two variables function on D. We say that

(a) f has a local maximum (minimum) at (a, b) if

$$f(x, y) \le f(a, b) \qquad \big(f(x, y) \ge f(a, b)\big)$$

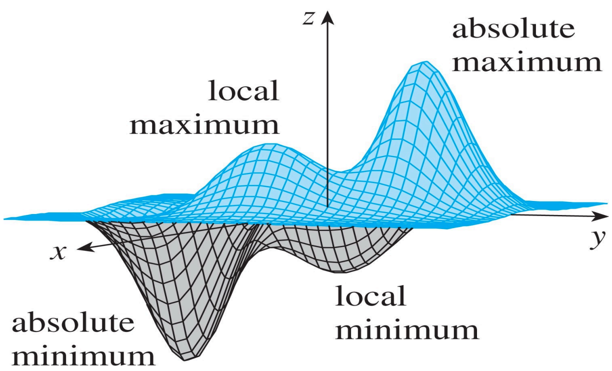
when (x, y) is near (a, b). [This means that $f(x, y) \le f(a, b)$ for all point (x, y) in some dist center (a, b)]. The number $f(a, b)$ is called a "*local maximum (minimum) value*".

(b) f has an absolute maximum (minimum) at (a,b)if

$$f(x,y) \le f(a,b) \qquad \big(f(x,y) \ge f(a,b)\big)$$

for all $(x,y) \in D$. The number $f(a,b)$ is called an "*absolute maximum (minimum) values*".

(c) The maximum and minimum values of f are called the "*extreme values of f*".



Question: How to find the extreme values of f?

Theorem 14.10.2. *If f has a local maximum or minimum at (a,b) and the first-order partial derivatives of f exists there, then $f_x(a,b) = 0$ and $f_y(a,b) = 0$.* $\quad (\nabla f(a,b) = \mathbf{0})$

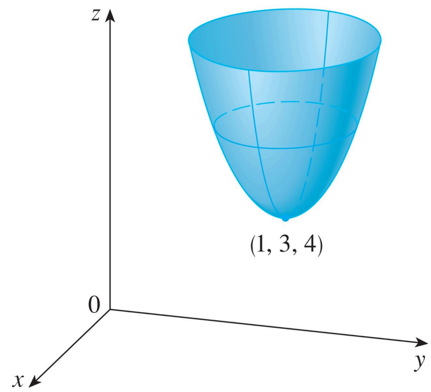
Proof. Let $g(x) = f(x,b)$. If f has a local maximum or minimum at (a,b), g has a local maximum or minimum at a. Thus, $0 = g'(a) = f_x(a,b)$. Similarly, $f_y(a,b) = 0$. □

Note. The geometric interpretation is that if the graph of f has a tangent plane at a local maximum or minimum, then the tangent plane must be horizontal.

Definition 14.10.3. We call that point (a,b) a "*critical point*" of f if either (1) $f_x(a,b) = 0$ and $f_y(a,b) = 0$ or (2) one of $f_x(a,b)$ and $f_y(a,b)$ does not exist.

Example 14.10.4. Let $f(x,y) = x^2 + y^2 - 2x - 6y + 14$. Find the critical point of f.

Proof. The partial derivatives $f_x(x,y) = 2x - 2$ and $f_y(x,y) = 2y - 6$. Therefore, $f_x(x,y) = 0$ when $x = 1$ and $f_y(x,y) = 0$ when $y = 3$. The point $(1,3)$ is a critical point of f. In fact, $f(x,y) = 4 + (x-1)^2 + (y-3)$ ahs a local and an absolute maximum at $(1,3)$. □

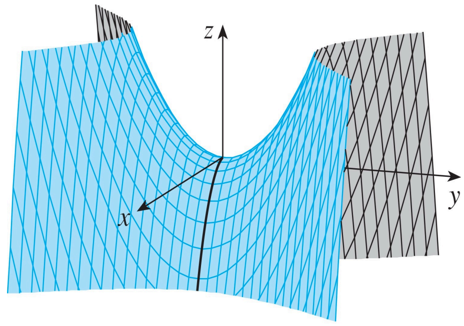


$z = x^2 + y^2 - 2x - 6y + 14$

Remark. The above theorem says that if f has a local maximum or minimum at (a, b), then (a, b) is a critical point of f. However, not all critical points give rise to maximum or minima.

Example 14.10.5. Find the extreme values of $f(x, y) = y^2 - x^2$.

