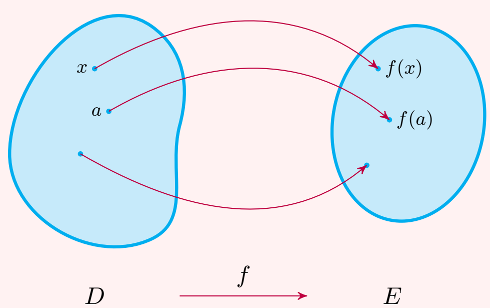
Contents

Chapter 1

Functions and Models

Function

A **function** f is a rule that assigns to each element x in a set D exactly one element, called $f(x)$, in a set E.
The set D is call the **domain** of f. The number $f(x)$ is the **value of f at** x and is read "f of x". The **range** of f is the set of all possible values of $f(x)$ as x varies throughout the domain. A symbol that represents an arbitrary number in the *domain* of f is called an **independent variable**. A symbol that represents a number in the *range* of f is called a **dependent variable**.



The **graph** of f is the set of ordered pairs

$$\left\{(x, f(x)) \,\middle|\, x \in D\right\}.$$

The Vertical Line Test

A curve in the xy-plane is the graph of a function of x if and only if no vertical line intersects the curve more than once.

Piecewise Function

A function is called a **piecewise function** or **piecewise-defined function** if it is defined by different formulas in different parts of their domains.

Step Function

A **step function** is a function whose graph looks like a series of steps. In other words, it is a piecewise constant function having only finitely many pieces.

Symmetry

A function f is called an **even function** if $f(-x) = f(x)$ for every number x in its domain.
A function f is called an **odd function** if $f(-x) = -f(x)$ for every number x in its domain.
A function f is called an **periodic function** if $f(x+p) = f(x)$ for every number x in its domain, where p is a positive constant. The smallest such number p is called the **period**.

Increasing and Decreasing

A function f is called an **increasing** on an interval I if

$$f(x_1) < f(x_2) \qquad \text{whenever } x_1 < x_2 \text{ in } I.$$

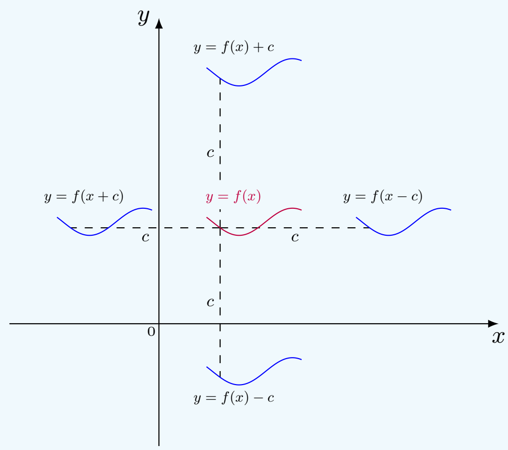
A function f is called an **decreasing** on an interval I if

$$f(x_1) > f(x_2) \qquad \text{whenever } x_1 < x_2 \text{ in } I.$$

Vertical and Horizontal Shifts

Suppose $c > 0$. To obtain the graph of

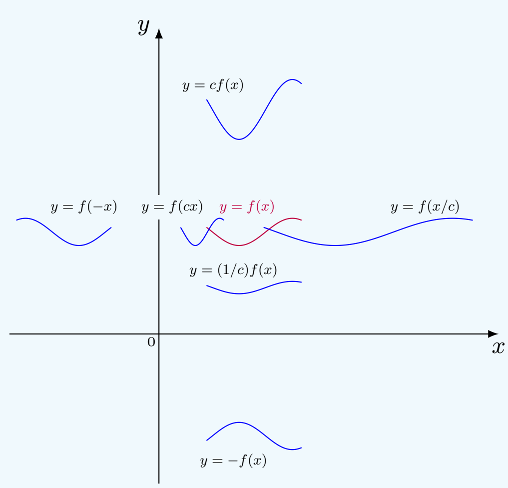
$$\begin{aligned}
y &= f(x) + c, && \text{shift the graph of } y = f(x) \text{ a distance } c \text{ units upward} \\
y &= f(x) - c, && \text{shift the graph of } y = f(x) \text{ a distance } c \text{ units downward} \\
y &= f(x - c), && \text{shift the graph of } y = f(x) \text{ a distance } c \text{ units to the right} \\
y &= f(x + c), && \text{shift the graph of } y = f(x) \text{ a distance } c \text{ units to the left}
\end{aligned}$$



Vertical and Horizontal Stretching and Reflecting

Suppose $c > 1$. To obtain the graph of

$y = cf(x)$,	stretch the graph of $y = f(x)$ vertically by a factor of c
$y = (1/c)f(x)$,	shrink the graph of $y = f(x)$ vertically by a factor of c
$y = f(cx)$,	shrink the graph of $y = f(x)$ horizontally by a factor of c
$y = f(x/c)$,	stretch the graph of $y = f(x)$ horizontally by a factor of c
$y = -f(x)$,	reflect the graph of $y = f(x)$ about the x-axis
$y = f(-x)$,	reflect the graph of $y = f(x)$ about the y-axis



Combinations of Functions

Two functions f and g can be combined to form new functions $f+g$, $f-g$, fg, and f/g in a manner similar to the way we add, subtract, multiply, and divide real numbers. The sum and difference functions are defined by

$$(f+g)(x) = f(x) + g(x), \qquad (f-g)(x) = f(x) - g(x).$$

If the domain of f is A and the domain of g is B, then the domain of $f+t$ is the intersection $A \cap B$. Similarly, the product and quotient functions are defined by

$$(fg)(x) = f(x)\,g(x), \qquad \left(\frac{f}{g}\right)(x) = \frac{f(x)}{g(x)}.$$

The domain of fg is $A \cap B$, but we can't divide by 0 and so the domain of f/g is $\{x \in A \cap B \colon g(x) \neq 0\}$.

Composition

Given two functions f and g, the **composite function** $f \circ g$ (also called the **composition** of f and g) is defined by

$$(f \circ g)(x) = f(g(x)).$$

In general, $f \circ g \neq g \circ f$.

One-to-One Function

A function f is called a **one-to-one function** if it never takes on the same value twice; that is,

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

Horizontal Line Test

A function is one-to-one if and only if no horizontal line intersects its graph more than once.

Justification

Suppose there is a horizontal line intersecting the graph of f more than once. Then, there exist x_1 and x_2, such that $f(x_1) = f(x_2)$. This contradicts to the hypothesis that the function f is one-to-one.

Linear Functions

The formula $y = mx + b$ defines a **linear function**, that is, the graph of the function is a line. Here m is the slope of the line and b is the y-intercept.

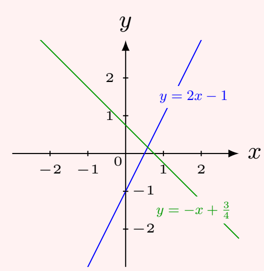
Polynomials

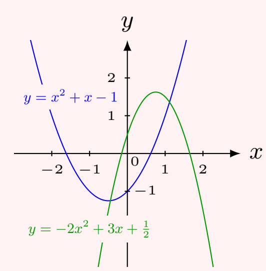
A function P is called a **polynomial** if

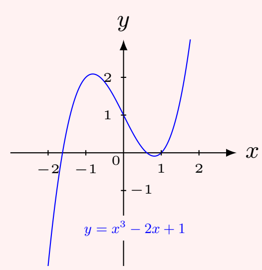
$$P(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0$$

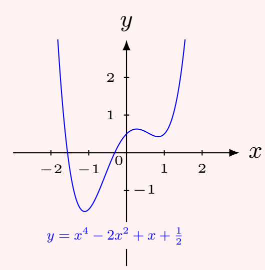
where n is a nonnegative integer and the numbers $a_0, a_1, a_2, \ldots, a_n$ are constants called the **coefficients** of the polynomial. The domain of any polynomial is $\mathbb{R} = (-\infty, \infty)$. If the leading coefficient $a_n \neq 0$, then the **degree** of the polynomial is n. In, particular,

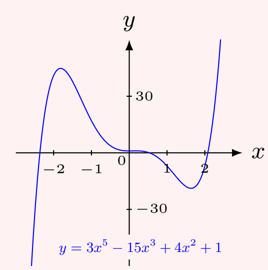
linear function:	$P(x) = ax + b, \quad a \neq 0;$
quadratic function:	$P(x) = ax^2 + bx + c, \quad a \neq 0;$
cubic function:	$P(x) = ax^3 + bx^2 + cx + d, \quad a \neq 0.$











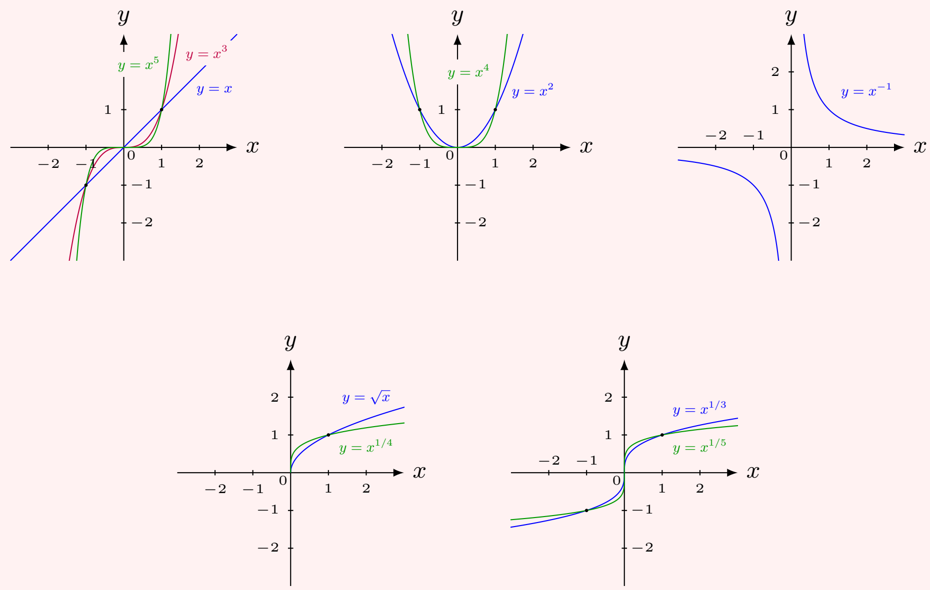
Power Functions

A function of the form $f(x) = x^a$, where a is a constant, is called a **power function**.

(i) When $a = n$, where n is a positive integer, the power function x^n is a special case of polynomials. Its domain is $\mathbb{R} = (-\infty, \infty)$.

(ii) When $a = 1/n$, where n is a positive integer, the power function $x^{1/n} = \sqrt[n]{x}$ is a **root function**. Its domain is $[0, \infty)$ if n is even; is $\mathbb{R} = (-\infty, \infty)$ if n is odd.

(iii) When $a = -1$, the power function $x^{-1} = 1/x$ is the **reciprocal function**, whose domain is $\mathbb{R} - \{0\} = (-\infty, 0) \cup (0, \infty)$.



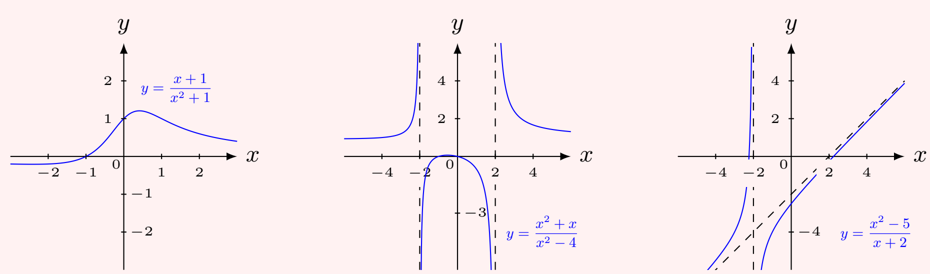
Rational Functions

A **rational function** f is a ratio of two polynomials:

$$f(x) = \frac{P(x)}{Q(x)}$$

where P and Q are polynomials. The domain consists of all values of x such that $Q(x) \neq 0$, that is,

$$\left\{ x \in \mathbb{R} \,\middle|\, Q(x) \neq 0 \right\}.$$



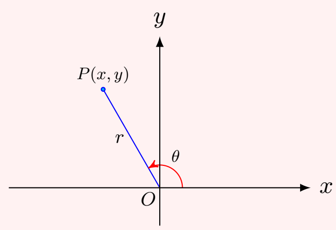
Algebraic Functions

A function f is called an **algebraic function** if it can be constructed using algebraic operations (such as addition, subtraction, multiplication, division, and taking roots) starting with polynomials.

Trigonometric Functions

For an angle θ, with $\theta \in [0, 2\pi)$, if we let $P(x, y)$ be the point on the terminal side of θ and if let r be the distance $|OP|$, the values of the sine and the cosine are defined as

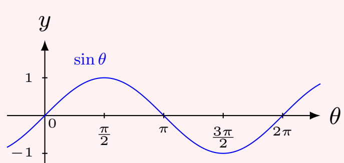
$$\sin\theta = \frac{y}{r}, \qquad \cos\theta = \frac{x}{r}.$$



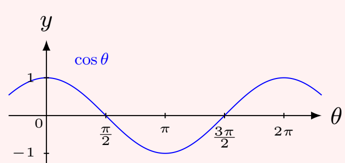
For θ outside $[0, 2\pi)$, the values of the sine and the cosine are defined in terms of the following periodic property:

$$\sin(\theta + 2\pi) = \sin\theta, \qquad \cos(\theta + 2\pi) = \cos\theta.$$

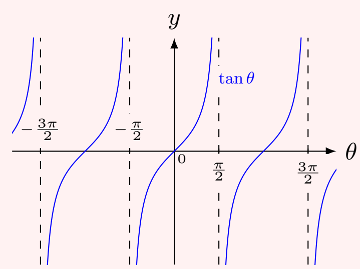
The tangent function is defined by $\tan\theta = \dfrac{\sin\theta}{\cos\theta}$. The cosecant, secant, and cotangent function are the reciprocals of the sine, cosine, and tangent functions.



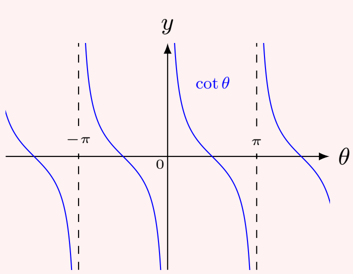
domain: $(-\infty, \infty)$
range: $[-1, 1]$



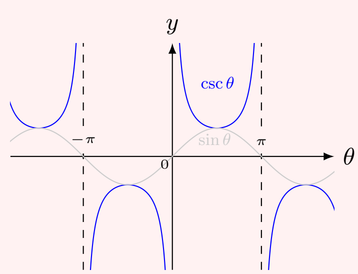
domain: $(-\infty, \infty)$
range: $[-1, 1]$



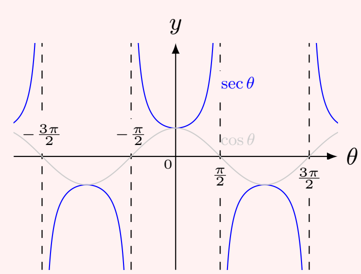
domain: $(-\infty, \infty) - \{\pm\frac{\pi}{2}, \pm\frac{3\pi}{2}, \dots\}$
range: $(-\infty, \infty)$



domain: $(-\infty, \infty) - \{\pm\pi, \pm 3\pi, \dots\}$
range: $(-\infty, \infty)$



domain: $(-\infty, \infty) - \{\pm\pi, \pm 3\pi, \dots\}$
range: $(-\infty, -1] \cup [1, \infty)$



domain: $(-\infty, \infty) - \{\pm\frac{\pi}{2}, \pm\frac{3\pi}{2}, \dots\}$
range: $(-\infty, -1] \cup [1, \infty)$

Inverse Function

Let f be a one-to-one function with domain A and range B. Then its **inverse function** f^{-1} has domain B and range A and is defined by

$$f^{-1}(y) = x \iff f(x) = y$$

for any y in B. Thus,

$$\begin{aligned} \text{domain of } f^{-1} &= \text{ range of } f, \\ \text{range of } f^{-1} &= \text{ domain of } f. \end{aligned}$$

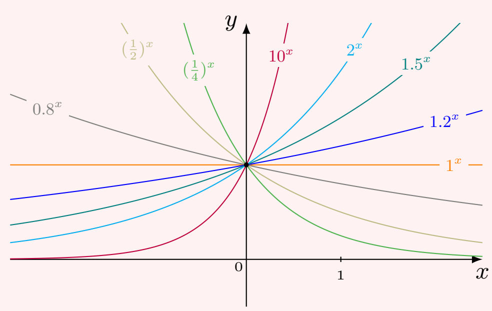
Cancellation Equations

The function f and its inverse function f^{-1} satisfy the following **cancellation equations**:

$$\begin{aligned} f^{-1}(f(x)) = x \quad & \text{for every } x \text{ in } A; \\ f\left(f^{-1}(x)\right) = x \quad & \text{for every } x \text{ in } B. \end{aligned}$$

Exponential Functions

An **exponential function** is a function of the following form $f(x) = a^x$, where x is a real variable, and $a > 0$ is a constant called the base of the function. It has domain $\mathbb{R}$ and range $(0, \infty)$.
The graph of the exponential function $f(x) = a^x$ depends the value of a.



Laws of Exponents

If a and b are positive numbers and x and y are any real numbers, then

1. $a^{x+y} = a^x a^y$ **2.** $a^{x-y} = \dfrac{a^x}{a^y}$ **3.** $(a^x)^y = a^{xy}$ **4.** $(ab)^x = a^x b^x$

How to Find the Inverse Function of a One-to-One Function f

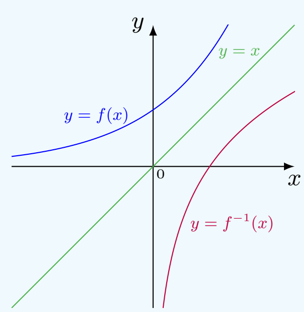
Step 1 Write $y = f(x)$.

Step 2 Solve this equation for x in terms of y (if possible).

Step 3 To express f^{-1} as a function of x, interchange x and y. The resulting equation is $y = f^{-1}(x)$.

Method for Obtaining the Graph of f^{-1} from the Graph of f

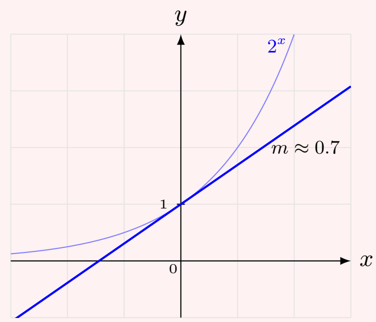
If f is a one-to-one function, then $f(a) = b$ if and only if $f^{-1}(b) = a$. So, the point (a, b) is on the graph of f if and only if the point (b, a) is on the graph of f^{-1}. Hence, the graph of f^{-1} is the reflection of the graph of f about the line $y = x$.

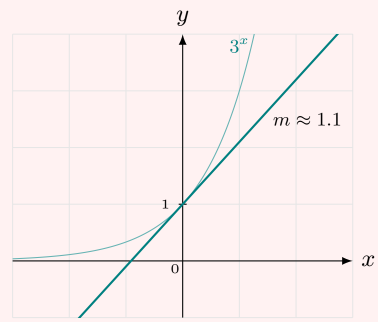


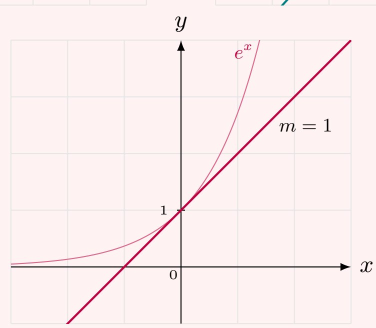
The Number e

As shown in the figures, the slopes of the tangent lines at $(0, 1)$ are $m \approx 0.7$ for $y = 2^x$ and $m \approx 1.1$ for $y = 3^x$. Thus, it is plausible that there is a constant, denoted by e, such that the slope of the tangent line at $(0, 1)$ is $m = 1$ for $y = e^x$. Further study reveals that the value of e, correct to five decimal places, is

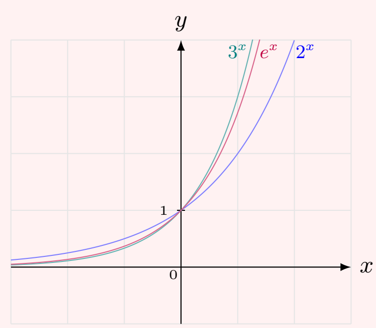
$$e \approx 2.71828.$$







The function $f(x) = e^x$ is the **natural exponential function**.

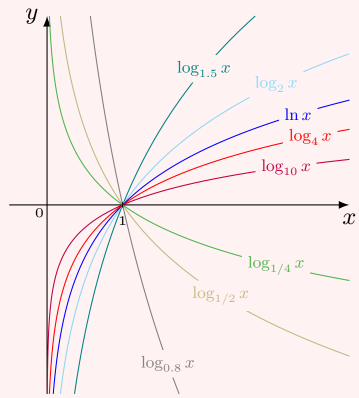


Logarithmic Functions

If $a > 0$ and $a \neq 1$, the exponential function $f(x) = a^x$ is either increasing or decreasing and so it is one-to-one by the Horizontal Line Test. It therefore has an inverse function f^{-1}, which is called the **logarithmic function** with base a and is denoted by $\log_a$. So,

$$\log_a x = y \iff a^y = x.$$

In particular, if $a = e$, the logarithmic function with base e is the natural logarithm, denoted as $\ln x$; if $a = 10$, the logarithmic function with base 10 is the common logarithm, denoted as $\log x$.



Properties of Logarithm

Laws of Logarithms

If x and y are positive numbers, then

1. $\log_a(xy) = \log_a x + \log_a y$.

2. $\log_a\left(\dfrac{x}{y}\right) = \log_a x - \log_a y$.

3. $\log_a(x^r) = r\log_a x$, where r is any real number.

Cancellation Equations for Exponents and Logarithms

$$\begin{aligned} \log_a(a^x) &= x \quad \text{for every } x \in \mathbb{R}; \\ a^{\log_a x} &= x \quad \text{for every } x \in (0, \infty). \end{aligned}$$

Change of Base Formula

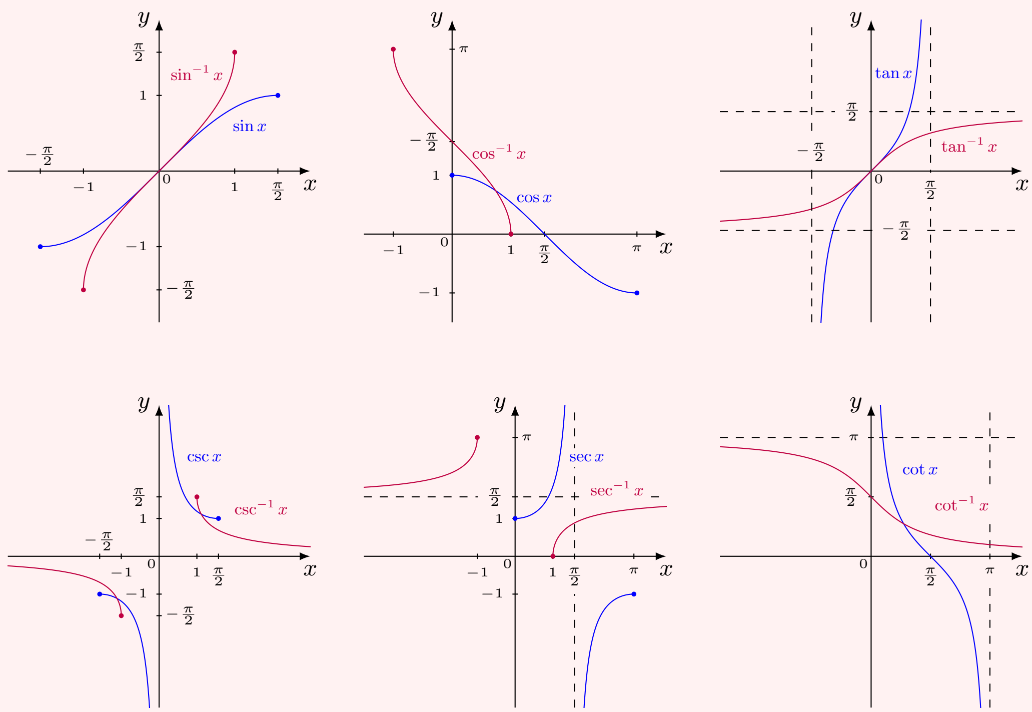
For any positive number a ($a \neq 1$), we have

$$\log_a x = \frac{\log_b x}{\log_b a},$$

where b is any positive number with $b \neq 1$.

Inverse Trigonometric Functions

$$
\begin{aligned}
\sin^{-1} x = y \ \ (|x| \le 1) \quad &\Longleftrightarrow \quad \sin y = x \text{ and } -\tfrac{\pi}{2} \le y \le \tfrac{\pi}{2} \\
\cos^{-1} x = y \ \ (|x| \le 1) \quad &\Longleftrightarrow \quad \cos y = x \text{ and } 0 \le y \le \pi \\
\tan^{-1} x = y \ \ (x \in \mathbb{R}) \quad &\Longleftrightarrow \quad \tan y = x \text{ and } -\tfrac{\pi}{2} < y < \tfrac{\pi}{2} \\
\csc^{-1} x = y \ \ (|x| \ge 1) \quad &\Longleftrightarrow \quad \csc y = x \text{ and } y \in [-\pi/2, 0) \cup (0, \pi/2] \\
\sec^{-1} x = y \ \ (|x| \ge 1) \quad &\Longleftrightarrow \quad \sec y = x \text{ and } y \in [0, \pi/2) \cup (\pi/2, \pi] \\
\cot^{-1} x = y \ \ (x \in \mathbb{R}) \quad &\Longleftrightarrow \quad \cot y = x \text{ and } y \in (0, \pi)
\end{aligned}
$$



Solving Inequalities by Using the Number Line

The method works for solving inequalities of the form $f(x) > 0$, or inequalities involving $\ge$, $<$, or $\le$, where f is a polynomial. It consists of the following major steps:
(1) Solve the equation $f(x) = 0$ to find all real roots.
(2) Specify the roots on the number line.
(3) Sketch the graph of $y = f(x)$ by starting from x near $-\infty$. If the order of f is odd, the graph of f is below the number line; if the order of f is even, the graph is above the number line.
(4) From the left to the right, the curve $y = f(x)$ intersects the number line at the roots, one after another.

(5) The curve $y = f(x)$ crosses the number line at roots when the multiplicities are odd and does not cross at roots when the multiplicities are even.
(6) The inequality $f(x) > 0$ holds on the interval over which the graph of $y = f(x)$ is above the number line.
The method can be modified for inequalities with f being rational functions or even transcendental functions.

Chapter 2

Limits and Derivatives

Secant line

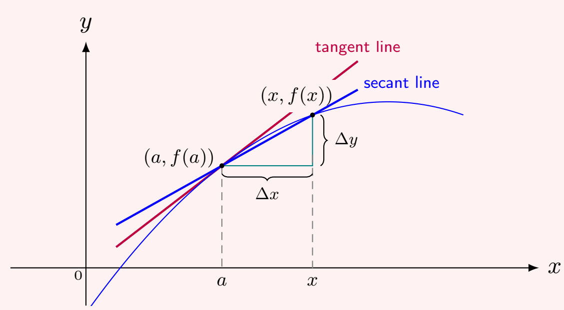
A *secant line* of a curve is a line that (locally) intersects two points on the curve. The *slope* of the secant line connecting $(x, f(x))$ and $(a, f(a))$ on a curve given by a function $y = f(x)$ is

$$A(x) = \frac{\Delta y}{\Delta x} = \frac{f(x) - f(a)}{x - a},$$

which describes the *average rate of change* of f from x to a. As the point x approaches a, the ratio (difference quotient) gives the *instantaneous slope* of the tangent line to $f(x)$ at point a.
If f represents a position function of a moving object (in this case, usually we use s to denote the function) while the independent variable is the time t, then the above difference quotient gives the average velocity:

$$\text{average velocity} = \frac{\text{change in position}}{\text{time elapsed}} = \frac{s(t) - s(a)}{t - a}.$$

The *instantaneous velocity* when $t = a$ is defined to be the limiting value of these average velocities over shorter and shorter time periods that start at $t = a$.



Limit of a Function at a (non-rigorous)

Suppose $f(x)$ is defined when x is near the number a. (This means that f is defined on some open interval that contains a, except possibly at a itself.) Then we write

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

and say "the limit of $f(x)$, as x approaches a, equals L", if we can make the values of $f(x)$ arbitrarily close to L (as close to L as we like) by taking x to be sufficiently close to a (on either side of a) but not equal to a.

The phrase "but not equal to a" means that in finding the limit of $f(x)$ as x approaches a, we never consider $x = a$. In fact, $f(x)$ need not even be defined when $x = a$. What matters is how f is defined *near* a.

Limit of a Function at a (rigorous)

Let f be a function defined on some open interval that contains the number a, except possibly at a itself. Then we say that the **limit of** $f(x)$ **as** x **approaches** a **is** L, and we write

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

if for every number $\varepsilon > 0$ there is a number $\delta > 0$ such that

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

One-Sided Limit

We write

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

and say the **left-hand limit of** $f(x)$ **as** x **approaches** a or the limit of $f(x)$ as x approaches a from the left is equal to L if we can make the values of $f(x)$ arbitrarily close to L by taking x to be sufficiently close to a and x less than a.

Left-Hand Limit

If for every number $\varepsilon > 0$ there is a number $\delta > 0$ such that

$$\text{if } a - \delta < x < a \quad \text{then} \quad |f(x) - L| < \varepsilon,$$

then we write

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

Similarly, we write

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

and say the **right-hand limit of $f(x)$ as x approaches** a or the limit of $f(x)$ as x approaches a from the right is equal to L if we can make the values of $f(x)$ arbitrarily close to L by taking x to be sufficiently close to a and x greater than a.

Right-Hand Limit

If for every number $\varepsilon > 0$ there is a number $\delta > 0$ such that

$$\text{if } a < x < a + \delta \quad \text{then} \quad |f(x) - L| < \varepsilon,$$

then we write

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

Left-Sided and Right-Sided Limits

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

Justification

Let $\varepsilon > 0$ be given.
If $\lim_{x \to a^-} f(x) = \lim_{x \to a^+} f(x) = L$, by the definition, there are $\delta_1, \delta_2 > 0$, such that

$$\begin{aligned} &\text{if } a - \delta_1 < x < a \quad \text{then} \quad |f(x) - L| < \varepsilon; \\ &\text{if } a < x < a + \delta_2 \quad \text{then} \quad |f(x) - L| < \varepsilon. \end{aligned}$$

Denote $\delta = \min\{\delta_1, \delta_2\}$. Then $\delta > 0$. It is clear that

$$0 < |x - a| < \delta \iff \text{either } a - \delta < x < a \text{ or } a < x < a + \delta.$$

Since

$$\begin{aligned} a - \delta < x < a &\Longrightarrow a - \delta_1 < x < a; \\ a < x < a + \delta &\Longrightarrow a < x < a + \delta_2, \end{aligned}$$

we have

$$\begin{aligned} 0 < |x - a| < \delta &\Longrightarrow \text{either } a - \delta_1 < x < a \text{ or } a < x < a + \delta_2 \\ &\Longrightarrow |f(x) - L| < \varepsilon. \end{aligned}$$

Hence, $\lim_{x \to a} f(x) = L$.
Conversely, if $\lim_{x \to a} f(x) = L$, then there is $\delta > 0$, such that

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

Since

$$a - \delta < x < a \Longrightarrow 0 < |x - a| < \delta;$$
$$a < x < a + \delta \Longrightarrow 0 < |x - a| < \delta,$$

we have that

$$a - \delta < x < a \Longrightarrow |f(x) - L| < \varepsilon;$$
$$a < x < a + \delta \Longrightarrow |f(x) - L| < \varepsilon.$$

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

Infinite Limits (non-rigorous)

Let f be a function defined on both sides of a, except possibly at a itself. Then

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

means that the values of $f(x)$ can be made arbitrarily large (as large as we please) by taking x sufficiently close to a, but not equal to a.
Similarly,

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

means that the values of $f(x)$ can be made arbitrarily large negative by taking x sufficiently close to a, but not equal to a.

Infinite Limits (rigorous)

Let f be a function defined on some open interval that contains the number a, except possibly at a itself. Then

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

means that for every positive number M there is a positive number δ such that

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

Similarly,

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

means that for every negative number N there is a positive number δ such that

$$\text{if } 0 < |x - a| < \delta \quad \text{then} \quad f(x) < N.$$

Limit Laws

Suppose $\lim_{x\to a} f(x)$ and $\lim_{x\to a} g(x)$ exist. Then

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

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

3. $\lim_{x\to a} [cf(x)] = c \lim_{x\to a} f(x)$, where c is any constant.

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

5. $\lim_{x\to a} \dfrac{f(x)}{g(x)} = \dfrac{\lim_{x\to a} f(x)}{\lim_{x\to a} g(x)}$ if $\lim_{x\to a} g(x) \neq 0$.

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

7. $\lim_{x\to a} c = c$, where c is any constant.

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

9. $\lim_{x\to a} x^n = a^n$, where n is a positive integer.

10. $\lim_{x\to a} \sqrt[n]{x} = \sqrt[n]{a}$, where n is a positive integer. (If n is even, we assume that $a > 0$.)

11. $\lim_{x\to a} \sqrt[n]{f(x)} = \sqrt[n]{\lim_{x\to a} f(x)}$, where n is a positive integer. [If n is even, we assume that $\lim_{x\to a} f(x) > 0$.]

Order Rule

If $f(x) \leq g(x)$ when x is near a (except possibly at a) and the limits of f and g both exist as x approaches a, then

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

Justification

Suppose $\lim_{x\to a} f(x) = L_1$ and $\lim_{x\to a} g(x) = L_2$. If $L_1 > L_2$, we let $\varepsilon = \frac{1}{2}(L_1 - L_2)$. Then $\varepsilon > 0$. By the definition, there are $\delta_1, \delta_2 > 0$, such that

$$\begin{aligned} &\text{if } 0 < |x-a| < \delta_1 \quad \text{then} \quad |f(x) - L_1| < \varepsilon; \\ &\text{if } 0 < |x-a| < \delta_2 \quad \text{then} \quad |g(x) - L_2| < \varepsilon. \end{aligned}$$

Denote $\delta = \min\{\delta_1, \delta_2\}$. Then $\delta > 0$. Hence,

$$\begin{aligned} 0 < |x-a| < \delta &\Longrightarrow 0 < |x-a| < \delta_1 \text{ and } 0 < |x-a| < \delta_2 \\ &\Longrightarrow |f(x) - L_1| < \varepsilon \text{ and } |g(x) - L_2| < \varepsilon \\ &\Longrightarrow -\tfrac{1}{2}(L_1 - L_2) < f(x) - L_1 < \tfrac{1}{2}(L_1 - L_2) \text{ and} \\ &\qquad -\tfrac{1}{2}(L_1 - L_2) < g(x) - L_2 < \tfrac{1}{2}(L_1 - L_2) \\ &\Longrightarrow f(x) > \tfrac{1}{2}(L_1 + L_2) \text{ and } g(x) < \tfrac{1}{2}(L_1 + L_2), \end{aligned}$$

so that if $0 < |x-a| < \delta$, then $f(x) > \frac{1}{2}(L_1 + L_2) > g(x)$. This contradicts to the hypothesis that $f(x) \leq g(x)$ when x is near a. Therefore, we must have $L_1 \leq L_2$, or

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

The Squeeze Theorem

If $f(x) \leq g(x) \leq h(x)$ when x is near a (except possibly at a) and

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

then

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

Justification

Let $\varepsilon > 0$ be given.
Since $\lim\limits_{x \to a} f(x) = \lim\limits_{x \to a} h(x) = L$, by the definition, there are $\delta_1, \delta_2 > 0$, such that

$$\begin{aligned} &\text{if } 0 < |x-a| < \delta_1 \quad \text{then} \quad |f(x) - L| < \varepsilon; \\ &\text{if } 0 < |x-a| < \delta_2 \quad \text{then} \quad |h(x) - L| < \varepsilon. \end{aligned}$$

Denote $\delta = \min\{\delta_1, \delta_2\}$. Then $\delta > 0$. Hence,

$$\begin{aligned} 0 < |x-a| < \delta &\Longrightarrow 0 < |x-a| < \delta_1 \text{ and } 0 < |x-a| < \delta_2 \\ &\Longrightarrow |f(x) - L| < \varepsilon \text{ and } |h(x) - L| < \varepsilon \\ &\Longrightarrow -\varepsilon < f(x) - L < \varepsilon \text{ and } -\varepsilon < h(x) - L < \varepsilon \\ &\Longrightarrow L - \varepsilon < f(x) \text{ and } h(x) < L + \varepsilon. \end{aligned}$$

Therefore, if $f(x) \leq g(x) \leq h(x)$ when x is near a (except possibly at a), then for $0 < |x-a| < \delta$,

$$L - \varepsilon < f(x) \leq g(x) \leq h(x) < L + \varepsilon,$$

so that $|g(x) - L| < \varepsilon$. By the definition, we have $\lim\limits_{x \to a} g(x) = L$.

Continuity of Function

Function Continuous at a Number a

A function f is **continuous at a number** a if

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

We say that f is **discontinuous at** a (or f has a **discontinuity** at a) if f is not continuous at a.

Function Continuous from One Side

A function f is **continuous from the right at a number** a if

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

and f is **continuous from the left at a number** a if

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

Function Continuous on an Interval

A function f is **continuous on an interval** if it is continuous at every number in the interval. (If f is defined only on one side of an endpoint of the interval, we understand *continuous* at the endpoint to mean *continuous from the right* or *continuous from the left*.)

Properties of Continuous Functions

Arithmetic Operations of Continuous Functions

If f and g are continuous at a and c is a constant, then the following functions are also continuous at a:

1. $f + g$ **2.** $f - g$ **3.** cf **4.** fg **5.** $\dfrac{f}{g}$ if $g(a) \neq 0$

Continuity of Elementary Functions

The following types of functions are continuous at every number in their domains:

polynomials   rational functions   root functions
trigonometric functions   inverse trigonometric functions
exponential functions   logarithmic functions

Limit of Composite Function

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,

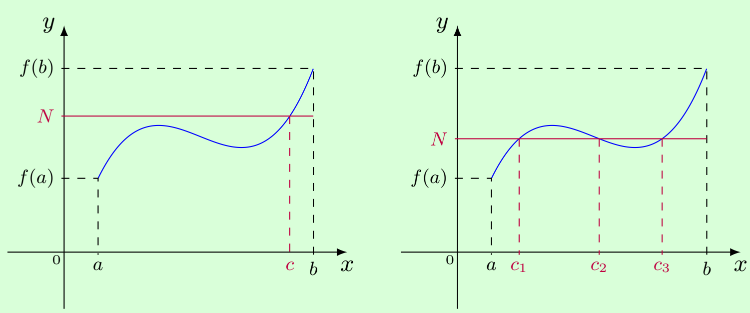
$$\lim_{x\to a} f(g(x)) = f\left(\lim_{x\to a} g(x)\right).$$

Continuity of Composite Function

If g is continuous at a and f is continuous at $g(a)$, then the composite function $f \circ g$ is continuous at a.

The Intermediate Value Theorem

Suppose that f is continuous on the closed interval $[a, b]$ and let N 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 $f(c) = N$.



Limits at Infinity (non-rigorous)

Let f be a function defined on some interval (a, ∞). Then

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

means that the values of $f(x)$ can be made arbitrarily close to L by taking x sufficiently large. Similarly, if f is a function defined on some interval $(-\infty, a)$. Then

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

means that the values of $f(x)$ can be made arbitrarily close to L by taking x sufficiently large negative.

Limits at Infinity (rigorous)

Suppose f is a function defined on some interval (a, ∞). Then

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

means that for every number $\varepsilon > 0$ there is a number M such that

$$\text{if } x > M \quad \text{then} \quad |f(x) - L| < \varepsilon.$$

Similarly, suppose f is a function defined on some interval $(-\infty, a)$. Then

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

means that for every number $\varepsilon > 0$ there is a number N such that

$$\text{if } x < N \quad \text{then} \quad |f(x) - L| < \varepsilon.$$

The Squeeze Theorem for Limits at Infinity

If $f(x) \leq g(x) \leq h(x)$ when $x \geq M$ for some constant M and

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

then

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

Limit Laws at Infinity

Suppose $\lim_{x \to \infty} f(x)$ and $\lim_{x \to \infty} g(x)$ exist. Then

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

2. $\lim_{x \to \infty} [f(x) - g(x)] = \lim_{x \to \infty} f(x) - \lim_{x \to \infty} g(x)$.

3. $\lim_{x \to \infty} [cf(x)] = c \lim_{x \to \infty} f(x)$, where c is any constant.

4. $\lim_{x\to\infty} [f(x)\, g(x)] = \lim_{x\to\infty} f(x) \cdot \lim_{x\to\infty} g(x)$.

5. $\lim_{x\to\infty} \dfrac{f(x)}{g(x)} = \dfrac{\lim_{x\to\infty} f(x)}{\lim_{x\to\infty} g(x)}$ if $\lim_{x\to\infty} g(x) \neq 0$.

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

7. $\lim_{x\to\infty} c = c$, where c is any constant.

8. $\lim_{x\to\infty} x = \infty$.

9. $\lim_{x\to\infty} x^n = \infty$, where n is a positive integer.

10. $\lim_{x\to\infty} \sqrt[n]{x} = \infty$, where n is a positive integer.

11. $\lim_{x\to\infty} \sqrt[n]{f(x)} = \sqrt[n]{\lim_{x\to\infty} f(x)}$, where n is a positive integer. [If n is even, we assume that $\lim_{x\to\infty} f(x) > 0$.]

Similar limits are also valid if we replace $x \to \infty$ by $x \to -\infty$.

Limit at Infinity of a Power Function

If $r > 0$ is a rational number, then

$$\lim_{x\to\infty} \frac{1}{x^r} = 0.$$

If $r > 0$ is a rational number such that x^r is defined for all x, then

$$\lim_{x\to-\infty} \frac{1}{x^r} = 0.$$

Asymptotes

Vertical Asymptote

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

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

Horizontal Asymptote

The line $y = b$ is called a **horizontal asymptote** of the curve $y = f(x)$ if either

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

or

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

Infinite Limit at Infinity

- $\lim_{x\to\infty} f(x) = \infty$ means that for every number $M > 0$ there is a number N such that

$$\text{if } x > N \quad \text{then} \quad f(x) > M.$$

- $\lim_{x\to\infty} f(x) = -\infty$ means that for every number $M > 0$ there is a number N such that

$$\text{if } x > N \quad \text{then} \quad f(x) < -M.$$

- $\lim_{x\to-\infty} f(x) = \infty$ means that for every number $M > 0$ there is a number N such that

$$\text{if } x < N \quad \text{then} \quad f(x) > M.$$

- $\lim_{x\to-\infty} f(x) = -\infty$ means that for every number $M > 0$ there is a number N such that

$$\text{if } x < N \quad \text{then} \quad f(x) < -M.$$

Tangent Line

The **tangent line** to the curve $y = f(x)$ at the point $P(a, f(a))$ is the line through P with slope

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

provided that this limit exists. In other words, the tangent line to $y = f(x)$ at the point $P(a, f(a))$ is the line through $(a, f(a))$ whose slope is equal to $f'(a)$, the derivative of f at a.

Normal Line

The normal line to a curve C at a point P is the line through P that is perpendicular to the tangent line at P.

Derivative at a Number

The **derivative of a function f at a number a**, denoted by $f'(a)$, is

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

provided that this limit, or equivalently,

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

exists.

Derivative as a Function

A function f is **differentiable** at a if $f'(a)$ exists. The derivative of f, denoted as f', is the function whose value at x is the limit

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

provided that this limit exists.
If f is differentiable at every number in a set $E \subset \mathbb{R}$, then we say that f is **differentiable on** E.

Relationship Between Continuous Functions and Differential Functions

If f is differentiable at a, then f is continuous at a. The converse is false; that is, there are functions that are continuous but not differentiable.

Justification

If f is differentiable at a, then $\lim\limits_{x\to a} \dfrac{f(x)-f(a)}{x-a} = f'(a)$ exists. Thus,

$$\lim_{x\to a} [f(x)-f(a)] = \lim_{x\to a} \frac{f(x)-f(a)}{x-a} \cdot (x-a) = \lim_{x\to a} \frac{f(x)-f(a)}{x-a} \cdot \lim_{x\to a} (x-a) = f'(a)\cdot 0 = 0.$$

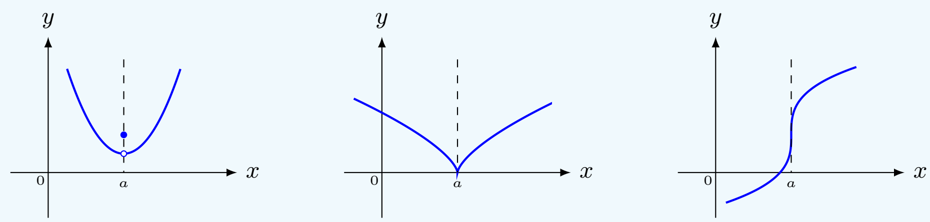
Hence $\lim\limits_{x\to a} f(x) = f(a)$, so that f is continuous at a.
Conversely, there are functions that are continuous but not differentiable. Indeed, consider the function $f(x) = |x|$. It is continuous everywhere. However, since

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

we know that $\lim\limits_{h\to 0} \dfrac{f(h)-f(0)}{h}$ does not exist. So, f is not differentiable at $x = 0$.

Examples of Non-Differentiable Functions

A non-differentiable situation can occur when a function has either a discontinuity, a cusp, or a vertical tangent line, as illustrated in the following figures.



Rate of Change

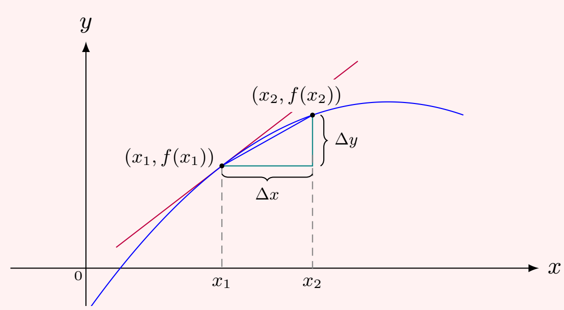
Suppose $y = f(x)$ is a quantity that depends on another quantity x. If x changes from x_1 to x_2, the difference quotient

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

is called the **average rate of change of** y **with respect to** x over the interval $[x_1, x_2]$. The limit of these average rates of change is called the **(instantaneous) rate of change of** y **with respect to** x at x_1:

$$\text{instantaneous rate of change} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}.$$

That is, the derivative $f'(x_1)$ is the instantaneous rate of change of $y = f(x)$ with respect to x when $x = x_1$.



In particular, if $s = f(t)$ is the position function of a particle that moves along a straight line, then $f'(a)$ is the rate of change of the displacement s with respect to the time t. In other words, $f'(a)$ is the velocity of the particle at time $t = a$. The speed of the particle is the absolute value of the velocity, that is, $|f'(a)|$.

Higher Derivatives

The **second derivative** of f is the derivative of the derivative of f:

$$f''(x) = \left(f'\right)'(x).$$

The **third derivative** of f is the derivative of the second derivative of f:

$$f'''(x) = \left(f''\right)'(x).$$

In general, the **n-th derivative** of f is the derivative of the $(n-1)$th derivative of f:

$$f^{(n)}(x) = \left(f^{(n-1)}\right)'(x).$$

Chapter 3

Differentiation Rules

Derivatives of Elementary Functions

1. $\dfrac{d}{dx}(c) = 0.$

2. $\dfrac{d}{dx}(x^r) = rx^{r-1}$, where r is any real number.

3. $\dfrac{d}{dx}(e^x) = e^x$, where

e is the number such that $\lim_{h\to 0} \dfrac{e^h - 1}{h} = 1$. Numerically, $e \approx 2.71828$.

4. $\dfrac{d}{dx}(a^x) = a^x \ln a$, where

$\ln a = \log_e a$ is the natural logarithm.

5. $\dfrac{d}{dx}(\ln |x|) = \dfrac{1}{x}.$

6. $\dfrac{d}{dx}(\log_a |x|) = \dfrac{1}{x \ln a}.$

Justification - Derivative of a Constant Function

By the definition,

$$\frac{d}{dx}(c) = \lim_{h\to 0} \frac{c-c}{h} = \lim_{h\to 0} 0 = 0.$$

Justification - Derivative of Power Functions

If n is an integer, then, for $x \neq a$, by summing the geometric series, we get

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

Thus,

$$\begin{aligned}\frac{d}{dx}(x^n) &= \lim_{h\to 0} \frac{(x+h)^n - x^n}{h} \\ &= \lim_{h\to 0} \left[(x+h)^{n-1} + (x+h)^{n-2}x + \cdots + (x+h)x^{n-2} + x^{n-1}\right] \\ &= \underbrace{x^{n-1} + x^{n-1} + \cdots + x^{n-1} + x^{n-1}}_{n \text{ terms}} = nx^{n-1}.\end{aligned}$$

For general real r, write $y = x^r$, so we have $\ln|y| = r\ln|x|$. Differentiating the equality (by applying the Chain Rule) gives

$$\frac{y'}{y} = r \cdot \frac{1}{x},$$

so that

$$\frac{d}{dx}(x^r) = y' = r \cdot \frac{y}{x} = rx^{r-1}.$$

Justification - Derivative of the Exponential Function e^x

By the definition,

$$\frac{d}{dx}(e^x) = \lim_{h\to 0} \frac{e^{x+h} - e^x}{h} = \lim_{h\to 0} \left[e^x \cdot \frac{e^h - 1}{h}\right] = e^x \cdot \lim_{h\to 0} \frac{e^h - 1}{h} = e^x \cdot 1 = e^x.$$

Here we have used the definition of the number e: e is the number such that $\lim_{h\to 0} \frac{e^h - 1}{h} = 1$.

Justification - Derivative of the General Exponential Function a^x

By the definition,

$$\frac{d}{dx}(a^x) = \lim_{h\to 0} \frac{a^{x+h} - a^x}{h} = \lim_{h\to 0} \left[a^x \cdot \frac{a^h - 1}{h}\right] = a^x \cdot \lim_{h\to 0} \left[\frac{e^{h\ln a} - 1}{h\ln a} \cdot \ln a\right] = a^x \ln a \cdot \lim_{h\to 0} \frac{e^{h\ln a} - 1}{h\ln a}.$$

By making $h\ln a = t$, we have

$$\lim_{h\to 0} \frac{e^{h\ln a} - 1}{h\ln a} = \lim_{t\to 0} \frac{e^t - 1}{t} = 1.$$

Here we have used the definition of the number e: e is the number such that $\lim_{h\to 0} \frac{e^h - 1}{h} = 1$. Hence,

$$\frac{d}{dx}(a^x) = a^x \ln a \cdot 1 = a^x \ln a.$$

Justification - Derivative of the Logarithmic Function $\ln|x|$

For $x > 0$, let $y = \ln x$. Then $e^y = x$. Differentiating the equation implicitly with respect to x gives

$$e^y \cdot y' = 1,$$

so that

$$\frac{d}{dx}(\ln x) = y' = 1/e^y = 1/x.$$

For any $x \neq 0$, let $y = \ln|x|$. Then

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

Thus,

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

Here we have applied the Chain Rule in the case of $x < 0$. Hence,

$$\frac{d}{dx}(\ln|x|) = 1/x.$$

Justification - Derivative of the General Logarithmic Function $\log_a|x|$

Since

$$\log_a|x| = \frac{\ln|x|}{\ln a},$$

by using the formula $(\ln|x|)' = 1/x$, we have

$$\frac{d}{dx}(\log_a|x|) = \left[\frac{\ln|x|}{\ln a}\right]' = \frac{1}{x \ln a}.$$

Differential Rules

Suppose f and g are differentiable and c is a constant.

1. **The Constant Multiple Rule**

$$\frac{d}{dx}\left[cf(x)\right] = c\frac{d}{dx}\left[f(x)\right].$$

2. The Sum Rule

$$\frac{d}{dx}[f(x)+g(x)] = \frac{d}{dx}[f(x)] + \frac{d}{dx}[g(x)].$$

3. The Difference Rule

$$\frac{d}{dx}[f(x)-g(x)] = \frac{d}{dx}[f(x)] - \frac{d}{dx}[g(x)].$$

4. The Product Rule

$$\frac{d}{dx}[f(x)\,g(x)] = f(x)\,\frac{d}{dx}[g(x)] + g(x)\,\frac{d}{dx}[f(x)].$$

5. The Quotient Rule

$$\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] = \frac{g(x)\,\frac{d}{dx}[f(x)] - f(x)\,\frac{d}{dx}[g(x)]}{[g(x)]^2}.$$

Justification - the Constant Multiple Rule

By the definition,

$$\frac{d}{dx}[cf(x)] = \lim_{h\to 0}\frac{cf(x+h)-cf(x)}{h} = c\cdot\lim_{h\to 0}\frac{f(x+h)-f(x)}{h} = cf'(x) = c\frac{d}{dx}[f(x)].$$

Justification - the Sum/Difference Rule

By the definition,

$$\begin{aligned}
\frac{d}{dx}[f(x)+g(x)] &= \lim_{h\to 0}\frac{[f(x+h)+g(x+h)]-[f(x)+g(x)]}{h}\\
&= \lim_{h\to 0}\frac{[f(x+h)-f(x)]+[g(x+h)-g(x)]}{h}\\
&= \lim_{h\to 0}\left[\frac{f(x+h)-f(x)}{h}+\frac{g(x+h)-g(x)}{h}\right]\\
&= \lim_{h\to 0}\frac{f(x+h)-f(x)}{h}+\lim_{h\to 0}\frac{g(x+h)-g(x)}{h}\\
&= \frac{d}{dx}[f(x)]+\frac{d}{dx}[g(x)].
\end{aligned}$$

Similarly,

$$\begin{aligned}\frac{d}{dx}\left[f(x)-g(x)\right] &= \lim_{h\to 0}\frac{[f(x+h)-g(x+h)]-[f(x)-g(x)]}{h}\\ &= \lim_{h\to 0}\frac{[f(x+h)-f(x)]-[g(x+h)-g(x)]}{h}\\ &= \lim_{h\to 0}\left[\frac{f(x+h)-f(x)}{h}-\frac{g(x+h)-g(x)}{h}\right]\\ &= \lim_{h\to 0}\frac{f(x+h)-f(x)}{h}-\lim_{h\to 0}\frac{g(x+h)-g(x)}{h}\\ &= \frac{d}{dx}\left[f(x)\right]-\frac{d}{dx}\left[g(x)\right].\end{aligned}$$

Justification - the Product Rule

By the definition,

$$\begin{aligned}\frac{d}{dx}\left[f(x)g(x)\right] &= \lim_{h\to 0}\frac{f(x+h)g(x+h)-f(x)g(x)}{h}\\ &= \lim_{h\to 0}\frac{f(x+h)g(x+h)-f(x+h)g(x)+f(x+h)g(x)-f(x)g(x)}{h}\\ &= \lim_{h\to 0}\left[f(x+h)\cdot\frac{g(x+h)-g(x)}{h}+\frac{f(x+h)-f(x)}{h}\cdot g(x)\right]\\ &= \lim_{h\to 0}f(x+h)\cdot\lim_{h\to 0}\frac{g(x+h)-g(x)}{h}+\lim_{h\to 0}\frac{f(x+h)-f(x)}{h}\cdot\lim_{h\to 0}g(x).\end{aligned}$$

If f is differentiable, f is continuous, so that

$$\lim_{h\to 0}f(x+h)=f(x).$$

It is clear that $\lim\limits_{h\to 0}g(x)=g(x)$. Hence,

$$\frac{d}{dx}\left[f(x)g(x)\right]=f(x)\frac{d}{dx}\left[g(x)\right]+g(x)\frac{d}{dx}\left[f(x)\right].$$

Justification - the Quotient Rule

By the definition,

$$\begin{aligned}\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] &= \lim_{h\to 0}\frac{\frac{f(x+h)}{g(x+h)}-\frac{f(x)}{g(x)}}{h} = \lim_{h\to 0}\frac{f(x+h)g(x)-g(x+h)f(x)}{h\cdot g(x+h)g(x)}\\ &= \lim_{h\to 0}\frac{f(x+h)g(x)-f(x)g(x)+f(x)g(x)-g(x+h)f(x)}{h\cdot g(x+h)g(x)}\\ &= \lim_{h\to 0}\left[\frac{\frac{f(x+h)-f(x)}{h}\cdot g(x)-f(x)\cdot\frac{g(x+h)-g(x)}{h}}{g(x+h)g(x)}\right].\end{aligned}$$

If g is differentiable, g is continuous, so that

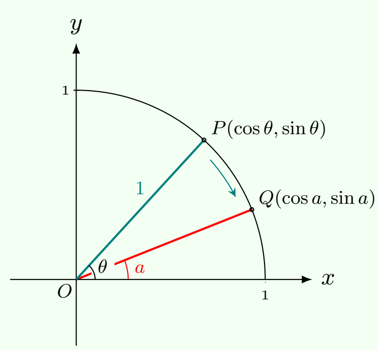
$$\lim_{h\to 0} g(x+h) = g(x).$$

Since $g(x)\neq 0$, we have $\lim_{h\to 0}\frac{1}{g(x+h)g(x)} = \frac{1}{[g(x)]^2}$. Hence,

$$\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] = \frac{g(x)\,\frac{d}{dx}\left[f(x)\right] - f(x)\,\frac{d}{dx}\left[g(x)\right]}{[g(x)]^2}.$$

Limits of the Sine and Cosine Functions - Continuity

$$\lim_{\theta\to a}\sin\theta = \sin a \quad\text{and}\quad \lim_{\theta\to a}\cos\theta = \cos a.$$



Limits Related to the Sine and Cosine Functions Near Zero

$$\lim_{\theta\to 0}\frac{\sin\theta}{\theta} = 1 \quad\text{and}\quad \lim_{\theta\to 0}\frac{\cos\theta - 1}{\theta} = 0.$$

Justification

The first limit can be shown by using the following diagram. For $0 < \theta < \frac{\pi}{2}$, we have

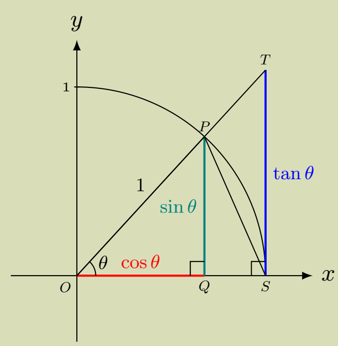
$$\text{Area } \triangle OSP < \text{Area sector } OSP < \text{Area } \triangle OST,$$

or

$$\tfrac{1}{2} \cdot \sin\theta \cdot < \pi 1^2 \left(\theta \cdot \tfrac{1}{2\pi}\right) < \tfrac{1}{2} \cdot 1 \cdot \tan\theta.$$

These inequalities give

$$\cos\theta < \frac{\sin\theta}{\theta} < 1.$$



Since $\lim_{\theta\to 0^+} \cos\theta = 1$, by the Squeeze Theorem, we obtain $\lim_{\theta\to 0^+} \dfrac{\sin\theta}{\theta} = 1$. Because $\sin\theta$ is an odd function, we further have

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

Hence, $\lim_{\theta\to 0} \dfrac{\sin\theta}{\theta} = 1$.
By using the half-angle formula $\cos h = 1 - 2\sin^2(h/2)$, we have

$$\lim_{\theta\to 0} \frac{\cos\theta - 1}{\theta} = \lim_{\theta\to 0} \frac{-2\sin^2(\theta/2)}{\theta} = \lim_{\theta\to 0} \frac{\sin^2(\theta/2)}{(\theta/2)^2} \cdot (-\theta/2) = 1^2 \cdot 0 = 0.$$

Derivatives of Trigonometric Functions

$$\frac{d}{dx}(\sin x) = \cos x \qquad \frac{d}{dx}(\csc x) = -\csc x\, \cot x$$
$$\frac{d}{dx}(\cos x) = -\sin x \qquad \frac{d}{dx}(\sec x) = \sec x\, \tan x$$
$$\frac{d}{dx}(\tan x) = \sec^2 x \qquad \frac{d}{dx}(\cot x) = -\csc^2 x$$

Chain Rule

If g is differentiable at x and f is differentiable at $g(x)$, then the composite function $F = f \circ g$ defined by $F(x) = f\left(g(x)\right)$ is differentiable at x and F' is given by the product

$$F'(x) = f'\left(g(x)\right) \cdot g'(x).$$

In Leibniz notation, if $y = f(u)$ and $u = g(x)$ are both differentiable functions, then

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}.$$

This is the so-called **Chain Rule**.

Justification

Suppose $u = g(x)$ is differentiable at a and $y = f(u)$ is differentiable at $b = g(a)$. If Δx is an increment in x and Δu and Δy are the corresponding increments in u and y, then, from the hypotheses,

$$\lim_{\Delta x \to 0} \frac{\Delta u}{\Delta x} = g'(a), \qquad \lim_{\Delta u \to 0} \frac{\Delta y}{\Delta u} = f'(b).$$

Thus,

$$\begin{aligned} \Delta u &= g'(a)\Delta x + \varepsilon_1 \Delta x, \quad \text{where } \varepsilon_1 \to 0 \text{ as } \Delta x \to 0; \\ \Delta y &= f'(b)\Delta u + \varepsilon_2 \Delta u, \quad \text{where } \varepsilon_2 \to 0 \text{ as } \Delta u \to 0. \end{aligned}$$

Hence,

$$\Delta y = [f'(b) + \varepsilon_2]\Delta u = [f'(b) + \varepsilon_2][g'(a) + \varepsilon_1]\Delta x,$$

so that

$$\frac{\Delta y}{\Delta x} = [f'(b) + \varepsilon_2][g'(a) + \varepsilon_1].$$

From above, we see that $\Delta u \to 0$ as $\Delta x \to 0$. So, both $\varepsilon_1 \to 0$ and $\varepsilon_2 \to 0$ as $\Delta x \to 0$. Therefore,

$$\begin{aligned} \frac{dy}{dx} &= \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} [f'(b) + \varepsilon_2][g'(a) + \varepsilon_1] \\ &= f'(b)g'(a) = f'(g(a))g'(a). \end{aligned}$$

This proves the Chain Rule.

Steps in Implicit Differentiation

Step 1 Replace y by $y(x)$ in the given equation which defines the function $y = y(x)$ implicitly.

Step 2 Differentiate both sides of the equation with respect to x, by applying the differential rules.

Step 3 Solve for $\dfrac{dy}{dx}$ from the resulting equation.

Derivatives of Inverse Trigonometric Functions

$$\frac{d}{dx}(\sin^{-1} x) = \frac{1}{\sqrt{1-x^2}} \qquad \frac{d}{dx}(\csc^{-1} x) = -\frac{1}{x\sqrt{x^2-1}}$$
$$\frac{d}{dx}(\cos^{-1} x) = -\frac{1}{\sqrt{1-x^2}} \qquad \frac{d}{dx}(\sec^{-1} x) = \frac{1}{x\sqrt{x^2-1}}$$
$$\frac{d}{dx}(\tan^{-1} x) = \frac{1}{1+x^2} \qquad \frac{d}{dx}(\cot^{-1} x) = -\frac{1}{1+x^2}$$

Steps in Logarithmic Differentiation

Step 1 Take natural logarithms of both sides of an equation $y = f(x)$ and use the Laws of Logarithms to simplify.

Step 2 Differentiate implicitly with respect to x.

Step 3 Solve the resulting equation for y'.

Limits to Define e

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

Law of Laminar Flow

When the flow of blood moves through a blood vessel, the velocity is given by

$$v = \frac{P}{4\eta l}(R^2 - r^2),$$

where

η : the viscosity of the blood
P : the pressure difference between the ends of the tube
l : the length of the tube
R : the outer radius of the tube
l : the inner radius of the tube

Law of Natural Growth or Decay

Suppose $y(t)$ is the value of a quantity y at time t. Suppose $y(t)$ satisfies the **differential equation**

$$\frac{dy}{dt} = ky,$$

where k is a constant. The equation is called the **law of natural growth** if $k > 0$ or the **law of natural decay** if $k < 0$.

Solution

The only solutions of the differential equation $\frac{dy}{dt} = ky$ are the exponential functions $y(t) = y(0)e^{kt}$.

Newton's Law of Cooling

If $T(t)$ is the temperature of the object at time t and T_s is the temperature of the surroundings, then T satisfies the differential equation

$$\frac{dT}{dt} = k(T - T_s),$$

where k is a constant.

Continuously Compounded Interest

If an amount A_0 is invested at an interest rate r and if interest is compounded n times a year, then in each compounding period the interest rate is r/n and there are nt compounding periods in t years, so the value of the investment is

$$A_0\left(1 + \frac{r}{n}\right)^{nt}.$$

If we let $n \to \infty$, then we will be compounding the interest continuously and the value of the investment will be

$$\begin{aligned}
A(t) &= \lim_{n\to\infty} A_0\left(1 + \frac{r}{n}\right)^{nt} \\
&= \lim_{n\to\infty} A_0\left[\left(1 + \frac{r}{n}\right)^{n/r}\right]^{rt} \\
&= A_0 \lim_{n\to\infty}\left[\left(1 + \frac{r}{n}\right)^{n/r}\right]^{rt} \\
&= A_0 \lim_{m\to\infty}\left[\left(1 + \frac{1}{m}\right)^{m}\right]^{rt} = A_0 e^{rt}.
\end{aligned}$$

Linear Approximation

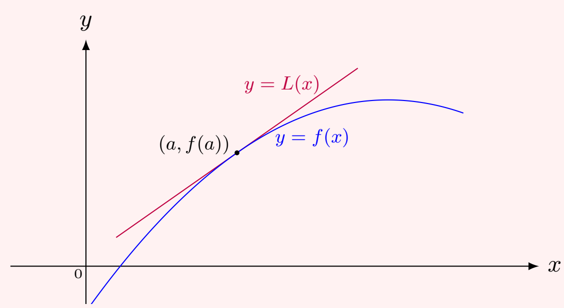
The tangent line at $x = a$ to the curve $y = f(x)$, that is,

$$y = L(x) = f(a) + f'(a)(x - a),$$

is call the **linearization** of f at a. The approximation

$$f(x) \approx f(a) + f'(a)(x - a)$$

is called the **linear approximation** or **tangent line approximation** of f at a.



Differential

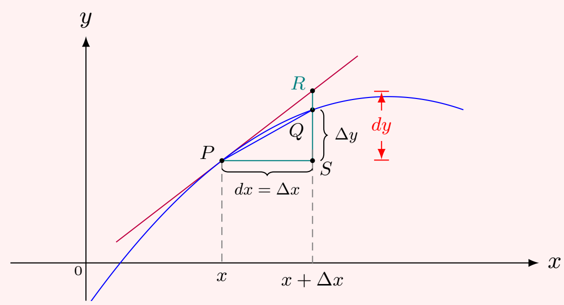
If $y = f(x)$, where f is a differentiable function, then the **differential** dx is an independent variable; that is, dx can be given the value of any real number. The **differential** dy is then defined in terms of dx by the equation

$$dy = f'(x)\, dx.$$

Geometrically, let $P(x, f(x))$ and $Q(x + \Delta x, f(x + \Delta x))$ be points on the graph of f and let $dx = \Delta x$. The corresponding change in y is

$$\Delta y = f(x + \Delta x) - f(x).$$

The slope of the tangent line PR is the derivative $f'(x)$. Thus the directed distance from S to R is $f'(x)\, dx = dy$. Therefore dy represents the amount that the tangent line rises or falls (the change in the linearization), whereas Δy represents the amount that the curve $y = f(x)$ rises or falls when x changes by an amount dx.

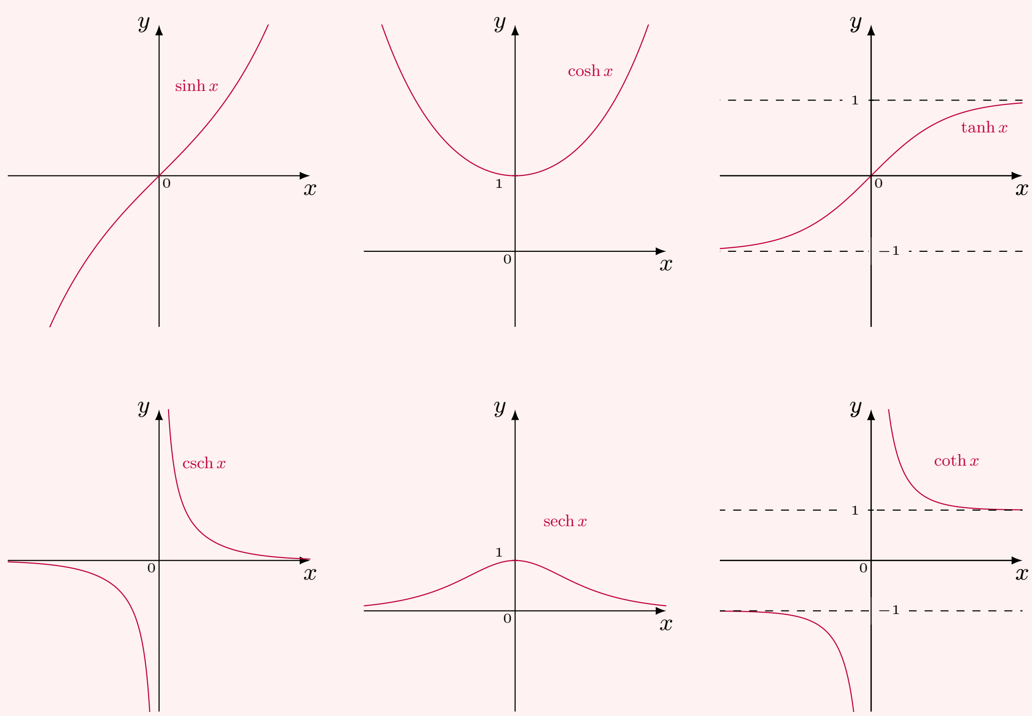


The Hyperbolic Functions

$$\sinh x = \frac{e^x - e^{-x}}{2} \qquad \operatorname{csch} x = \frac{1}{\sinh x}$$

$$\cosh x = \frac{e^x + e^{-x}}{2} \qquad \operatorname{sech} x = \frac{1}{\cosh x}$$

$$\tanh x = \frac{\sinh x}{\cosh x} \qquad \coth x = \frac{\cosh x}{\sinh x}$$



Hyperbolic Identities

$$\sinh(-x) = -\sinh x \qquad \cosh(-x) = \cosh x$$

$$\cosh^2 x - \sinh^2 x = 1 \qquad 1 - \tanh^2 x = \operatorname{sech}^2 x$$

$$\sinh(x+y) = \sinh x\,\cosh y + \cosh x\,\sinh y$$

$$\cosh(x+y) = \cosh x\,\cosh y + \sinh x\,\sinh y$$

Derivatives of Hyperbolic Functions

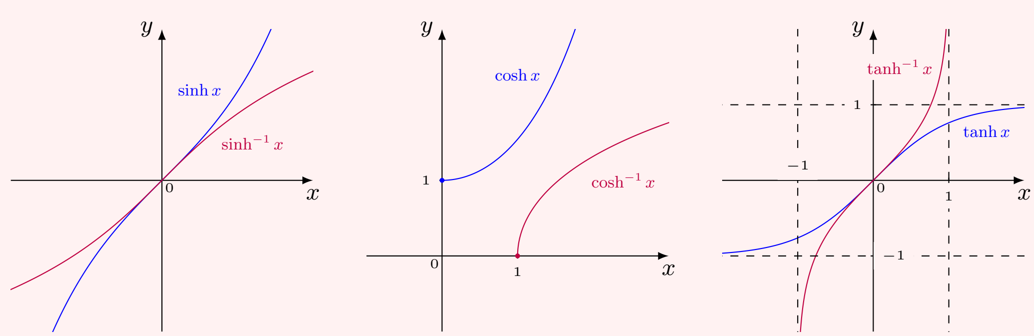
$$\frac{d}{dx}(\sinh x) = \cosh x \qquad \frac{d}{dx}(\operatorname{csch} x) = -\operatorname{csch} x\,\coth x$$

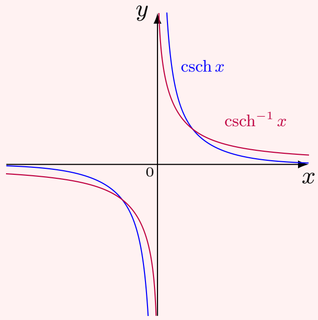
$$\frac{d}{dx}(\cosh x) = \sinh x \qquad \frac{d}{dx}(\operatorname{sech} x) = -\operatorname{sech} x\,\tanh x$$

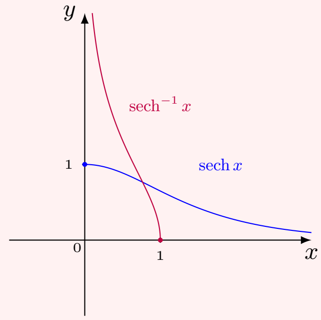
$$\frac{d}{dx}(\tanh x) = \operatorname{sech}^2 x \qquad \frac{d}{dx}(\coth x) = -\operatorname{csch}^2 x$$

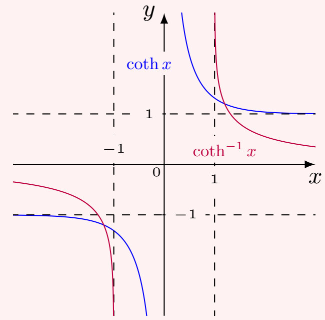
Inverse Hyperbolic Functions

$$\begin{aligned}
\sinh^{-1} x = y \;\; (x \in \mathbb{R}) &\iff \sinh y = x\\
\cosh^{-1} x = y \;\; (x \geq 1) &\iff \cosh y = x \text{ and } y \geq 0\\
\tanh^{-1} x = y \;\; (-1 < x < 1) &\iff \tanh y = x\\
\operatorname{csch}^{-1} x = y \;\; (x \neq 0) &\iff \operatorname{csch} y = x \text{ and } y \neq 0\\
\operatorname{sech}^{-1} x = y \;\; (0 < x \leq 1) &\iff \operatorname{sech} y = x \text{ and } y \geq 0\\
\coth^{-1} x = y \;\; (|x| > 1) &\iff \coth y = x \text{ and } y \neq 0
\end{aligned}$$









Explicit Formulas of Inverse Hyperbolic Functions

$$\sinh^{-1} x = \ln\left(x + \sqrt{x^2+1}\right) \qquad x \in \mathbb{R}$$

$$\cosh^{-1} x = \ln\left(x + \sqrt{x^2-1}\right) \qquad x \geq 1$$

$$\tanh^{-1} x = \tfrac{1}{2} \ln\left(\frac{1+x}{1-x}\right) \qquad -1 < x < 1$$

$$\operatorname{csch}^{-1} x = \ln\left(\frac{1}{x} + \frac{\sqrt{x^2+1}}{|x|}\right) \qquad x \neq 0$$

$$\operatorname{sech}^{-1} x = \ln\left(\frac{1+\sqrt{1-x^2}}{x}\right) \qquad 0 < x \leq 1$$

$$\coth^{-1} x = \tfrac{1}{2} \ln\left(\frac{x+1}{x-1}\right) \qquad |x| > 1$$

Derivatives of Inverse Hyperbolic Functions

$$\frac{d}{dx}(\sinh^{-1} x) = \frac{1}{\sqrt{1+x^2}} \qquad \frac{d}{dx}(\operatorname{csch}^{-1} x) = -\frac{1}{x\sqrt{x^2+1}}$$

$$\frac{d}{dx}(\cosh^{-1} x) = \frac{1}{\sqrt{x^2-1}} \qquad \frac{d}{dx}(\operatorname{sech}^{-1} x) = -\frac{1}{x\sqrt{1-x^2}}$$

$$\frac{d}{dx}(\tanh^{-1} x) = \frac{1}{1-x^2} \qquad \frac{d}{dx}(\coth^{-1} x) = \frac{1}{1-x^2}$$

Chapter 4

Applications of Differentiation

Extreme Values

Suppose f is a function defined in its domain D.

Absolute Extreme Values

Let c be a number in the domain D of a function f. Then $f(c)$ is the **absolute maximum value** of f on D if $f(c) \geq f(x)$ for all x in D; the **absolute minimum value** of f on D if $f(c) \leq f(x)$ for all x in D.

Local Extreme Values

Let c be a number in the domain D of a function f. Then $f(c)$ is the **local maximum value** of f if $f(c) \geq f(x)$ when x is near c; the **local minimum value** of f if $f(c) \leq f(x)$ when x is near c.

The Extreme Value Theorem

If f is continuous on a closed interval $[a, b]$, then f attains an absolute maximum value $f(c)$ and an absolute minimum value $f(d)$ at some numbers c and d in $[a, b]$.

Critical Number

A **critical number** of a function f is a number c in the domain of f such that either $f'(c) = 0$ or $f'(c)$ does not exist.

Fermat's Theorem

If f has a local maximum or minimum at c, and if $f'(c)$ exists, then $f'(c) = 0$. In other words, c must be a critical number of f.

Justification

Without loss of generality, suppose f has a local maximum at c, that is,

$$f(x) \leq f(c) \qquad \text{for } x \text{ sufficiently close to } c.$$

This implies that if h is sufficiently close to 0, with h being positive or negative, then

$$f(x+h) - f(c) \leq 0.$$

Thus, if $h > 0$ and h is sufficiently small, we have

$$\frac{f(x+h) - f(c)}{h} \leq 0.$$

Since f is differentiable at c, we have

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

Similarly, if $h < 0$ and h is sufficiently small, we have

$$\frac{f(x+h) - f(c)}{h} \geq 0.$$

Since f is differentiable at c, we have

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

Since both $f'(c) \leq 0$ and $f'(c) \geq 0$ hold, we get $f'(c) = 0$.
We have proved Fermat's Theorem for the case of a local maximum. The case of a local minimum can be proved in a similar manner.

The Closed Interval Method

To find the absolute maximum and minimum values of a continuous function f on a closed interval $[a, b]$:

1. Find the values of f at the critical numbers of f in (a, b).

2. Find the values of f at the endpoints of the interval.

3. The largest of the values from Steps 1 and 2 is the absolute maximum value; the smallest of these values is the absolute minimum value.

Rolle's Theorem

Suppose f is a function that satisfies the following:

1. f is continuous on the closed interval $[a, b]$.
2. f is differentiable on the open interval (a, b).
3. $f(a) = f(b)$.

Then there is a number c in (a, b) such that $f'(c) = 0$.

Justification

There are three cases:
CASE I: $f(x) = k$, where k is a constant.
Then $f'(x) = 0$, so the number c can be taken to be any number in (a, b).
CASE II: $f(x) > f(a)$, for some x in (a, b).
By the Extreme Value Theorem, f has a maximum value somewhere in (a, b). Since $f(a) = f(b)$, it must attain this maximum value at a number c in the open interval (a, b). Then f has a local maximum at c. Since f is differentiable at c, $f'(c) = 0$ by Fermat's Theorem.
CASE III: $f(x) < f(a)$, for some x in (a, b).
By the Extreme Value Theorem, f has a minimum value in (a, b) and, since $f(a) = f(b)$, it attains this minimum value at a number c in (a, b). Again $f'(c) = 0$ by Fermat's Theorem.

The Mean Value Theorem

Suppose f is a function that satisfies the following:

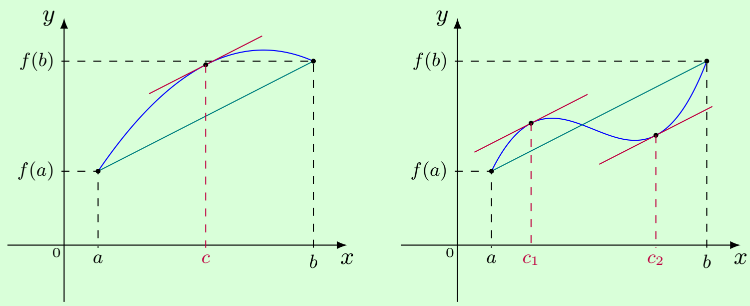
1. f is continuous on the closed interval $[a, b]$.
2. f is differentiable on the open interval (a, b).

Then there is a number c in (a, b) such that

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

or, equivalently,

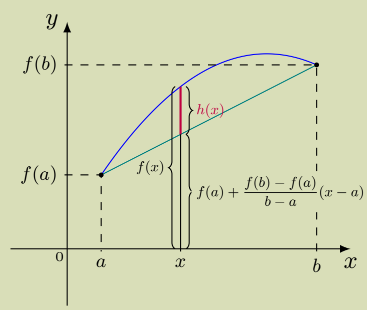
$$f(b) - f(a) = f'(c)(b - a).$$



Justification

An equation of the line segment connecting $(a, f(a))$ and $(b, f(b))$ is $y = f(a) - \dfrac{f(b) - f(a)}{b - a}(x - a)$. As shown in the figure, at x with $a \leq x \leq b$, consider the auxiliary function

$$h(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b - a}(x - a).$$



We can verify that h satisfies the three hypotheses of Rolle's Theorem.

1. h is continuous on the closed interval $[a, b]$ since it is the sum of f and a first-degree polynomial, both of which are continuous.

2. h is differentiable on the open interval (a, b) because both f and the first-degree polynomial are differentiable. In fact, we have

$$h'(x) = f'(x) - \frac{f(b) - f(a)}{b - a}.$$

3. Since

$$h(a) = f(a) - f(a) - \frac{f(b) - f(a)}{b - a}(a - a) = 0,$$
$$h(b) = f(b) - f(a) - \frac{f(b) - f(a)}{b - a}(b - a) = f(b) - f(a) - [f(b) - f(a)] = 0,$$

we have $h(a) = h(b)$.

Hence, by Rolle's Theorem, there is a number c in (a, b) such that $h'(c) = 0$. Therefore,

$$0 = h'(c) = f'(c) - \frac{f(b) - f(a)}{b - a},$$

so that

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

Function with Zero Derivative

If $f'(x) = 0$ for all x in an interval (a, b), then f is constant on (a, b).