Proof. The partial derivatives $f_x = -2x$ and $f_y = 2y$. Then $f_x = 0$ when $x = 0$ and $f_y = 0$ when $y = 0$. The point $(0, 0)$ is a critical point of f. But $f(0, 0)$ is neither a local maximum nor a local minimum.
Indeed, on the x-axis, $f(x, y) = -x^2 < 0$ if $x \neq 0$ and on the y-axis, $f(x, y) = y^2$ if $y \neq 0$. □



$z = y^2 - x^2$

Note. Near the origin the graph has the shape of a saddle and so $(0, 0)$ is called a "*saddle point*" of f.

❑ <u>Second Derivative Test</u>

Theorem 14.10.6. *Suppose that f_{xx}, f_{xy}, f_{yx} and f_{yy} are continuous near (a, b) and $f_x(a, b) = f_y(a, b) = 0$ (that is, (a, b) is a critical point of f). Let*

$$D = D(a, b) = f_{xx}(a, b)f_{yy}(a, b) - [f_{xy}(a, b)]^2.$$

(a) If $D > 0$ and $f_{xx}(a, b) > 0$, then $f(a, b)$ is a local minimum.

(b) If $D > 0$ and $f_{xx}(a, b) < 0$, then $f(a, b)$ is a local maximum.

(c) If $D < 0$ and $f(a, b)$ is not a local maximum or minimum.

Note. (1) In case(c), (a, b) is called a "*saddle point*" of f.

(2) If $D = 0$, the test is inconclusive, f could have a local maximum or local minimum at (a, b), or (a, b) could be a saddle point of f.

(3)

$$D = \begin{vmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{vmatrix} = f_{xx}f_{yy} - (f_{xy})^2.$$

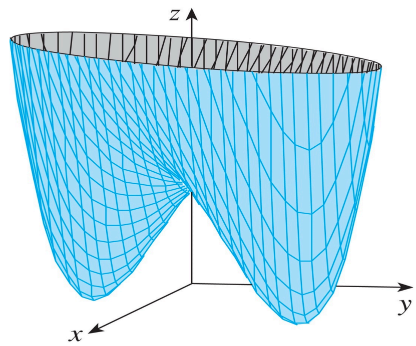
Example 14.10.7. Find the local maximum and minimum values and saddle points of $f(x, y) = x^4 + y^4 - 4xy + 1$.

Proof. The first and second partial derivatives of f are $f_x = 4x^3 - 4y$, $f_y = 4y^3 - 4x$, $f_{xx} = 12x^2$, $f_{xy} = -4 = f_{yx}$ and $f_{yy} = 12y^2$. Then $f_x = 0$ when $x^3 = y$ and $f_y = 0$ when $y^3 = x$. We can solve the critical points of f are $(0, 0)$, $(1, 1)$ and $(-1, -1)$, and

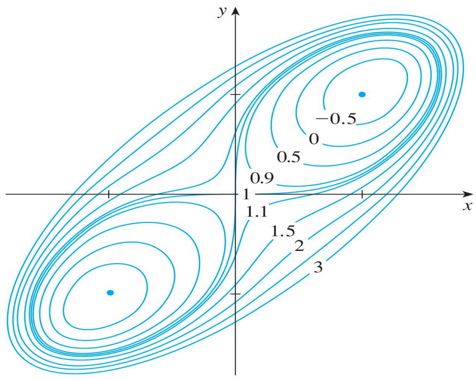
$$D(x, y) = f_{xx}f_{yy} - (f_{xy})^2 = 144x^2y^2 - 16.$$

- At $(0, 0)$, $D(0.0) = -16 < 0$. Then f has neither a local maximum nor a local minimum at $(0, 0)$.

- At $(1,1)$, $D(1,1) = 128 > 0$ and $f_{xx}(1,1) = 12 > 0$. Then $f(1,1) = -1$ is a local minimum of f.
- At $(-1,-1)$, $D(-1,-1) = 128 > 0$ and $f_{xx}(-1,-1) = 12 > 0$. Then $f(-1,-1) = -1$ is a local minimum of f.



$z = x^4 + y^4 - 4xy + 1$



□

Example 14.10.8. Find and classify the critical points of the function $f(x,y) = 10x^2y - 5x^2 - 4y^2 - x^4 - 2y^4$. Also find the highest points on the graph of f.