Justification

For any two numbers x_1, x_2 in (a, b), by the Mean Value Theorem, there is a number $x^* \in (x_1, x_2) \subset (a, b)$ such that

$$f(x_1) - f(x_2) = f'(x^*)(x_1 - x_2).$$

Since $f'(x) = 0$ for all x in an interval (a, b), we have $f'(x^*) = 0$. Thus, $f(x_1) = f(x_2)$. This shows that at any two numbers in (a, b), the values of f are the same. So, f is constant on (a, b).

Functions with the Same Derivative

If $f'(x) = g'(x)$ for all x in an interval (a, b), then $f - g$ is constant on (a, b); that is, $f(x) = g(x) + c$, where c is a constant.

Justification

Let $h(x) = f(x) - g(x)$. By the hypothesis, $h'(x) = 0$ for all x in an interval (a, b). For any two numbers x_1, x_2 in (a, b), by the Mean Value Theorem, there is a number $x^* \in (x_1, x_2) \subset (a, b)$ such that

$$h(x_1) - h(x_2) = h'(x^*)(x_1 - x_2) = 0 \cdot (x_1 - x_2) = 0.$$

Thus, $h(x_1) = h(x_2)$. This shows that at any two numbers in (a, b), the values of h are the same. So, $h(x) = c$ on (a, b), where c is a constant. Therefore, $f(x) = g(x) + c$.

Increasing/Decreasing Test

(a) If $f'(x) > 0$ on an interval, then f is increasing on that interval.
(b) If $f'(x) < 0$ on an interval, then f is decreasing on that interval.

Justification

Suppose $f'(x) > 0$ on (a, b). We take any two numbers $x_1, x_2 \in (a, b)$ such that with $x_1 < x_2$. By the Mean Value Theorem, there is $c \in (x_1, x_2)$ such that

$$f(x_1) - f(x_2) = f'(c)(x_1 - x_2) < 0,$$

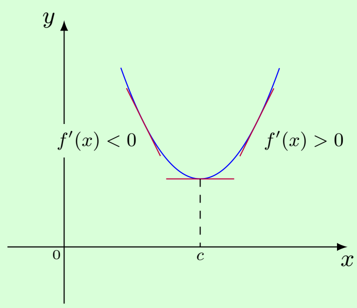
which implies that $f(x_1) < f(x_2)$. Hence, f is increasing on (a, b).
Similarly, if $f'(x) < 0$ on (a, b), and if we take any two numbers $x_1, x_2 \in (a, b)$ such that with $x_1 < x_2$, then, by the Mean Value Theorem, there is $c \in (x_1, x_2)$ such that

$$f(x_1) - f(x_2) = f'(c)(x_1 - x_2) > 0,$$

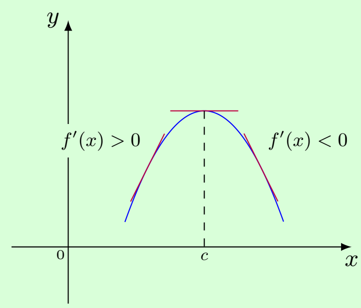
which implies that $f(x_1) > f(x_2)$. Hence, f is decreasing on (a, b).

The First Derivative Test

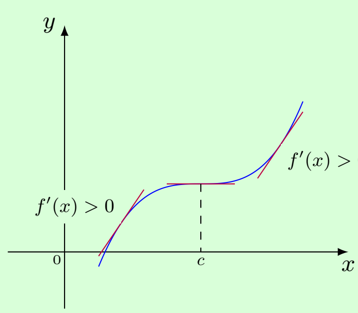
Suppose that c is a critical number of a continuous function f.
(a) If f' changes from positive to negative at c, then f has a local maximum at c.
(b) If f' changes from negative to positive at c, then f has a local minimum at c.
(c) If f' does not change sign at c, then f has no local maximum or minimum at c.



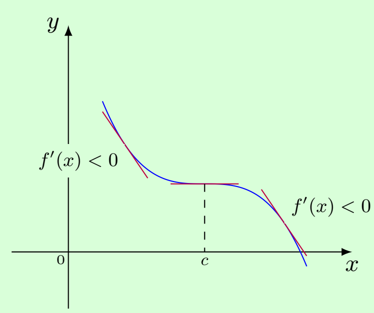
f' changes from positive to negative at c



f' changes from negative to positive at c



No maximum or minimum



No maximum or minimum

Justification

(a) Suppose f' changes from positive to negative at c, in other words, for some $\delta > 0$,

$$f'(x) > 0 \text{ for } c - \delta < x < c;$$
$$f'(x) < 0 \text{ for } c < x < c + \delta.$$

From the Increasing/Decreasing Test, we know that f is increasing on $(c - \delta, c)$ and decreasing on $(c, c + \delta)$. Since f is continuous, we have

$$f(x) < f(c) \text{ for } c - \delta < x < c;$$
$$f(c) > f(x) \text{ for } c < x < c + \delta.$$

It follows that $f(c) > f(x)$ on $(c - \delta, c + \delta)$. Hence, f has a local maximum at c.
(b) If f' changes from negative to positive at c, then $-f$ changes from positive to negative at c. From part (a), $-f$ has a local maximum at c. Hence, f has a local minimum at c.
(c) If f' does not change sign at c, without loss of generality, we suppose $f'(x) > 0$ in $(c - \delta, c + \delta)\backslash\{c\}$ for some $\delta > 0$. Then, by the Increasing/Decreasing Test, f is increasing on $(c - \delta, c + \delta)$. So, $f(c)$ is neither a local maximum nor minimum.

The Second Derivative Test

Suppose f'' 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.

Justification

(a) Suppose $f'(c) = 0$ and $f''(c) > 0$. Since f'' is continuous near c, there is $\delta > 0$ such that $f''(x) > 0$ in $(c - \delta, c + \delta)$. Thus, by the Mean Value Theorem, for $c - \delta < x < c$, there is a number ξ satisfying $x < \xi < c$ such that

$$f'(x) - f'(c) = f''(\xi)(x - c).$$

This equality implies that for $c - \delta < x < c$, $f'(x) < 0$. Similarly, for $c < x < c + \delta$, there is a number η satisfying $c < \eta < x$ such that

$$f'(x) - f'(c) = f''(\eta)(x - c).$$

This equality implies that for $c < x < c + \delta$, $f'(x) > 0$. Hence, f' changes from negative to positive at c. By the First Derivative Test, then f has a local minimum at c.
(b) Let $g(x) = -f(x)$. If $f'(c) = 0$ and $f''(c) < 0$, then $g'(c) = 0$ and $g''(c) > 0$. From part (a), g has a local minimum at c. So, f has a local maximum at c.

Concavity

If the graph of f lies above all of its tangents on an interval I, then it is called **concave upward** on I. If the graph of f lies below all of its tangents on I, it is called **concave downward** on I.

Inflection Point

A point P on a curve $y = f(x)$ is called an inflection point if f is continuous there and the curve changes from concave upward to concave downward or from concave downward to concave upward at P.

Concavity Test

(a) If $f''(x) > 0$ for all x in I, then the graph of f is concave upward on I.
(b) If $f''(x) < 0$ for all x in I, then the graph of f is concave downward on I.

Justification

(a) Let a be a point in the open interval I where $f''(x) > 0$. We shall show that graph $y = f(x)$ lies above the tangent line at $(a, f(a))$. The tangent line in question has the equation

$$y = T(x) = f(a) + f'(a)(x - a).$$

Consider the function $g(x) = f(x) - T(x)$. It is easy to see that $g'(a) = f'(a) - f'(a) = 0$, so $x = a$ is a critical number of g. Moreover,

$$g''(a) = f''(a) > 0.$$

Thus, by the Second Derivative Test, g has a local minimum at $x = a$. This proves that the curve $y = f(x)$ lies above the tangent line $y = T(x)$.
(b) Let $g(x) = -f(x)$. Then $g''(x) > 0$ for all x in I. From part (a), g is concave upward on I. So, f is concave downward on I.

Indeterminate Forms

The following are typical **indeterminate forms**:

$$\left.\begin{array}{l}\text{indeterminate form of type } \frac{0}{0}\\ \text{indeterminate form of type } \frac{\infty}{\infty}\end{array}\right\}\text{two standard indeterminate forms}$$

$$\text{indeterminate form of type } 0\cdot\infty\ \big\}\text{indeterminate product}$$

$$\text{indeterminate form of type } \infty-\infty\ \big\}\text{indeterminate difference}$$

$$\left.\begin{array}{l}\text{indeterminate form of type } 0^0\\ \text{indeterminate form of type } \infty^0\\ \text{indeterminate form of type } 1^\infty\end{array}\right\}\text{indeterminate powers}$$

Here the form $\frac{0}{0}$ means a limit of the form

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

where $f(x) \to 0$ and $g(x) \to 0$ as $x \to a$. The other indeterminate forms arise similarly in the corresponding symbolic expressions.

L'Hospital's Rule

Suppose f and g are differentiable and $g'(x) \neq 0$ on an open interval I that contains a (except possibly at a). Suppose that

$$\lim_{x\to a} f(x) = 0 \quad \text{and} \quad \lim_{x\to a} g(x) = 0$$

or that

$$\lim_{x\to a} f(x) = \pm\infty \quad \text{and} \quad \lim_{x\to a} g(x) = \pm\infty$$

(In other words, we have an indeterminate form of type $\frac{0}{0}$ or $\frac{\infty}{\infty}$.) Then

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

if the limit on the right side exists (or is ∞ or $-\infty$).

Guidelines for Sketching a Curve

By following the guidelines, you are able to make a sketch that displays the most important aspects of a given function $y = f(x)$.

A. Domain: Domain D is the set of values of x for which $f(x)$ is defined.

B. Intercepts:

The y-intercept is $f(0)$. The x-intercepts are the real solutions of the equation $f(x) = 0$.

C. Symmetry:

(i) If $f(-x) = f(x)$ for all $x \in D$, then f is an <u>even function</u> and the curve is symmetric about the y-axis.

(ii) If $f(-x) = -f(x)$ for all $x \in D$, then f is an <u>odd function</u> and the curve is symmetric about the origin.

(iii) If $f(x+p) = f(x)$ for all $x \in D$, where p is a positive constant, then f is a <u>periodic function</u> and the smallest such number p is the period.

D. Asymptotes:

(i) Horizontal asymptote: The line $y = L$ is a <u>horizontal asymptote</u> of the curve $y = f(x)$ if either

$$\lim_{x\to\infty} f(x) = L \quad \text{or} \quad \lim_{x\to-\infty} f(x) = L.$$

(ii) Vertical asymptote: The line $x = a$ is a <u>vertical asymptote</u> of the curve $y = f(x)$ if at least one of the following statements is true:

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

E. Intervals of Increase or Decrease:

Apply the <u>Increasing/Decreasing Test</u>:

(a) If $f'(x) > 0$ on an interval, then f is increasing on that interval.

(b) If $f'(x) < 0$ on an interval, then f is decreasing on that interval.

F. Local Maximum and Minimum Values:
First find the critical numbers, the numbers c where $f'(c) = 0$ or $f'(c)$ does not exist. Then apply the First Derivative Test or the Second Derivative Test to determine whether $f(c)$ is a local maximum or minimum value.
(i) The First Derivative Test: Suppose that c is a critical number of a continuous function f.
(a) If f' changes from positive to negative at c, then f has a local maximum at c.
(b) If f' 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.
(ii) The Second Derivative Test: Suppose f'' 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.
G. Concavity and Points of Inflection:
(i) Apply the Concavity Test:
(a) If $f''(x) > 0$ on an interval, then the graph of f is concave upward on that interval.
(b) If $f''(x) < 0$ on an interval, then the graph of f is concave downward on that interval.
(ii) Points of Inflection: At these points, the curve changes the direction of concavity.
H. Sketch the Curve: Sketch the curve $y = f(x)$ by using the information in items **A-G**.

Slant Asymptote

The line $y = ax + b$ is called a **slant asymptote** of the curve $y = f(x)$ if either

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

or

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

How to Find a Slant Asymptote

Since $\lim\limits_{x\to\infty} [f(x) - (ax+b)] = 0$, we have

$$a = \lim_{x\to\infty} \frac{f(x)}{x}, \qquad b = \lim_{x\to\infty} [f(x) - ax]\,.$$

Alternatively, if f is a rational function of the form

$$\frac{a_{n+1}x^{n+1} + \cdots + a_1 x + a_0}{b_n x^n + \cdots + b_1 x + b_0},$$

one may use long division to obtain the slant asymptotes of f.
Similarly, one can find a slant asymptote as $x \to -\infty$.

Cost and Profit - Application

If $C(x)$, the **cost function**, is the cost of producing x units of a certain product, then the **marginal cost** is the rate of change of C with respect to x. In other words, the marginal cost function is the derivative, $C'(x)$, of the cost function.
If $p(x)$ is the price per unit that the company can charge if it sells x units. Then p is called the **demand function** (or **price function**) and we would expect it to be a decreasing function of x. If x units are sold and the price per unit is $p(x)$, then the total revenue is

$$R(x) = xp(x)$$

and R is called the **revenue function**. The derivative R' of the revenue function is called the **marginal revenue function** and is the rate of change of revenue with respect to the number of units sold.
If x units are sold, then the total profit is

$$P(x) = R(x) - C(x)$$

and P is called the **profit function**. The marginal profit function is P', the derivative of the profit function.

Newton's Method

To find approximations to a solution r of the equation $f(x) = 0$, we start with a first approximation x_1, which is obtained usually by guessing. Consider the tangent line to the curve $y = f(x)$ at the point $(x_1, f(x_1))$. An equation of the line is

$$y - f(x_1) = f'(x_1)(x - x_1),$$

whose x-intercept gives the next approximation x_2, obtained by putting $y = 0$ in the last equation:

$$0 - f(x_1) = f'(x_1)(x_2 - x_1) \Longrightarrow x_2 - x_1 = -\frac{f(x_1)}{f'(x_1)} \Longrightarrow x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}.$$

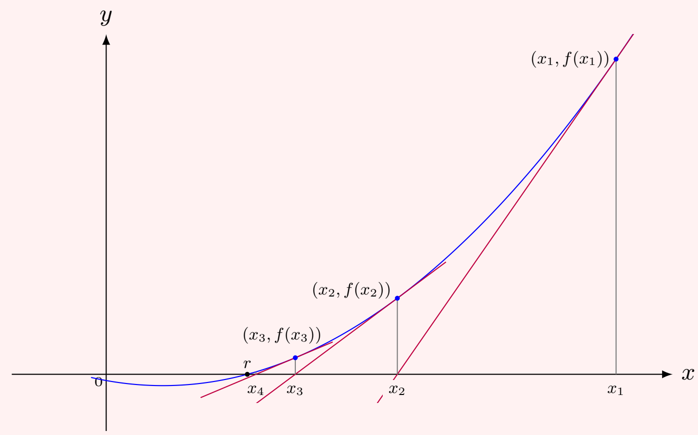
The formula for x_2 is meaningful if $f'(x_1) \neq 0$.
We repeat this procedure with x_1 replaced by the second approximation x_2, using the tangent line at $(x_2, f(x_2))$. This gives a third approximation:

$$x_3 = x_2 - \frac{f(x_2)}{f'(x_2)}.$$

If we keep repeating this process, we generate a sequence of approximations $x_1, x_2, x_3, x_4, \ldots$. In general, if x_n is the nth approximation and $f'(x_n) \neq 0$, then the next approximation is given by

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

This recursive formula is called **Newton's approximation**.



Convergence of Newton's Method

It is known that if x_1 is chosen to be sufficiently close to r, and if $f'(x_n) \neq 0$ for $n = 1, 2, 3, \ldots,$ then the sequence $\{x_n\}$ converges to r:

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

Antiderivative

A function F is called an **antiderivative** of f on an interval I if $F'(x) = f(x)$ for all x in I.

General Antiderivative

If F is an antiderivative of f on an interval I, then the most general antiderivative of f on I is

$$F(x) + C,$$

where C is an arbitrary constant.

Justification

Suppose G is any antiderivative of f so that $G'(x) = f(x)$. Thus, we have $G'(x) = F'(x)$, or $[G(x) - F(x)]' = 0$. So, $G(x) - F(x) = C$, where C is a constant. Hence $G(x) = F(x) + C$.
On the other hand, for any constant C, the function $F(x) + C$ is an antiderivative of f, since

$$[F(x) + C]' = F'(x) = f(x).$$

Therefore, the most general antiderivative of f is $F(x) + C$, where C is an arbitrary constant.

Table of Antidifferentiation Formulas

Function	Particular antiderivative	Function	Particular antiderivative
$cf(x)$	$cF(x)$	$\sec^2 x$	$\tan x$
$f(x)+g(x)$	$F(x)+G(x)$	$\sec x\,\tan x$	$\sec x$
$x^n \quad (n \neq -1)$	$\dfrac{x^{n+1}}{n+1}$	$\dfrac{1}{\sqrt{1-x^2}}$	$\sin^{-1} x$
$\dfrac{1}{x}$	$\ln \lvert x \rvert$	$\dfrac{1}{1+x^2}$	$\tan^{-1} x$
e^x	e^x	$\cosh x$	$\sinh x$
$\cos x$	$\sin x$	$\sinh x$	$\cosh x$
$\sin x$	$-\cos x$		

Chapter 5

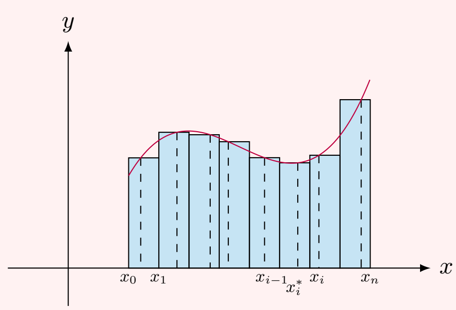
Integrals

Definite Integral

If f is a function defined for $a \leq x \leq b$, we divide the interval $[a, b]$ into n subintervals of equal width $\Delta x = (b-a)/n$. We let $x_0(=a), x_1, x_2, \ldots, x_n(=b)$ be the endpoints of these subintervals and we let $x_1^*, x_2^*, \ldots, x_n^*$ be any **sample points** in these subintervals, so x_i^* lies in the ith subinterval $[x_{i-1}, x_i]$. Then the **definite integral of f from a to b** is

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*)\Delta x$$

provided that this limit exists and gives the same value for all possible choices of sample points. If it does exist, we say that f is **integrable** on $[a, b]$.



More precise, the last limit means that

For every number $\varepsilon > 0$ there is a integer N such that

$$\left| \sum_{i=1}^{n} f(x_i^*)\Delta x - \int_a^b f(x)\,dx \right| < \varepsilon$$

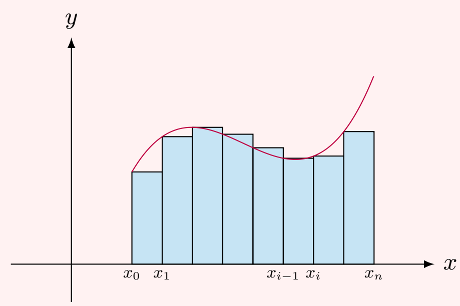
for every integer $n > N$ and for every choice of x_i^* in $[x_{i-1}, x_i]$.

In the definition, the sum $\sum_{i=1}^{n} f(x_i^*)\Delta x$ is called a **Riemann sum**, the function f is the **integrand**, a is the **lower limit**, and b is the **upper limit**.
Geometrically, if f is a nonnegative and continuous, the definite integral is the area of the region that lies under the graph of f.

Left Endpoint Approximation

If $x_i^* = x_{i-1}$, $i = 1, 2, \ldots, n$, then the Riemann sum is called a **left endpoint approximation**:

$$\int_a^b f(x)\,dx \approx L_n = \Delta x\left[f(x_0) + f(x_1) + f(x_2) + \cdots + f(x_{n-1})\right].$$

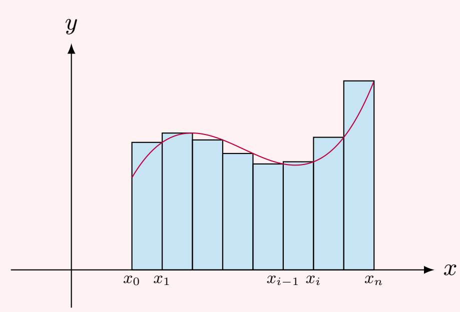


Left Endpoint Approximation L_n

Right Endpoint Approximation

If $x_i^* = x_i$, $i = 1, 2, \ldots, n$, then the Riemann sum is called a **right endpoint approximation**:

$$\int_a^b f(x)\,dx \approx R_n = \Delta x\left[f(x_1) + f(x_2) + f(x_3) + \cdots + f(x_n)\right].$$

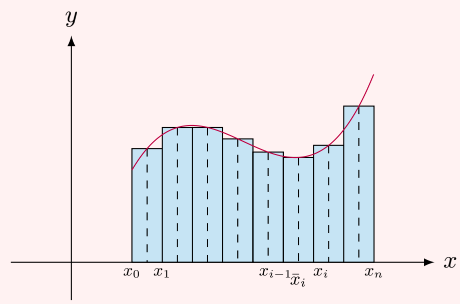


Right Endpoint Approximation R_n

Midpoint Approximation

If $x_i^* = \frac{1}{2}(x_{i-1} + x_i) = \bar{x}_i$, $i = 1, 2, \ldots, n$, then the Riemann sum is called a **midpoint rule** (or **midpoint approximation**):

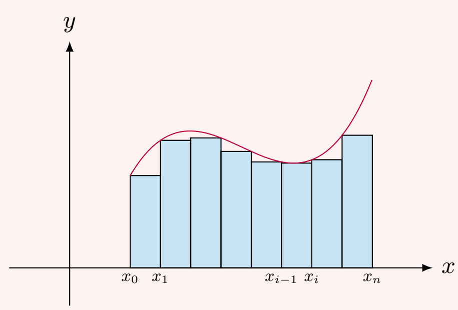
$$\int_a^b f(x)\,dx \approx M_n = \Delta x\left[f(\bar{x}_1) + f(\bar{x}_2) + f(\bar{x}_3) + \cdots + f(\bar{x}_n)\right].$$



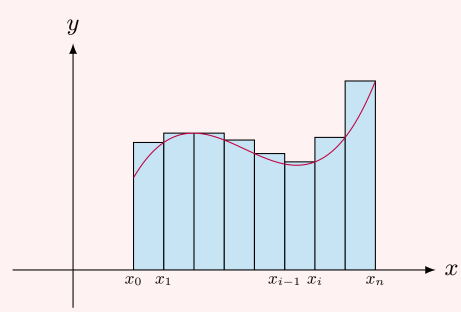
Midpoint Approximation M_n

Lower and Upper Sums

If the sample points x_i^* are chosen so that $f(x_i^*)$ is the minimum (or maximum) value of f on the ith subinterval $[x_{i-1}, x_i]$, then the corresponding approximation $\displaystyle\sum_{i=1}^{n} f(x_i^*)\Delta x$ is called the **lower** (or **upper**) **sum**.



a lower sum



an upper sum

Integrability of Continuous Function

If f is continuous on $[a, b]$, or if f has only a finite number of jump discontinuities, then f is integrable on $[a, b]$; that is,the definite integral $\int_a^b f(x)\,dx$ exists.

Conventions

(i) If $a > b$, define $\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx$.

(ii) Define $\int_a^a f(x)\,dx = 0$.

Properties of the Integral

Suppose f and g are integrable from a to b.

1. $\int_a^b c\,dx = c(b-a)$, where c is any constant.

2. $\int_a^b [f(x)+g(x)]\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx$.

3. $\int_a^b cf(x)\,dx = c\int_a^b f(x)\,dx$, where c is any constant.

4. $\int_a^b [f(x)-g(x)]\,dx = \int_a^b f(x)\,dx - \int_a^b g(x)\,dx$.

5. $\int_a^c f(x)\,dx + \int_c^b f(x)\,dx = \int_a^b f(x)\,dx$.

6. If $f(x) \geq 0$ for $a \leq x \leq b$, then $\int_a^b f(x)\,dx \geq 0$.

7. If $f(x) \geq g(x)$ for $a \leq x \leq b$, then $\int_a^b f(x)\,dx \geq \int_a^b g(x)\,dx$.

8. If $m \leq f(x) \leq M$ for $a \leq x \leq b$, then

$$m(b-a) \leq \int_a^b f(x)\,dx \leq M(b-a).$$

Justification - Property 1

By the definition,

$$\int_a^b c\,dx = \lim_{n\to\infty} \sum_{i=1}^n c\Delta x = c\lim_{n\to\infty} \sum_{i=1}^n \Delta x = c\lim_{n\to\infty} (b-a) = c(b-a).$$

Justification - Property 2

By the definition,

$$\begin{aligned}\int_a^b [f(x)+g(x)]\ dx &= \lim_{n\to\infty}\sum_{i=1}^{n}[f(x_i^*)+g(x_i^*)]\,\Delta x\\ &= \lim_{n\to\infty}\sum_{i=1}^{n} f(x_i^*)\Delta x + \lim_{n\to\infty}\sum_{i=1}^{n} g(x_i^*)\Delta x\\ &= \int_a^b f(x)\,dx + \int_a^b g(x)\,dx.\end{aligned}$$

Justification - Property 3

By the definition,

$$\int_a^b cf(x)\,dx = \lim_{n\to\infty}\sum_{i=1}^{n} cf(x_i^*)\Delta x = c\lim_{n\to\infty}\sum_{i=1}^{n} f(x_i^*)\Delta x = c\int_a^b f(x)\,dx.$$

Justification - Property 4

By the definition,

$$\begin{aligned}\int_a^b [f(x)-g(x)]\ dx &= \lim_{n\to\infty}\sum_{i=1}^{n}[f(x_i^*)-g(x_i^*)]\,\Delta x\\ &= \lim_{n\to\infty}\sum_{i=1}^{n} f(x_i^*)\Delta x - \lim_{n\to\infty}\sum_{i=1}^{n} g(x_i^*)\Delta x\\ &= \int_a^b f(x)\,dx - \int_a^b g(x)\,dx.\end{aligned}$$

Justification - Property 6

If $f(x)\geq 0$ for $a\leq x\leq b$, then

$$\sum_{i=1}^{n} f(x_i^*)\Delta x \geq 0.$$

Thus,

$$\int_a^b f(x)\,dx = \lim_{n\to\infty}\sum_{i=1}^{n} f(x_i^*)\Delta x \geq 0.$$

Justification - Property 7

Let $h(x) = f(x) - g(x)$. Then $h(x) \geq 0$ for $a \leq x \leq b$. From Property 6,

$$\int_a^b [f(x) - g(x)]\, dx = \int_a^b h(x)\, dx \geq 0.$$

From Property 4, we have

$$\int_a^b f(x)\, dx - \int_a^b g(x)\, dx = \int_a^b [f(x) - g(x)]\, dx.$$

Thus,

$$\int_a^b f(x)\, dx \geq \int_a^b g(x)\, dx.$$

Justification - Property 8

If $m \leq f(x) \leq M$ for $a \leq x \leq b$, from Property 7,

$$\int_a^b m\, dx \leq \int_a^b f(x)\, dx \leq \int_a^b m\, dx.$$

From Property 1, we have

$$\int_a^b m\, dx = m(b-a), \qquad \int_a^b M\, dx = M(b-a).$$

Thus,

$$m(b-a) \leq \int_a^b f(x)\, dx \leq M(b-a).$$

The Fundamental Theorem of Calculus

Part 1

If f is continuous on $[a, b]$, then the function g defined by

$$g(x) = \int_a^x f(t)\, dt \qquad a \leq x \leq b$$

is continuous on $[a, b]$ and differentiable on (a, b), and $g'(x) = f(x)$.

Part 2

If f is continuous on $[a,b]$, then

$$\int_a^b f(t)\,dt = F(b) - F(a), \qquad a \le x \le b$$

where F is any antiderivative of f, that is, a function such that $F' = f$.

Justification - Part 1

If x and $x+h$ are in (a,b), then, for $h \neq 0$, we have

$$\begin{aligned}
\frac{g(x+h)-g(x)}{h} &= \frac{\displaystyle\int_a^{x+h} f(t)\,dt - \int_a^x f(t)\,dt}{h} \\
&= \frac{\displaystyle\left(\int_a^x f(t)\,dt + \int_x^{x+h} f(t)\,dt\right) - \int_a^x f(t)\,dt}{h} = \frac{\displaystyle\int_x^{x+h} f(t)\,dt}{h}.
\end{aligned}$$

For now let's assume that $h > 0$. Since f is continuous on $[x, x+h]$, the Extreme Value Theorem says that there are numbers u and v in $[x, x+h]$ such that $f(u) = m$ and $f(v) = M$, where m and M are the absolute minimum and maximum values of f on $[x, x+h]$. From Property 8 of integrals, we have

$$f(u)h = mh \le \int_x^{x+h} f(t)\,dt \le Mh = f(v)h,$$

so that, by dividing the inequalities by h,

$$f(u) \le \frac{g(x+h)-g(x)}{h} \le f(v).$$

The last two inequalities can be proved in a similar manner for the case where $h < 0$.
Since u and v lie between x and $x+h$, we know that both $u \to x$ and $v \to x$ as $h \to 0$. Thus, because f is continuous at x, we have

$$\lim_{h\to 0} f(u) = \lim_{u\to x} f(u) = f(x) \qquad \text{and} \qquad \lim_{h\to 0} f(v) = \lim_{v\to x} f(v) = f(x).$$

From the Squeeze Theorem, we have

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

When $x = a$ or b, the above limit can be interpreted as a one-sided limit. The existence of one-sided derivatives at $x = a$ and $x = b$ implies that the function g is continuous at these endpoints.

Justification - Part 2

Let $g(x) = \int_a^x f(t)\,dt$. From Part 1, we have $g'(x) = f(x)$, that is, g is an antiderivative of f. If F is any antiderivative of f, then

$$[g(x) - F(x)]' = g'(x) - F'(x) = f(x) - f(x) = 0.$$

Thus, $g(x) = F(x) + C$, where C is a constant. This gives

$$\int_a^x f(t)\,dt = F(x) + C.$$

By putting $x = a$ and $x = b$ into the last equation, we get

$$0 = F(a) + C \qquad \text{and} \qquad \int_a^b f(t)\,dt = F(b) + C.$$

Combining these two equations, we have

$$\int_a^b f(t)\,dt = F(b) - F(a).$$

Indefinite Integral

$\int f(x)\,dx = F(x)$ means $F'(x) = f(x)$.

Table of Indefinite Integrals

$$\int cf(x)\,dx = c\int f(x)\,dx \qquad \int [f(x)+g(x)]\,dx = \int f(x)\,dx + \int g(x)\,dx$$

$$\int k\,dx = kx + C$$

$$\int x^n\,dx = \frac{x^{n+1}}{n+1} + C \quad (n \neq -1) \qquad \int \frac{1}{x}\,dx = \ln|x| + C$$

$$\int e^x\,dx = e^x + C \qquad \int a^x\,dx = \frac{a^x}{\ln a} + C$$

$$\int \sin x\,dx = -\cos x + C \qquad \int \cos x\,dx = \sin x + C$$

$$\int \sec^2 x\,dx = \tan x + C \qquad \int \csc^2 x\,dx = -\cot x + C$$

$$\int \sec x\,\tan x\,dx = \sec x + C \qquad \int \csc x\,\cot x\,dx = -\csc x + C$$

$$\int \frac{1}{x^2+1}\,dx = \tan^{-1} x + C \qquad \int \frac{1}{\sqrt{1-x^2}}\,dx = \sin^{-1} x + C$$

$$\int \sinh x\,dx = \cosh x + C \qquad \int \cosh x\,dx = \sinh x + C$$

Net Change Theorem

The integral of a rate of change is the **net change**:

$$\int_a^b F'(x)\,dx = F(b) - F(a).$$

Justification

From Part 2 of the Fundamental Theorem of Calculus,

$$\int_a^b f(x)\,dx = F(b) - F(a),$$

where F is an antiderivative of f that satisfies $F'(x) = f(x)$. Thus, we have

$$\int_a^b F'(x)\,dx = F(b) - F(a).$$

The Substitution Rule

If $u = g(x)$ is a differentiable function whose range is an interval I and f is continuous on I, then

$$\int f(g(x))\, g'(x)\, dx = \int f(u)\, du.$$

Justification

By the Chain Rule,

$$\frac{d}{dx}\left[F(g(x))\right] = F'(g(x))g'(x).$$

This means that $F(g(x))$ is an antiderivative of the function $F'(g(x))g'(x)$. So,

$$\int F'(g(x))g'(x)\, dx = F(g(x)) + C.$$

Since

$$\int F'(u)\, du = F(u) + C,$$

by using $u = g(x)$, we get

$$\int F'(g(x))g'(x)\, dx = F(u) + C = \int F'(u)\, du.$$

Writing $F' = f$ gives

$$\int f(g(x))g'(x)\, dx = \int f(u)\, du.$$

The Substitution Rule for Definite Integrals

If g' is continuous on $[a, b]$ and f is continuous on the range of $u = g(x)$, then

$$\int_a^b f(g(x))\, g'(x)\, dx = \int_{g(a)}^{g(b)} f(u)\, du.$$

Justification

Let F be an antiderivative of f. By the Chain Rule,

$$\frac{d}{dx}\left[F(g(x))\right] = F'(g(x))g'(x).$$

This means that $F(g(x))$ is an antiderivative of the function $F'(g(x))g'(x)$. By Part 2 of the Fundamental Theorem of Calculus,

$$\int_a^b f(g(x))g'(x)\,dx = F(g(x))\Big]_{x=a}^{b} = F(g(b)) - F(g(a)).$$

Again by Part 2 of the Fundamental Theorem of Calculus,

$$\int_{g(a)}^{g(b)} f(u)\,du = F(u)\Big]_{u=g(a)}^{g(b)} = F(g(b)) - F(g(a)).$$

Hence,

$$\int_a^b f(g(x))\,g'(x)\,dx = \int_{g(a)}^{g(b)} f(u)\,du.$$

Integrals of Symmetric Functions

Suppose f is continuous on $[-a, a]$.

(a) If f is even, $f(-x) = f(x)$, then $\displaystyle\int_{-a}^{a} f(x)\,dx = 2\int_0^a f(x)\,dx$.

(b) If f is odd, $f(-x) = -f(x)$, then $\displaystyle\int_{-a}^{a} f(x)\,dx = 0$.

Justification

We write the integral as

$$\int_{-a}^{a} f(x)\,dx = \int_{-a}^{0} f(x)\,dx + \int_0^a f(x)\,dx = -\int_0^{-a} f(x)\,dx + \int_0^a f(x)\,dx.$$

In the first integral on the right side we make the substitution $u = -x$. Then $du = -dx$ and when $x = -a$, $u = a$. Thus

$$-\int_0^{-a} f(x)\,dx = -\int_0^a f(-u)(-du) = \int_0^a f(-u)\,du.$$

Hence,

$$\int_{-a}^{a} f(x)\,dx = \int_0^a f(-u)\,du + \int_0^a f(x)\,dx.$$

(a) If f is even, $f(-x) = f(x)$, then

$$\int_{-a}^{a} f(x)\,dx = \int_0^a f(u)\,du + \int_0^a f(x)\,dx = 2\int_0^a f(x)\,dx.$$

(b) If f is odd, $f(-x) = -f(x)$, then

$$\int_{-a}^{a} f(x)\,dx = -\int_0^a f(u)\,du + \int_0^a f(x)\,dx = 0.$$

Chapter 6

Applications of Integration

Area Bounded by Two Curves Along the x-Axis

Let A be the area of the region that lies between two curves $y = y_T = f(x)$ and $y = y_B = g(x)$ and between the vertical lines $x = a$ and $x = b$, where f and g are continuous functions and $f(x) \geq g(x)$ for all x in $[a, b]$. Then

$$A = \int_a^b [f(x) - g(x)]\, dx = \int_a^b (y_T - y_B)\, dx.$$

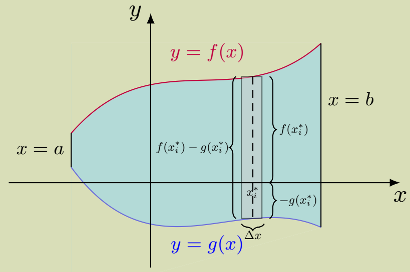
In general, the area between the curves $y = f(x)$ and $y = g(x)$ and between $x = a$ and $x = b$ is

$$A = \int_a^b |f(x) - g(x)|\, dx.$$

Justification

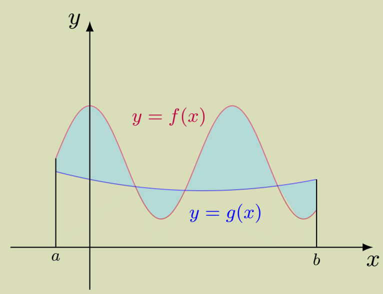
Let S be the region that lies between two curves $y = y_T = f(x)$ and $y = y_B = g(x)$ and between the vertical lines $x = a$ and $x = b$. We divide S into n strips of equal width and then approximate the ith strip by a rectangle with base Δx and height $f(x_i^*) - g(x_i^*)$, where the $x_i^* \in [x_{i-1}, x_i]$. Then the area A of the region S is

$$\begin{aligned} A = \lim_{n\to\infty} \sum_{i=1}^{n} [f(x_i^*) - g(x_i^*)]\, \Delta x &= \int_a^b [f(x) - g(x)]\, dx \\ &= \int_a^b (y_T - y_B)\, dx. \end{aligned}$$



In general, the area consists of two types of contributions: one from the region where $f(x) \geq g(x)$ and the other from the region where $g(x) \geq f(x)$. Thus, by the additivity of definite integral, we have

$$A = \int_a^b |f(x) - g(x)| \, dx.$$



Area Bounded by Two Curves Along the y-Axis

Let A be the area of the region that lies between two curves $x = x_R = f(y)$ and $x = x_L = g(y)$ and between the vertical lines $y = c$ and $y = d$, where f and g are continuous functions and $f(y) \geq g(y)$ for all x in $[c, d]$. Then

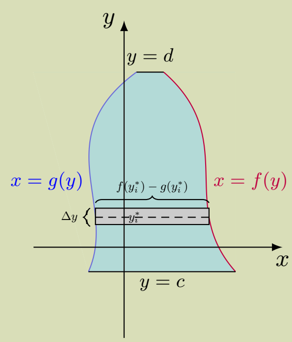
$$A = \int_c^d [f(y) - g(y)] \, dy = \int_c^d (x_R - x_L) \, dy.$$

Justification

Let S be the region that lies between two curves $x = x_R = f(y)$ and $x = x_L = g(y)$ and between the vertical lines $y = c$ and $y = d$. We divide S into n strips of equal width and then approximate the ith strip by a rectangle with base $f(y_i^*) - g(y_i^*)$ and height Δy, where the $y_i^* \in [y_{i-1}, y_i]$. Then the area A

of the region S is

$$A = \lim_{n\to\infty} \sum_{i=1}^{n} [f(y_i^*) - g(y_i^*)]\ \Delta y = \int_c^d [f(y) - g(y)]\ dy$$
$$= \int_c^d (x_R - x_L)\, dy.$$



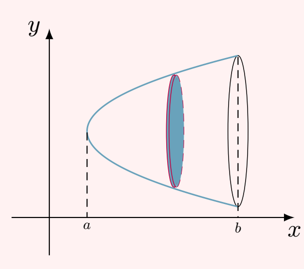
Volume

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

$$V = \lim_{n\to\infty} \sum_{i=1}^{n} A(x_i^*)\Delta x = \int_a^b A(x)\, dx.$$

Similarly, if S is a solid that lies between $y = c$ and $y = d$ and if the cross-sectional area of S through y perpendicular to the y-axis is $A(y)$, where A is a continuous function, then the volume of S is

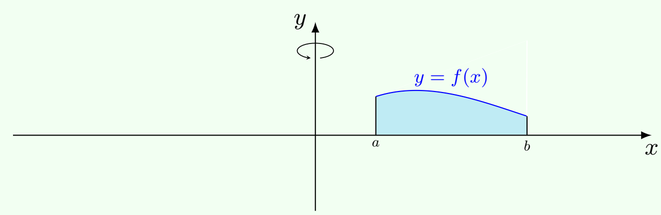
$$V = \lim_{n\to\infty} \sum_{i=1}^{n} A(y_i^*)\Delta y = \int_c^d A(y)\, dy.$$



Volumes by Cylindrical Shells

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

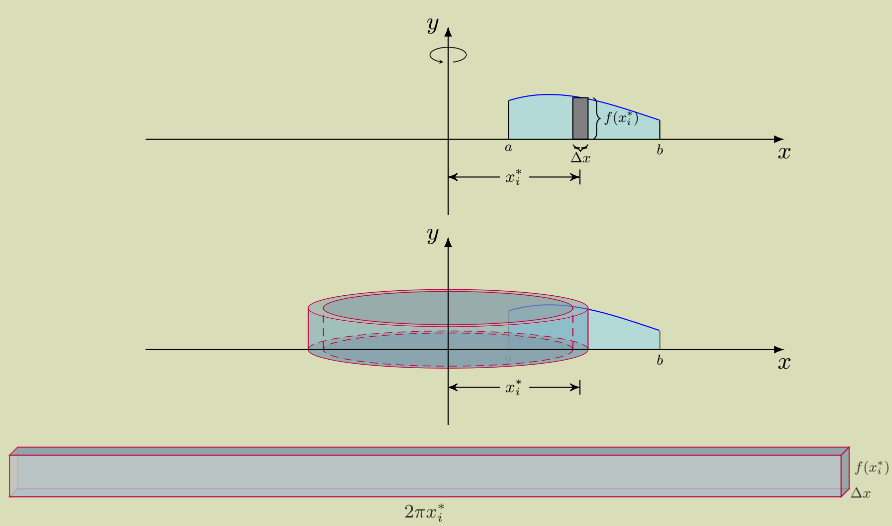
$$V = \int_a^b 2\pi x f(x)\,dx, \qquad \text{where } 0 \le a < b.$$



Justification

We divide the interval $[a, b]$ into n subintervals $[x_{i-1}, x_i]$ of equal width Δx and let x_i^* be a point in the ith subinterval. If the rectangle with base $[x_{i-1}, x_i]$ and height $f(x_i^*)$ is rotated about the y-axis, then the result is a cylindrical shell with average radius x_i^*, height $f(x_i^*)$, and thickness Δx, so its volume is

$$V_i = (\text{circumference}) \times (\text{height}) \times (\text{thickness}) = (2\pi x_i^*) \cdot f(x_i^*) \cdot \Delta x.$$



Therefore an approximation to the volume V of S is given by the sum of the volumes of these shells:

$$V \approx \sum_{i=1}^{n} V_i = \sum_{i=1}^{n} 2\pi x_i^* f(x_i^*) \Delta x.$$

This gives the volume:

$$V = \lim_{n\to\infty} \sum_{i=1}^{n} 2\pi x_i^* f(x_i^*)\Delta x = \int_a^b 2\pi x f(x)\, dx,$$

where the terms in the formula can be interpreted as

x :	radius of a typical cylinder shell
$2\pi x$:	circumference
$f(x)$:	height
dx :	thickness of the shell

$$\int_a^b \underbrace{(2\pi x)}_{\text{circumference}} \quad \underbrace{[f(x)]}_{\text{height}} \quad \underbrace{(dx)}_{\text{thickness}}$$

Work Done in Moving an Object Along a Line - Application

Suppose that the object moves along the x-axis in the positive direction, from $x = a$ to $x = b$, and at each point x between a and b a force $f(x)$ acts on the object, where f is a continuous function. Then the total work by the force is

$$W = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*)\Delta x = \int_a^b f(x)\, dx.$$

Justification

Suppose that the object moves along the x-axis in the positive direction, from $x = a$ to $x = b$, and at each point x between a and b a force $f(x)$ acts on the object, where f is a continuous function. We divide the interval $[a, b]$ into n subintervals with endpoints $x_0, x_1, \ldots, x_n$ and equal width Δx. We choose a sample point x_i^* in the ith subinterval $[x_{i-1}, x_i]$. Then the force at that point is $f(x_i^*)$. If n is large, then Δx is small, and since f is continuous, the values of f don't change very much over the interval $[x_{i-1}, x_i]$. In other words, f is almost constant on the interval and so the work W_i that is done in moving the particle from x_{i-1} to x_i is approximately given by

$$W_i \approx f(x_i^*) \times \Delta x$$

Thus we can approximate the total work by

$$W \approx \sum_{i=1}^{n} f(x_i^*)\Delta x.$$

Therefore we define the work done in moving the object from a to b as the limit of this quantity as $n \to \infty$. Since the right side is a Riemann sum, we recognize its limit as being a definite integral:

$$W = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*)\Delta x = \int_a^b f(x)\, dx.$$

Average Value of of a Function

The **average value of** f on the interval $[a, b]$ is

$$f_{\text{ave}} = \frac{1}{b-a}\int_a^b f(x)\,dx,$$

provided that the integral exists.

Justification

To compute the average value of a function $y = f(x)$ on $[a, b]$, we start by dividing the interval $[a, b]$ into n equal subintervals, each with length $\Delta x = (b - a)/n$. Then we choose points $x_1^*, \ldots, x_n^*$ in successive subintervals and calculate the average of the numbers $f(x_1^*), \ldots, f(x_n^*)$:

$$\frac{f(x_1^*) + \cdots + f(x_n^*)}{n}.$$

Since $n = (b - a)/\Delta x$, the average value becomes

$$\frac{f(x_1^*) + \cdots + f(x_n^*)}{(b-a)/\Delta x} = \frac{1}{b-a}\left[f(x_1^*)\Delta x + \cdots + f(x_n^*)\Delta x\right] = \frac{1}{b-a}\sum_{i=1}^{n} f(x_i^*)\Delta x,$$

so that

$$\lim_{n\to\infty}\frac{1}{b-a}\sum_{i=1}^{n} f(x_i^*)\Delta x = \frac{1}{b-a}\int_a^b f(x)\,dx.$$

Therefore, we define the average value of f on $[a, b]$ as

$$f_{\text{ave}} = \frac{1}{b-a}\int_a^b f(x)\,dx.$$

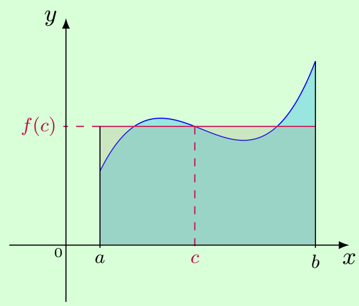
The Mean Value Theorem for Integrals

If f is continuous on $[a, b]$, then there exists a number c in $[a, b]$ such that

$$f(c) = f_{\text{ave}} = \frac{1}{b-a}\int_a^b f(x)\,dx,$$

that is

$$\int_a^b f(x)\,dx = f(c)(b-a).$$



Justification

Let $g(t) = \displaystyle\int_a^t f(x)\,dx$. Since f is continuous on $[a, b]$, g is continuous on $[a, b]$ and differentiable in (a, b). By the Mean Value Theorem, there is a number c in (a, b) such that

$$g(b) - g(a) = g'(c)(b - a).$$

By using

$$g(b) = \int_a^b f(x)\,dx, \qquad g(a) = 0, \qquad g'(c) = f(c),$$

we get

$$\int_a^b f(x)\,dx = f(c)(b - a).$$

Chapter 7

Techniques of Integration

Integration by Parts

Integration by Parts

For indefinite integrals, we have

$$\int f(x)\, g'(x)\, dx = f(x)g(x) - \int g(x)\, f'(x)\, dx$$

or equivalently

$$\int u\, dv = uv - \int v\, du.$$

Integration by Parts for Definite Integrals

For definite integrals, we have

$$\int_a^b f(x)\, g'(x)\, dx = f(x)g(x)\Big]_a^b - \int_a^b g(x)\, f'(x)\, dx.$$

Justification

From the Product Rule, we have

$$[f(x)\, g(x)]' = f(x)\, g'(x) + g(x)\, f'(x).$$

In the notation for indefinite integrals, we have

$$\int [f(x)\, g'(x) + g(x)\, f'(x)]\; dx = \int [f(x)\, g(x)]'\; dx,$$

or

$$\int f(x)\,g'(x)\,dx = \int [f(x)\,g(x)]'\,dx - \int g(x)\,f'(x)\,dx,$$

which gives the formula of integration by parts for indefinite integrals

$$\int f(x)\,g'(x)\,dx = f(x)\,g(x) - \int g(x)\,f'(x)\,dx$$

and the formula for definite integrals

$$\int_a^b f(x)\,g'(x)\,dx = \int_a^b [f(x)\,g(x)]'\,dx - \int_a^b g(x)\,f'(x)\,dx.$$

By applying the Fundamental Theorem of Calculus, the last one gives the formula of integration by parts for definite integrals

$$\int_a^b f(x)\,g'(x)\,dx = f(x)g(x)\Big]_a^b - \int_a^b g(x)\,f'(x)\,dx.$$

Strategy for Evaluating $\int \sin^m x \cos^n x\,dx$

(a) If the power of cosine is odd ($n = 2k+1$), save one cosine factor and use $\cos^2 x = 1 - \sin^2 x$ to express the remaining factors in terms of sine:

$$\begin{aligned}\int \sin^m x\,\cos^{2k+1} x\,dx &= \int \sin^m x\,(\cos^2 x)^k\,\cos x\,dx\\ &= \int \sin^m x\,(1-\sin^2 x)^k\,\cos x\,dx.\end{aligned}$$

Then use the substitution $u = \sin x$ to reduce the last integral to an integral with polynomial integrand.

(b) If the power of sine is odd ($m = 2k+1$), save one sine factor and use $\sin^2 x = 1 - \cos^2 x$ to express the remaining factors in terms of cosine:

$$\begin{aligned}\int \sin^{2k+1} x\,\cos^n x\,dx &= \int (\sin^2 x)^k\,\cos^n x\,\sin x\,dx\\ &= \int (1-\cos^2 x)^k\,\cos^n x\,\sin x\,dx.\end{aligned}$$

Then use the substitution $u = \cos x$ to reduce the last integral to an integral with polynomial integrand.

(c) If the powers of both sine and cosine are odd, use either (a) or (b).

(d) If the powers of both sine and cosine are even, use the half-angle identities,

$$\sin^2 x = \tfrac{1}{2}(1-\cos 2x), \qquad \cos^2 x = \tfrac{1}{2}(1+\cos 2x).$$

It is sometimes helpful to use the identity

$$\sin x \, \cos x = \tfrac{1}{2} \sin 2x.$$

Strategy for Evaluating $\int \tan^m x \sec^n x \, dx$

(a) If the power of secant is even ($n = 2k$), save a factor of $\sec^2 x$ and use $\sec^2 x = 1 + \tan^2 x$ to express the remaining factors in terms of $\tan x$:

$$\begin{aligned}\int \tan^m x \, \sec^{2k} x \, dx &= \int \tan^m x \, (\sec^2 x)^{k-1} \, \sec^2 x \, dx \\ &= \int \tan^m x \, (1 + \tan^2 x)^{k-1} \, \sec^2 x \, dx.\end{aligned}$$

Then use the substitution $u = \tan x$ to reduce the last integral to an integral with polynomial integrand.

(b) If the power of tangent is odd ($m = 2k+1$), save a factor of $\sec x \, \tan x$ and use $\tan^2 x = \sec^2 x - 1$ to express the remaining factors in terms of $\sec x$:

$$\begin{aligned}\int \tan^{2k+1} x \, \sec^n x \, dx &= \int (\tan^2 x)^k \, \sec^{n-1} x \, \sec x \, \tan x \, dx \\ &= \int (\sec^2 x - 1)^k \, \sec^{n-1} x \, \sec x \, \tan x \, dx.\end{aligned}$$

Then use the substitution $u = \sec x$ to reduce the last integral to an integral with polynomial integrand.

Evaluating $\int \sec x \, dx$

By using the substitution $t = \sec x + \tan x$, we can get

$$\int \sec x \, dx = \ln |\sec x + \tan x| + C.$$

Evaluating $\int \csc x \, dx$

By using the substitution $t = \csc x - \cot x$, we can get

$$\int \csc x \, dx = \ln |\csc x - \cot x| + C.$$

Evaluating $\int \sin mx \, \cos nx \, dx$, $\int \sin mx \, \sin nx \, dx$, and $\int \cos mx \, \cos nx \, dx$

Use the identities:

$$\sin\alpha \, \cos\beta = \tfrac{1}{2}\left[\sin(\alpha-\beta)+\sin(\alpha+\beta)\right]$$
$$\sin\alpha \, \sin\beta = \tfrac{1}{2}\left[\cos(\alpha-\beta)-\cos(\alpha+\beta)\right]$$
$$\cos\alpha \, \cos\beta = \tfrac{1}{2}\left[\cos(\alpha-\beta)+\cos(\alpha+\beta)\right]$$

Table of Trigonometric Substitutions

Expression	Substitution	Identity
$\sqrt{a^2-x^2}$	$x = a\sin\theta, \quad -\frac{\pi}{2} \le \theta \le \frac{\pi}{2}$	$1-\sin^2\theta = \cos^2\theta$
$\sqrt{a^2+x^2}$	$x = a\tan\theta, \quad -\frac{\pi}{2} < \theta < \frac{\pi}{2}$	$1+\tan^2\theta = \sec^2\theta$
$\sqrt{x^2-a^2}$	$x = a\sec\theta, \quad 0 \le \theta < \frac{\pi}{2}$ or $\pi \le \theta < \frac{3\pi}{2}$	$\sec^2\theta - 1 = \tan^2\theta$

Integration for Rational Functions by Partial Fractions

For an improper rational function $P(x)/Q(x)$, where P and Q are polynomials with $\deg P \ge \deg Q$, one can use long division to reduce it to the form

$$\frac{P(x)}{Q(x)} = S(x) + \frac{R(x)}{Q(x)},$$

where $S(x)$ and $R(x)$ are polynomials and $R(x)/Q(x)$ is a proper rational function with $\deg R < \deg Q$. Hence, for integrating rational functions, it is essential to how to integrate proper rational functions. For any proper rational function, we can integrate it by using partial fractions in one of the four cases:

Case 1 - The denominator $Q(x)$ is a product of distinct linear factors

We write

$$Q(x) = (a_1x+b_1)(a_2x+b_2)\cdots(a_kx+b_k).$$

In this case, a proper rational function $R(x)/Q(x)$ can be expressed as

$$\frac{R(x)}{Q(x)} = \frac{A_1}{a_1x+b_1} + \frac{A_2}{a_2x+b_2} + \cdots + \frac{A_k}{a_kx+b_k}.$$

For each fraction, we can make a substitution $t = ax+b$ to integrate:

$$\int \frac{A}{ax+b}\,dx = \int \frac{A}{t}\cdot\frac{1}{a}\,dt = \frac{A}{a}\ln|t| + C.$$

Case 2 - $Q(x)$ is a product of linear factors and some of them are repeated

Suppose the function (a_1x+b_1) is repeated r times. Then instead of the single term $\dfrac{A_1}{a_1x+b_1}$, one needs to use the form

$$\frac{A_1}{a_1x+b_1}+\frac{A_2}{(a_1x+b_1)^2}+\cdots+\frac{A_r}{(a_1x+b_1)^r}.$$

For each fraction of the form $\dfrac{A}{(ax+b)^m}$, we can make a substitution to integrate. In fact, if $m=1$, we can integrate the fraction as in Case 1. If $m>1$, by making the substitution $t=ax+b$, we get

$$\int\frac{A}{(ax+b)^m}\,dx=\int\frac{A}{t^m}\cdot\frac{1}{a}\,dt=\frac{A}{a}\cdot\frac{1}{-m+1}t^{-m+1}+C.$$

Case 3 - $Q(x)$ contains irreducible quadratic factors and none of them is repeated

Suppose ax^2+bx+c is an irreducible quadratic factor, with $b^2-4ac<0$. Then the partial fraction for the proper rational function $R(x)/Q(x)$ will have a term of the form

$$\frac{Ax+B}{ax^2+bx+c}$$

whose integration can be re-written as

$$\begin{aligned}\int\frac{Ax+B}{ax^2+bx+c}\,dx&=\frac{A}{2a}\int\frac{(ax^2+bx+c)'}{ax^2+bx+c}\,dx+\left(B-\frac{bA}{2a}\right)\int\frac{1}{ax^2+bx+c}\,dx\\&=\frac{A}{2a}\ln|ax^2+bx+c|+\left(B-\frac{bA}{2a}\right)\int\frac{1}{ax^2+bx+c}\,dx.\end{aligned}$$

For the last integral, by using the substitution $t=x+\frac{b}{2a}$, we can reduce the last integral to the form

$$\int\frac{1}{t^2+\alpha^2}\,dt=\frac{1}{\alpha}\tan^{-1}\left(\frac{t}{\alpha}\right)+C.$$

Case 4 - $Q(x)$ contains a repeated irreducible quadratic factor

Suppose Q contains the factor $(ax^2+bx+c)^r$, where $b^2-4ac<0$. Then instead of the single factor $\dfrac{A_1x+B_1}{ax^2+bx+c}$, one needs to use the form

$$\frac{A_1x+B_1}{ax^2+bx+c}+\frac{A_2x+B_2}{(ax^2+bx+c)^2}+\cdots+\frac{A_rx+B_r}{(ax^2+bx+c)^r}.$$

We re-write the integral of a typical term as

$$\begin{aligned}&\int \frac{Ax+B}{(ax^2+bx+c)^m}\,dx\\&= \frac{A}{2a}\int \frac{(ax^2+bx+c)'}{(ax^2+bx+c)^m}\,dx + \left(B-\frac{bA}{2a}\right)\int \frac{1}{(ax^2+bx+c)^m}\,dx.\end{aligned}$$

In fact, if $m=1$, we can integrate the fraction as in Case 3. If $m>1$, the first integral on the right-hand side is

$$\int \frac{(ax^2+bx+c)'}{(ax^2+bx+c)^m}\,dx = -\frac{1}{(m-1)(ax^2+bx+c)^{m-1}} + C.$$

For the second one, by using the substitution $t = x + \frac{b}{2a}$, we can reduce it to the form

$$\int \frac{1}{(t^2+\alpha^2)^m}\,dt.$$

By integrating by parts, we have

$$\begin{aligned}&\int \frac{1}{(t^2+\alpha^2)^{m-1}}\,dt\\&= \frac{t}{(t^2+\alpha^2)^{m-1}} - \int t\cdot(-m+1)(t^2+\alpha^2)^{-m}\cdot 2t\,dt\\&= \frac{t}{(t^2+\alpha^2)^{m-1}} + 2(m-1)\int (t^2+\alpha^2)^{-m}(t^2+\alpha^2-\alpha^2)\,dt\\&= \frac{t}{(t^2+\alpha^2)^{m-1}} + 2(m-1)\int \frac{1}{(t^2+\alpha^2)^{m-1}}\,dt - 2(m-1)\alpha^2\int \frac{1}{(t^2+\alpha^2)^m}\,dt.\end{aligned}$$

So, we have the recursive formula

$$\int \frac{1}{(t^2+\alpha^2)^m}\,dt = \frac{1}{2(m-1)\alpha^2}\frac{t}{(t^2+\alpha^2)^{m-1}} + \frac{2m-3}{2(m-1)\alpha^2}\int \frac{1}{(t^2+\alpha^2)^{m-1}}\,dt.$$

Using this formula recursively reduces the integral to the integral

$$\int \frac{1}{t^2+\alpha^2}\,dt = \frac{1}{\alpha}\tan^{-1}\left(\frac{t}{\alpha}\right) + C.$$

Table of Integration Formulas

1. $\int x^n\,dx = \frac{x^{n+1}}{n+1} + C \quad (n \neq -1)$

2. $\int \frac{1}{x}\,dx = \ln|x| + C$

3. $\int e^x\,dx = e^x + C$

4. $\int a^x\,dx = \frac{a^x}{\ln a} + C$

5. $\int \sin x\,dx = -\cos x + C$

6. $\int \cos x\,dx = \sin x + C$

7. $\int \sec^2 x\,dx = \tan x + C$

8. $\int \csc^2 x\,dx = -\cot x + C$

9. $\int \sec x\,\tan x\,dx = \sec x + C$

10. $\int \csc x\,\cot x\,dx = -\csc x + C$

11. $\int \sec x\,dx = \ln|\sec x + \tan x| + C$

12. $\int \csc x\,dx = \ln|\csc x - \cot x| + C$

13. $\int \tan x\,dx = \ln|\sec x| + C$

14. $\int \cot x\,dx = \ln|\sin x| + C$

15. $\int \sinh x\,dx = \cosh x + C$

16. $\int \cosh\,dx = \sinh x + C$

17. $\int \frac{1}{x^2+a^2}\,dx = \frac{1}{a}\tan^{-1}\left(\frac{x}{a}\right) + C$

18. $\int \frac{1}{\sqrt{a^2-x^2}}\,dx = \sin^{-1}\left(\frac{x}{|a|}\right) + C$

19. $\int \frac{1}{x^2-a^2}\,dx = \frac{1}{2a}\ln\left|\frac{x-a}{x+a}\right| + C$

20. $\int \frac{1}{\sqrt{x^2 \pm a^2}}\,dx = \ln\left|x + \sqrt{x^2 \pm a^2}\right| + C$

Table of Integrals

Basic Forms

1. $\int u\,dv = uv - \int v\,du$

2. $\int u^n\,du = \frac{u^{n+1}}{n+1} + C, \quad n \neq -1$

3. $\int \frac{du}{u} = \ln|u| + C$

4. $\int e^u\,du = e^u + C$

5. $\int a^u\,dx = \frac{a^u}{\ln a} + C$

6. $\int \sin u\,du = -\cos u + C$

7. $\int \cos u\,du = \sin u + C$

8. $\int \sec^2 u\,du = \tan u + C$

9. $\int \csc^2 u\,dx = -\cot u + C$

10. $\int \sec u\,\tan u\,du = \sec u + C$

11. $\int \csc u\,\cot u\,du = -\csc u + C$

12. $\int \tan u\,du = \ln|\sec u| + C$

13. $\int \cot u\,du = \ln|\sin u| + C$

14. $\int \sec u\,du = \ln|\sec u + \tan u| + C$

15. $\int \csc u\,du = \ln|\csc u - \cot u| + C$

16. $\int \frac{du}{\sqrt{a^2-u^2}} = \sin^{-1}\frac{x}{a} + C, \quad a > 0$

17. $\int \frac{du}{a^2+u^2} = \frac{1}{a}\tan^{-1}\frac{u}{a} + C$

18. $\int \frac{du}{u\sqrt{u^2-a^2}} = \frac{1}{a}\sec^{-1}\frac{u}{a} + C$

19. $\int \frac{du}{a^2-u^2} = \frac{1}{2a}\ln\left|\frac{u+a}{u-a}\right| + C$

20. $\int \frac{du}{u^2-a^2} = \frac{1}{2a}\ln\left|\frac{u-a}{u+a}\right| + C$

Table of Integrals

Forms Involving $\sqrt{a^2+u^2}$, $a>0$

21. $\displaystyle\int \sqrt{a^2+u^2}\,du = \frac{u}{2}\sqrt{a^2+u^2} + \frac{a^2}{2}\ln\left(u+\sqrt{a^2+u^2}\right) + C$

22. $\displaystyle\int u^2\sqrt{a^2+u^2}\,du = \frac{u}{8}(a^2+2u^2)\sqrt{a^2+u^2} - \frac{a^4}{8}\ln\left(u+\sqrt{a^2+u^2}\right) + C$

23. $\displaystyle\int \frac{\sqrt{a^2+u^2}}{u}\,du = \sqrt{a^2+u^2} - a\ln\left|\frac{a+\sqrt{a^2+u^2}}{u}\right| + C$

24. $\displaystyle\int \frac{\sqrt{a^2+u^2}}{u^2}\,du = -\frac{\sqrt{a^2+u^2}}{u} + \ln\left(u+\sqrt{a^2+u^2}\right) + C$

25. $\displaystyle\int \frac{du}{\sqrt{a^2+u^2}} = \ln\left(u+\sqrt{a^2+u^2}\right) + C$

26. $\displaystyle\int \frac{u^2\,du}{\sqrt{a^2+u^2}} = \frac{u}{2}\sqrt{a^2+u^2} - \frac{a^2}{2}\ln\left(u+\sqrt{a^2+u^2}\right) + C$

27. $\displaystyle\int \frac{du}{u\sqrt{a^2+u^2}} = -\frac{1}{a}\ln\left|\frac{\sqrt{a^2+u^2}+a}{u}\right| + C$

28. $\displaystyle\int \frac{du}{u^2\sqrt{a^2+u^2}} = -\frac{\sqrt{a^2+u^2}}{a^2u} + C$

29. $\displaystyle\int \frac{du}{(a^2+u^2)^{3/2}} = \frac{u}{a^2\sqrt{a^2+u^2}} + C$

Table of Integrals

Forms Involving $\sqrt{a^2-u^2}$, $a>0$

30. $\displaystyle\int \sqrt{a^2-u^2}\,du = \frac{u}{2}\sqrt{a^2-u^2} + \frac{a^2}{2}\sin^{-1}\frac{u}{a} + C$

31. $\displaystyle\int u^2\sqrt{a^2-u^2}\,du = \frac{u}{8}(2u^2-a^2)\sqrt{a^2-u^2} + \frac{a^4}{8}\sin^{-1}\frac{u}{a} + C$

32. $\displaystyle\int \frac{\sqrt{a^2-u^2}}{u}\,du = \sqrt{a^2-u^2} - a\ln\left|\frac{a+\sqrt{a^2-u^2}}{u}\right| + C$

33. $\displaystyle\int \frac{\sqrt{a^2-u^2}}{u^2}\,du = -\frac{\sqrt{a^2-u^2}}{u} - \sin^{-1}\frac{u}{a} + C$

34. $\displaystyle\int \frac{u^2\,du}{\sqrt{a^2-u^2}} = -\frac{u}{2}\sqrt{a^2-u^2} + \frac{a^2}{2}\sin^{-1}\frac{u}{a} + C$

35. $\displaystyle\int \frac{du}{u\sqrt{a^2-u^2}} = -\frac{1}{a}\ln\left|\frac{a+\sqrt{a^2-u^2}}{u}\right| + C$

36. $\displaystyle\int \frac{du}{u^2\sqrt{a^2-u^2}} = -\frac{\sqrt{a^2-u^2}}{a^2u} + C$

37. $\displaystyle\int (a^2-u^2)^{3/2}\,du = -\frac{u}{8}(2u^2-5a^2)\sqrt{a^2-u^2} + \frac{3a^4}{8}\sin^{-1}\frac{u}{a} + C$

38. $\displaystyle\int \frac{du}{(a^2-u^2)^{3/2}} = \frac{u}{a^2\sqrt{a^2-u^2}} + C$

Table of Integrals

Forms Involving $\sqrt{u^2 - a^2}$, $a > 0$

39. $\displaystyle\int \sqrt{u^2 - a^2}\, du = \frac{u}{2}\sqrt{u^2 - a^2} - \frac{a^2}{2} \ln \left|u + \sqrt{u^2 - a^2}\right| + C$

40. $\displaystyle\int u^2\sqrt{u^2 - a^2}\, du = \frac{u}{8}(2u^2 - a^2)\sqrt{u^2 - a^2} - \frac{a^4}{8} \ln \left|u + \sqrt{u^2 - a^2}\right| + C$

41. $\displaystyle\int \frac{\sqrt{u^2 - a^2}}{u}\, du = \sqrt{u^2 - a^2} - a\cos^{-1} \frac{a}{|u|} + C$

42. $\displaystyle\int \frac{\sqrt{u^2 - a^2}}{u^2}\, du = -\frac{\sqrt{u^2 - a^2}}{u} + \ln \left|u + \sqrt{u^2 - a^2}\right| + C$

43. $\displaystyle\int \frac{du}{\sqrt{u^2 - a^2}} = \ln \left|u + \sqrt{u^2 - a^2}\right| + C$

44. $\displaystyle\int \frac{u^2\, du}{\sqrt{u^2 - a^2}} = \frac{u}{2}\sqrt{u^2 - a^2} + \frac{a^2}{2} \ln \left|u + \sqrt{u^2 - a^2}\right| + C$

45. $\displaystyle\int \frac{du}{u^2\sqrt{u^2 - a^2}} = \frac{\sqrt{u^2 - a^2}}{a^2 u} + C$

46. $\displaystyle\int \frac{du}{(u^2 - a^2)^{3/2}} = -\frac{u}{a^2\sqrt{u^2 - a^2}} + C$

Table of Integrals

Forms Involving $a + bu$

47. $\displaystyle\int \frac{u\,du}{a+bu} = \frac{1}{b^2}\left(a+bu-a\ln|a+bu|\right)+C$

48. $\displaystyle\int \frac{u^2\,du}{a+bu} = \frac{1}{2b^3}\left[(a+bu)^2-4a(a+bu)+2a^2\ln|a+bu|\right]+C$

49. $\displaystyle\int \frac{du}{u(a+bu)} = \frac{1}{a}\ln\left|\frac{u}{a+bu}\right|+C$

50. $\displaystyle\int \frac{du}{u^2(a+bu)} = -\frac{1}{au}+\frac{b}{a^2}\ln\left|\frac{a+bu}{u}\right|+C$

51. $\displaystyle\int \frac{u\,du}{(a+bu)^2} = \frac{a}{b^2(a+bu)}+\frac{1}{b^2}\ln|a+bu|+C$

52. $\displaystyle\int \frac{du}{u(a+bu)^2} = \frac{1}{a(a+bu)}-\frac{1}{a^2}\ln\left|\frac{a+bu}{u}\right|+C$

53. $\displaystyle\int \frac{u^2\,du}{(a+bu)^2} = \frac{1}{b^3}\left(a+bu-\frac{a^2}{a+bu}-2a\ln|a+bu|\right)+C$

54. $\displaystyle\int u\sqrt{a+bu}\,du = \frac{2}{15b^2}(3bu-2a)(a+bu)^{3/2}+C$

55. $\displaystyle\int \frac{u\,du}{\sqrt{a+bu}} = \frac{2}{3b^2}(bu-2a)\sqrt{a+bu}+C$

56. $\displaystyle\int \frac{u^2\,du}{\sqrt{a+bu}} = \frac{2}{15b^3}(8a^2+3b^2u^2-4abu)\sqrt{a+bu}+C$

57. $\displaystyle\int \frac{du}{u\sqrt{a+bu}} = \begin{cases} \dfrac{1}{\sqrt{a}}\ln\left|\dfrac{\sqrt{a+bu}-\sqrt{a}}{\sqrt{a+bu}+\sqrt{a}}\right|+C, & \text{if } a>0, \\[2ex] \dfrac{2}{\sqrt{-a}}\tan^{-1}\sqrt{\dfrac{a+bu}{-a}}+C, & \text{if } a<0 \end{cases}$

58. $\displaystyle\int \frac{\sqrt{a+bu}}{u}\,du = 2\sqrt{a+bu}+a\int\frac{du}{u\sqrt{a+bu}}$

59. $\displaystyle\int \frac{\sqrt{a+bu}}{u^2}\,du = -\frac{\sqrt{a+bu}}{u}+\frac{b}{2}\int\frac{du}{u\sqrt{a+bu}}$

60. $\displaystyle\int u^n\sqrt{a+bu}\,du = \frac{2}{b(2n+3)}\left[u^n(a+bu)^{3/2}-na\int u^{n-1}\sqrt{a+bu}\,du\right]$

61. $\displaystyle\int \frac{u^n\,du}{\sqrt{a+bu}} = \frac{2u^n\sqrt{a+bu}}{b(2n+1)}-\frac{2na}{b(2n+1)}\int\frac{u^{n-1}\,du}{\sqrt{a+bu}}$

62. $\displaystyle\int \frac{du}{u^n\sqrt{a+bu}} = -\frac{\sqrt{a+bu}}{a(n-1)u^{n-1}}-\frac{b(2n-3)}{2a(n-1)}\int\frac{du}{u^{n-1}\sqrt{a+bu}}$

Table of Integrals

Trigonometric Forms

63. $\int \sin^2 u\, du = \frac{1}{2}u - \frac{1}{4}\sin 2u + C$

64. $\int \cos^2 u\, du = \frac{1}{2}u + \frac{1}{4}\sin 2u + C$

65. $\int \tan^2 u\, du = \tan u - u + C$

66. $\int \cot^2 u\, du = -\cot u - u + C$

67. $\int \sin^3 u\, du = -\frac{1}{3}(2 + \sin^2 u)\cos u + C$

68. $\int \cos^3 u\, du = \frac{1}{3}(2 + \cos^2 u)\sin u + C$

69. $\int \tan^3 u\, du = \frac{1}{2}\tan^2 u + \ln|\cos u| + C$

70. $\int \cot^3 u\, du = -\frac{1}{2}\cot^2 u - \ln|\sin u| + C$

71. $\int \sec^3 u\, du = \frac{1}{2}\sec u\,\tan u + \frac{1}{2}\ln|\sec u + \tan u| + C$

72. $\int \csc^3 u\, du = -\frac{1}{2}\csc u\,\cot u + \frac{1}{2}\ln|\csc u - \cot u| + C$

73. $\int \sin^n u\, du = -\frac{1}{n}\sin^{n-1} u\,\cos u + \frac{n-1}{n}\int \sin^{n-2} u\, du$

74. $\int \cos^n u\, du = \frac{1}{n}\cos^{n-1} u\,\sin u + \frac{n-1}{n}\int \cos^{n-2} u\, du$

75. $\int \tan^n u\, du = \frac{1}{n-1}\tan^{n-1} u - \int \tan^{n-2} u\, du$

76. $\int \cot^n u\, du = \frac{-1}{n-1}\cot^{n-1} u - \int \cot^{n-1} u\, du$

77. $\int \sec^n u\, du = \frac{1}{n-1}\tan u\,\sec^{n-2} u + \frac{n-2}{n-1}\int \sec^{n-2} u\, du$

78. $\int \csc^n u\, du = \frac{-1}{n-1}\cot u\,\csc^{n-2} u + \frac{n-2}{n-1}\int \csc^{n-2} u\, du$

79. $\int \sin au\,\sin bu\, du = \frac{\sin(a-b)u}{2(a-b)} - \frac{\sin(a+b)u}{2(a+b)} + C$

80. $\int \cos au\,\cos bu\, du = \frac{\sin(a-b)u}{2(a-b)} + \frac{\sin(a+b)u}{2(a+b)} + C$

81. $\int \sin au\,\cos bu\, du = -\frac{\cos(a-b)u}{2(a-b)} - \frac{\cos(a+b)u}{2(a+b)} + C$

82. $\int u\sin u\, du = \sin u - u\cos u + C$

83. $\int u\cos u\, du = \cos u + u\sin u + C$

84. $\int u^n \sin u\, du = -u^n\cos u + n\int u^{n-1}\cos u\, du$

85. $\int u^n \cos u\, du = u^n\sin u - n\int u^{n-1}\sin u\, du$

86.
$$\int \sin^n u\,\cos^m u\, du = -\frac{\sin^{n-1} u\,\cos^{m+1} u}{n+m} + \frac{n-1}{n+m}\int \sin^{n-2} u\,\cos^m u\, du$$
$$= \frac{\sin^{n+1} u\,\cos^{m-1} u}{n+m} + \frac{m-1}{n+m}\int \sin^n u\,\cos^{m-2} u\, du$$

Table of Integrals

Inverse Trigonometric Forms

87. $\int \sin^{-1} u\, du = u\sin^{-1} u + \sqrt{1-u^2} + C$

88. $\int \cos^{-1} u\, du = u\cos^{-1} u - \sqrt{1-u^2} + C$

89. $\int \tan^{-1} u\, du = u\tan^{-1} u - \frac{1}{2}\ln(1+u^2) + C$

90. $\int u\sin^{-1} u\, du = \frac{2u^2-1}{4}\sin^{-1} u + \frac{u\sqrt{1-u^2}}{4} + C$

91. $\int u\cos^{-1} u\, du = \frac{2u^2-1}{4}\cos^{-1} u - \frac{u\sqrt{1-u^2}}{4} + C$

92. $\int u\tan^{-1} u\, du = \frac{u^2+1}{2}\tan^{-1} u - \frac{u}{2} + C$

93. $\int u^n\sin^{-1} u\, du = \frac{1}{n+1}\left[u^{n+1}\sin^{-1} u - \int \frac{u^{n+1}\, du}{\sqrt{1-u^2}}\right], \quad n \neq -1$

94. $\int u^n\cos^{-1} u\, du = \frac{1}{n+1}\left[u^{n+1}\cos^{-1} u + \int \frac{u^{n+1}\, du}{\sqrt{1-u^2}}\right], \quad n \neq -1$

95. $\int u^n\tan^{-1} u\, du = \frac{1}{n+1}\left[u^{n+1}\tan^{-1} u - \int \frac{u^{n+1}\, du}{1+u^2}\right], \quad n \neq -1$

Table of Integrals

Exponential and Logarithmic Forms

96. $\displaystyle\int ue^{au}\,du = \frac{1}{a^2}(au-1)e^{au} + C$

97. $\displaystyle\int u^n e^{au}\,du = \frac{1}{a}\,u^n e^{au} - \frac{n}{a}\int u^{n-1}e^{au}\,du$

98. $\displaystyle\int e^{au}\,\sin bu\,du = \frac{e^{au}}{a^2+b^2}(a\sin bu - b\cos bu) + C$

99. $\displaystyle\int e^{au}\,\cos bu\,du = \frac{e^{au}}{a^2+b^2}(a\cos bu + b\sin bu) + C$

100. $\displaystyle\int \ln u\,du = u\ln u - u + C$

101. $\displaystyle\int u^n \ln u\,du = \frac{u^{n+1}}{(n+1)^2}\,[(n+1)\ln u - 1]$

102. $\displaystyle\int \frac{1}{u\ln u}\,du = \ln|\ln u| + C$

Table of Integrals

Hyperbolic Forms

103. $\displaystyle\int \sinh u\,du = \cosh u + C$

104. $\displaystyle\int \cosh u\,du = \sinh u + C$

105. $\displaystyle\int \tanh u\,du = \ln\cosh u + C$

106. $\displaystyle\int \coth u\,du = \ln|\sinh u| + C$

107. $\displaystyle\int \operatorname{sech} u\,du = \tanh^{-1}|\sinh u| + C$

108. $\displaystyle\int \operatorname{csch} u\,du = \ln|\tanh \tfrac{1}{2}u| + C$

109. $\displaystyle\int \operatorname{sech}^2 u\,du = \tanh u + C$

110. $\displaystyle\int \operatorname{csch}^2 u\,du = -\coth u + C$

111. $\displaystyle\int \operatorname{sech} u\,\tanh u\,du = -\operatorname{sech} u + C$

112. $\displaystyle\int \operatorname{csch} u\,\coth u\,du = -\operatorname{csch} u + C$

Table of Integrals

Forms Involving $\sqrt{2au-u^2}$, $a>0$

113. $\displaystyle\int \sqrt{2au-u^2}\,du = \frac{u-a}{2}\sqrt{2au-u^2} + \frac{a^2}{2}\cos^{-1}\left(\frac{a-u}{a}\right) + C$

114. $\displaystyle\int u\sqrt{2au-u^2}\,du = \frac{2u^2-au-3a^2}{6}\sqrt{2au-u^2} + \frac{a^3}{2}\cos^{-1}\left(\frac{a-u}{a}\right) + C$

115. $\displaystyle\int \frac{\sqrt{2au-u^2}}{u}\,du = \sqrt{2au-u^2} + a\cos^{-1}\left(\frac{a-u}{a}\right) + C$

116. $\displaystyle\int \frac{\sqrt{2au-u^2}}{u^2}\,du = -\frac{2\sqrt{2au-u^2}}{u} - \cos^{-1}\left(\frac{a-u}{a}\right) + C$

117. $\displaystyle\int \frac{du}{\sqrt{2au-u^2}} = \cos^{-1}\left(\frac{a-u}{a}\right) + C$

118. $\displaystyle\int \frac{u\,du}{\sqrt{2au-u^2}} = -\sqrt{2au-u^2} + a\cos^{-1}\left(\frac{a-u}{a}\right) + C$

119. $\displaystyle\int \frac{u^2\,du}{\sqrt{2au-u^2}} = -\frac{u+3a}{2}\sqrt{2au-u^2} + \frac{3a^2}{2}\cos^{-1}\left(\frac{a-u}{a}\right) + C$

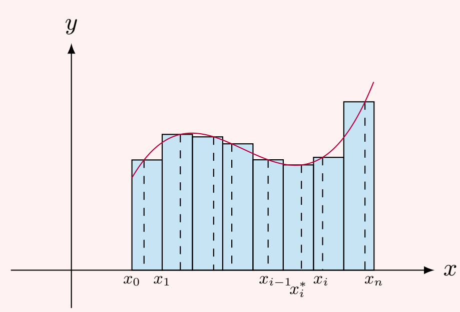
120. $\displaystyle\int \frac{du}{u\sqrt{2au-u^2}} = -\frac{\sqrt{2au-u^2}}{au} + C$

Definite Integral

If f is a function defined for $a \le x \le b$, we divide the interval $[a,b]$ into n subintervals of equal width $\Delta x = (b-a)/n$. We let $x_0(=a), x_1, x_2, \ldots, x_n(=b)$ be the endpoints of these subintervals and we let $x_1^*, x_2^*, \ldots, x_n^*$ be any **sample points** in these subintervals, so x_i^* lies in the ith subinterval $[x_{i-1}, x_i]$. Then the **definite integral of f from a to b** is

$$\int_a^b f(x)\,dx = \lim_{n\to\infty}\sum_{i=1}^{n} f(x_i^*)\Delta x$$

provided that this limit exists and gives the same value for all possible choices of sample points. If it does exist, we say that f is **integrable** on $[a,b]$.



More precise, the last limit means that

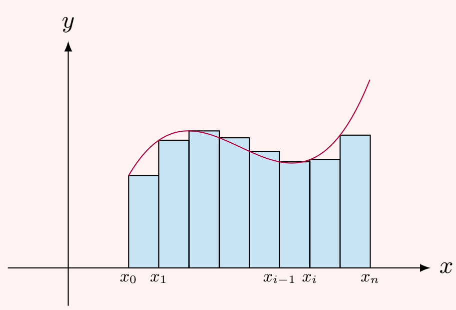
For every number $\varepsilon > 0$ there is a integer N such that
$$\left| \sum_{i=1}^{n} f(x_i^*)\Delta x - \int_a^b f(x)\,dx \right| < \varepsilon$$
for every integer $n > N$ and for every choice of x_i^* in $[x_{i-1}, x_i]$.