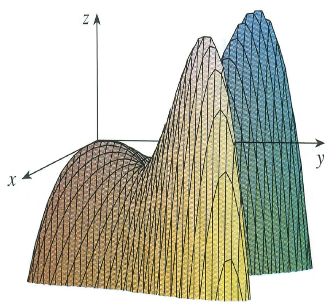
Proof. The first and second partial derivatives of f are

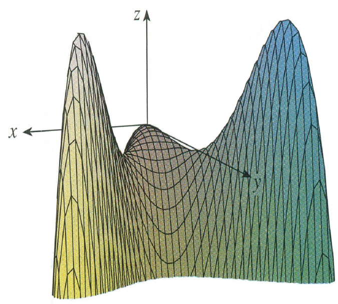
$f_x = 20xy - 10x - 4x^3,\ f_y = 10x^2 - 8y - 8y^3,\ f_{xx} = 20y - 10 - 12x^2,\ f_{xy} = f_{yx} = 20x,\ f_{yy} = -8 - 24y^2.$

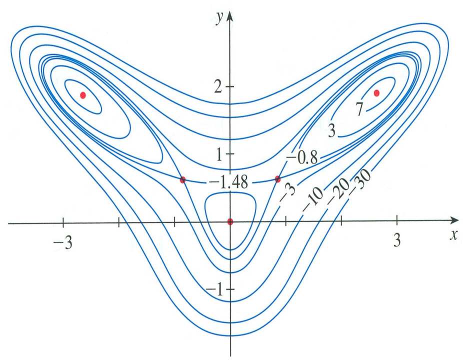
To find the critical points of f by solving $f_x = 0$ and $f_y = 0$, we have $(x,y) = (0,0),\ (\pm 2.64, 1.90),\ (\pm 0.86, 0.65)$.

Critical point	Value of f	f_{xx}	D	Conclusion
$(0,0)$	0	-10	80	local maximum
$(\pm 2.64, 1.90)$	8.50	-55.93	2488.72	local maximum
$(\pm 0.86, 0.65)$	-1.48	-5.87	-187.64	saddle point

The highest points on the graph of f are $(\pm 2.64, 1.90, 8.50)$.







□

Example 14.10.9. Find the shortest distance from the point $(1, 0, -2)$ to the plane $x+2y+z = 4$.

Proof. Let (x, y, z) be a point on the plane $x + 2y + z = 4$. The distance from (x, y, z) to $(1, 0, -2)$ is

$$d(x, y, z) = \sqrt{(x-1)^2 + y^2 + (z+2)^2}.$$

Taking $z = 4 - x - 2y$, then $d = \sqrt{(x-1)^2 + y^2 + (-x-2y+6)^2}$. Consider $f(x, y) = d^2(x, y) = (x-1)^2 + y^2 + (-x-2y+6)^2$. The first and second partial derivatives of f are

$$f_x = 4x + 4y - 14,\ f_y = 4x + 10y - 24,\ f_{xx} = 4,\ f_{xy} = f_{yx} = 4,\ f_{yy} = 10.$$

To find the critical point of f by solving $f_x = 0$ and $f_y = 0$, the point $(x, y) = (\frac{11}{6}, \frac{5}{3})$ is the only critical point of f. Also, $D = 4 \cdot 10 - 4^2 = 24 > 0$ and $f_{xx} = 4 > 0$. By the second derivatives test, $f(x, y)$ has a local minimium at $(\frac{11}{6}, \frac{5}{3})$. Then $d(\frac{11}{6}, \frac{5}{3}) = \frac{5}{\sqrt{6}}$. In fact, it is the absolute minimum.

□

Example 14.10.10. A rectangle box without a lid is to be made from $12m^2$ of cardboard. Find the maximum volume of such a box.

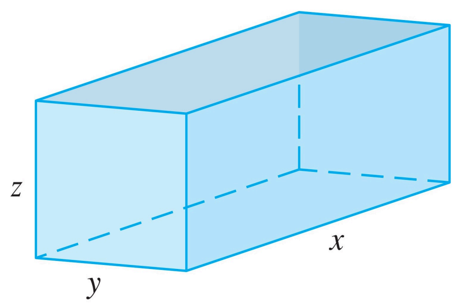
Proof. Let x, y and z be the length, width and height of the box. Then the volume of the box is $V(x, y, z) = xyz$ and the area of the four sides and the bottom is $2xz + 2yz + xy = 12$. Hence $z = \frac{12 - xy}{2(x+y)}$ and we can rewrite the volume function

$$V(x, y) = \frac{12xy - x^2y^2}{2(x+y)}.$$

Consider

$$\frac{\partial V}{\partial x} = \frac{y^2(12 - 2xy - x^2)}{2(x+y)^2} \quad \text{and} \quad \frac{\partial V}{\partial y} = \frac{x^2(12 - 2xy - y^2)}{2(x+y)^2}.$$

The critical point of V is $(2, 2)$. We can use the second derivative test to check that V has a local maximum at $(2, 2, 1)$. Then the maximum volume of the box is $4m^3$.