In the definition, the sum $\sum_{i=1}^{n} f(x_i^*)\Delta x$ is called a **Riemann sum**, the function f is the **integrand**, a is the **lower limit**, and b is the **upper limit**.
Geometrically, if f is a nonnegative and continuous, the definite integral is the area of the region that lies under the graph of f.

Left Endpoint Approximation

If $x_i^* = x_{i-1}$, $i = 1, 2, \ldots, n$, then the Riemann sum is called a **left endpoint approximation**:

$$\int_a^b f(x)\,dx \approx L_n = \Delta x \left[f(x_0) + f(x_1) + f(x_2) + \cdots + f(x_{n-1}) \right].$$

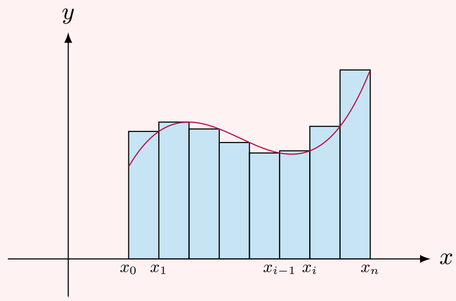


Left Endpoint Approximation L_n

Right Endpoint Approximation

If $x_i^* = x_i$, $i = 1, 2, \ldots, n$, then the Riemann sum is called a **right endpoint approximation**:

$$\int_a^b f(x)\,dx \approx R_n = \Delta x \left[f(x_1) + f(x_2) + f(x_3) + \cdots + f(x_n) \right].$$

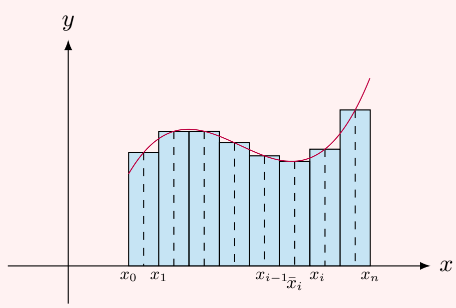


Right Endpoint Approximation R_n

Midpoint Approximation

If $x_i^* = \frac{1}{2}(x_{i-1} + x_i) = \bar{x}_i$, $i = 1, 2, \ldots, n$, then the Riemann sum is called a **midpoint rule** (or **midpoint approximation**):

$$\int_a^b f(x)\,dx \approx M_n = \Delta x\,[f(\bar{x}_1) + f(\bar{x}_2) + f(\bar{x}_3) + \cdots + f(\bar{x}_n)]\,.$$

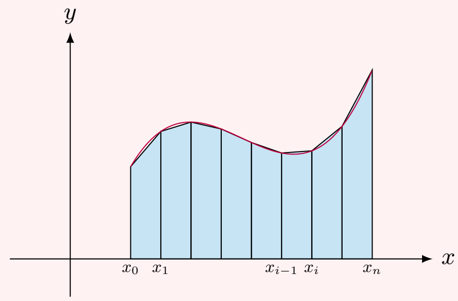


Midpoint Approximation M_n

Trapezoidal Approximation

If we use $\frac{1}{2}[f(x_{i-1}) + f(x_i)] \approx f(x_i^*)$, then the Riemann sum is called a **trapezoidal rule** (or **trapezoidal approximation**). This is because the term $\frac{1}{2}[f(x_{i-1}) + f(x_i)]\,\Delta x$ equals the area of the trapezoid with the vertices $(x_{i-1}, 0)$, $(x_i, 0)$, $(x_i, f(x_i))$, and $(x_{i-1}, f(x_{i-1}))$. In this case, the sum is

$$\int_a^b f(x)\,dx \approx T_n = \frac{\Delta x}{2}\,[f(x_0) + 2f(x_1) + 2f(x_2) + \cdots + 2f(x_{n-1}) + f(x_n)]\,.$$



Trapezoidal Approximation T_n

Simpson's Approximation

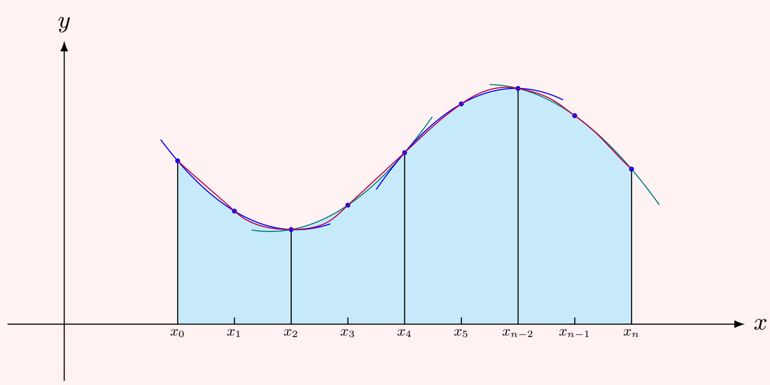
We divide $[a, b]$ into an **even** number n of subintervals of equal length $\Delta x = \dfrac{b-a}{n}$ by using $n+1$ points

$$x_0 = a, \quad x_1 = a + \Delta x, \quad x_2 = a + 2\Delta x, \quad \ldots, \quad x_n = a + n\Delta x = b.$$

For each consecutive pair of subintervals, we apply the the area under the parabola through the points $(x_i, f(x_i))$, $(x_{i+1}, f(x_{i+1}))$, and $(x_{i+2}, f(x_{i+2}))$, where $i = 0, 2, 4, \ldots, n-2$. Adding all the areas gives

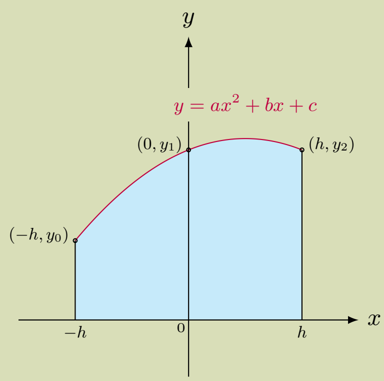
$$\begin{aligned}
&\int_a^b f(x)\,dx \\
&\approx \tfrac{1}{3}h(y_0 + 4y_1 + y_2) + \tfrac{1}{3}h(y_2 + 4y_3 + y_4) + \cdots + \tfrac{1}{3}h(y_{n-2} + 4y_{n-1} + y_n) \\
&= \tfrac{1}{3}h(y_0 + 4y_1 + 2y_2 + 4y_3 + 2y_4 + \cdots + 2y_{n-2} + 4y_{n-1} + y_n) = S_n.
\end{aligned}$$

This is called a **Simpson's rule** (or **Simpson's approximation**).



Justification - Part 1

Let us first derive a formula for the area under a parabola of equation $y = ax^2 + bx + c$ passing through the three points $(-h, y_0)$, $(0, y_1)$, and (h, y_2).



The area is

$$A = \int_{-h}^{h} (ax^2 + bx + c)\,dx = \left[\tfrac{1}{3}ax^3 + \tfrac{1}{2}bx^2 + cx\right]_{x=-h}^{h}$$
$$= \tfrac{2}{3}ah^3 + 2ch = \tfrac{1}{3}h(2ah^2 + 6c).$$

Justification - Part 2

Since the parabola passes through $(-h, y_0)$, $(0, y_1)$, and (h, y_2), we have

$$\begin{aligned} y_0 &= a(-h)^2 + b(-h) + c, \\ y_1 &= c, \\ y_2 &= ah^2 + bh + c. \end{aligned}$$

Since

$$y_0 + 4y_1 + y_2 = \left[a(-h)^2 + b(-h) + c\right] + 4c + (ah^2 + bh + c) = 2ah^2 + 6c,$$

which is the same the area under the parabola, so

$$A = \tfrac{1}{3}h(y_0 + 4y_1 + y_2).$$

This formula shows that the area under the parabola depends only on the values of y_0, y_1, y_2, and h. It is clear that shifting this parabola horizontally does not change the area under it.

Justification - Part 3

For an integral $\displaystyle\int_a^b f(x)\,dx$, we divide $[a,b]$ into an **even** number n of subintervals of equal length $\Delta x = \dfrac{b-a}{n}$ by using $n+1$ points

$$x_0 = a, \quad x_1 = a + \Delta x, \quad x_2 = a + 2\Delta x, \quad \ldots, \quad x_n = a + n\Delta x = b.$$

For each consecutive pair of subintervals, we apply the the area under the parabola. Adding all the areas gives Simpson's approximation:

$$\begin{aligned}
&\int_a^b f(x)\,dx \\
&\approx \tfrac{1}{3}h(y_0 + 4y_1 + y_2) + \tfrac{1}{3}h(y_2 + 4y_3 + y_4) + \cdots + \tfrac{1}{3}h(y_{n-2} + 4y_{n-1} + y_n) \\
&= \tfrac{1}{3}h(y_0 + 4y_1 + 2y_2 + 4y_3 + 2y_4 + \cdots + 2y_{n-2} + 4y_{n-1} + y_n) = S_n.
\end{aligned}$$

Error Bounds for the Trapezoidal and Midpoint Rules

Suppose $|f''(x)| \le K$ for $a \le x \le b$. If E_T and E_M are the errors in the Trapezoidal and Midpoint Rules, then

$$|E_T| \le \frac{K(b-a)^3}{12n^2} \quad \text{and} \quad |E_M| \le \frac{K(b-a)^3}{24n^2}.$$

Error Bound for Simpson's Rule

Suppose $|f^{(4)}(x)| \le K$ for $a \le x \le b$. If E_S is the error involved in using Simpson's Rule, then

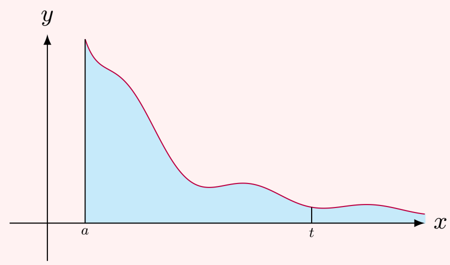
$$|E_S| \le \frac{K(b-a)^5}{180n^4}.$$

Improper Integral of Type 1

(a) If $\int_a^t f(x)\,dx$ exists for every number $t \ge a$, then

$$\int_a^\infty f(x)\,dx = \lim_{t\to\infty} \int_a^t f(x)\,dx,$$

provided this limit exists (as a finite number).



(b) If $\int_t^b f(x)\,dx$ exists for every number $t \leq b$, then

$$\int_{-\infty}^{b} f(x)\,dx = \lim_{t\to-\infty} \int_t^b f(x)\,dx,$$

provided this limit exists (as a finite number).

The improper integrals $\int_a^{\infty} f(x)\,dx$ and $\int_{-\infty}^{b} f(x)\,dx$ are called **convergent** if the corresponding limit exists and **divergent** if the limit does not exist.

(c) If $\int_a^{\infty} f(x)\,dx$ and $\int_{-\infty}^{a} f(x)\,dx$ are convergent, then we define

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{a} f(x)\,dx + \int_a^{\infty} f(x)\,dx.$$

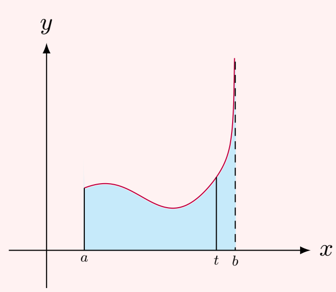
In part (c) any real number a can be used.

Improper Integral of Type 2

(a) If f is continuous on $[a, b)$ and is discontinuous at b, then

$$\int_a^b f(x)\,dx = \lim_{t\to b^-} \int_a^t f(x)\,dx,$$

provided this limit exists (as a finite number).



(b) If f is continuous on $(a, b]$ and is discontinuous at a, then

$$\int_a^b f(x)\,dx = \lim_{t \to a^+} \int_t^b f(x)\,dx,$$

provided this limit exists (as a finite number).
The improper integral $\int_a^b f(x)\,dx$ is called **convergent** if the corresponding limit exists and **divergent** if the limit does not exist.
(c) If f has a discontinuity at c, where $a < c < b$, and both $\int_a^c f(x)\,dx$ and $\int_c^b f(x)\,dx$ are convergent, then we define

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$$

Comparison Theorem

Suppose that f and g are continuous functions with $f(x) \geq g(x) \geq 0$ for $x \geq a$.
(a) If $\displaystyle\int_a^\infty f(x)\,dx$ is convergent, then $\displaystyle\int_a^\infty g(x)\,dx$ is convergent.
(b) If $\displaystyle\int_a^\infty g(x)\,dx$ is divergent, then $\displaystyle\int_a^\infty f(x)\,dx$ is divergent.
A similar Theorem is true for Type 2 integrals.

Justification

(a) If f and is continuous with $f(x) \geq 0$ for $x \geq a$, then $\displaystyle\int_a^t f(x)\,dx$ is increasing in t. Since $\displaystyle\int_a^\infty f(x)\,dx$ is convergent, we know that there is a constant M such that

$$\int_a^t f(x)\,dx < M, \qquad \text{for all } t \geq a.$$

The hypothesis $f(x) \geq g(x) \geq 0$ for $x \geq a$ implies that

$$\int_a^t g(x)\,dx \leq \int_a^t f(x)\,dx < M, \qquad \text{for all } t \geq a.$$

Again, since $\displaystyle\int_a^t g(x)\,dx$ is increasing in t, the inequality implies that $\displaystyle\int_a^\infty g(x)\,dx$ is convergent.
(b) If $\displaystyle\int_a^\infty g(x)\,dx$ is divergent, we know that $\displaystyle\int_a^t f(x)\,dx$ is unbounded as $t \to \infty$ since g is non-negative.
The hypothesis $f(x) \geq g(x) \geq 0$ for $x \geq a$ implies that $\displaystyle\int_a^t f(x)\,dx$ is also unbounded as $t \to \infty$. So, $\displaystyle\int_a^\infty f(x)\,dx$ is divergent.

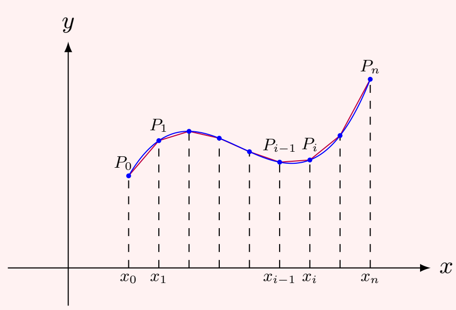
Chapter 8

Further Applications of Integration

Arc Length

Suppose s curve is defined by $y = f(x)$, where f is continuous and $a \leq x \leq b$. We approximate C by a collection of line segments, by dividing the interval $[a, b]$ into n subintervals with end points $x_0, x_1, \ldots, x_n$ and equal width Δx. Denote the endpoints of the line segments to be $P_0, P_1, \ldots, P_n$, where P_i lines on the curve C with coordinates $(x_i, f(x_i))$. Then, the length L os the curve C is defined as

$$L = \lim_{n\to\infty} \sum_{i=1}^{n} |P_{i-1}P_i|.$$



Arc Length of a Smooth Curve

If f' is continuous on $[a,b]$, then the length of the curve $y=f(x)$, $a \le x \le b$, is

$$L=\int_a^b \sqrt{1+\left(\frac{dy}{dx}\right)^2}\,dx=\int_a^b \sqrt{1+[f'(x)]^2}\,dx.$$

Similarly, If g' is continuous on $[c,d]$, then the length of the curve $x=g(y)$, $c \le y \le d$, is

$$L=\int_c^d \sqrt{1+\left(\frac{dx}{dy}\right)^2}\,dy=\int_c^d \sqrt{1+[g'(y)]^2}\,dy.$$

The Arc Length Function

If f' is continuous on $[a,b]$, the distance along the curve from the initial point $(a,f(a))$ to the point $(x,f(x))$ defines the **arc length function**:

$$s(x)=\int_a^x \sqrt{1+[f'(t)]^2}\,dt.$$

Justification

Let $\Delta y_i = y_i - y_{i-1} = f(x_i)-f(x_{i-1})$. Then

$$|P_{i-1}P_i| = \sqrt{(x_i-x_{i-1})^2+(y_i-y_{i-1})^2}=\sqrt{(\Delta x)^2+(\Delta y_i)^2}.$$

By the Mean Value Theorem,

$$\Delta y_i = f(x_i)-f(x_{i-1}) = f'(x_i^*)(x_i-x_{i-1}) = f'(x_i^*)\Delta x,$$

where x_i^* is between x_{i-1} and x_i. Thus,

$$\begin{aligned}|P_{i-1}P_i| &= \sqrt{(\Delta x)^2+(\Delta y_i)^2}=\sqrt{(\Delta x)^2+[f'(x_i^*)\Delta x]^2}\\ &= \sqrt{1+[f'(x_i^*)]}\Delta x.\end{aligned}$$

Hence,

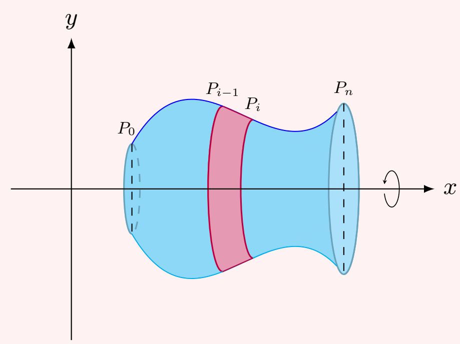
$$\begin{aligned}L &= \lim_{n\to\infty}\sum_{i=1}^n |P_{i-1}P_i| \lim_{n\to\infty}\sum_{i=1}^n \sqrt{1+[f'(x_i^*)]}\Delta x\\ &= \int_a^b \sqrt{1+[f'(x)]^2}\,dx.\end{aligned}$$

Surface Area of Revolution

Consider the surface is obtained by rotating the curve defined by $y = f(x)$, where f is positive and $a \leq x \leq b$. In order to define its surface area, we divide the interval $[a, b]$ into n subintervals with endpoints $x_0, x_1, \ldots, x_n$ and equal width Δx. The part of the surface between x_{i-1} and x_i is approximated by taking the line segment $P_{i-1}P_i$ and rotating it about the x-axis, where $P_i(x_i, f(x_i))$ lies on the curve. The result is a band with slant height $|P_{i-1}P_i|$. Therefore, in the case where f is positive we define the surface area obtained by rotating the curve $y = f(x)$, $a \leq x \leq b$, about the x-axis as

$$S = \lim_{n\to\infty} \sum_{i=1}^{n} S_i,$$

where S_i is the area of the band with slant height $|P_{i-1}P_i|$.



Formula for Surface Area of Revolution

If f' is continuous on $[a, b]$, then the surface area of the surface obtained by rotating the curve $y = f(x)$, $a \leq x \leq b$, about the x-axis, is

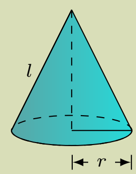
$$S = \int_a^b 2\pi y(x)\sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx = \int_a^b 2\pi f(x)\sqrt{1 + [f'(x)]^2}\, dx.$$

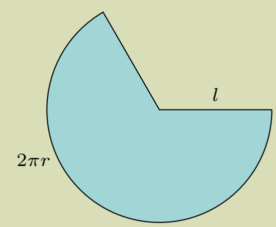
Similarly, if g' is continuous on $[c, d]$, then the surface area of the surface obtained by rotating the curve $x = g(y)$, $c \leq y \leq d$, about the y-axis, is

$$S = \int_c^d 2\pi x(y)\sqrt{1 + \left(\frac{dx}{dy}\right)^2}\, dy = \int_c^d 2\pi g(y)\sqrt{1 + [g'(y)]^2}\, dy.$$

Justification - Part 1

Let us first derive a formula for the area of a circular cone with base radius r and slant height l.





In fact, we can flat it to form a sector of a circle with radius l and central angle $\theta = 2\pi r/l$. Since it is a part of a disk of radius l, we have the area of the sector to be

$$A = \tfrac{1}{2}l^2\theta = \tfrac{1}{2}l^2 \cdot (2\pi r/l) = \pi r l.$$

Justification - Part 2

Thus, the area of the band, or frustum of a cone, with slant height l and upper and lower radii r_1 and r_2 is

$$A = \pi r_2(l_1 + l) - \pi r_1 l_1 = \pi\left[(r_2 - r_1)l_1 + r_2 l\right].$$

From similar triangles we have

$$\frac{l_1}{r_1} = \frac{l_1 + l}{2_2}$$

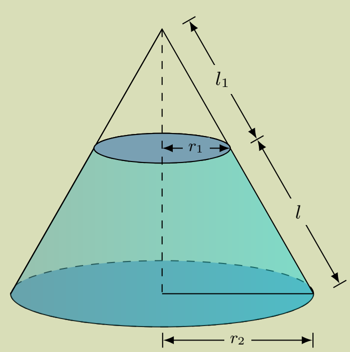
which gives

$$r_2 l_1 = r_1 l_1 + r_1 l \quad \text{or} \quad (r_2 - r_1)l_1 = r_1 l.$$

Hence,

$$A = \pi\left[r_1 l + r_2 l\right] = 2\pi r l = (\text{average circumference}) \times (\text{slant height}),$$

where $r = \frac{1}{2}(r_1 + r_2)$ is the average radius of the band.



Justification - Part 3

Suppose f' is continuous on $[a, b]$. For a band with slant height $l = |P_{i-1}P_i|$ and average radius $r = \frac{1}{2}(f(x_{i-1}) + f(x_i))$, its surface area is

$$S_i = 2\pi \cdot \tfrac{1}{2}(f(x_{i-1}) + f(x_i)) \cdot |P_{i-1}P_i|.$$

By the Pythagorean Theorem and the Mean Value Theorem,

$$\begin{aligned} |P_{i-1}P_i| &= \sqrt{(x_i - x_{i-1})^2 + [f(x_i) - f(x_{i-1})]^2} \\ &= \sqrt{(\Delta x)^2 + [f'(x_i^*)\Delta x]^2} = \sqrt{1 + [f'(x_i^*)]}\Delta x. \end{aligned}$$

Thus, by approximating $\frac{1}{2}(f(x_{i-1}) + f(x_i))$ using $f(x_i^*)$, we get

$$S_i \approx 2\pi \cdot f(x_i^*)\sqrt{1 + [f'(x_i^*)]}\Delta x.$$

Therefore,

$$\begin{aligned} S = \lim_{n\to\infty} \sum_{i=1}^{n} S_i &= \lim_{n\to\infty} \sum_{i=1}^{n} 2\pi \cdot f(x_i^*)\sqrt{1 + [f'(x_i^*)]}\Delta x \\ &= \int_a^b 2\pi f(x)\sqrt{1 + [f'(x)]^2}\, dx. \end{aligned}$$

General Formula for Surface Area of Revolution

For rotations about the x-axis and the y-axis, respectively, the surface area formula becomes

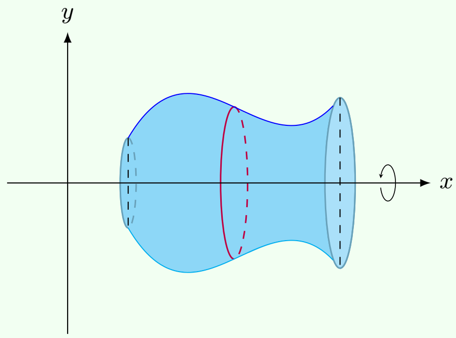
$$S = \int 2\pi y\, ds \quad \text{or} \quad S = \int 2\pi x\, ds,$$

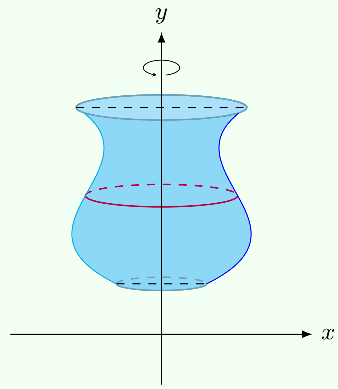
where

$$ds = \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx = \sqrt{1 + \left(\frac{dx}{dy}\right)^2}\, dy.$$

In general, the surface area of rotation is

$$\int \underbrace{(2\pi r)}_{\text{circumference}} \quad \underbrace{(ds)}_{\text{arc length}}$$





Hydrostatic Pressure and Force - Application

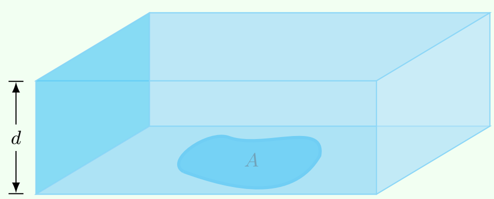
Suppose that a thin horizontal plate with area A square meters is submerged in a fluid of density ρ kilograms per cubic meter at a depth d meters below the surface of the fluid. The fluid directly above the plate has volume $V = Ad$, so its mass is $m = \rho V = \rho Ad$. The force exerted by the fluid on the plate is therefore

$$F = mg = \rho g A d.$$

where g is the acceleration due to gravity. The pressure P on the plate is defined to be the force per unit area:

$$P = \frac{F}{A} = \rho g d = \delta d,$$

where δ is called the **weight density** of the fluid.



Moments and Centers of Mass for Discrete System

For a system of n particles with masses $m_1, m_2, \ldots, m_n$ located at the points $x_1, x_2, \ldots, x_n$ on the x-axis, the **moment of the system about the origin** is

$$M = \sum_{i=1}^{n} m_i x_i,$$

and the center of mass of the system is located at

$$\bar{x} = \frac{M}{m} = \frac{\displaystyle\sum_{i=1}^{n} m_i x_i}{\displaystyle\sum_{i=1}^{n} m_i}.$$

Moments and Centers of Mass for Planar System

For a system of n particles with masses $m_1, m_2, \ldots, m_n$ located at the points $(x_1, y_1), (x_2, y_2) \ldots, (x_n, y_n)$ in the xy-plane, the **moment of the system about the y-axis** is

$$M_y = \sum_{i=1}^{n} m_i x_i,$$

and the **moment of the system about the x-axis** is

$$M_x = \sum_{i=1}^{n} m_i y_i.$$

The center of mass of the system is located at $(\bar{x}, \bar{y})$, where

$$\bar{x} = \frac{M_y}{m} = \frac{\displaystyle\sum_{i=1}^{n} m_i x_i}{\displaystyle\sum_{i=1}^{n} m_i}, \qquad \bar{y} = \frac{M_x}{m} = \frac{\displaystyle\sum_{i=1}^{n} m_i y_i}{\displaystyle\sum_{i=1}^{n} m_i}.$$

Moments and Center of Mass for Planar Region

Consider a flat plate with uniform density ρ that occupies a region $\mathscr{R}$ of the plane. Suppose the region $\mathscr{R}$ lies between the lines $x = a$ and $x = b$, above the x-axis, and beneath the graph of f, where f is a continuous function. Then the mass of the flat plate is

$$m = \rho A = \rho \int_a^b f(x)\, dx.$$

The **center of mass** of the plate (or the **centroid** of $\mathscr{R}$) is located at the point $(\bar{x}, \bar{y})$, where

$$\bar{x} = \frac{M_y}{m} = \frac{\displaystyle\int_a^b x f(x)\, dx}{\displaystyle\int_a^b f(x)\, dx}, \qquad \bar{y} = \frac{M_x}{m} = \frac{\displaystyle\int_a^b \tfrac{1}{2} \left[f(x)\right]^2\, dx}{\displaystyle\int_a^b f(x)\, dx}.$$

Centroid for Planar Region Between Two Curves

If the region $\mathscr{R}$ lies between two curves $y = f(x)$ and $y = g(x)$, where $f(x) \geq g(x)$ for $a \leq x \leq b$, then the centroid of $\mathscr{R}$ is $(\bar{x}, \bar{y})$, where

$$\bar{x} = \frac{1}{A}\int_a^b x\,[f(x) - g(x)]\,dx, \qquad \bar{y} = \frac{1}{A}\int_a^b \tfrac{1}{2}\left\{[f(x)]^2 - [g(x)]^2\right\}dx,$$

where $A = \displaystyle\int_a^b [f(x) - g(x)]\,dx$ is the area of $\mathscr{R}$.

Justification - Part 1

We divide the interval $[a, b]$ into n subintervals with endpoints $x_0, x_1, \ldots, x_n$ and equal width Δx. Denote $\bar{x}_i = \frac{1}{2}(x_{i-1} + x_i)$ to be the midpoint of the ith subinterval. Then the centroid of the ith approximating rectangle is $(\bar{x}_i, \frac{1}{2}f(\bar{x}_i))$. Its area is $f(\bar{x}_i)\Delta x$, so its mass is $\rho f(\bar{x}_i)\Delta x$. Thus, for the ith approximating rectangle, the moment about the y-axis is

$$\rho f(\bar{x}_i)\Delta x \cdot \bar{x}_i$$

Adding these moments, we obtain the moment of the polygonal approximation to $\mathscr{R}$. By taking the limit as $n \to \infty$ we obtain the moment of $\mathscr{R}$ itself about the y-axis:

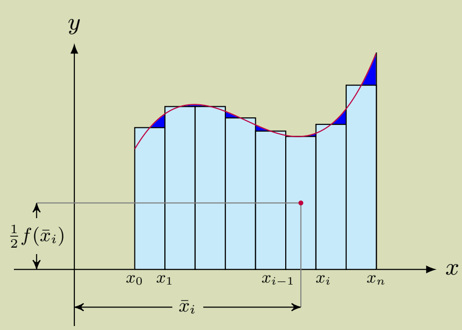
$$M_y = \lim_{n\to\infty}\sum_{i=1}^{n} \rho\bar{x}_i f(\bar{x}_i)\Delta x = \rho\int_a^b x f(x)\,dx.$$

In a similar fashion, the moment of the ith approximating rectangle about the x-axis is

$$\rho f(\bar{x}_i)\Delta x \cdot \tfrac{1}{2}f(\bar{x}_i)$$

so we get the moment of $\mathscr{R}$ itself about the x-axis:

$$M_x = \lim_{n\to\infty}\sum_{i=1}^{n} \rho \cdot \tfrac{1}{2}\left[f(\bar{x}_i)\right]^2\Delta x = \rho\int_a^b \tfrac{1}{2}\left[f(x)\right]^2\,dx.$$



Justification - Part 2

The mass of the flat plate is

$$m = \rho A = \rho \int_a^b f(x)\,dx.$$

So, the **center of mass** of the plate (or the **centroid** of $\mathscr{R}$) is located at the point $(\bar{x}, \bar{y})$, where

$$\bar{x} = \frac{M_y}{m} = \frac{\displaystyle\int_a^b x f(x)\,dx}{\displaystyle\int_a^b f(x)\,dx}, \qquad \bar{y} = \frac{M_x}{m} = \frac{\displaystyle\int_a^b \tfrac{1}{2}\left[f(x)\right]^2\,dx}{\displaystyle\int_a^b f(x)\,dx}.$$

Theorem of Pappus

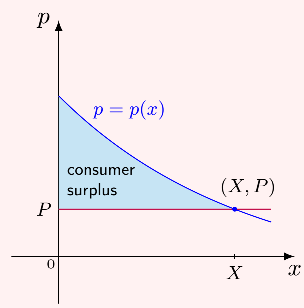
Let $\mathscr{R}$ be a plane region that lies entirely on one side of a line l in the plane. If $\mathscr{R}$ is rotated about l, then the volume of the resulting solid is the product of the area A of $\mathscr{R}$ and the distance d traveled by the centroid of $\mathscr{R}$.

Consumer Surplus - Application

The demand function $p(x)$ is the price that a company has to charge in order to sell x units of a commodity. If X is the amount of the commodity that is currently available, then $P = p(X)$ is the current selling price. Then the integral

$$\int_0^X \left[p(x) - P\right]\,dx$$

is call the **consumer surplus** for the commodity.



Justification

We divide the interval $[0, X]$ into n subintervals, each of length $\Delta x = X/n$, and let $x_i^* = x_i$ be the right endpoint of the ith subinterval. If, after the first x_{i-1} units were sold, a total of only x_i units had been available and the price per unit had been set at $p(x_i)$ dollars, then the additional Δx units could have been sold (but no more). The consumers who would have paid $p(x_i)$ dollars placed a high value on the product; they would have paid what it was worth to them. So in paying only P dollars they have saved an amount of

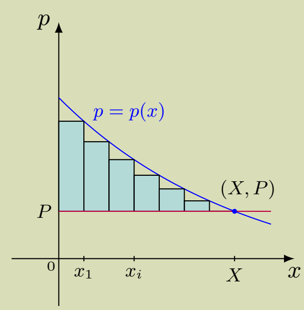
$$(\text{savings per unit}) \times (\text{number of units}) = [p(x_i) - P]\,\Delta x$$

By adding the savings, we get the total savings:

$$\sum_{i=1}^{n} [p(x_i) - P]\,\Delta x.$$

If we let $n \to \infty$, the Riemann sum approaches the integral

$$\int_0^X [p(x) - P]\ dx.$$



Poiseuille's Law - Application

The law of laminar flow gives the velocity v of blood that flows along a blood vessel with radius R and length l at a distance r from the central axis:

$$v(r) = \frac{P}{4\eta l}(R^2 - r^2),$$

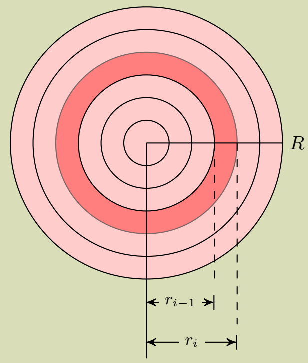
where P is the pressure difference between the ends of the vessel and η is the viscosity of the blood. Then the **flux** (or discharge), which is the volume of blood that passes a cross-section per unit time, is

$$F = \frac{\pi P R^4}{8\eta l}.$$

It shows that the flux is proportional to the fourth power of the radius of the blood vessel. It is called **Poiseuille's Law**.

Justification

We divide the interval $[0, R]$ into n subintervals, each of length $\Delta r = R/n$, with endpoints $r_0, r_1, \ldots, r_n$.



The approximate area of the ring (or washer) with inner radius r_{i-1} and outer radius r_i is

$$\pi r_i^2 - \pi r_{i-1}^2 = \pi(r_i + r_{i-1})(r_i - r_{i-1}) \approx 2\pi r_i \Delta r.$$

If Δr is small, then the velocity is almost constant throughout this ring and can be approximated by $v(r_i)$. Thus the volume of blood per unit time that flows across the ring is approximately

$$(2\pi r_i \Delta r)\, v(r_i) = 2\pi r_i v(r_i)\, \Delta r.$$

By adding the volumes, we get the total volume of blood that flows across a cross-section per unit time is about:

$$\sum_{i=1}^{n} 2\pi r_i v(r_i)\, \Delta r.$$

If we let $n \to \infty$, the Riemann sum approaches the integral

$$\begin{aligned} F &= \int_0^R 2\pi r v(r)\, dr = \int_0^R 2\pi r \cdot \frac{P}{4\eta l}(R^2 - r^2)\, dr \\ &= \frac{\pi P}{2\eta l}\left[\frac{1}{2}R^2 r^2 - \frac{1}{4}r^4\right]_{r=0}^{R} = \frac{\pi P R^4}{8\eta l}. \end{aligned}$$

Cardiac Output - Application

The **cardiac output** of the heart is the volume of blood pumped by the heart per unit time, that is, the rate of flow into the aorta. To measure the cardiac output, dye is injected into the right atrium and flows through the heart into the aorta and the concentration of the dye is measured.
Suppose $c(t)$ is the concentration of the dye at time t. If $[0, T]$ is the time period in which the dye

exists in the heart, then the cardiac output is given by

$$F = \frac{A}{\int_0^T c(t)\,dt},$$

where the amount of dye A is known and the integral can be approximated from the concentration readings.

Justification

If we divide $[0, T]$ into subintervals of equal length Δt, then the amount of dye that flows past the measuring point during the subinterval from $t = t_{i-1}$ to $t = t_i$ is approximately

$$(\text{concentration})(\text{volume}) = c(t_i)(F\Delta t),$$

where F is the rate of flow that we are trying to determine. Thus the total amount of dye is

$$\sum_{i=1}^{n} c(t_i)F\,\Delta t = F\sum_{i=1}^{n} c(t_i)\,\Delta t.$$

If we let $n \to \infty$, the Riemann sum approaches the integral

$$F\int_0^T c(t)\,dt.$$

Therefore,

$$F = \frac{A}{\int_0^T c(t)\,dt}.$$

Probability Density Function

Every continuous random variable, X, has a **probability density function**, $f(x)$, which satisfies the following conditions

1. $f(x) \geq 0$ for all x;

2. $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

The probability that X lies between a and b is

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx.$$

The **mean** of any probability density function f is defined to be

$$\mu = \int_{-\infty}^{\infty} xf(x)\,dx.$$

The **median** of a probability density function is the number m such that

$$\int_m^\infty f(x)\,dx = \tfrac{1}{2}.$$

Normal Distribution

A **normal distribution** is a random variable X that has the following probability density function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}.$$

The mean for this function is μ. The positive constant σ is called the **standard deviation**, which measures how spread out the values of X are.

Chapter 9

Differential Equations

Differential Equation

A **differential equation** is an equation that contains an unknown function and one or more of its derivatives. The **order** of a differential equation is the order of the highest derivative that occurs in the equation.

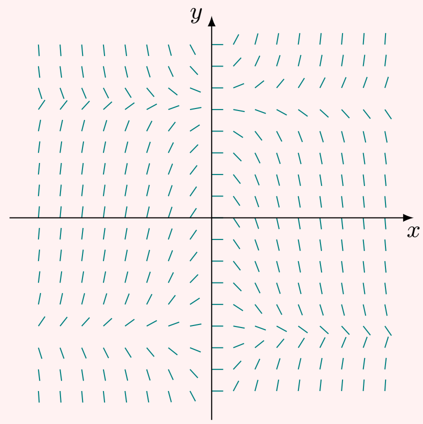
A function f is called a **solution** of a differential equation if the equation is satisfied when $y = f(x)$ and its derivatives are substituted into the equation.

A condition of the form $y(x_0) = y_0$ is called an **initial condition**. The problem of finding a solution of the differential equation that satisfies the initial condition is called an initial-value problem.

For the differential equation $y' = F(y)$, if $F(c) = 0$, the constant function $y(x) = c$ is called an **equilibrium solution**.

Direction Field

A **direction field** (or **slope filed**) is a graphical representation of the solutions of a first-order differential equation $y' = F(x, y)$. It consists of the line segments that indicate the direction in which a solution curve is heading.



Euler's Method

Approximate values for the solution of the initial-value problem $y' = F(x, y)$, $y(x_0) = y_0$, with step size h, at $x_n = x_{n-1} + h$, are

$$y_n = y_{n-1} + hF(x_{n-1}, y_{n-1}) \qquad n = 1, 2, 3, \ldots$$

Separable Equation

A **separable equation** is a first-order differential equation of the form

$$\frac{dy}{dx} = g(x)f(y),$$

or equivalently (if $f(y) \neq 0$),

$$\frac{dy}{dx} = \frac{g(x)}{h(y)}.$$

Solution of Separable Equation

By rewriting the equation as

$$h(y)\,dy = g(x)\,dx,$$

we integrate both side to have the implicit solution of the separable equation:

$$\int h(y)\,dy = \int g(x)\,dx.$$

Law of Natural Growth

The first-order differential equation

$$\frac{dP}{dt} = kP$$

is called the **law of natural growth**, where

$$k = \frac{1}{P}\frac{dP}{dt}$$

is the **relative growth rate**.

Solution of the Initial-Value Problem

The solution of the initial-value problem

$$\frac{dP}{dt} = kP \qquad P(0) = P_0$$

is

$$P(t) = P_0 e^{kt}.$$

Justification

The differential equation is separable. By rewriting the equation as

$$\frac{dP}{P} = k\,dt$$

we integrate both side to have the implicit solution of the equation:

$$\ln|P| = kt + C.$$

The initial condition $P(0) = P_0$ implies $C = \ln|P_0|$. Thus,

$$\ln|P| = kt + \ln|P_0|$$

or equivalently,

$$\frac{P}{P_0} = e^{kt}.$$

So, we have

$$P(t) = P_0 e^{kt}.$$

Logistic Differential Equation

The first-order differential equation

$$\frac{dP}{dt} = kP\left(1 - \frac{P}{M}\right)$$

is called the **logistic differential equation**, where the constant M is called the **carrying capacity**.

Solution of the Initial-Value Problem

The solution of the initial-value problem

$$\frac{dP}{dt} = kP\left(1 - \frac{P}{M}\right), \qquad P(0) = P_0$$

is

$$P(t) = \frac{M}{1 + Ae^{-kt}}, \qquad \text{where } A = \frac{M - P_0}{P_0}.$$

Justification

The differential equation is separable. By rewriting the equation as

$$\frac{dP}{P\left(1 - \frac{P}{M}\right)} = k\,dt$$

or

$$\left(\frac{1}{P} - \frac{1}{P - M}\right) dP = k\,dt$$

we integrate both side to have the implicit solution of the equation:

$$\ln|P| - \ln|P - M| = kt + C.$$

The initial condition $P(0) = P_0$ implies $C = \ln|P_0| - \ln|P_0 - M|$. Thus,

$$\ln|P| - \ln|P - M| = kt + \ln|P_0| - \ln|P_0 - M|$$

or equivalently,

$$\frac{P}{P_0} = e^{kt} \cdot \frac{P - M}{P_0 - M}.$$

Solving for P gives

$$P(t) = \frac{M}{1 + Ae^{-kt}}, \qquad \text{where } A = \frac{M - P_0}{P_0}.$$

First-Order Linear Differential Equation

The **first-order linear differential equation** has the form

$$\frac{dy}{dx} + P(x)y = Q(x).$$

Integrating Factor

To solve the first-order linear differential equation, multiply both sides by the **integrating factor** $I(x) = e^{\int P(x)\,dx}$ and integrate both sides.

General Solution

The general solution of the first-order linear differential equation is

$$y(x) = \frac{1}{I(x)}\left[\int I(x)Q(x)\,dx + C\right].$$

Justification

By multiplying the integrating factor $I(x) = e^{\int P(x)\,dx}$ to both sides of the differential equation $\frac{dy}{dx} + P(x)y = Q(x)$, we have

$$e^{\int P(x)\,dx}\frac{dy}{dx} + e^{\int P(x)\,dx}P(x)y = I(x)Q(x).$$

This equation can be written as

$$\left(e^{\int P(x)\,dx}y\right)' = I(x)Q(x).$$

Integrating the last equation gives

$$e^{\int P(x)\,dx}y = \int I(x)Q(x) + C,$$

or equivalently,

$$y = \frac{1}{I(x)}\left[\int I(x)Q(x)\,dx + C\right].$$

Predator-Prey Equations

The **predator-prey equations** (or the **Lotka-Volterra equations**) are a system of two differential equations of the form

$$\frac{dR}{dt} = kR - aRW, \qquad \frac{dW}{dt} = -rW + bRW,$$

where k, r, a, and b are positive constants.

Equilibrium Solutions

By solving the system

$$kR - aRW = 0, \qquad -rW + bRW = 0,$$

we have two **equilibrium solutions** (or **equilibrium points**):

$$(R, W) = (0, 0), \qquad (R, W) = (r/b, k/a).$$

These are special solutions of the predator-prey equations.

Phase Portrait

A **phase portrait** is a geometric representation of the solution curves (called **phase trajectories**) of a dynamical system in the **phase plane**.

Chapter 10

Parametric Equations and Polar Coordinates

Parametric Equations and Parametric Curve

Suppose that x and y are both given as functions of a third variable t (called a **parameter**) by the **parametric equations**

$$x = f(t), \qquad y = g(t), \qquad a \leq t \leq b.$$

As t varies, the point $(x, y) = (f(t), g(t))$ varies and traces out a curve, which we call a **parametric curve**. The points $(f(a), g(a))$ and $(f(b), g(b))$ are called the **initial point** and **terminal point** of the curve.

Tangent to General Parametric Curve

Suppose a parametric curve is given by the parametric equations

$$x = f(t), \qquad y = g(t).$$

If f and g are differentiable functions, and if $f'(t) \neq 0$, then the Chain Rule gives the slope of the tangent line at $(x, y) = (f(t), g(t))$:

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt} = \frac{g'(t)}{f'(t)}.$$

Horizontal Tangent

The parametric curve has a horizontal tangent at $(x, y) = (f(t), g(t))$ if

$$\frac{dy}{dt} = g'(t) = 0, \qquad \frac{dx}{dt} = f'(t) \neq 0.$$

Vertical Tangent

The parametric curve has a vertical tangent at $(x, y) = (f(t), g(t))$ if

$$\frac{dx}{dt} = f'(t) = 0, \qquad \frac{dy}{dt} = g'(t) \neq 0.$$

Area Under General Parametric Curve

Suppose a parametric curve is given by the parametric equations

$$x = f(t), \qquad y = g(t), \qquad \alpha \leq t \leq \beta.$$

If the curve is traced out once when t varies from α to β, then the area under the curve is, by using the Substitution Rule,

$$A = \int_a^b y\,dx = \int_\alpha^\beta g(t)\,f'(t)\,dt.$$

Arc Length for General Parametric Curve

Suppose a parametric curve is given by the parametric equations

$$x = f(t), \qquad y = g(t), \qquad \alpha \leq t \leq \beta$$

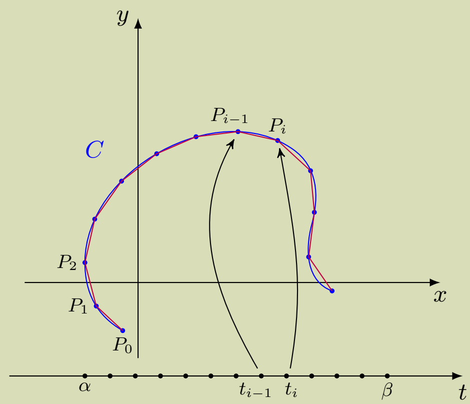
where f' and g' are continuous and the parametric curve is traversed exactly once as t increases from α to β. Then the arc length is

$$L = \int_\alpha^\beta \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\,dt = \int_\alpha^\beta \sqrt{[f'(t)]^2 + [g'(t)]^2}\,dt.$$

Justification - Part 1

We divide the parameter interval $[\alpha, \beta]$ into n subintervals of equal width Δt. If $t_0, t_1, t_2, \ldots, t_n$ are the endpoints of these subintervals, then $x_i = f(t_i)$ and $y_i = g(t_i)$ are the coordinates of points $P_i(x_i, y_i)$ that lie on C and the polygon with vertices $P_0, P_1, \ldots, P_n$ approximates C.
The length L of C is defined as the limit of the lengths of these approximating polygons as $n \to \infty$:

$$L = \lim_{n\to\infty} \sum_{i=1}^{n} |P_{i-1}P_i|.$$



Justification - Part 2

By the Mean Value Theorem, there are t_i^* and t_i^{**} in (t_{i-1}, t_i) such that

$$\Delta x_i = f(t_i) - f(t_{i-1}) = f'(t_i^*)\Delta t,$$
$$\Delta y_i = g(t_i) - g(t_{i-1}) = g'(t_i^{**})\Delta t.$$

Thus,

$$\begin{aligned} |P_{i-1}P_i| &= \sqrt{(\Delta x_i)^2 + (\Delta y_i)^2} = \sqrt{[f'(t_i^*]\Delta t)^2 + [g'(t_i^{**}]\Delta t)^2} \\ &= \sqrt{[f'(t_i^*)]^2 + [g'(t_i^{**})]^2}\,\Delta t, \end{aligned}$$

and so

$$L = \lim_{n\to\infty} \sum_{i=1}^{n} \sqrt{[f'(t_i^*)]^2 + [g'(t_i^{**})]^2}\,\Delta t.$$

The last sum resembles a Riemann sum for the function $\sqrt{[f'(t)]^2 + [g'(t)]^2}$ but it is not exactly a Riemann sum because $t_i^* \neq t_i^{**}$ in general. Nevertheless, if f' and g' are continuous, it can be shown that the limit is the same as if t_i^* and t_i^{**} were equal, namely,

$$L = \int_{\alpha}^{\beta} \sqrt{[f'(t)]^2 + [g'(t)]^2}\,dt.$$

Surface Area of Revolution by General Parametric Curve

Suppose a parametric curve is given by the parametric equations

$$x = f(t), \qquad y = g(t), \qquad \alpha \leq t \leq \beta$$

where f' and g' are continuous and $g(t) \geq 0$. Then the area of surface of revolution about the x-axis is

$$S = \int_{\alpha}^{\beta} 2\pi y \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt = \int_{\alpha}^{\beta} 2\pi y(t) \sqrt{[f'(t)]^2 + [g'(t)]^2}\, dt.$$

Justification

The general formula for the area of surface of revolution about the x-axis is

$$S = \int 2\pi y\, ds.$$

For the parametric curve we have

$$ds = \sqrt{[f'(t)]^2 + [g'(t)]^2}\, dt.$$

Thus, by the Substitution Rule, we get

$$S = \int_{\alpha}^{\beta} 2\pi y(t) \sqrt{[f'(t)]^2 + [g'(t)]^2}\, dt.$$

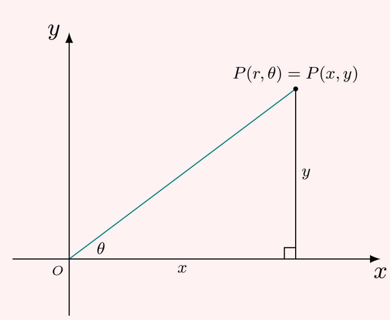
Cartesian Coordinates and Polar Coordinates

A point (x, y) in Cartesian coordinates can be expressed in polar coordinates by

$$x = r\cos\theta, \qquad y = r\sin\theta.$$

Conversely, we can express r and θ in terms of x and y:

$$r^2 = x^2 + y^2, \qquad \tan\theta = \frac{y}{x}.$$



Polar Curves

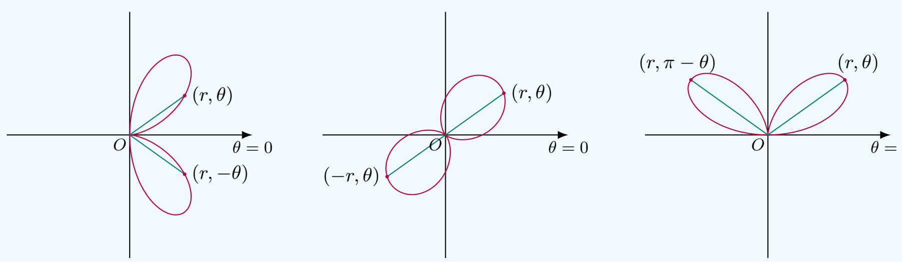
The graph of a polar equation $r = f(\theta)$, or more generally $F(r, \theta) = 0$, consists of all points P that have at least one polar representation (r, θ) whose coordinates satisfy the equation.

How to Determine Symmetry of a Polar Curve

(a) If a polar equation is unchanged when θ is replaced by $-\theta$, the curve is symmetric about the polar axis.

(b) If the equation is unchanged when r is replaced by $-r$, or when θ is replaced by $\theta + \pi$, the curve is symmetric about the pole.

(c) If the equation is unchanged when θ is replaced by $\pi - \theta$, the curve is symmetric about the vertical line $\theta = \pi/2$.



Tangent to Polar Curve

Suppose a polar curve is given by the equation $r = f(\theta)$. Then we can write its parametric equations as

$$x = r\cos\theta = f(\theta)\cos\theta, \qquad y = r\sin\theta = f(\theta)\sin\theta.$$

If f is differentiable, and if $\dfrac{dx}{d\theta} = f'(\theta)\cos\theta - f(\theta)\sin\theta \neq 0$, then the Chain Rule gives the slope of the tangent line at $(x, y) = (f(\theta)\cos\theta, f(\theta)\sin\theta)$:

$$\frac{dy}{dx} = \frac{dy/d\theta}{dx/d\theta} = \frac{f'(\theta)\sin\theta + f(\theta)\cos\theta}{f'(\theta)\cos\theta - f(\theta)\sin\theta}.$$

Horizontal Tangent

The parametric curve has a horizontal tangent at $(x, y) = (f(\theta)\cos\theta, f(\theta)\sin\theta)$ if

$$\frac{dy}{d\theta} = f'(\theta)\sin\theta + f(\theta)\cos\theta = 0, \qquad \frac{dx}{d\theta} = f'(\theta)\cos\theta - f(\theta)\sin\theta \neq 0.$$

Vertical Tangent

The parametric curve has a vertical tangent at $(x, y) = (f(\theta)\cos\theta, f(\theta)\sin\theta)$ if

$$\frac{dx}{d\theta} = f'(\theta)\cos\theta - f(\theta)\sin\theta = 0, \qquad \frac{dy}{d\theta} = f'(\theta)\sin\theta + f(\theta)\cos\theta \neq 0.$$

Area of Polar Region

Let $\mathscr{R}$ be a polar region bounded by the polar curve $r = f(\theta)$ and by the rays $\theta = \alpha$ to $\theta = \beta$, where f is a positive continuous function and the angles satisfy $0 < \beta - \alpha \leq 2\pi$. Then the area of $\mathscr{R}$ is

$$A = \int_a^b \tfrac{1}{2} r^2 \, d\theta = \int_a^b \tfrac{1}{2} [f(\theta)]^2 \, d\theta.$$

Justification

For the polar curve $r = f(\theta)$, we write its parametric equations as

$$x = r\cos\theta = f(\theta)\cos\theta, \qquad y = r\sin\theta = f(\theta)\sin\theta.$$

We divide the interval $[\alpha, \beta]$ into subintervals of equal width $\Delta\theta$ with endpoints $\theta_0, \theta_1, \theta_2, \ldots, \theta_n$. If we choose θ_i^* in the ith subinterval $[\theta_{i-1}, \theta_i]$, then the area ΔA_i of the ith region is approximated by the area of the sector of a circle with central angle $\Delta\theta$ and radius $f(\theta_i^*)$. The area can be approximated by

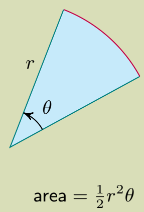
$$\begin{aligned}\Delta A_i &\approx \pi(\text{radius})^2 \cdot \frac{(\text{central angle})}{2\pi} = \tfrac{1}{2}(\text{radius})^2 \times (\text{central angle}) \\ &= \tfrac{1}{2}\left[f(\theta_i^*)\right]^2 \Delta\theta.\end{aligned}$$

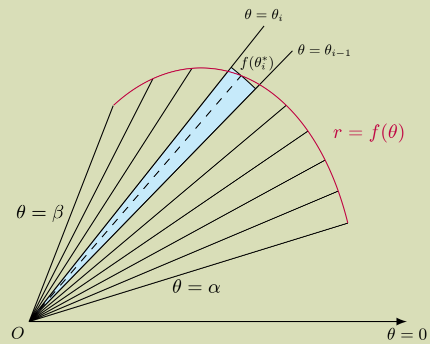
So, an approximation to the total area A of $\mathscr{R}$ is

$$A \approx \sum_{i=1}^{n} \tfrac{1}{2}\left[f(\theta_i^*)\right]^2 \Delta\theta.$$

Hence, we obtain the formula for the area A of the polar region $\mathscr{R}$

$$A = \lim_{n\to\infty} \sum_{i=1}^{n} \tfrac{1}{2}\left[f(\theta_i^*)\right]^2 \Delta\theta = \int_\alpha^\beta \tfrac{1}{2}\left[f(\theta)\right]^2 \, d\theta.$$

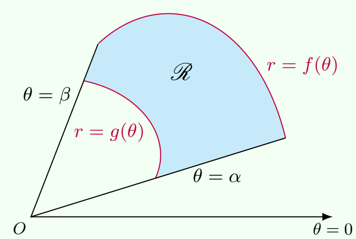




Area Between Two Polar Curves

Let $\mathscr{R}$ be a polar region bounded by the polar curves $r = f(\theta)$, $r = g(\theta)$, and by the rays $\theta = \alpha$ to $\theta = \beta$, where $f(\theta) \geq g(\theta) \geq 0$ and the angles satisfy $0 < \beta - \alpha \leq 2\pi$. Then the area A of the polar region $\mathscr{R}$ is

$$A = \frac{1}{2}\int_{\alpha}^{\beta} \left([f(\theta)]^2 - [g(\theta)]^2\right) d\theta$$



Arc Length of Polar Curve

Suppose a polar curve is given by the equation $r = f(\theta)$. If f' is continuous, then the arc length of the polar curve from $\theta = \alpha$ to $\theta = \beta$ is

$$L = \int_{\alpha}^{\beta} \sqrt{[f(\theta)]^2 + [f'(\theta)]^2}\, d\theta = \int_{\alpha}^{\beta} \sqrt{r^2 + (dr/d\theta)^2}\, d\theta.$$

Justification

We write its parametric equations as

$$x = r\cos\theta = f(\theta)\cos\theta, \qquad y = r\sin\theta = f(\theta)\sin\theta.$$

Since

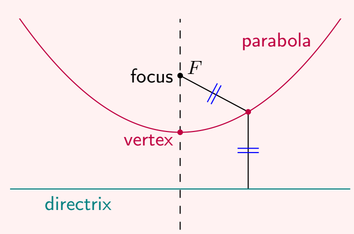
$$\begin{aligned}\left(\frac{dx}{d\theta}\right)^2 + \left(\frac{dy}{d\theta}\right)^2 &= [f'(\theta)\cos\theta - f(\theta)\sin\theta]^2 + [f'(\theta)\sin\theta + f(\theta)\cos\theta]^2 \\ &= [f(\theta)]^2 + [f'(\theta)]^2,\end{aligned}$$

the length is

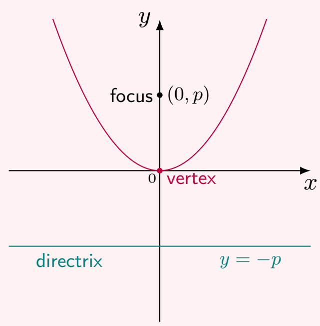
$$L = \int_\alpha^\beta \sqrt{\left(\frac{dx}{d\theta}\right)^2 + \left(\frac{dy}{d\theta}\right)^2}\, d\theta = \int_\alpha^\beta \sqrt{[f(\theta)]^2 + [f'(\theta)]^2}\, d\theta.$$

Parabola

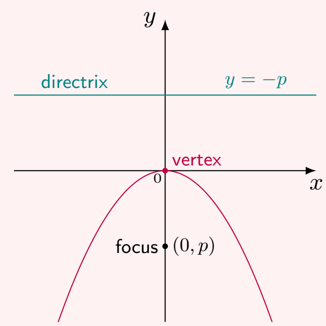
A **parabola** is the set of points in a plane that are equidistant from a fixed point F (called the **focus**) and a fixed line (called the **directrix**). The point halfway between the focus and the directrix lies on the parabola; it is called the **vertex**.



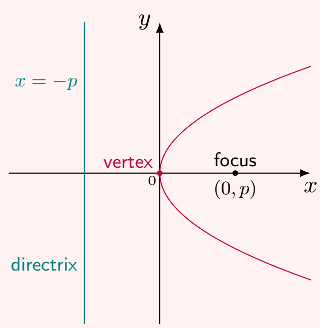
Standard Parabolas



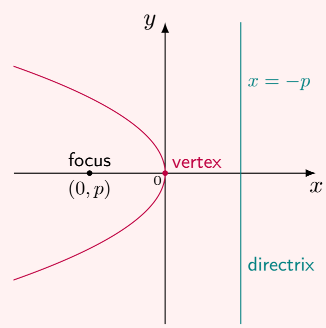
(a) $x^2 = 4py$, $p > 0$



(b) $x^2 = 4py$, $p < 0$



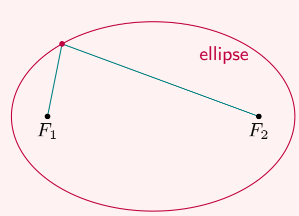
(c) $y^2 = 4px$, $p > 0$



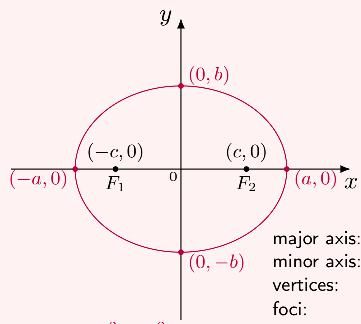
(d) $y^2 = 4px$, $p < 0$

Ellipse

An **ellipse** is the set of points in a plane the sum of whose distances from two fixed points F_1 and F_2 (called the **foci**).

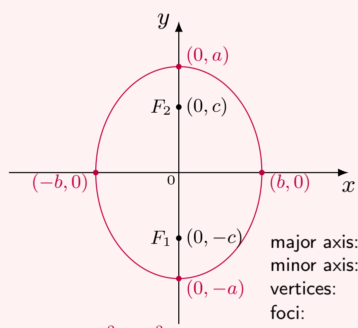


Standard Ellipses



major axis: from $(-a,0)$ to $(a,0)$
minor axis: from $(0,-b)$ to $(0,b)$
vertices: $(\pm a,0)$
foci: $(\pm c,0)$, where $c=\sqrt{a^2-b^2}$

(a) $\dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}=1,\ a\geq b>0$

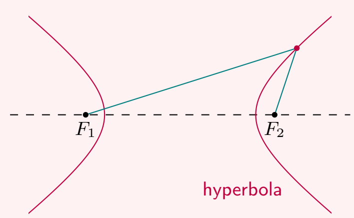


major axis: from $(0,-a)$ to $(0,a)$
minor axis: from $(-b,0)$ to $(b,0)$
vertices: $(0,\pm a)$
foci: $(0,\pm c)$, where $c=\sqrt{a^2-b^2}$

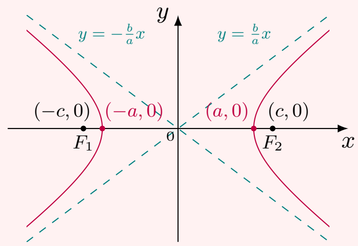
(b) $\dfrac{x^2}{b^2}+\dfrac{y^2}{a^2}=1,\ a\geq b>0$

Hyperbola

A **hyperbola** is the set of all points in a plane the difference of whose distances from two fixed points F_1 and F_2 (called the **foci**).

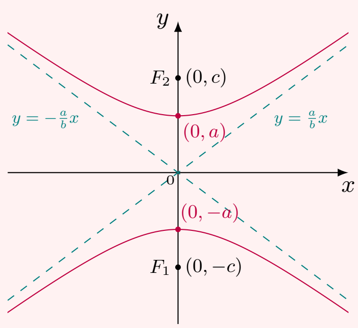


Standard Hyperbolas



asymptotes: $y = \pm\frac{b}{a}x$
vertices: $(\pm a, 0)$
foci: $(\pm c, 0)$, where $c = \sqrt{a^2 + b^2}$

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



asymptotes: $y = \pm\frac{a}{b}x$
vertices: $(0, \pm a)$
foci: $(0, \pm c)$, where $c = \sqrt{a^2 + b^2}$

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

Conic Curves

Let F be a fixed point (called the **focus**) and l be a fixed line (called the **directrix**) in a plane. Let e be a fixed positive number (called the **eccentricity**). The set of all points P in the plane such that

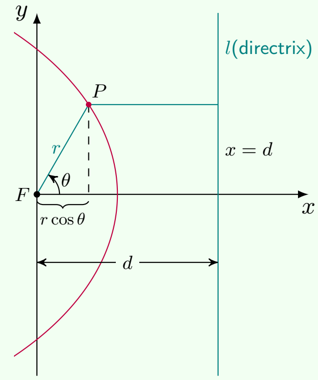
$$\frac{|PF|}{|Pl|} = e$$

is a conic section. The conic is

(a) an ellipse if $e < 1$

(b) a parabola if $e = 1$

(c) a hyperbola if $e > 1$



Justification - Part 1

Let us place the focus F at the origin and the directrix parallel to the y-axis and d units to the right. Thus the directrix has equation $x = d$ and is perpendicular to the polar axis. If the point P has polar coordinates (r, θ), we see that $|PF| = r$ and $|Pl| = d - r\cos\theta$. Thus, the condition $|PF|/|Pl| = e$ gives

$$r = e(d - r\cos\theta).$$

Solving the equation for r gives

$$r = \frac{ed}{1 + e\cos\theta}.$$

In fact, if the directrix is chosen to be perpendicular to the polar axis as $x = \pm d$, or parallel to the polar axis as $y = \pm d$, then the polar equation of the conic is given by one of the four forms:

$$r = \frac{ed}{1 \pm e\cos\theta} \quad \text{and} \quad r = \frac{ed}{1 \pm e\sin\theta}$$

Justification - Part 2

If the eccentricity is $e = 1$, then $|PF| = |Pl|$ and so the given condition simply becomes the definition of a parabola.

Let us consider the conic curve given by the equation $r = e(d - r\cos\theta)$. If we square both sides of this polar equation and convert to rectangular coordinates, we get

$$x^2 + y^2 = e^2(d - x)^2 = e^2(d^2 - 2dx + x^2),$$

or

$$(1 - e^2)x^2 + 2de^2x + y^2 = e^2d^2.$$

After completing the square, we have

$$\left(x + \frac{e^2d}{1 - e^2}\right)^2 + \frac{y^2}{1 - e^2} = \frac{e^2d^2}{(1 - e^2)^2}.$$

Suppose $e < 1$. Then the last equation can be re-written as

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

where

$$h = -\frac{e^2 d}{1-e^2}, \qquad a^2 = \frac{e^2 d^2}{(1-e^2)^2}, \qquad b^2 = \frac{e^2 d^2}{1-e^2}.$$

This shows that the curve is an ellipse.
Similarly, if $e > 1$, the equation becomes

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

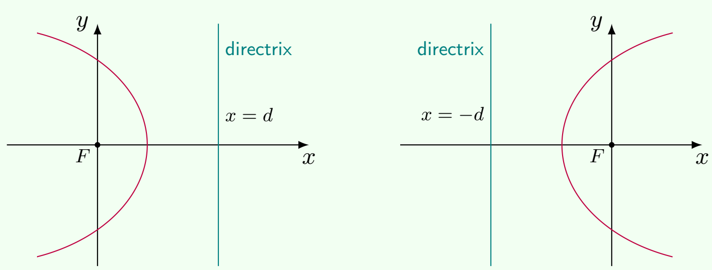
which is a hyperbola.
The above argument works for the other three cases of conic curves.

Polar Equations for Conic Curves

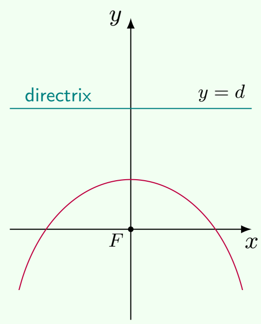
A polar equation of the form

$$r = \frac{ed}{1 \pm e\cos\theta} \quad \text{or} \quad r = \frac{ed}{1 \pm e\sin\theta}$$

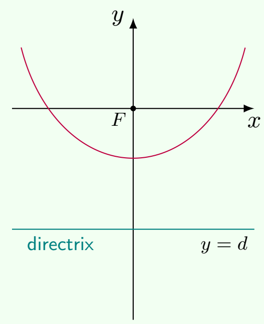
represents a conic section with eccentricity e. The conic is an ellipse if $e < 1$, a parabola if $e = 1$, or a hyperbola if $e > 1$.



(a) $r = \dfrac{ed}{1 + e\cos\theta}$ (b) $r = \dfrac{ed}{1 - e\cos\theta}$



(c) $r = \dfrac{ed}{1 + e\sin\theta}$



(d) $r = \dfrac{ed}{1 - e\sin\theta}$

Kepler's Laws

1. A planet revolves around the sun in an elliptical orbit with the sun at one focus.
2. The line joining the sun to a planet sweeps out equal areas in equal times.
3. The square of the period of revolution of a planet is proportional to the cube of the length of the major axis of its orbit.

Polar Equations for Conic Curves (in terms of semimajor axis)

The polar equation of an ellipse with focus at the origin, semimajor axis a, eccentricity e, and directrix $x = d$ can be written in the form

$$r = \frac{a(1-e^2)}{1 + e\cos\theta}.$$

Justification

An equation of the ellipse is

$$r = \frac{ed}{1 + e\cos\theta}.$$

Since

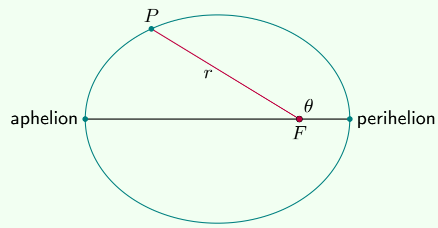
$$a^2 = \frac{e^2d^2}{(1-e^2)^2} \Longrightarrow d^2 = \frac{a^2(1-e^2)^2}{e^2} \Longrightarrow d = \frac{a(1-e^2)}{e}$$

we have

$$r = \frac{a(1-e^2)}{1 + e\cos\theta}.$$

Perihelion and Aphelion

The positions of a planet that are closest to and farthest from the sun are called its **perihelion** and **aphelion**, respectively.



The perihelion distance from a planet to the sun is $a(1-e)$ and the aphelion distance is $a(1+e)$.

Justification

In the equation

$$r = \frac{a(1-e^2)}{1+e\cos\theta},$$

we put $\theta = 0$ and $\theta = \pi$, respectively, to have

$$\text{perihelion distance} = \frac{a(1-e^2)}{1+e\cos 0} = \frac{a(1-e)(1+e))}{1+e} = a(1-e),$$
$$\text{aphelion distance} = \frac{a(1-e^2)}{1+e\cos\pi} = \frac{a(1-e)(1+e))}{1-e} = a(1+e).$$

Chapter 11

Infinite Sequences and Series

Limit of a Sequence (non-rigorous)

A sequence $\{a_n\}$, which is an ordered list of numbers, has the **limit** L and we write

$$\lim_{x\to a} a_n = L \quad \text{or} \quad a_n \to L \text{ as } n \to \infty$$

if we can make the terms a_n as close to L as we like by taking n sufficiently large. If $\lim_{n\to\infty} a_n$ exists, we say the sequence **converges** (or is **convergent**). Otherwise, we say the sequence **diverges** (or is **divergent**).

Limit of a Sequence (rigorous)

A sequence $\{a_n\}$ has the **limit** L and we write

$$\lim_{x\to a} a_n = L \quad \text{or} \quad a_n \to L \text{ as } n \to \infty$$

if for every number $\varepsilon > 0$ there is a corresponding integer N such that

$$\text{if } n > N \quad \text{then} \quad |a_n - L| < \varepsilon.$$

Relationship with the Limit of a Function

If $\lim\limits_{x\to\infty} f(x) = L$ and $f(n) = a_n$ when n is an integer, then $\lim\limits_{n\to\infty} a_n = L$. The same result holds when $L = \infty$.

Justification

Let $\varepsilon > 0$ be given. If $\lim_{x\to\infty} f(x) = L$, there is N such that

$$\text{if } x > N \quad \text{then} \quad |f(x) - L| < \varepsilon.$$

Thus, if $n > N$, then $|a_n - L| = |f(n) - L| < \varepsilon$. This means that $\lim_{x\to a} a_n = L$.
If $\lim_{x\to\infty} f(x) = \infty$, then, for any M, there is N such that

$$\text{if } x > N \quad \text{then} \quad f(x) > M.$$

Thus, if $n > N$, then $a_n = f(n) > M$. This means that $\lim_{x\to a} a_n = \infty$.

Sequence Diverges to infinity

$\lim_{x\to a} a_n = \infty$ means that for every positive number M there is an integer N such that

$$\text{if } n > N \quad \text{then} \quad a_n > M.$$

Limit Laws

If $\{a_n\}$ and $\{b_n\}$ are convergent sequences and c is a constant, then

1. $\lim_{n\to\infty} (a_n + b_n) = \lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n$.

2. $\lim_{n\to\infty} (a_n - b_n) = \lim_{n\to\infty} a_n - \lim_{n\to\infty} b_n$.

3. $\lim_{n\to\infty} ca_n = c \lim_{n\to\infty} a_n$. In particular, $\lim_{n\to\infty} c = c$.

4. $\lim_{n\to\infty} (a_n b_n) = \lim_{n\to\infty} a_n \cdot \lim_{n\to\infty} b_n$.

5. $\lim_{n\to\infty} \dfrac{a_n}{b_n} = \dfrac{\lim_{n\to\infty} a_n}{\lim_{n\to\infty} b_n}$ if $\lim_{n\to\infty} b_n \neq 0$.

6. $\lim_{n\to\infty} a_n^p = \left[\lim_{n\to\infty} a_n\right]^p$ if $p > 0$ and $a_n > 0$.

Justification - Law 1

Let $\varepsilon > 0$ be given. Suppose $\lim_{n\to\infty} a_n = a$ and $\lim_{n\to\infty} b_n = b$. Then, there are N_1, N_2, such that

$$\begin{aligned} &\text{if } n > N_1 \quad \text{then} \quad |a_n - a| < \varepsilon/2; \\ &\text{if } n > N_2 \quad \text{then} \quad |b_n - b| < \varepsilon/2. \end{aligned}$$

Let $N = \max\{N_1, N_2\}$. Then $N \geq N_1$ and N_2. If $n > N$, then

$$\begin{aligned} |(a_n + b_n) - (a + b)| &= |(a_n - a) + (b_n - b)| \\ &\leq |a_n - a| + |b_n - b| && \text{(triangle inequality)} \\ &< \varepsilon/2 + \varepsilon/2 = \varepsilon. \end{aligned}$$

Therefore, $\lim\limits_{n\to\infty} (a_n + b_n) = a + b = \lim\limits_{n\to\infty} a_n + \lim\limits_{n\to\infty} b_n$.

Justification - Law 2

Let $\varepsilon > 0$ be given. Suppose $\lim\limits_{n\to\infty} a_n = a$ and $\lim\limits_{n\to\infty} b_n = b$. Then, there are N_1, N_2, such that

$$\begin{aligned} &\text{if } n > N_1 \quad \text{then} \quad |a_n - a| < \varepsilon/2; \\ &\text{if } n > N_2 \quad \text{then} \quad |b_n - b| < \varepsilon/2. \end{aligned}$$

Let $N = \max\{N_1, N_2\}$. Then $N \geq N_1$ and N_2. If $n > N$, then

$$\begin{aligned} |(a_n - b_n) - (a - b)| &= |(a_n - a) - (b_n - b)| \\ &\leq |a_n - a| + |b_n - b| && \text{(triangle inequality)} \\ &< \varepsilon/2 + \varepsilon/2 = \varepsilon. \end{aligned}$$

Therefore, $\lim\limits_{n\to\infty} (a_n - b_n) = a - b = \lim\limits_{n\to\infty} a_n - \lim\limits_{n\to\infty} b_n$.

Justification - Law 3

It is clear that Law 3 holds if $c = 0$.
Suppose $c \neq 0$. Let $\varepsilon > 0$ be given. Suppose $\lim\limits_{n\to\infty} a_n = a$. Then, there is N, such that

$$\text{if } n > N \quad \text{then} \quad |a_n - a| < \varepsilon/|c|.$$

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

$$\begin{aligned} |ca_n - ca| &= |c| \cdot |a_n - a| \\ &< |c| \cdot \varepsilon/|c| = \varepsilon. \end{aligned}$$

Therefore, if $c \neq 0$, we also have $\lim\limits_{n\to\infty} cc_a = ca = c \lim\limits_{n\to\infty} a_n$.

Justification - Law 4

Let $\varepsilon > 0$ be given. Suppose $\lim_{n\to\infty} a_n = a$ and $\lim_{n\to\infty} b_n = b$. Then, there are N_1, N_2, and N_3, such that

$$\begin{aligned} &\text{if } n > N_1 \quad \text{then} \quad |b_n - b| < 1 \Longrightarrow |b_n| < |b| + 1;\\ &\text{if } n > N_2 \quad \text{then} \quad |a_n - a| < \varepsilon/(2|b| + 2);\\ &\text{if } n > N_3 \quad \text{then} \quad |b_n - b| < \varepsilon/(2|a| + 1). \end{aligned}$$

Let $N = \max\{N_1, N_2, N_3\}$. Then $N \geq N_1$, N_2, and N_3. If $n > N$, then

$$\begin{aligned} |a_n b_n - ab| &= |(a_n - a)b_n + a(b_n - b)|\\ &\leq |a_n - a| \cdot |b_n| + |a| \cdot |b_n - b| \qquad \text{(triangle inequality)}\\ &< \varepsilon/(2|b| + 2) \cdot (|b| + 1) + |a| \cdot \varepsilon/(2|a| + 1) < \varepsilon. \end{aligned}$$

Therefore, $\lim_{n\to\infty} (a_n b_n) = ab = \lim_{n\to\infty} a_n \cdot \lim_{n\to\infty} b_n$.

Justification - Law 5

Let $\varepsilon > 0$ be given. Suppose $\lim_{n\to\infty} a_n = a$ and $\lim_{n\to\infty} b_n = b \neq 0$. Then, there are N_1, N_2, and N_3, such that

$$\begin{aligned} &\text{if } n > N_1 \quad \text{then} \quad |b_n - b| < |b|/2\\ &\qquad\qquad\qquad\qquad \Longrightarrow |b_n| > |b| - |b_n - b| > |b| - |b|/2 = |b|/2\\ &\qquad\qquad\qquad\qquad \Longrightarrow 1/|b_n| < 2/|b|;\\ &\text{if } n > N_2 \quad \text{then} \quad |a_n - a| < (|b|/4)\varepsilon;\\ &\text{if } n > N_3 \quad \text{then} \quad |b_n - b| < |b|^2\varepsilon/(4|a| + 1). \end{aligned}$$

Let $N = \max\{N_1, N_2, N_3\}$. Then $N \geq N_1$, N_2, and N_3. If $n > N$, then

$$\begin{aligned} \left|\frac{a_n}{b_n} - \frac{a}{b}\right| &= \left|\frac{ba_n - ab_n}{bb_n}\right| = \frac{|b(a_n - a) - a(b_n - b)|}{|bb_n|}\\ &\leq |a_n - a| \cdot (1/|b_n|) + |b_n - b| \cdot (|a|/|bb_n|) \qquad \text{(triangle inequality)}\\ &< (|b|/4)\varepsilon \cdot (2/|b|) + (|a|/|b|) \cdot |b|^2\varepsilon/(4|a| + 1) \cdot (2/|b|) < \varepsilon. \end{aligned}$$

Therefore, $\lim_{n\to\infty} \frac{a_n}{b_n} = \frac{a}{b} = \frac{\lim_{n\to\infty} a_n}{\lim_{n\to\infty} b_n}$.

Justification - Law 6

This is a special case for sequential limit of continuous function. In fact, we know that the power function $f(x) = x^p$ is continuous on $[0, \infty)$ if $p > 0$.

The Squeeze Theorem

If $a_n \leq b_n \leq c_n$ for $n \geq n_0$ and $\lim\limits_{n\to\infty} a_n = \lim\limits_{n\to\infty} c_n = L$, then $\lim\limits_{n\to\infty} b_n = L$.

Justification

Let $\varepsilon > 0$ be given. Since $\lim\limits_{n\to\infty} a_n = L$, there is N_1 such that

$$\text{if } n > N_1 \quad \text{then} \quad |a_n - L| < \varepsilon.$$

Since

$$|a_n - L| < \varepsilon \iff -\varepsilon < a_n - L < \varepsilon \iff -\varepsilon + L < a_n < \varepsilon + L,$$

we know that

$$\text{if } n > N_1 \quad \text{then} \quad -\varepsilon + L < a_n < \varepsilon + L.$$

Similarly, since $\lim\limits_{n\to\infty} c_n = L$, there is N_2 such that

$$\text{if } n > N_2 \quad \text{then} \quad -\varepsilon + L < c_n < \varepsilon + L.$$

Let $N = \max\{n_0, N_1, N_2\}$. Then $N \geq n_0, N_1$, and N_2. Thus, if $n > N$, we have

$$a_n \leq b_n \leq c_n, \qquad -\varepsilon + L < a_n < \varepsilon + L, \qquad -\varepsilon + L < c_n < \varepsilon + L.$$

Hence, if $n > N$, then

$$-\varepsilon + L < a_n \leq b_n \leq c_n < \varepsilon + L,$$

which imply

$$|b_n - L| < \varepsilon.$$

Therefore, $\lim\limits_{n\to\infty} b_n = L$.

Zero Limit

If $\lim\limits_{n\to\infty} |a_n| = 0$, then $\lim\limits_{n\to\infty} a_n = 0$.

Justification

Let $\varepsilon > 0$ be given. If $\lim\limits_{n\to\infty} |a_n| = 0$, there is N, such that

$$\text{if } n > N \quad \text{then} \quad \big||a_n|\big| < \varepsilon.$$

Or equivalently,

$$\text{if } n > N \quad \text{then} \quad |a_n| < \varepsilon.$$

Hence, $\lim\limits_{n\to\infty} a_n = 0$.

Sequential Limit for Continuous Function

If $\lim_{n\to\infty} a_n = L$ and the function f is continuous at L, then

$$\lim_{n\to\infty} f(a_n) = f(L).$$

Justification

Let $\varepsilon > 0$ be given. Since f is continuous at L, there is $\delta > 0$, such that

$$\text{if } |x - L| < \delta \quad \text{then} \quad |f(x) - f(L)| < \varepsilon.$$

If $\lim_{n\to\infty} a_n = L$, there is N, such that

$$\text{if } n > N \quad \text{then} \quad |a_n - L| < \delta.$$

Hence, if $n > N$, then

$$|f(a_n) - f(L)| < \varepsilon.$$

Therefore, $\lim_{n\to\infty} f(a_n) = f(L)$.