□

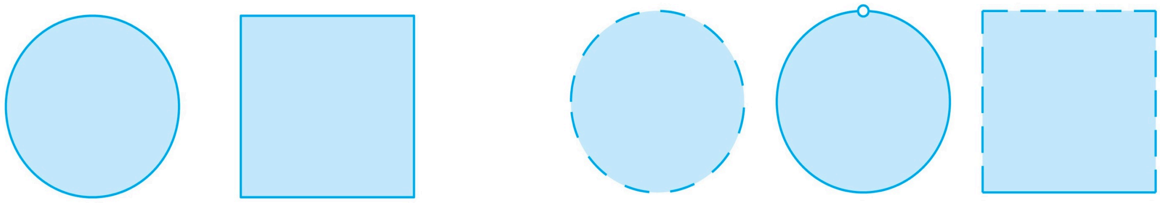
❑ Absolute Maximum and Minimum Values

Question: Under what conditions does a function $f(x, y)$ have (absolute) extreme values?

Recall that, for a single variable function $f(x)$, we have the "*Extreme Value Theorem*" that if f is continuous on a closed interval $[a, b]$, then f has an absolute maximum value and an absolute minimum value.

Question: How about two variables function $f(x, y)$?

Heuristically, corresponding to the "closed interval" in $\mathbb{R}$, a "close set" in $\mathbb{R}^2$ is a set contains all its boundary points. Also, a bounded set in $\mathbb{R}^2$ is a set that is contained within some disk.



(a) Closed sets

(b) Sets that are not closed

■ Extreme Value Theorem

Theorem 14.10.11. *If f is continuous on a closed and bounded set D in $\mathbb{R}^2$, then f attains an absolute maximum value $f(x_1, y_1)$ and an absolute minimum value $f(x_2, y_2)$ at some point (x_1, y_1) and (x_2, y_2) in D.*

Note. If $f(x, y)$ has an extreme value at (x_1, y_1), then (x_1, y_1) is either a critical point of f or a boundary point of D.

Question: How to find the absolute maximum value or minimum value of a continuous function $f(x, y)$ on a closed and bounded set D?

■ Strategy:

(1) Find the values of f at the critical point of f in D.

(2) Find the extreme value of f on the boundary of D.

(3) Check the values in (1) and (2). The largest value is the absolute maximum value and the smallest value is the absolute minimum.

Example 14.10.12. Find the absolute maximum and minimum values of the function $f(x, y) = x^2 - 2xy + 2y$ on the rectangle $D = \{(x, y) \mid 0 \le x \le 3,\ 0 \le y \le 2\}$.

Proof. Since f is a polynomial on the closed and bounded set D, there exists absolute maximum and minimum values in D.

First of all, we find the critical points of f in the interior of D. The partial derivatives of f are $f_x = 2x - 2y$ and $f_y = -2x + 2$. Hence, $(1, 1)$ is a critical point of f in D and $\boxed{f(1,1)=1}$.

Next, we consider the candinates of extreme point on the boundary D. The boundary of D consists of four lines L_1, L_2, L_3 and L_4.

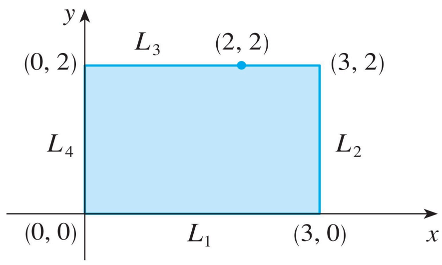
- For $(x, y) \in L_1$, $0 \le x \le 3$ and $y = 0$, $f(x, 0) = x^2$ is increasing. On L_1, f has a local maximum $\boxed{f(3,0)=9}$ and a local minimum $\boxed{f(0,0)=0}$.

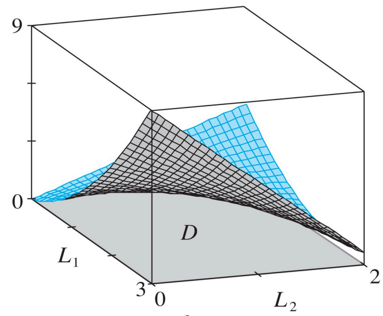
- For $(x, y) \in L_2$, $x = 3$ and $0 \le y \le 2$, $f(3, y) = -4y + 9$ is decreasing. On L_2, f has a local maximum $\boxed{f(3,0)=9}$ and a local minimum $\boxed{f(3,2)=1}$.

- For $(x, y) \in L_3$, $0 \le x \le 3$ and $y = 2$, $f(x, 2) = x^2 - 4x + 4 = (x - 2)^2$. On L_3, f has a loca maximum $\boxed{f(0,2)=4}$ and a local minimum $\boxed{f(2,2)=0}$.