Limit of Exponential Sequence

The sequence r^n is convergent if $-1 < r \le 1$ and divergent for all other values of r:

$$\lim_{n\to\infty} r^n = \begin{cases} 0, & \text{if } -1 < r < 1, \\ 1, & \text{if } r = 1. \end{cases}$$

Justification

It is clear that the limit holds if $r = 0$ or $r = 1$.
Suppose $-1 < r < 1$ with $r \neq 0$. Let $\varepsilon > 0$ be given. If $n > \dfrac{\ln \varepsilon}{\ln |r|}$, then

$$|r^n - 0| = |r|^n < |r|^{\frac{\ln \varepsilon}{\ln |r|}} = |r|^{\log_{|r|} \varepsilon} = \varepsilon.$$

Hence, in this case, we have $\lim_{n\to\infty} r^n = 0$.
For $r = -1$, r^n is 1 if n is even; -1 if n is odd. So, the sequence $\{r^n\}$ does not approach any finite number. Thus, $\lim_{n\to\infty} r^n$ diverges.
For $r > 1$, we have $0 < 1/r < 1$, so that $\lim_{n\to\infty} (1/r)^n = 0$. Thus, for sufficiently large n, r^n can be larger than any finite number. Hence, $\lim_{n\to\infty} r^n$ diverges.
For $r < -1$, we have $-1 < 1/r < 0$, so that $\lim_{n\to\infty} (1/r)^n = 0$. Thus, for sufficiently large n, r^n can be larger than any finite number if n is even; smaller than any number if n is odd. So, the sequence $\{r^n\}$ does not approach any finite number. Hence, $\lim_{n\to\infty} r^n$ diverges.

Monotonic Sequence

A sequence $\{a_n\}$ is called **increasing** if $a_n < a_{n+1}$ for all $n \geq 1$, that is, $a_1 < a_2 < a_3 < \cdots$. It is called **decreasing** if $a_n > a_{n+1}$ for all $n \geq 1$. A sequence is **monotonic** if it is either increasing or decreasing.

Bounded Sequence

A sequence $\{a_n\}$ is **bounded above** if there is a number M such that

$$a_n \leq M \quad \text{for all } n \geq 1.$$

It is **bounded below** if there is a number m such that

$$m \leq a_n \quad \text{for all } n \geq 1.$$

If it is bounded above and below, then $\{a_n\}$ is a **bounded sequence**.

Completeness Axiom

If a nonempty set S of real numbers that has an upper bound M, that is $x \leq M$ for all x in S, then S has a **least upper bound** b. This means that b is an upper bound for S, but if M is any other upper bound, then $b \leq M$. This is also called the **least-upper-bound property**.

Monotonic Sequence Theorem

Every bounded, monotonic sequence is convergent.

Justification

Suppose $\{a_n\}$ is a bounded increasing sequence. By the least-upper-bound property, there is a least upper bound b for the sequence, that is

$$a_n \leq b, \quad \text{for all } n.$$

Let $\varepsilon > 0$ be given. Since $b - \varepsilon$ is not an upper bound of the sequence, there is a positive integer N such that

$$b - \varepsilon < a_N.$$

Since $\{a_n\}$ is increasing, we know that if $n > N$, then $a_N < a_n$. Hence, if $n > N$,

$$b - \varepsilon < a_n < b < b + \varepsilon,$$

so that $|a_n - b| < \varepsilon$. Therefore, $\lim_{n\to\infty} a_n = b$, that is, $\{a_n\}$ is convergent.
Similarly, $\{a_n\}$ is convergent if $\{a_n\}$ is a bounded decreasing sequence.

Convergent Series

Given a series $\sum_{i=1}^{\infty} a_i = a_1 + a_2 + a_3 + \cdots$, let s_n denote its nth partial sum:

$$s_n = \sum_{i=1}^{n} a_i = a_1 + a_2 + \cdots + a_n.$$

If the sequence $\{s_n\}$ is convergent and $\lim_{n\to\infty} s_n = s$ exists as a real number, then the series $\sum a_n$ is called **convergent** and we write

$$a_1 + a_2 + \cdots + a_n + \cdots = s \quad \text{or} \quad \sum_{i=1}^{n} a_i = s.$$

The number s is called the **sum** of the series. If the sequence $\{s_n\}$ is divergent, then the series is called **divergent**.

Geometric Series

The geometric series

$$\sum_{n=0}^{\infty} ar^n = a + ar + ar^2 + \cdots \qquad (a \neq 0)$$

is convergent if $|r| < 1$ and its sum is

$$\sum_{n=0}^{\infty} ar^n = \frac{a}{1-r} \quad |r| < 1.$$

If $|r| \geq 1$, the geometric sequence is divergent.

Justification

The nth partial sum is

$$s_n = \sum_{i=0}^{n-1} ar^i = a + ar + \cdots + ar^{n-1} = \frac{a(1-r^n)}{1-r}.$$

If $|r| < 1$, we know that $\lim\limits_{n\to\infty} s_n = \lim\limits_{n\to\infty} \dfrac{a(1-r^n)}{1-r} = \dfrac{a}{1-r}$. Thus, the geometric series is convergent if $|r| < 1$.
If $r \leq -1$ or $r > 1$, $\lim\limits_{n\to\infty} r^n$ diverges, so the limit $\lim\limits_{n\to\infty} s_n$ does not exist. When $r = 1$,

$$s_n = \underbrace{a + a + \cdots + a}_{n \text{ terms}} = na$$

diverges as $n \to \infty$. Hence, the geometric series is divergent if $|r| \geq 1$.

Divergence of Harmonic Series

The harmonic series
$$\sum_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots$$
is divergent.

Justification

It is sufficient to show that the partial sums $\{s_k\}$ form an unbounded sequence. In fact,

$$\begin{aligned}
s_2 &= 1 + \tfrac{1}{2}, \\
s_4 &= 1 + \tfrac{1}{2} + (\tfrac{1}{3} + \tfrac{1}{4}) > 1 + \tfrac{1}{2} + (\tfrac{1}{4} + \tfrac{1}{4}) = 1 + \tfrac{2}{2}, \\
s_8 &= 1 + \tfrac{1}{2} + (\tfrac{1}{3} + \tfrac{1}{4}) + (\tfrac{1}{5} + \tfrac{1}{6} + \tfrac{1}{7} + \tfrac{1}{8}) \\
&> 1 + \tfrac{1}{2} + (\tfrac{1}{4} + \tfrac{1}{4}) + (\tfrac{1}{8} + \tfrac{1}{8} + \tfrac{1}{8} + \tfrac{1}{8}) \\
&= 1 + \tfrac{1}{2} + \tfrac{1}{2} + \tfrac{1}{2} = 1 + \tfrac{3}{2}, \\
s_{16} &= 1 + \tfrac{1}{2} + (\tfrac{1}{3} + \tfrac{1}{4}) + (\tfrac{1}{5} + \cdots + \tfrac{1}{8}) + (\tfrac{1}{9} + \cdots + \tfrac{1}{16}) \\
&> 1 + \tfrac{1}{2} + (\tfrac{1}{4} + \tfrac{1}{4}) + (\tfrac{1}{8} + \cdots + \tfrac{1}{8}) + (\tfrac{1}{16} + \cdots + \tfrac{1}{16}) \\
&= 1 + \tfrac{1}{2} + \tfrac{1}{2} + \tfrac{1}{2} + \tfrac{1}{2} = 1 + \tfrac{4}{2}.
\end{aligned}$$

In general, we have
$$s_{2^n} > 1 + \frac{n}{2},$$
which implies that the partial sums $\{s_k\}$ form an unbounded sequence.

A Necessary Condition for Convergence

If the series $\displaystyle\sum_{n=1}^{\infty} a_n$ is convergent, then $\displaystyle\lim_{n\to\infty} a_n = 0$.

Test for Divergence

If $\displaystyle\lim_{n\to\infty} a_n$ does not exist or $\displaystyle\lim_{n\to\infty} a_n \neq 0$, then the series $\displaystyle\sum_{n=1}^{\infty} a_n$ is divergent.

Justification

If $\sum_{n=1}^{\infty} a_n$ is convergent, by the definition, $\lim_{k\to\infty} s_k = s$ exists, where $s_k = \sum_{n=1}^{k} a_n$. Thus,

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} (s_{n+1} - s_n) = \lim_{n\to\infty} s_{n+1} - \lim_{n\to\infty} s_n = s - s = 0.$$

Properties of Infinite Series

If $\sum a_n$ and $\sum b_n$ are convergent series, then so are the series $\sum ca_n$, $\sum(a_n + b_n)$, and $\sum(a_n - b_n)$, and

(i) $\sum_{n=1}^{\infty} ca_n = c\sum_{n=1}^{\infty} a_n$

(ii) $\sum_{n=1}^{\infty}(a_n + b_n) = \sum_{n=1}^{\infty} a_n + \sum_{n=1}^{\infty} b_n$

(iii) $\sum_{n=1}^{\infty}(a_n - b_n) = \sum_{n=1}^{\infty} a_n - \sum_{n=1}^{\infty} b_n$

Justification - Part (i)

Let

$$s_n = \sum_{i=1}^{n} a_i, \qquad s = \sum_{n=1}^{\infty} a_n.$$

The nth partial sum for the series $\sum ca_n$ is

$$u_n = \sum_{i=1}^{n} ca_i.$$

Thus,

$$\lim_{n\to\infty} u_n = \lim_{n\to\infty} \sum_{i=1}^{n} ca_i = \lim_{n\to\infty} c\sum_{i=1}^{n} a_i = c\lim_{n\to\infty} \sum_{i=1}^{n} a_i = c\lim_{n\to\infty} s_n = cs.$$

Therefore $\sum ca_n$ is convergent and its sum is

$$\sum_{n=1}^{\infty} ca_n = cs = c\sum_{n=1}^{\infty} a_n.$$

Justification - Part (ii) and (iii)

Let

$$s_n = \sum_{i=1}^{n} a_i, \qquad s = \sum_{n=1}^{\infty} a_n, \qquad t_n = \sum_{i=1}^{n} b_i, \qquad t = \sum_{n=1}^{\infty} b_n$$

The nth partial sums for the series $\sum(a_n + b_n)$ and $\sum(a_n - b_n)$ are

$$u_n^{\pm} = \sum_{i=1}^{n}(a_i \pm b_i).$$

Thus,

$$\begin{aligned}\lim_{n\to\infty} u_n^{\pm} &= \lim_{n\to\infty} \sum_{i=1}^{n}(a_i \pm b_i) = \lim_{n\to\infty} \left(\sum_{i=1}^{n} a_i \pm \sum_{i=1}^{n} b_i\right) \\ &= \lim_{n\to\infty} \sum_{i=1}^{n} a_i \pm \lim_{n\to\infty} \sum_{i=1}^{n} a_i = \lim_{n\to\infty} s_n \pm \lim_{n\to\infty} t_n = s \pm t.\end{aligned}$$

Therefore $\sum(a_n + b_n)$ and $\sum(a_n - b_n)$ are convergent and the sums are

$$\sum_{n=1}^{\infty}(a_n \pm b_n) = s \pm t = \sum_{n=1}^{\infty} a_n \pm \sum_{n=1}^{\infty} b_n.$$

The Integral Test

Suppose f is a continuous, positive, decreasing function on $[1, \infty)$ and let $a_n = f(n)$. Then the series $\sum_{n=1}^{\infty} a_n$ is convergent if and only if the improper integral $\int_1^{\infty} f(x)\,dx$ is convergent. In other words:

(i) If $\displaystyle\int_1^{\infty} f(x)\,dx$ is convergent, then $\displaystyle\sum_{n=1}^{\infty} a_n$ is convergent.

(ii) If $\displaystyle\int_1^{\infty} f(x)\,dx$ is divergent, then $\displaystyle\sum_{n=1}^{\infty} a_n$ is divergent.

Justification - Part (i)

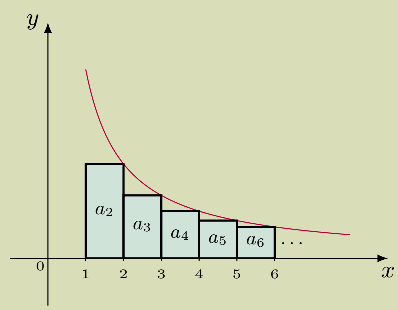
Suppose $\displaystyle\int_1^{\infty} f(x)\,\mathrm{d}x$ converges. To show that $\displaystyle\sum_{n=1}^{\infty} f(n)$ converges, since $f \geq 0$, we only need to show that $\displaystyle\sum_{k=1}^{n} f(k)$ is bounded with respect to n.

For each positive integer n, put $h(x) = f(n+1)$ for $x \in [n, n+1)$. Then h is a function defined on

$[1, \infty)$. Since f is decreases monotonically, we know that $h \leq f$. Hence

$$\begin{aligned}\sum_{k=1}^{n} f(k) = f(1) + \sum_{k=2}^{n} f(k) &\leq f(1) + \int_1^n h(x)\,\mathrm{d}x \\ &\leq f(1) + \int_1^n f(x)\,\mathrm{d}x \leq f(1) + \int_1^\infty f(x)\,\mathrm{d}x,\end{aligned}$$

so that $\sum_{k=1}^{n} f(k)$ is bounded.

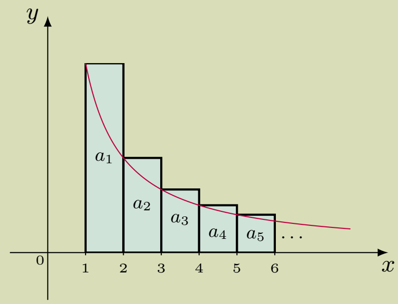


Justification - Part (ii)

Suppose $\int_1^\infty f(x)\,dx$ is divergent. We show that $\sum_{n=1}^{\infty} f(n)$ converges by contradiction.
For each positive integer n, put $g(x) = f(n)$ for $x \in [n, n+1)$. Then g is a function defined on $[1, \infty)$. Since f is decreases monotonically, it is clear that $g \geq f$. For any b, we choose an integer n such that $n \geq b$. Then

$$\int_1^b f(x)\,\mathrm{d}x \leq \int_1^n f(x)\,\mathrm{d}x \leq \int_1^n g(x)\,\mathrm{d}x = \sum_{k=1}^{n-1} f(k) \leq \sum_{k=1}^{\infty} f(k),$$

so that $\int_1^b f(x)\,\mathrm{d}x$ is bounded, which implies that the improper integral is convergent. This contradicts to the hypothesis that $\int_1^\infty f(x)\,dx$ is divergent.



Estimation Theorem for the Integral Test

Suppose $f(k) = a_k$, where f is a continuous, positive, decreasing function for $x \geq n$ and $\sum a_n$ is convergent. Denote $R_n = a_{n+1} + a_{n+2} + \cdots$ to be the remainder. Then

$$\int_{n+1}^{\infty} f(x)\,dx \leq R_n \leq \int_{n}^{\infty} f(x)\,dx,$$

or equivalently,

$$s_n + \int_{n+1}^{\infty} f(x)\,dx \leq \sum_{n=1}^{\infty} a_n \leq s_n + \int_{n}^{\infty} f(x)\,dx.$$

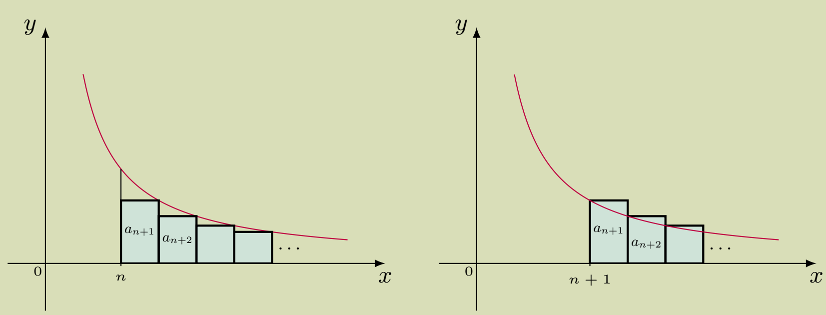
Justification

Since f is decreasing on $[n, \infty)$, by comparing the areas of the rectangles with the area under $y = f(x)$ for $x > n$, we see that

$$R_n = a_{n+1} + a_{n+2} + \cdots \leq \int_{n}^{\infty} f(x)\,dx.$$

Similarly, we have that

$$R_n = a_{n+1} + a_{n+2} + \cdots \geq \int_{n+1}^{\infty} f(x)\,dx.$$



The p-Test

The p-series $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^p}$ is convergent if $p > 1$ and divergent if $p \leq 1$.

Justification

If $p < 0$, then $\displaystyle\lim_{n\to\infty} \frac{1}{n^p} = \infty$. If $p = 0$, then $\displaystyle\lim_{n\to\infty} \frac{1}{n^p} = 1$. In either case $\displaystyle\lim_{n\to\infty} \frac{1}{n^p} \neq 0$, so the p-series diverges by the Test for Divergence.
If $p > 0$, then the function $f(x) = 1/x^p$ is clearly continuous, positive, and decreasing on $[1, \infty)$. For the improper integral $\displaystyle\int_1^{\infty} \frac{1}{x^p}\,dx$, we know that it converges if $p > 1$ and diverges if $p \leq 1$. It follows from the Integral Test that the p-series $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^p}$ converges if $p > 1$ and diverges if $0 < p \leq 1$.

The Comparison Test

Suppose that $\sum a_n$ and $\sum b_n$ are series with positive terms.
(i) If $\sum b_n$ is convergent and $a_n \leq b_n$ for all n, then $\sum a_n$ is also convergent.
(ii) If $\sum b_n$ is divergent and $a_n \geq b_n$ for all n, then $\sum a_n$ is also divergent.

Justification - Part (i)

Let

$$s_n = \sum_{i=1}^{n} a_i, \qquad t_n = \sum_{i=1}^{n} b_i, \qquad t = \sum_{n=1}^{\infty} b_n$$

Since both series have positive terms, the sequences $\{s_n\}$ and $\{t_n\}$ are increasing. Because $t_n \to t$, we have $t_n \leq t$ for all n. Since $a_i \leq b_i$, we have $s_n \leq t_n$. Thus, $s_n \leq t$ for all n. This means that $\{s_n\}$ is increasing and bounded above and therefore converges by the Monotonic Sequence Theorem. Thus $\sum a_n$ converges.

Justification - Part (ii)

Suppose $\sum b_n$ is divergent. We show that $\sum a_n$ is divergent by contradiction.
If $\sum a_n$ is convergent, because $b_n \leq a_n$, from Part (i), we know that $\sum b_n$ is convergent, a contradiction.

The Limit Comparison Test

Suppose that $\sum a_n$ and $\sum b_n$ are series with positive terms. If

$$\lim_{n\to\infty} \frac{a_n}{b_n} = c,$$

where c is a finite number and $c > 0$, then either both series converge or both diverge.

Justification

Let m and M be positive numbers such that $m < c < M$. Because $a_n/b_n \to c$ as $n \to \infty$, there is an integer N such that

$$m < a_n/b_n < M \qquad \text{for } n > N,$$

or equivalently,

$$mb_n < a_n < Mb_n \qquad \text{for } n > N$$

If $\sum b_n$ converges, so does $\sum Mb_n$. Thus $\sum a_n$ converges by part (i) of the Comparison Test. If $\sum b_n$ diverges, so does $\sum mb_n$ and part (ii) of the Comparison Test shows that $\sum a_n$ diverges.

How to Estimate the Remainder

If we have used the Comparison Test to show that a series $\sum a_n$ converges by comparison with a series $\sum b_n$, then we may be able to estimate the sum $\sum a_n$ by comparing remainders.
In fact, since $a_n \leq b_n$ for all n, we have

$$R_n = s - s_n = a_{n+1} + a_{n+2} + \cdots \leq b_{n+1} + b_{n+2} + \cdots = t - t_n = T_n.$$

Thus, if we can compute or estimate the remainder T_n, then we know that R_n is smaller than T_n.
In particular, if $\sum b_n$ is a geometric series, then T_n is the sum of a geometric series and we can sum it exactly. If the Integral Test applies to the series $\sum b_n$, then $T_n \leq \displaystyle\int_n^\infty f(x)\,dx$.

Alternating Series Test

If the alternating series

$$\sum_{n=1}^{\infty} (-1)^{n-1} b_n = b_1 - b_2 + b_3 - b_4 + b_5 - b_6 + \cdots, \quad b_n > 0$$

satisfies
(i) $b_{n+1} \leq b_n$ for all n
(ii) $\lim_{n\to\infty} b_n = 0$
then the series is convergent.

Justification

First consider the even partial sums:

$$\begin{aligned}
s_2 &= b_1 - b_2 \geq 0, && \text{(since } b_1 \geq b_2\text{)} \\
s_4 &= s_2 + (b_3 - b_4) \geq s_2, && \text{(since } b_3 \geq b_4\text{)} \\
&\vdots \\
s_{2n} &= s_{2n-2} + (b_{2n-1} - b_{2n}) \geq s_{2n-2} && \text{(since } b_{2n-1} \geq b_{2n}\text{)}
\end{aligned}$$

Also,

$$s_{2n} = b_1 - (b_2 - b_3) - (b_4 - b_5) - \cdots - (b_{2n-2} - b_{2n-1}) - b_{2n} \leq b_1.$$

So, $\{s_{2n}\}$ is increasing and bounded above. Hence, it is convergent by the Monotonic Sequence Theorem. Denote $s = \lim_{n\to\infty} s_{2n}$. Now we compute the limit of the odd partial sums:

$$\lim_{n\to\infty} s_{2n+1} = \lim_{n\to\infty} (s_{2n} + b_{2n+1}) = \lim_{n\to\infty} s_{2n} + \lim_{n\to\infty} b_{2n+1} = s + 0 = s.$$

Since both the even and odd partial sums converge to s, we have $\lim_{n\to\infty} s_n = s$ and so the series is convergent.

Alternating Series Estimation Theorem

If $s = \sum_{n=1}^{\infty}(-1)^n b_n$ is the sum of an alternating series that satisfies
(i) $b_{n+1} \leq b_n$ for all n
(ii) $\lim_{n\to\infty} b_n = 0$
then

$$|R_n| = |s - s_n| \leq b_{n+1}.$$

Justification

As in the proof of Alternating Series Test, $\{s_{2n}\}$ is increasing and bounded and converges to the sum of the series. Similarly, we have

$$\begin{aligned}
s_3 &= b_1 - b_2 + b_3 = b_1 - (b_2 - b_3) \leq b_1 = s_1, && \text{(since } b_2 \geq b_3\text{)} \\
s_5 &= s_3 - (b_4 - b_5) \leq s_3, && \text{(since } b_4 \geq b_5\text{)} \\
&\vdots \\
s_{2n+1} &= s_{2n-1} - (b_{2n} - b_{2n+1}) \leq s_{2n-1} && \text{(since } b_{2n} \geq b_{2n+1}\text{)}
\end{aligned}$$

Also,

$$s_{2n+1} = (b_1 - b_2) + (b_3 - b_4) + \cdots + (b_{2n-1} - b_{2n}) + b_{2n+1} \geq b_1 - b_2.$$

So, the odd partial sums $\{s_{2n+1}\}$ is decreasing and bounded below. Because we know that both $\{s_{2n}\}$

and $\{s_{2n+1}\}$ converge to s, we get

$$s_{2n} \leq s \leq s_{2n+1}, \qquad \text{for all } n.$$

It follows that

$$|s - s_n| \leq |s_{n+1} - s_n| = b_{n+1}.$$

Absolutely Convergent and Conditionally Convergent

Absolutely Convergent

A series $\sum a_n$ is called **absolutely convergent** if the series of absolute values $\sum |a_n|$ is convergent.

Conditionally Convergent

A series $\sum a_n$ is called **conditionally convergent** if it is convergent but not absolutely convergent.

Convergence of Absolutely Convergent Series

If a series $\sum a_n$ is absolutely convergent, then it is convergent.

Justification

It is clear that

$$0 \leq a_n + |a_n| \leq 2|a_n|, \qquad \text{for all } n.$$

If $\sum a_n$ is absolutely convergent, then $\sum 2|a_n|$ is convergent. By the Comparison Test, $\sum(a_n + |a_n|)$ is convergent. Thus,

$$\sum a_n = \sum (a_n + |a_n|) - \sum |a_n|$$

is convergent.

Rearrangement of an Infinite Series

By a rearrange-ment of an infinite series $\sum a_n$ we mean a series obtained by simply changing the order of the terms.

Rearrangement of an Absolutely Convergent Series

If $\sum a_n$ is an absolutely convergent series with sum s, then any rearrangement of $\sum a_n$ has the same sum s.

Rearrangement of a Conditionally Convergent Series

If $\sum a_n$ is a conditionally convergent series and r is any real number whatsoever, then there is a rearrangement of $\sum a_n$ that has a sum equal to r.

The Ratio Test

(i) If $\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| = L < 1$, then the series $\sum_{n=1}^{\infty} a_n$ is absolutely convergent (and therefore convergent).

(ii) If $\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| = L > 1$ or $\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| = \infty$, then the series $\sum_{n=1}^{\infty} a_n$ is divergent.

(iii) If $\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| = 1$, the Ratio Test is inconclusive.

Justification - Part (i)

For $\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| = L < 1$, we take any number r such that $L < r < 1$. Thus, by the definition, there is N such that whenever $n \geq N$,

$$\left|\left|\frac{a_{n+1}}{a_n}\right| - L\right| < r - L.$$

So, whenever $n \geq N$,

$$\left|\frac{a_{n+1}}{a_n}\right| < r.$$

or equivalently,

$$|a_{n+1}| < r|a_n|.$$

Putting n successively equal to $N, N+1, N+2, \ldots$ in the inequality, we obtain

$$\begin{aligned}
|a_{N+1}| &< r|a_N|,\\
|a_{N+2}| &< r|a_{N+1}| < r^2|a_N|,\\
|a_{N+3}| &< r|a_{N+2}| < r^3|a_N|,\\
&\vdots
\end{aligned}$$

so that

$$|a_{N+k}| < r^k|a_N|, \qquad \text{for all } k \geq 1.$$

Since $|r| = r < 1$, the geometric series $\sum_{k=1}^{\infty} |a_N| r^k$ is convergent. By the Comparison Test, the series $\sum_{n=N+1}^{\infty} |a_n|$ is convergent. Hence, $\sum_{n=1}^{\infty} a_n$ is absolutely convergent.

Justification - Part (ii)

To show $\sum_{n=1}^{\infty} a_n$ is divergent, by the Test for Divergence, it is sufficient to show that $\lim_{n\to\infty} a_n \neq 0$.

If $\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| = L > 1$, then, by the definition, there is N such that whenever $n \geq N$,

$$\left|\left|\frac{a_{n+1}}{a_n}\right| - L\right| < (L-1).$$

Thus, whenever $n \geq N$,

$$-(L-1) < \left|\frac{a_{n+1}}{a_n}\right| - L,$$

or

$$\left|\frac{a_{n+1}}{a_n}\right| > 1.$$

Hence, for $n \geq N$,

$$|a_n| > a_N,$$

which implies that $\lim_{n\to\infty} a_n \neq 0$. By the Test for Divergence, the series $\sum_{n=1}^{\infty} a_n$ is divergent.

Similarly as above, if $\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| = \infty$, we can also show that $\lim_{n\to\infty} a_n \neq 0$.

Justification - Part (iii)

Consider the p-series $\sum_{n=1}^{\infty}\frac{1}{n}$ with $p = 1$ and $\sum_{n=1}^{\infty}\frac{1}{n^2}$ with $p = 2 > 1$. The former one is divergent and the latter one is convergent. For each series, we have

$$\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| = 1.$$

Hence, the Ratio Test is inconclusive if $\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| = 1$.

The Root Test

(i) If $\lim_{n\to\infty}\sqrt[n]{|a_n|} = L < 1$, then the series $\sum_{n=1}^{\infty} a_n$ is absolutely convergent (and therefore convergent).

(ii) If $\lim_{n\to\infty}\sqrt[n]{|a_n|} = L > 1$ or $\lim_{n\to\infty}\sqrt[n]{|a_n|} = \infty$, then the series $\sum_{n=1}^{\infty} a_n$ is divergent.

(iii) If $\lim_{n\to\infty} \sqrt[n]{|a_n|} = 1$, the Root Test is inconclusive.

Justification - Part (i)

If $\lim_{n\to\infty} \sqrt[n]{|a_n|} = L < 1$, we take any number r such that $L < r < 1$. Thus, by the definition, there is N such that whenever $n \geq N$,

$$\left| \sqrt[n]{|a_n|} - L \right| < (r - L).$$

So, whenever $n \geq N$,

$$\sqrt[n]{|a_n|} - L < (r - L),$$

or

$$\sqrt[n]{|a_n|} < r.$$

Hence, for $n \geq N$,

$$|a_n| < r^n.$$

The geometric series $\sum r^n$ is convergent because the ratio satisfies $|r| < 1$. By the Comparison Test, the series $\sum_{n=1}^{\infty} |a_n|$ converges, so $\sum_{n=1}^{\infty} a_n$ is absolutely convergent.

Justification - Part (ii)

To show $\displaystyle\sum_{n=1}^{\infty} a_n$ is divergent, by the Test for Divergence, it is sufficient to show that $\lim_{n\to\infty} a_n \neq 0$.

If $\lim_{n\to\infty} \sqrt[n]{|a_n|} = L > 1$, then, by the definition, there is N such that whenever $n \geq N$,

$$\left| \sqrt[n]{|a_n|} - L \right| < (L - 1).$$

Thus, whenever $n \geq N$,

$$-(L - 1) < \sqrt[n]{|a_n|} - L,$$

or

$$\sqrt[n]{|a_n|} > 1.$$

Hence, for $n \geq N$,

$$|a_n| > 1,$$

which implies that $\lim_{n\to\infty} a_n \neq 0$.

Similarly as above, if $\lim_{n\to\infty} \sqrt[n]{|a_n|} = \infty$, we can also show that $\lim_{n\to\infty} a_n \neq 0$.

Justification - Part (iii)

Consider the p-series $\sum_{n=1}^{\infty} \frac{1}{n}$ with $p = 1$ and $\sum_{n=1}^{\infty} \frac{1}{n^2}$ with $p = 2 > 1$. The former one is divergent and the latter one is convergent. For each series, we have

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

Hence, the Root Test is inconclusive if $\lim_{n\to\infty} \sqrt[n]{|a_n|} = 1$.

Power Series

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's are constants called the **coefficients** of the series. Generally, 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.

Convergence of Power Series

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 converges if $|x-a| < R$ and diverges if $|x-a| > R$. The number R is called the **radius of convergence of the power series**.

Radius of Convergence

The number R in case (iii) is called the **radius of convergence of the power series**. By convention, the radius of convergence is $R = 0$ in case (i) and $R = \infty$ in case (ii). The **interval of convergence** of a power series is the interval that consists of all values of x for which the series converges.

How to Determine the Interval of Convergence

Let $\sum_{n=0}^{\infty} c_n(x-a)^n$ be a given power series.

Step 1. Compute the radius of convergence R by using the Ratio Test (or sometimes the Root Test) by

$$R = \frac{1}{\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right|} = \lim_{n\to\infty}\left|\frac{a_n}{a_{n+1}}\right| \quad \text{or} \quad R = \frac{1}{\lim_{n\to\infty}\sqrt[n]{|a_n|}}.$$

Step 2. The Ratio and Root Tests always fail when x is an endpoint of the interval of convergence. Apply some *other* test to determine whether the series is convergent or divergent at an endpoint.

Differentiation and Integration of Power Series

If the power series $\sum_{n=0}^{\infty} c_n(x-a)^n$ has radius of convergence $R > 0$, then the function f defined by

$$f(x) = c_0 + c_1(x-a) + c_2(x-a)^2 + \cdots = \sum_{n=0}^{\infty} c_n(x-a)^n$$

is differentiable (and therefore continuous) on the interval $(a-R, a+R)$ and

(i) $f'(x) = c_1 + 2c_2(x-a) + 3c_3(x-a)^2 + \cdots = \sum_{n=1}^{\infty} nc_n(x-a)^{n-1}$.

(ii) $\int f(x)\,dx = C + c_0(x-a) + c_1\frac{(x-a)^2}{2} + c_2\frac{(x-a)^3}{3} + \cdots = C + \sum_{n=0}^{\infty} c_n\frac{(x-a)^{n+1}}{n+1}$.

The radii of convergence of the power series in Equations (i) and (ii) are both R.

Important Maclaurin Series and Their Radii of Convergence

Power Series	Radius of Convergence	Interval of Convergence
$\dfrac{1}{1-x} = \sum_{n=0}^{\infty} x^n = 1 + x + x^2 + x^3 + \cdots$	$R = 1$	$(-1, 1)$
$e^x = \sum_{n=0}^{\infty} \dfrac{x^n}{n!} = 1 + \dfrac{x}{1!} + \dfrac{x^2}{2!} + \dfrac{x^3}{3!} + \cdots$	$R = \infty$	$(-\infty, \infty)$
$\sin x = \sum_{n=0}^{\infty} (-1)^n \dfrac{x^{2n+1}}{(2n+1)!} = x - \dfrac{x^3}{3!} + \dfrac{x^5}{5!} - \dfrac{x^7}{7!} + \cdots$	$R = \infty$	$(-\infty, \infty)$
$\cos x = \sum_{n=0}^{\infty} (-1)^n \dfrac{x^{2n}}{(2n)!} = 1 - \dfrac{x^2}{2!} + \dfrac{x^4}{4!} - \dfrac{x^6}{6!} + \cdots$	$R = \infty$	$(-\infty, \infty)$
$\tan^{-1} x = \sum_{n=0}^{\infty} (-1)^n \dfrac{x^{2n+1}}{2n+1} = x - \dfrac{x^3}{3} + \dfrac{x^5}{5} - \dfrac{x^7}{7} + \cdots$	$R = 1$	$[-1, 1]$
$\ln(1+x) = \sum_{n=1}^{\infty} (-1)^{n-1} \dfrac{x^n}{n} = x - \dfrac{x^2}{2} + \dfrac{x^3}{3} - \dfrac{x^4}{4} + \cdots$	$R = 1$	$(-1, 1]$
$(1+x)^k = \sum_{n=0}^{\infty} \binom{k}{n} x^n = 1 + kx + \dfrac{k(k-1)}{2!}x^2 + \dfrac{k(k-1)(k-2)}{3!}x^3 + \cdots$	$R = 1$	at least $(-1, 1)$

Power Series Expansion

If f has a power series representation (expansion) at a, that is, if

$$f(x) = \sum_{n=0}^{\infty} c_n (x-a)^n, \quad |x-a| < R,$$

then its coefficients are given by the formula

$$c_n = \frac{f^{(n)}(a)}{n!}.$$

Justification

We know that it is legitimate to differentiate the infinite series term by term at any number x satisfying $|x-a| < R$. Thus, for each n, $n = 0, 1, 2, \ldots$, we differentiate and substitute $x = a$ to have

$$f^{(n)}(a) = 2 \cdot 3 \cdot 4 \cdots n c_n = n! c_n,$$

so that

$$c_n = \frac{f^{(n)}(a)}{n!}.$$

Taylor Series and Maclaurin Series

For a given function f, the power series

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots$$

is called the **Taylor series of the function** f **at** a (or **about** a or **centered at** a). For the special case $a = 0$ the Taylor series becomes

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!}x^n = f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \cdots$$

which is called the **Maclaurin series of the function** f.
Let T_n denote the nth partial sum of the Taylor series of the function f at a:

$$\begin{aligned} T_n(x) &= \sum_{i=1}^{n} \frac{f^{(i)}(a)}{i!}(x-a)^i \\ &= f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n. \end{aligned}$$

It is a polynomial of degree n called the **nth-degree Taylor polynomial of** f **at** a. The term $R_n(x) = f(x) - T_n(x)$ is called the **remainder** of the Taylor series.

Convergence of the Taylor Series

If $\lim_{n\to\infty} R_n(x) = 0$ for $|x-a| < R$, then f is equal to the sum of its Taylor series on the interval $|x-a| < R$.

Taylor's Inequality

If $|f^{(n+1)}(x)| \le M$ for $|x-a| \le d$, then the remainder $R_n(x)$ of the Taylor series satisfies the inequality

$$|R_n(x)| \le \frac{M}{(n+1)!}|x-a|^{n+1} \quad \text{for } |x-a| \le d,$$

which implies that $\lim_{n\to\infty} R_n(x) = 0$ for $|x-a| \le d$.

Product of Power Series

Suppose the power series $\sum_{n=0}^{\infty} a_n x^n$ has radius of convergence R_1 and the power series $\sum_{n=0}^{\infty} b_n x^n$ has radius of convergence R_2. Then whenever both of these power series convergent we have that

$$\left(\sum_{n=0}^{\infty} a_n x^n\right)\left(\sum_{n=0}^{\infty} b_n x^n\right) = \sum_{n=0}^{\infty}\left(\sum_{j=0}^{n} a_j b_{n-j}\right) x^n$$
$$= \sum_{n=0}^{\infty} \left(a_0 b_n + a_1 b_{n-1} + \cdots + a_{n-1} b_1 + a_n b_0\right) x^n.$$

This power series has a radius of convergence R such that $R \geq \min\{R_1, R_2\}$.

How to Compute the Product of Power Series

The product can be found by a vertical set-up. For instance, for the power series representations

$$\begin{aligned} e^x &= 1 + x + \tfrac{1}{2}x^2 + \tfrac{1}{6}x^3 + \cdots, \\ \sin x &= x - \tfrac{1}{3!}x^3 + \cdots, \end{aligned}$$

to compute $e^x \sin x$, we collect like terms just as for polynomials

$$\begin{array}{rllll} & 1 + x + & \tfrac{1}{2}x^2 + \tfrac{1}{6}x^3 + \cdots & & \\ \times & x & - \tfrac{1}{6}x^3 + \cdots & & \\ \hline & x + & x^2 + \tfrac{1}{2}x^3 + \tfrac{1}{6}x^4 + \cdots & & \\ + & & - \tfrac{1}{6}x^3 - \tfrac{1}{6}x^4 - \cdots & & \\ \hline & x + & x^2 + \tfrac{1}{3}x^3 + \cdots & & \end{array}$$

Thus,

$$e^x \sin x = x + x^2 + \tfrac{1}{3}x^3 + \cdots.$$

Quotient of Power Series

Suppose the power series $\sum_{n=0}^{\infty} a_n x^n$ has radius of convergence R_1 and the power series $\sum_{n=0}^{\infty} b_n x^n$ has radius of convergence R_2. If $b_0 \neq 0$, then in a neighbourhood of $x = 0$, the quotient has a power series representation:

$$\frac{\sum_{n=0}^{\infty} a_n x^n}{\sum_{n=0}^{\infty} b_n x^n} = \sum_{n=0}^{\infty} c_n x^n,$$

where

$$c_0 = \frac{a_0}{b_0}, \qquad c_n = \frac{1}{b_0}\left[a_n - \sum_{k=1}^{n} b_n c_{n-k}\right] \quad \text{for } n = 1, 2, 3, \ldots.$$

This power series has a radius of convergence R such that $R \geq \min\{R_1, R_2, x_0\}$, where x_0 is the zero of the function represented by the power series in the denominator nearest to $x = 0$.

How to Compute the Quotient of Power Series

The quotient can be found by a vertical set-up. For instance, for the power series representations

$$\begin{aligned} \sin x &= x - \tfrac{1}{3!}x^3 + \tfrac{1}{5!}x^5 - \cdots, \\ \cos x &= 1 - \tfrac{1}{2!}x^2 + \tfrac{1}{4!}x^4 - \cdots, \end{aligned}$$

to compute $\tan x = \sin x / \cos x$, we use a procedure like long division

$$\begin{array}{r|l}
 & x + \tfrac{1}{3}x^3 + \tfrac{2}{15}x^5 + \cdots \\ \cline{2-2}
1 - \tfrac{1}{2}x^2 + \tfrac{1}{24}x^4 - \cdots & x - \tfrac{1}{6}x^3 + \tfrac{1}{120}x^5 - \cdots \\
 & x - \tfrac{1}{2}x^3 + \tfrac{1}{24}x^5 - \cdots \\ \cline{2-2}
 & \qquad \tfrac{1}{3}x^3 - \tfrac{1}{30}x^5 + \cdots \\
 & \qquad \tfrac{1}{3}x^3 - \tfrac{1}{6}x^5 + \cdots \\ \cline{2-2}
 & \qquad\qquad \tfrac{2}{15}x^5 + \cdots \\
 & \qquad\qquad \tfrac{2}{15}x^5 + \cdots \\ \cline{2-2}
 & \qquad\qquad\qquad \cdots
\end{array}$$

Thus,

$$\tan x = x + \tfrac{1}{3}x^3 + \tfrac{2}{15}x^5 + \cdots.$$

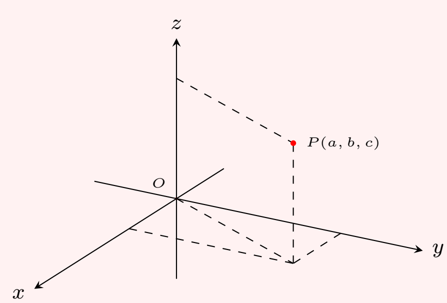
Chapter 12

Vectors and the Geometry of Space

Three-Dimensional Coordinate Systems

We first choose a fixed point O (the origin) and three directed lines through O that are perpendicular to each other, called the **coordinate axes** and labeled the x-axis, y-axis, and z-axis. Usually we think of the x- and y-axes as being horizontal and the z-axis as being vertical. The direction of the z-axis is determined by the **right-hand rule**. We represent any point in space by an ordered triple (a, b, c) of real numbers.

The xy-plane is the plane that contains the x- and y-axes; the yz-plane contains the y- and z-axes; the xz-plane contains the x- and z-axes. They are so-called the three **coordinate planes**.