- For $(x, y) \in L_4$, $x = 0$ and $0 \le y \le 2$, $f(0, y) = 2y$ is increasing. On L_4, f has a local maximum $\boxed{f(0,2)=4}$ and a local minimum $\boxed{f(0,0)=0}$.

Hence, f has an absolute maximum value $f(3, 0) = 9$ and an absolute minimum value $f(0, 0) = f(2, 2) = 0$.





$f(x, y) = x^2 - 2xy + 2y$

□

14.11 Lagrange Multipliers

In the present section, we will study the Lagrange's method to maximize or minimize a general function $f(\mathbf{x})$ subject to a constraint (or side condition) of the form $g(\mathbf{x}) = k$. The method works for n variables functions but we will only consider 2 or 3 variables functions in this section.

■ Geometric basis of Lagrange's method (for two variables functions)

Let $f(x, y)$ and $g(x, y)$ be two differentiable functions. The goal is to find the maximum (or minimum) of $f(x, y)$ subject to the constraint $g(x, y) = k$. For (x, y) satisfies $g(x, y) = k$, the point (x, y) lies on the level curve of $g(x, y)$ with the value k.

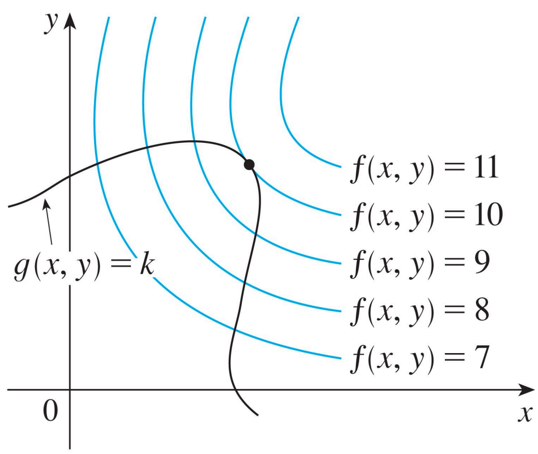
We want to find a point(s) (x_0, y_0) on the level curve $C = \{(x, y) \mid g(x, y) = k\}$ such that

$$f(x_0, y_0) \ge f(x, y) \quad \text{for all } (x, y) \in C. \tag{14.4}$$

Suppose that $(x_0, y_0) \in C$ satisfying (14.4) and $f(x_0, y_0) = M$. Then (x_0, y_0) is also on the level curve $C_1 = \{(x, y) \mid f(x, y) = M\}$. Moreover, since (x_0, y_0) is the maximum point, *the two level curve C and C_1 must be tangent each other at (x_0, y_0)*.

Since C and C_1 are level curves of g and f respectively, the gradient vectors $\nabla g \perp C$ and $\nabla f \perp C_1$. Then $\nabla g(x_0, y_0)$ is parallel to $\nabla f(x_0, y_0)$. Therefore, there exists a number λ ("*Lagrange multiplier*") such that

$$\nabla f(x_0, y_0) = \lambda \nabla g(x_0, y_0).$$



Conclusion: The candidnate point(s) where the extreme values occur must satisfy

$$\begin{cases} \nabla f(x,y) = \lambda \nabla g(x,y) & \text{for some number } \lambda \\ g(x,y) = k \end{cases}$$

■ **Lagrange methods for three variables functions**

For finding the extreme values of $f(x,y,z)$ subject to the constraint $g(x,y,z) = k$, by the same argument as above, if the maximum value of f is $f(x_0,y_0,z_0) = M$ where (x_0,y_0,z_0) lies on the level surface $S = \{(x,y,z) \mid g(x,y,z) = k\}$. Then the level surface $\{(x,y,z) \mid f(x,y,z) = M\}$ is tangent to S at (x_0,y_0,z_0). We have

$$\nabla f(x_0,y_0,z_0) \mathbin{/\!/} \nabla g(x_0,y_0,z_0).$$

(Intuitive veiwpoint) Let S be the level surface with equation $g(x,y,z) = k$. For every curve $\mathbf{r}(t) =< x(t), y(t), z(t) >$ lie on S, the tangent vector $\mathbf{r}'(t) \perp \nabla g\big(\mathbf{r}(t)\big)$ for every t.

Suppose that f has an extreme value at $P(x_0,y_0,z_0) \in S$ and $\mathbf{r}(t)$ is a curve on S passing P, say $\mathbf{r}(t_0) =< x_0, y_0, z_0 >$. Consider the function $h(t) = f\big(\mathbf{r}(t)\big)$ which has maximum value at t_0. Then $0 = h'(t_0) = \nabla f\big(\mathbf{r}(t_0)\big) \cdot \mathbf{r}'(t_0)$. We have $\nabla f\big(\mathbf{r}(t_0)\big) \perp \mathbf{r}'(t_0)$. Also, $\mathbf{r}'(t_0) \perp \nabla g\big(\mathbf{r}(t_0)\big)$. Then $\nabla f(x_0,y_0,z_0) \mathbin{/\!/} \nabla g(x_0,y_0,z_0)$. This implies that

$$\nabla f(x_0,y_0,z_0) = \lambda \nabla g(x_0,y_0,z_0) \qquad \text{for some number } \lambda.$$

This number λ is called a "*Lagrange multiplier*".

❑ Method of Lagrange Multiplier

To find the maximu and minimum values of $f(x,y,z)$ subject to the constraint $g(x,y,z) = k$ (assume that these extreme value exist and $\nabla g \neq \mathbf{0}$ on the surface $g(x,y,z) = k$). We solve this problem by following the below steps.

(a) Find all values of x, y, z and λ such that

$$\nabla f(x,y,z) = \lambda \nabla g(x,y,z) \qquad \text{and} \qquad g(x,y,z) = k.$$

(b) Evaluate f at all the points (x,y,z) that result from Step(a). The largest of these values is the maximum value of f; the smallest is the minimum value of f.

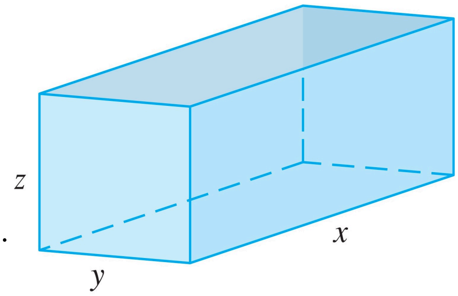
Example 14.11.1. A rectangle box without a lid is to be made from $12m^2$ of cardboard. Find the maximum volume of such a box.

Proof. Let the length, width and height of the box be x, y and z. Then the volume of the box is

$$V(x,y,z) = xyz.$$

The area of the four sides and the bottom is

$$g(x, y, z) = 2xz + 2yz + xy = 12.$$



To find the maximum of V subject to the constraint $g(x, y, z) = 12$. The gradient vector of V and g are

$$\nabla V =< yz, xz, xy > \quad \text{and} \quad \nabla g =< y{+}2z, x{+}2z, 2x{+}2y > .$$

Consider

$$\begin{cases} \nabla V = \lambda \nabla g \\ g(x, y, z) = 12 \end{cases} \Rightarrow \begin{cases} yz = \lambda(y + 2z) \\ xz = \lambda(x + 2z) \\ xy = \lambda(2x + 2y) \\ 2xz + 2yz + xy = 12 \end{cases} \Rightarrow \begin{cases} xyz = \lambda(xy + 2xz) & (1) \\ xyz = \lambda(xy + 2yz) & (2) \\ xyz = \lambda(2xz + 2yz) & (3) \\ 2xz + 2yz + xy = 12 & (4) \end{cases}$$

The number $\lambda \neq 0$; otherwise, we obtain $xy = xz = yz = 0$ and hence $g(x, y, z) = 0$ which contradicts the constraint. Also, Euqations(1),(2), (3) imply that

$$2xz + xy = 2yz + xy = 2xz + 2yz \quad \Rightarrow \quad xz = yz.$$

This says that either $x = y$ or $z = 0$.

(i) If $z = 0$, then $xy = 0$ and hence $x = y = 0$ which contradicts $g(x, y, z) = 12$.

(ii) If $x = y$ and $z \neq 0$, then $2xz + x^2 = 4xz$ and then $x = 2z = y$. Also, from Equation(4), we obtain $x = y = 2$ and $z = 1$.

The maximum volume of the box is $4m^3$. □

Example 14.11.2. Find teh extreme values of the function $f(x, y) = x^2 + 2y^2$ on the circle $x^2 + y^2 = 1$.

Proof. Let $g(x, y) = x^2 + y^2$. Then

$$\nabla f(x, y) =< 2x, 4y > \quad \text{and} \quad \nabla g(x, y) =< 2x, 2y > .$$

Consider

$$\begin{cases} \nabla f = \lambda \nabla g \\ g(x, y) = 1 \end{cases} \Rightarrow \begin{cases} 2x = 2\lambda x & (1) \\ 4y = 2\lambda y & (2) \\ x^2 + y^2 = 1 & (3) \end{cases}$$

By Equation(1), either $\lambda = 1$ or $x = 0$.

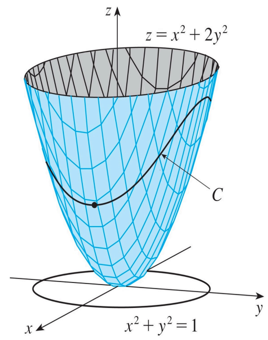
(i) If $\lambda = 1$, by Equation(2), $y = 0$. Then $x = \pm 1$ by Equation(3).

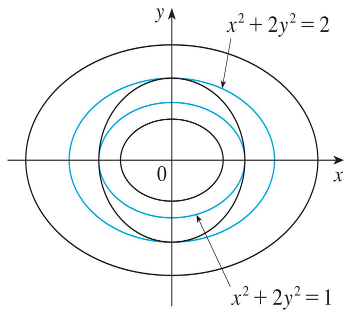
(ii) If $x = 0$, then $y = \pm 1$ by Equation(3) and $\lambda = 2$ by Equation(2).

Consider

$$\underbrace{f(1,0) = 1,\ f(-1,0) = 1}_{minimum} \text{ and } \underbrace{f(0,1) = 2,\ f(0,-1) = 2}_{maximum}.$$

The maximum value of f on the circle $x^2 + y^2 = 1$ is $f(0, \pm 1) = 2$ and the minimum value is $f(\pm 1, 0) = 1$.





□

Example 14.11.3. Find the extreme values of $f(x, y) = x^2 + 2y^2$ on the disk $x^2 + y^2 \leq 1$.

Proof. (1) Find the extreme values of f inside the disk $x^2 + y^2 \leq 1$. Consider $f_x = 2x = 0$ and $f_y = 4y = 0$. Then the critical point of f is $(0, 0)$. Moreover, $f_{xx} = 2$, $f_{xy} = f_{yx} = 0$ and $f_{yy} = 4$ and hence $D = f_{xx} f_{yy} - (f_{xy})^2 = 8 > 0$. Also, $f_{xx} > 0$. By the second derivative test, $f(0, 0)$ is a local minimum.

(2) Combining with the previous example, $f(0, 0) = 0$, $f(\pm 1, 0) = 1$ and $f(0, \pm 1) = 2$. Hence, the maximum value of f on the disk $x^2 + y^2 \leq 1$is $f(0, \pm 1) = 2$ and the minimum value is $f(0, 0) = 0$.

□

Example 14.11.4. Find the points on the sphere $x^2 + y^2 + z^2 = 4$ that are closest to and farthest from the point $(3, 1, -1)$

Proof. Let $f(x, y, z) = (x - 3)^2 + (y - 1)^2 + (z + 1)$ and $g(x, y, z) = x^2 + y^2 + z^2$. Then

$$\nabla f =< 2(x - 3), 2(y - 1), 2(z + 1) > \quad \text{and} \quad \nabla g =< 2x, 2y, 2z > .$$

Consider

$$\begin{cases} \nabla f = \lambda \nabla g \\ g(x, y, z) = 4 \end{cases} \Rightarrow \begin{cases} 2x - 6 = 2\lambda 2x \\ 2y - 2 = 2\lambda y \\ 2z + 1 = 2\lambda z \\ x^2 + y^2 + z^2 = 4 \end{cases} \Rightarrow \begin{cases} (1 - \lambda)x = 3 & (1) \\ (1 - \lambda)y = 1 & (2) \\ (1 - \lambda)z = -1 & (3) \\ 2xz + 2yz + xy = 12 & (4) \end{cases}$$

Clearly, $\lambda \neq 1$, $x \neq 0$, $y \neq 0$ and $z \neq 0$. Consider

$$\frac{(1)}{(2)} \Rightarrow \frac{x}{y} = 3 \Rightarrow x = 3y \quad \text{and} \quad \frac{(2)}{(3)} \Rightarrow \frac{y}{z} = -1 \Rightarrow z = -y.$$

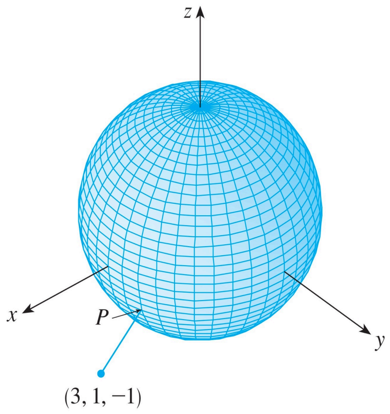
By (4), we have

$$(3y)^2 + y^2 + (-y)^2 = 4 \Rightarrow y = \pm\frac{2}{\sqrt{11}}.$$

Then

$$(x, y, z) = (\frac{6}{\sqrt{11}}, \frac{2}{\sqrt{11}}, -\frac{2}{\sqrt{11}}) \quad \text{or} \quad (-\frac{6}{\sqrt{11}}, -\frac{2}{\sqrt{11}}, \frac{2}{\sqrt{11}}).$$

Taking these two poinits into $f(x, y, z)$ the closest point is $(\frac{6}{\sqrt{11}}, \frac{2}{\sqrt{11}}, -\frac{2}{\sqrt{11}})$ and the farthest point is $(-\frac{6}{\sqrt{11}}, -\frac{2}{\sqrt{11}}, \frac{2}{\sqrt{11}})$.



□

Remark. In the example, the line passes through the origin and the point $(3, 1, -1)$ has parametric equation $x = 3t$, $y = t$ and $z = -t$. The line intersection the sphere $x^2 + y^2 + z^2 = 4$ when $t = \pm\frac{2}{\sqrt{11}}$. Then we can also solve the closest and the farthest points.

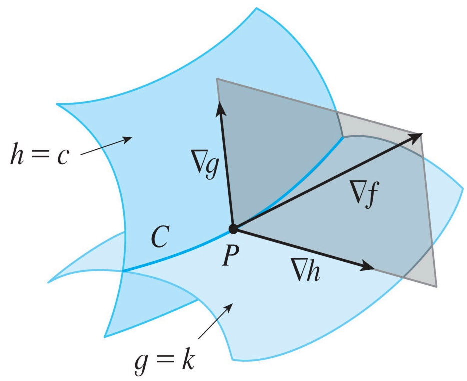
❑ Two Constraints

Find the maximum and minimum values of $f(x, y, z)$ subject to two constraints $g(x, y, z) = k$ and $h(x, y, z) = c$.

Let C be the intersection of the two level surfaces $g(x, y, z) = k$ and $h(x, y, z) = c$. Find $P(x_0, y_0, z_0) \in C$ such that $f(x_0, y_0, z_0)$ ahs extreme value along C.

To find the level surface $S = \{(x, y, z) \mid f(x, y, z) = M\}$ which tangnet to C. Then , at the intersection of C and S, $\nabla f \perp C$. We have

$$\nabla f(x_0, y_0, z_0) = \lambda \nabla g(x_0, y_0, z_0) + \mu \nabla h(x_0, y_0, z_0).$$



Example 14.11.5. Find the maximum value of the function $f(x, y, z) = x + 2y + 3z$ on the curve of intersection of the plane $x - y + z = 1$ and the cylinder $x^2 + y^2 = 1$.

Proof. Let $g(x, y, z) = x - y + z$ and $h(x, y, z) = x^2 + y^2$ Then

$$\nabla f =< 1, 2, 3 >, \quad \nabla g =< 1, -1, 1 > \quad \text{and} \quad \nabla h =< 2x, 2y, 0 > .$$

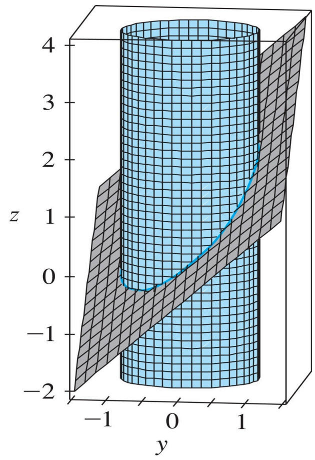
Consider

$$\begin{cases} <1,2,3> = \lambda <1,-1,1> + \mu <2x,2y,0> \\ x-y+z=1 \\ x^2+y^2=1 \qquad (*) \end{cases} \Rightarrow \begin{cases} 1=\lambda+2\mu x \\ 2=-\lambda+2\mu y \\ 3=\lambda \end{cases} \Rightarrow \begin{cases} \lambda=3 \\ x=-\dfrac{1}{\mu} \\ y=\dfrac{5}{2\mu} \end{cases}$$

Taking into $(*)$, we have $\mu = \pm\dfrac{\sqrt{29}}{2}$. Hence,

$(x,y,z) = (\dfrac{2}{\sqrt{29}}, -\dfrac{5}{\sqrt{29}}, 1+\dfrac{7}{\sqrt{29}})$ or $(\dfrac{2}{-\sqrt{29}}, \dfrac{5}{\sqrt{29}}, 1-\dfrac{7}{\sqrt{29}})$.

Therefore, the maximum value of f is $3+\sqrt{29}$.



□