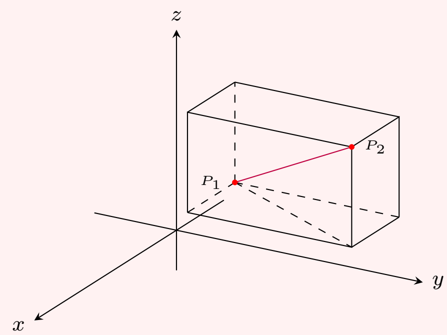


The Cartesian product $\mathbb{R} \times \mathbb{R} \times \mathbb{R} = \{(x, y, z) \,|\, x, y, z \in \mathbb{R}\}$ is the set of all ordered triples of real numbers and is denoted by $\mathbb{R}^3$. We have given a one-to-one correspondence between points P in space and ordered triples (a, b, c) in $\mathbb{R}^3$. It is called a three-dimensional rectangular coordinate system.

Distance Formula in Three Dimensions

The distance $|P_1P_2|$ between the points $P_1(x_1, y_1, z_1)$ and $P_1(x_2, y_2, z_2)$ is

$$|P_1P_2| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - y_1)^2}.$$

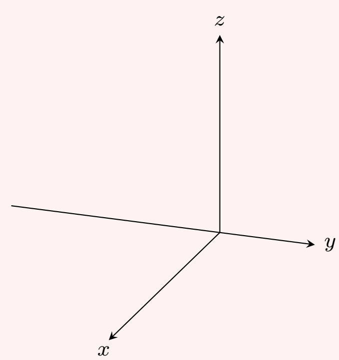


Equation of a Plane

In 3D, a linear equation of the form

$$ax + by + cz = r$$

is a plane. In particular, if k is a constant, then $x = k$ represents a plane parallel to the yz-plane, $y = k$ is a plane parallel to the xz-plane, and $z = k$ is a plane parallel to the xy-plane.



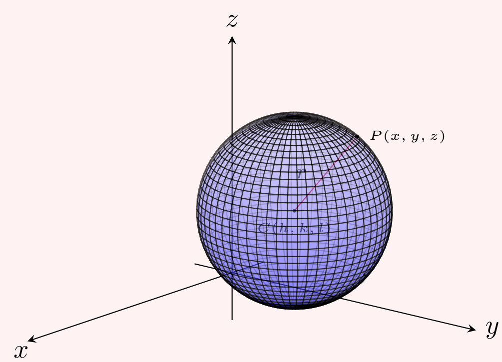
Equation of a Sphere

An equation of a sphere with center $C(h, k, l)$ and radius r is

$$(x - h)^2 + (y - k)^2 + (z - l)^2 = r^2.$$

In particular, if the center is the origin O, then an equation of the sphere is

$$x^2 + y^2 + z^2 = r^2.$$



Vector

The term **vector** is used to indicate a quantity that has both magnitude and direction. A vector is often represented by an arrow or a directed line segment. The length of the arrow represents the magnitude of the vector and the arrow points in the direction of the vector. We denote a vector by printing a letter in boldface $\mathbf{v}$ or by putting an arrow above the letter $\vec{v}$.

Two vectors are **equal** if they have the same magnitude and the same direction. Equal vectors may start at different positions. Note that when the vectors are equal, the directed line segments are **parallel**.

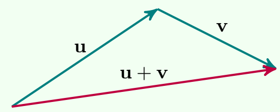
A **scalar** is a single real number, in other words, a scalar is a quantity that only has magnitude but no direction.

Vector Addition

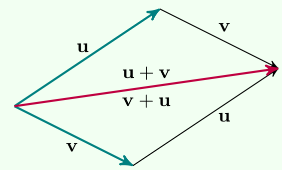
If $\mathbf{u}$ and $\mathbf{v}$ are vectors positioned so the initial point of $\mathbf{v}$ is at the terminal point of $\mathbf{u}$, then the sum $\mathbf{u} + \mathbf{v}$ is the vector from the initial point of $\mathbf{u}$ to the terminal point of $\mathbf{v}$.

Geometric Construction

The sum of the vectors $\mathbf{u}$ and $\mathbf{v}$ can be constructed either by the Triangle Law or by the Parallelogram Law.



Triangle Law



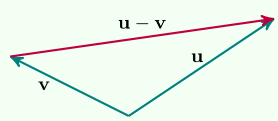
Parallelogram Law

Scalar Multiplication

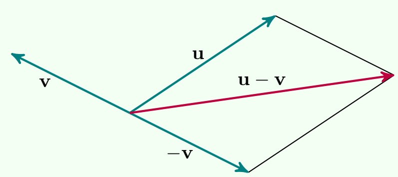
If c is a scalar and $\mathbf{v}$ is a vector, then the scalar multiple $c\mathbf{v}$ is the vector whose length is $|c|$ times the length of $\mathbf{v}$ and whose direction is the same as $\mathbf{v}$ if $c > 0$ and is opposite to $\mathbf{v}$ if $c < 0$. If $c = 0$ or $\mathbf{v} = \mathbf{0}$, then $c\,\mathbf{v} = \mathbf{0}$.

Vector Subtraction

The difference of the vectors $\mathbf{u}$ and $\mathbf{v}$, $\mathbf{u} - \mathbf{v}$, can be constructed either by the Triangle Law or by the Parallelogram Law.



Triangle Law



Parallelogram Law

Vector Components

If we place the initial point of a vector $\mathbf{a}$ at the origin of a rectangular coordinate system, then the terminal point of $\mathbf{a}$ has coordinates of the form (a_1, a_2) or (a_1, a_2, a_3), depending on whether our coordinate system is two- or three-dimensional. These coordinates are called the **components** of $\mathbf{a}$ and we write

$$\mathbf{a} = \langle a_1, a_2 \rangle \quad \text{or} \quad \mathbf{a} = \langle a_1, a_2, a_3 \rangle.$$

Vector Representation

Given the points $A(x_1, y_1)$ and $B(x_2, y_2)$, the vector $\mathbf{a}$ with representation $\overrightarrow{AB}$ is

$$\mathbf{a} = \langle x_2 - x_1, y_2 - y_1 \rangle.$$

In three dimensions, given the points $A(x_1, y_1, z_1)$ and $B(x_2, y_2, z_2)$, the vector $\mathbf{a}$ with representation $\overrightarrow{AB}$ is

$$\mathbf{a} = \langle x_2 - x_1, y_2 - y_1, z_2 - z_1 \rangle.$$

Vector Length

The **length** (or **magnitude**) of the two-dimensional vector $\mathbf{a} = \langle a_1, a_2 \rangle$ is

$$|\,\mathbf{a}\,| = \sqrt{a_1^2 + a_2^2}.$$

Similarly, for the three-dimensional vector $\mathbf{a} = \langle a_1, a_2, a_3 \rangle$,

$$|\,\mathbf{a}\,| = \sqrt{a_1^2 + a_2^2 + a_3^2}.$$

Operations on Components

If $\mathbf{a} = \langle a_1, a_2 \rangle$ and $\mathbf{b} = \langle b_1, b_2 \rangle$, then

$$\begin{aligned} \mathbf{a} + \mathbf{b} &= \langle a_1 + b_1, a_2 + b_2 \rangle, \\ \mathbf{a} - \mathbf{b} &= \langle a_1 - b_1, a_2 - b_2 \rangle, \\ c\,\mathbf{a} &= \langle ca_1, ca_2 \rangle. \end{aligned}$$

Similarly, for three-dimensional vectors,

$$\begin{aligned} \langle a_1, a_2, a_3 \rangle + \langle b_1, b_2, b_3 \rangle &= \langle a_1 + b_1, a_2 + b_2, a_3 + b_3 \rangle, \\ \langle a_1, a_2, a_3 \rangle - \langle b_1, b_2, b_3 \rangle &= \langle a_1 - b_1, a_2 - b_2, a_3 - b_3 \rangle, \\ c\,\langle a_1, a_2, a_3 \rangle &= \langle ca_1, ca_2, ca_3 \rangle. \end{aligned}$$

Properties of Vectors

If $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ are vectors and c and d are scalars, then

1. $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$ (Commutativity)
2. $\mathbf{a} + (\mathbf{b} + \mathbf{c}) = (\mathbf{a} + \mathbf{b}) + \mathbf{c}$ (Associativity of Vector Addition)
3. $\mathbf{a} + \mathbf{0} = \mathbf{a}$ (Zero Vector)
4. $\mathbf{a} + (-\mathbf{a}) = \mathbf{0}$ (Negative Vector)

5. $c\,(\mathbf{a}+\mathbf{b}) = c\,\mathbf{a} + c\,\mathbf{b}$ (Distributivity of Vector Addition)

6. $(c+d)\,\mathbf{a} = c\,\mathbf{a} + d\,\mathbf{a}$ (Distributivity of Scalar Addition)

7. $(cd)\,\mathbf{a} = c\,(d\,\mathbf{a})$ (Associativity of Scalar Multiplication)

8. $1\,\mathbf{a} = \mathbf{a}$ (Scalar One)

Standard Basis Vectors

In two dimensions,

$$\mathbf{i} = \langle 1, 0 \rangle, \qquad \mathbf{j} = \langle 0, 1 \rangle.$$

Any vector $\mathbf{a} = \langle a_1, a_2 \rangle$ can be expressed as $\mathbf{a} = a_1\,\mathbf{i} + a_2\,\mathbf{j}$.
Similarly, in three dimensions,

$$\mathbf{i} = \langle 1, 0, 0 \rangle, \qquad \mathbf{j} = \langle 0, 1, 0 \rangle, \qquad \mathbf{j} = \langle 0, 0, 1 \rangle.$$

Any vector $\mathbf{a} = \langle a_1, a_2, a_3 \rangle$ can be expressed as $\mathbf{a} = a_1\,\mathbf{i} + a_2\,\mathbf{j} + a_3\,\mathbf{k}$.

Normalization

If $\mathbf{a} \neq \mathbf{0}$, then

$$\mathbf{u} = \frac{1}{|\,\mathbf{a}\,|}\,\mathbf{a}$$

is the unit vector that has the same direction as $\mathbf{a}$.

Dot Product

In two dimensions, if $\mathbf{a} = \langle a_1, a_2 \rangle$ and $\mathbf{b} = \langle b_1, b_2 \rangle$, then the **dot product** (or **inner product**) of $\mathbf{a}$ and $\mathbf{b}$ is the number $\mathbf{a} \cdot \mathbf{b}$ given by

$$\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2.$$

Similarly, in three dimensions, if $\mathbf{a} = \langle a_1, a_2, a_3 \rangle$ and $\mathbf{b} = \langle b_1, b_2, b_3 \rangle$, then

$$\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2 + a_3 b_3.$$

Properties of the Dot Product

If $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ are vectors and c is a scalar, then

1. $\mathbf{a} \cdot \mathbf{a} = |\,\mathbf{a}\,|^{\mathbf{2}}$
2. $\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$
3. $\mathbf{a} \cdot (\mathbf{b} + \mathbf{c}) = \mathbf{a} \cdot \mathbf{b} + \mathbf{a} \cdot \mathbf{c}$
4. $(c\,\mathbf{a}) \cdot \mathbf{b} = c\,(\mathbf{a} \cdot \mathbf{b}) = \mathbf{a} \cdot (c\,\mathbf{b})$
5. $\mathbf{0} \cdot \mathbf{a} = 0$

Justification

We only prove the properties in 3D. The proof for 2D is similar.

1. $\mathbf{a}\cdot\mathbf{a} = a_1^2 + a_2^2 + a_3^2 = |\,\mathbf{a}\,|^{\mathbf{2}}$.

2. $\mathbf{a}\cdot\mathbf{b} = a_1b_1 + a_2b_2 + a_3b_3 = \mathbf{b}\cdot\mathbf{a}$.

3.
$$\begin{aligned}\mathbf{a}\cdot(\mathbf{b}+\mathbf{c}) &= \langle a_1, a_2, a_3\rangle\cdot\langle b_1+c_1, b_2+c_2, b_3+c_3\rangle\\ &= a_1(b_1+c_1) + a_2(b_2+c_2) + a_3(b_3+c_3)\\ &= (a_1b_1 + a_2b_2 + a_3b_3) + (a_1c_1 + a_2c_2 + a_3c_3)\\ &= \langle a_1, a_2, a_3\rangle\cdot\langle b_1, b_2, b_3\rangle + \langle a_1, a_2, a_3\rangle\cdot\langle c_1, c_2, c_3\rangle\\ &= \mathbf{a}\cdot\mathbf{b} + \mathbf{a}\cdot\mathbf{c}.\end{aligned}$$

4. Since

$$\begin{aligned}(c\,\mathbf{a})\cdot\mathbf{b} &= \langle ca_1, ca_2, ca_3\rangle\cdot\langle b_1, b_2, b_3\rangle = ca_1b_1 + ca_2b_2 + ca_3b_3,\\ c\,(\mathbf{a}\cdot\mathbf{b}) &= c\,(a_1b_1 + a_2b_2 + a_3b_3) = ca_1b_1 + ca_2b_2 + ca_3b_3,\\ \mathbf{a}\cdot(c\,\mathbf{b}) &= \langle a_1, a_2, a_3\rangle\cdot\langle cb_1, cb_2, cb_3\rangle = ca_1b_1 + ca_2b_2 + ca_3b_3,\end{aligned}$$

we have
$$(c\,\mathbf{a})\cdot\mathbf{b} = c\,(\mathbf{a}\cdot\mathbf{b}) = \mathbf{a}\cdot(c\,\mathbf{b}).$$

5. $\mathbf{0}\cdot\mathbf{a} = 0\,a_1 + 0\,a_2 + 0\,a_3 = 0$.

Angle Between the Vectors

If θ is the angle between the vectors $\mathbf{a}$ and $\mathbf{b}$, then

$$\mathbf{a}\cdot\mathbf{b} = |\,\mathbf{a}\,|\,|\,\mathbf{b}\,|\,\cos\theta.$$

Angle Between the Vectors

If θ is the angle between the nonzero vectors $\mathbf{a}$ and $\mathbf{b}$, then

$$\cos\theta = \frac{\mathbf{a}\cdot\mathbf{b}}{|\,\mathbf{a}\,|\,|\,\mathbf{b}\,|}.$$

Orthogonality

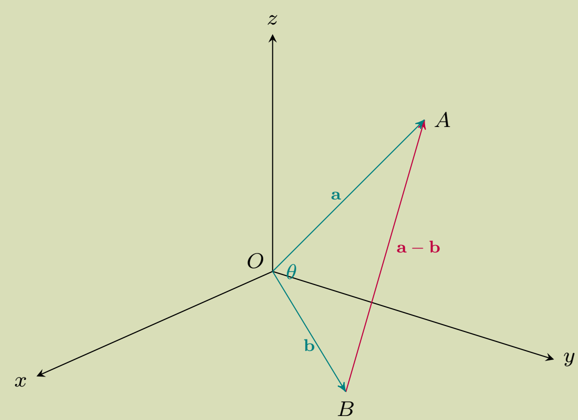
Two vectors $\mathbf{a}$ and $\mathbf{b}$ are orthogonal if and only if $\mathbf{a}\cdot\mathbf{b} = 0$.

Justification

As shown in the figure, we apply the Law of Cosines to triangle OAB and get
$$|AB|^2 = |OA|^2 + |OB|^2 - 2|OA|\,|OB|\,\cos\theta,$$

or

$$|\mathbf{a}-\mathbf{b}|^{\mathbf{2}} = |\mathbf{a}|^{\mathbf{2}} + |\mathbf{b}|^{\mathbf{2}} - \mathbf{2}|\mathbf{a}||\mathbf{b}|\cos\theta.$$



From Properties 1, 2, and 3 of the dot product, we can rewrite the left side of the last equation as follows:

$$\begin{aligned}|\mathbf{a}-\mathbf{b}|^{\mathbf{2}} &= (\mathbf{a}-\mathbf{b})\cdot(\mathbf{a}-\mathbf{b})\\ &= \mathbf{a}\cdot\mathbf{a}-\mathbf{a}\cdot\mathbf{b}-\mathbf{b}\cdot\mathbf{a}+\mathbf{b}\cdot\mathbf{b}\\ &= |\mathbf{a}|^2 - 2\mathbf{a}\cdot\mathbf{b} + |\mathbf{b}|^2.\end{aligned}$$

Thus, we have

$$|\mathbf{a}|^{\mathbf{2}} + |\mathbf{b}|^{\mathbf{2}} - \mathbf{2}|\mathbf{a}||\mathbf{b}|\cos\theta = |\mathbf{a}|^{\mathbf{2}} - \mathbf{2a}\cdot\mathbf{b} + |\mathbf{b}|^{\mathbf{2}}.$$

which gives

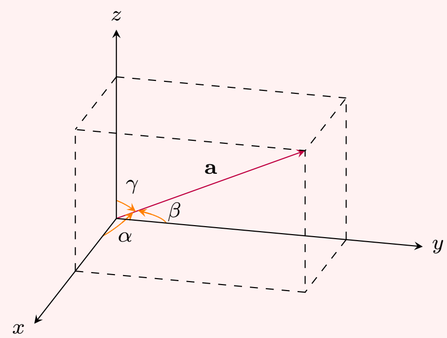
$$2\mathbf{a}\cdot\mathbf{b} = 2|\mathbf{a}||\mathbf{b}|\cos\theta,$$

or

$$\mathbf{a}\cdot\mathbf{b} = |\mathbf{a}||\mathbf{b}|\cos\theta.$$

Direction Angles

In three dimensions, the **direction angles** of a nonzero vector $\mathbf{a}$ are the angles α, β, and γ (in the interval $[0,\pi]$ that $\mathbf{a}$ makes with the positive x-, y-, and z-axes.



Direction Cosines

The direction cosines of the vector $\mathbf{a} = \langle a_1, a_2, a_3 \rangle$ are

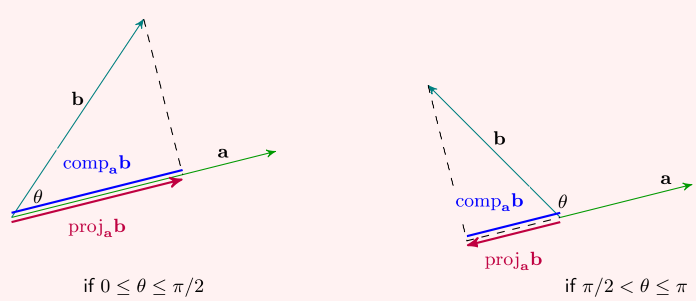
$$\cos\alpha = \frac{a_1}{|\,\mathbf{a}\,|}, \quad \cos\beta = \frac{a_2}{|\,\mathbf{a}\,|}, \quad \cos\gamma = \frac{a_3}{|\,\mathbf{a}\,|}.$$

Since $|\,\mathbf{a}\,| = \sqrt{a_1^2 + a_2^2 + a_3^2}$, we see that

$$\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1.$$

Vector Projections

The **scalar projection of b onto a** (also called the **component of b along a**), dented by $\mathrm{comp}_{\mathbf{a}}\mathbf{b}$, is defined to be the signed magnitude of the vector projection, which is the number $|\mathbf{b}|\cos\theta$, where θ is the angle between $\mathbf{a}$ and $\mathbf{b}$.



The **vector projection of a on b**, denoted by $\mathrm{proj}_{\mathbf{a}}\mathbf{b}$ is a vector whose magnitude is the scalar projection of $\mathbf{a}$ on $\mathbf{b}$.

Scalar Projection and Vector Projection

Scalar projection of $\mathbf{b}$ onto $\mathbf{a}$: $\quad \text{comp}_{\mathbf{a}}\mathbf{b} = \dfrac{\mathbf{a}\cdot\mathbf{b}}{|\,\mathbf{a}\,|}$

Vector projection of $\mathbf{b}$ onto $\mathbf{a}$: $\quad \text{proj}_{\mathbf{a}}\mathbf{b} = \left(\dfrac{\mathbf{a}\cdot\mathbf{b}}{|\,\mathbf{a}\,|}\right)\dfrac{\mathbf{a}}{|\,\mathbf{a}\,|}$

Justification

From the diagrams, for $0 \le \theta \le \pi$, the equation

$$\mathbf{a}\cdot\mathbf{b} = |\,\mathbf{a}\,||\,\mathbf{b}\,|\cos\theta = |\,\mathbf{a}\,|\,(|\,\mathbf{b}\,|\cos\theta)$$

shows that the dot product of $\mathbf{a}$ and $\mathbf{b}$ can be interpreted as the length of $\mathbf{a}$ times the scalar projection of $\mathbf{b}$ onto $\mathbf{a}$. Thus,

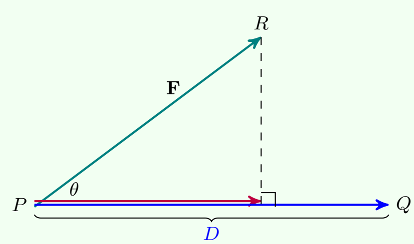
$$\text{comp}_{\mathbf{a}}\mathbf{b} = |\,\mathbf{b}\,|\cos\theta = \frac{\mathbf{a}\cdot\mathbf{b}}{|\,\mathbf{a}\,|}.$$

Since $\mathbf{a}/|\,\mathbf{a}\,|$ is the unit vector that has the same direction as $\mathbf{a}$, we have

$$\text{proj}_{\mathbf{a}}\mathbf{b} = \text{comp}_{\mathbf{a}}\mathbf{b} = \left(\frac{\mathbf{a}\cdot\mathbf{b}}{|\,\mathbf{a}\,|}\right)\frac{\mathbf{a}}{|\,\mathbf{a}\,|}.$$

Work Done by a Constant Force

We know that the work done by a constant force F in moving an object through a distance d as $W = Fd$, if the force is directed along the line of motion. Suppose, however, that the constant force is a vector $\mathbf{F} = \overrightarrow{PR}$ pointing in some other direction, as shown in the figure.



If the force moves the object from P to Q, then the **displacement vector** is $\mathbf{D} = \overrightarrow{PQ}$. The work done by this force is the product of the component of the force along $\mathbf{D}$ and the distance moved:

$$W = (|\,\mathbf{F}\,|\cos\theta)\,|\,\mathbf{D}\,| = \mathbf{F}\cdot\mathbf{D}.$$

That is, the work done by a constant force $\mathbf{F}$ is the dot product $\mathbf{F}\cdot\mathbf{D}$, where $\mathbf{D}$ is the displacement vector.

Cross Product

If $\mathbf{a} = \langle a_1, a_2, a_3 \rangle$ and $\mathbf{b} = \langle b_1, b_2, b_3 \rangle$, then the **cross product** of $\mathbf{a}$ and $\mathbf{b}$ is the vector

$$\mathbf{a} \times \mathbf{b} = \langle a_2 b_3 - a_3 b_2, a_3 b_1 - a_1 b_3, a_1 b_2 - a_2 b_1 \rangle,$$

or equivalently,

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}.$$

Length of the Cross Product

If θ is the angle between $\mathbf{a}$ and $\mathbf{b}$, $0 \leq \theta \leq \pi$, then

$$|\,\mathbf{a} \times \mathbf{b}\,| = |\,\mathbf{a}\,|\,|\,\mathbf{b}\,| \sin\theta.$$

Justification

From the definitions of the cross product and length of a vector, we have

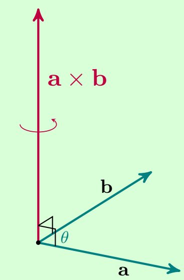
$$\begin{aligned}
|\,\mathbf{a} \times \mathbf{b}\,|^2 &= (a_2 b_3 - a_3 b_2)^2 + (a_3 b_1 - a_1 b_3)^2 + (a_1 b_2 - a_2 b_1)^2 \\
&= a_2^2 b_3^2 - 2a_2 a_3 b_2 b_3 + a_3^2 b_2^2 + a_3^2 b_1^2 - 2a_1 a_3 b_1 b_3 + a_1^2 b_3^2 + a_1^2 b_2^2 - 2a_1 a_2 b_1 b_2 + a_2^2 b_1^2 \\
&= (a_1^2 + a_2^2 + a_3^2)(b_1^2 + b_2^2 + b_3^2) - (a_1 b_1 + a_2 b_2 + a_3 b_3)^2 \\
&= |\,\mathbf{a}\,|^2 |\,\mathbf{b}\,|^2 - (\mathbf{a} \cdot \mathbf{b})^2 \\
&= |\,\mathbf{a}\,|^2 |\,\mathbf{b}\,|^2 - (|\,\mathbf{a}\,|\,|\,\mathbf{b}\,| \cos\theta)^2 \\
&= |\,\mathbf{a}\,|^2 |\,\mathbf{b}\,|^2 (1 - \cos^2\theta) \\
&= |\,\mathbf{a}\,|^2 |\,\mathbf{b}\,|^2 \sin^2\theta.
\end{aligned}$$

Since $\sin\theta \geq 0$ for $0 \leq \theta \leq \pi$, taking square roots gives

$$|\,\mathbf{a} \times \mathbf{b}\,| = |\,\mathbf{a}\,|\,|\,\mathbf{b}\,| \sin\theta.$$

Orthogonality

The vector $\mathbf{a} \times \mathbf{b}$ is orthogonal to both $\mathbf{a}$ and $\mathbf{b}$ and its direction is given by the right-hand rule, as shown in the diagram.



Justification

To show that $\mathbf{a}\times\mathbf{b}$ is orthogonal to $\mathbf{a}$, we compute their dot product as follows:

$$\begin{aligned}(\mathbf{a}\times\mathbf{b})\cdot\mathbf{a} &= \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} a_1 - \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} a_2 + \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} a_3 \\ &= (a_2b_3 - b_2a_3)a_1 - (a_1b_3 - b_1a_3)a_2 + (a_1b_2 - b_1a_2)a_3 \\ &= a_1a_2b_3 - a_1b_2a_3 - a_1a_2b_3 + b_1a_2a_3 + a_1b_2a_3 - b_1a_2a_3 = 0.\end{aligned}$$

A similar computation shows that $(\mathbf{a}\times\mathbf{b})\cdot\mathbf{b} = 0$. Therefore $(\mathbf{a}\times\mathbf{b})$ is orthogonal to both $\mathbf{a}$ and $\mathbf{b}$. The argument for the right-hand rule is non-trivial and will be skipped.

Parallel Vectors

Two nonzero vectors $\mathbf{a}$ and $\mathbf{b}$ are parallel if and only if $\mathbf{a}\times\mathbf{b} = 0$.

Justification

Two nonzero vectors $\mathbf{a}$ and $\mathbf{b}$ are parallel if and only if $\theta = 0$ or $\theta = \pi$. In either case, $\sin\theta = 0$ so that $|\,\mathbf{a}\times\mathbf{b}\,| = |\,\mathbf{a}\,|\,|\,\mathbf{b}\,|\,\sin\theta = 0$. Thus, $\mathbf{a}\times\mathbf{b} = \mathbf{0}$.

Cross Products of the Standard Basis Vectors

$$\begin{array}{lll} \mathbf{i}\times\mathbf{j} = \mathbf{k} & \mathbf{j}\times\mathbf{k} = \mathbf{i} & \mathbf{k}\times\mathbf{i} = \mathbf{j} \\ \mathbf{j}\times\mathbf{i} = -\mathbf{k} & \mathbf{k}\times\mathbf{j} = -\mathbf{i} & \mathbf{i}\times\mathbf{k} = -\mathbf{j} \end{array}$$

Justification

By the definition of the cross product,

$$\mathbf{i}\times\mathbf{j} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix} = 0\,\mathbf{i} - 0\,\mathbf{j} + 1\,\mathbf{k} = \mathbf{k}.$$

The other equalities can be verified similarly.

Properties of the Cross Product

If $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ are vectors and c is a scalar, then

1. $\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}$

2. $(c\,\mathbf{a}) \times \mathbf{b} = c\,(\mathbf{a} \times \mathbf{b}) = \mathbf{a} \times (c\,\mathbf{b})$

3. $\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}$

4. $(\mathbf{a} + \mathbf{b}) \times \mathbf{c} = \mathbf{a} \times \mathbf{c} + \mathbf{b} \times \mathbf{c}$

5. $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}$

6. $\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\,\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\,\mathbf{c}$

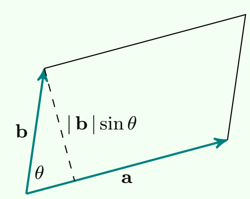
Justification

All these properties can be proved by writing the vectors in terms of their components and using the definition of a cross product. For example,

$$\begin{aligned}\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) &= a_1(b_2c_3 - b_3c_2) + a_2(b_3c_1 - b_1c_3) + a_3(b_1c_2 - b_2c_1)\\ &= a_1b_2c_3 - a_1b_3c_2 + a_2b_3c_1 - a_2b_1c_3 + a_3b_1c_2 - a_3b_2c_1\\ &= (a_2b_3 - a_3b_2)c_1 + (a_3b_1 - a_1b_3)c_2 + (a_1b_2 - a_2b_1)c_3 = (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}.\end{aligned}$$

Cross Product - Area of the Parallelogram

The area of the parallelogram determined by $\mathbf{a}$ and $\mathbf{b}$ is the magnitude of their cross product $A = |\,\mathbf{a}\times\mathbf{b}\,|$.



Justification

If $\mathbf{a}$ and $\mathbf{b}$ are represented by directed line segments with the same initial point, then the parallelogram determined by $\mathbf{a}$ and $\mathbf{b}$ has base $|\,\mathbf{a}\,|$, altitude $|\,\mathbf{b}\,|\sin\theta$, so the area is

$$A = |\,\mathbf{a}\,|\,(|\,\mathbf{b}\,|\sin\theta) = |\,\mathbf{a} \times \mathbf{b}\,|.$$

Scalar Triple Product

If $\mathbf{a} = \langle a_1, a_2, a_3\rangle$, $\mathbf{b} = \langle b_1, b_2, b_3\rangle$, and $\mathbf{c} = \langle c_1, c_2, c_3\rangle$, then their **scalar triple product** is

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}.$$

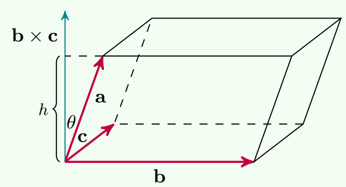
Justification

By the definitions of the cross product and the dot product,

$$\begin{aligned}\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) &= a_1(b_2c_3-b_3c_2)+a_2(b_3c_1-b_1c_3)+a_3(b_1c_2-b_2c_1)\\ &= \begin{vmatrix} a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\\ c_1 & c_2 & c_3\end{vmatrix}.\end{aligned}$$

Scalar Triple Product - Volume of the Parallelepiped

The volume of the parallelepiped determined by the vectors **a**, **b**, and **c** is the magnitude of their scalar triple product $V = |\,\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})\,|$.



Justification

As shown in the figure, for the parallelepiped determined by the vectors **a**, **b**, and **c**, the area of the base parallelogram is $A = |\,\mathbf{b}\times\mathbf{c}\,|$. If θ is the angle between **a** and $\mathbf{b}\times\mathbf{c}$, then the height h of the parallelepiped is $h = |\,\mathbf{a}\,|\cos\theta$. Here we must use $|\cos\theta|$ instead of $\cos\theta$ in case $\theta > \pi/2$. Therefore the volume of the parallelepiped is

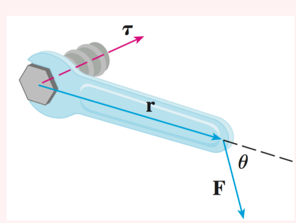
$$V = Ah = |\,\mathbf{b}\times\mathbf{c}\,|\,|\,\mathbf{a}\,|\,|\cos\theta| = |\,\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})\,|\,.$$

Torque - Application

Consider a force **F** acting on a rigid body at a point given by a position vector **r**. The **torque** (relative to the origin) is defined to be the cross product of the position and force vectors

$$\boldsymbol{\tau} = \mathbf{r}\times\mathbf{F}.$$

It measures the tendency of the body to rotate about the origin. The direction of the torque vector indicates the axis of rotation.



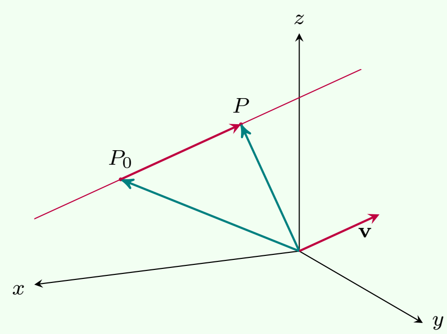
Line

Vector Equation

Let $P_0(x_0, y_0, z_0)$ be a fixed point and $\mathbf{v} = \langle a, b, c\rangle$ be a given nonzero vector. Then the line through the point P_0 and parallel to the $\mathbf{v}$ (called the **direction vector**), has the following **vector equation**:

$$\mathbf{r} = \mathbf{r}_0 + t\,\mathbf{v},$$

where $\mathbf{r}_0 = \langle x_0, y_0, z_0\rangle$ is the position vector of the point P_0 and $\mathbf{r} = \langle x, y, z\rangle$ is the position vector of an arbitrary point P on the line.



Parametric Equations

By writing the components of the vector equation of the line that passes through the point $P_0(x_0, y_0, z_0)$ and is parallel to the nonzero vector $\mathbf{v} = \langle a, b, c\rangle$, we get **parametric equations** of the line:

$$x = x_0 + at, \qquad y = y_0 + bt, \qquad z = z_0 + ct$$

Symmetric Equations

By eliminating the parameter t in parametric equations of the line that passes through the point $P_0(x_0, y_0, z_0)$ and is parallel to the nonzero vector $\mathbf{v} = \langle a, b, c\rangle$, we get **symmetric equations** of the line:

$$\frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c}$$

when a, b, and c are nonzero.
If $a = 0$, the equations of the line become:

$$x = x_0, \qquad \frac{y - y_0}{b} = \frac{z - z_0}{c}$$

Similarly, we could get the equations when b or c is zero.

Line Segment

The line segment from $\mathbf{r}_0$ to $\mathbf{r}_1$ is given by the vector equation

$$\mathbf{r}(t) = (1-t)\,\mathbf{r}_0 + t\,\mathbf{r}_1, \quad 0 \le t \le 1.$$

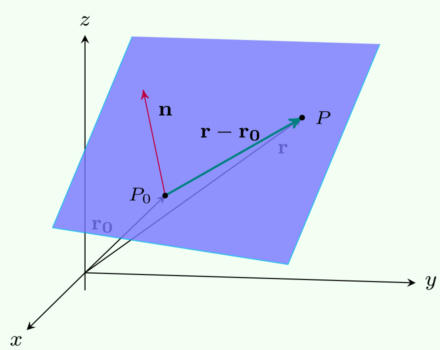
Plane

Vector Equation

Suppose $P_0(x_0, y_0, z_0)$ is a point in the plane and $\mathbf{n}$ is a **normal vector** of plane. Then the plane has the following **vector equation**:

$$\mathbf{n} \cdot (\mathbf{r} - \mathbf{r_0}) = 0,$$

where $\mathbf{r}_0 = \langle x_0, y_0, z_0 \rangle$ is the position vector of the point P_0 and $\mathbf{r} = \langle x, y, z \rangle$ is the position vector of an arbitrary point P on the line.



Scalar Equation

By writing the components of the vector equation of the plane that passes through the point $P_0(x_0, y_0, z_0)$ with normal vector $\mathbf{n} = \langle a, b, c \rangle$, we get **scalar equation** of the plane:

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0$$

General Equation

The general equation of a plane is

$$ax + by + cz + d = 0.$$

Parallel Planes

Two planes are parallel if their normal vectors are parallel.

Distance from a Point to a Plane

The distance from the point (x_1, y_1, z_1) to the plane $ax + by + cz + d = 0$ is

$$D = \frac{|ax_1 + by_1 + cz_1 + d|}{\sqrt{a^2 + b^2 + c^2}}.$$

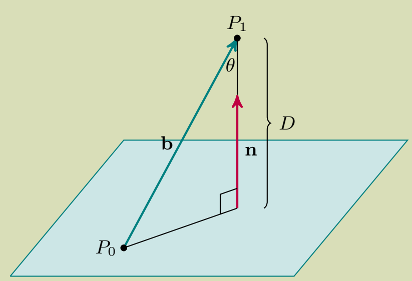
Justification

Let $P_0 = (x_0, y_0, z_0)$ be any point in the given plane and let $\mathbf{b}$ be the vector corresponding to $\overrightarrow{P_0P_1}$. Then

$$\mathbf{b} = \langle x_1 - x_0, y_1 - y_0, z_1 - z_0 \rangle.$$

The distance D from P_1 to the plane is equal to the absolute value of the scalar projection of $\mathbf{b}$ onto the normal vector $\mathbf{n} = \langle a, b, c \rangle$. Thus

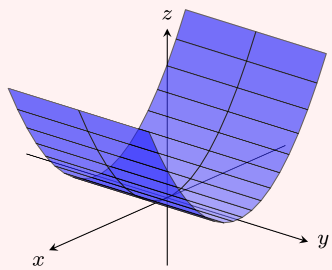
$$\begin{aligned}
D = |\text{comp}_{\mathbf{n}}\mathbf{b}| &= \frac{|\,\mathbf{n} \cdot \mathbf{b}\,|}{|\,\mathbf{n}\,|} \\
&= \frac{|a(x_1 - x_0) + b(y_1 - y_0) + c(z_1 - z_0)|}{\sqrt{a^2 + b^2 + c^2}} \\
&= \frac{|(ax_1 + by_1 + cz_1) - (ax_0 + by_0 + cz_0)|}{\sqrt{a^2 + b^2 + c^2}} = \frac{|ax_1 + by_1 + cz_1 + d|}{\sqrt{a^2 + b^2 + c^2}}.
\end{aligned}$$

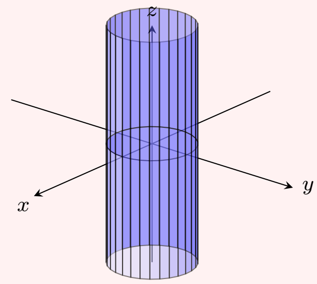


Cylinder

A **cylinder** is a surface that consists of all lines (called **rulings**) that are parallel to a given line and pass through a given plane curve.
For example, the graph of the surface $z = x^2$ is a **parabolic cylinder**, made up of infinitely many

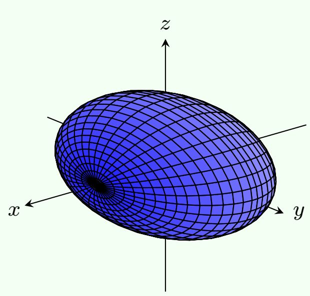
shifted copies of the same parabola. Here the rulings of the cylinder are parallel to the y-axis. The graph of the surface $x^2 + y^2 = 1$ is a **circular cylinder**, made up of infinitely many shifted copies of the same circles. Here the rulings of the cylinder are parallel to the z-axis.
If one of the variables x, y, or z is missing from the equation of a surface, then the surface is a cylinder.





Quadric Surfaces

Ellipsoid

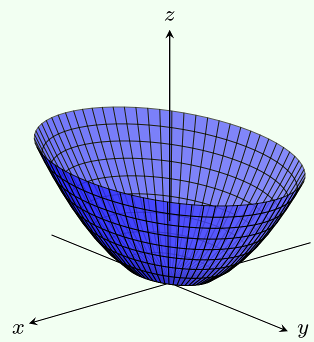


Equation: $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c^2} = 1$

All traces are ellipses.

If $a = b = c$, the ellipsoid is a sphere.

Elliptic Paraboloid



Equation: $\dfrac{z}{c} = \dfrac{x^2}{a^2} + \dfrac{y^2}{b^2}$

Horizontal traces are ellipses.

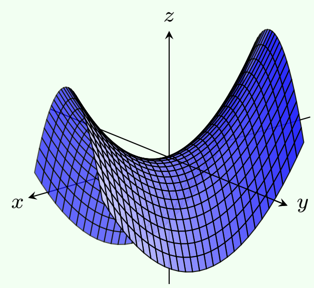
Vertical traces are parabolas.

The variable raised to the first power

indicates the axis of the paraboloid.

The case where $c > 0$ is illustrated.

Hyperbolic Paraboloid



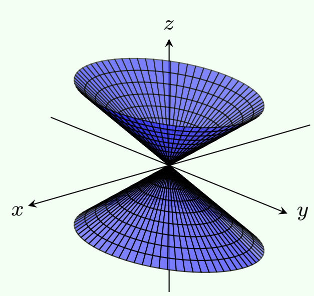
Equation: $\dfrac{z}{c} = \dfrac{x^2}{a^2} - \dfrac{y^2}{b^2}$

Horizontal traces are hyperbolas.

Vertical traces are parabolas.

The case where $c > 0$ is illustrated.

Cone



Equation: $\dfrac{z^2}{c^2} = \dfrac{x^2}{a^2} + \dfrac{y^2}{b^2}$

Horizontal traces are ellipses.

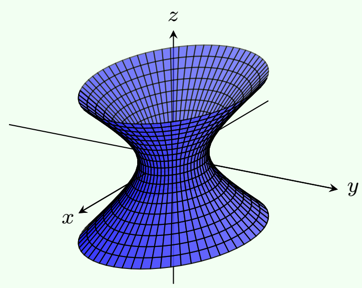
Vertical traces in the planes

$x = k$ and $y = k$

are hyperbolas if $k \neq 0$

but are pairs of lines if $k = 0$.

Hyperboloid of One Sheet



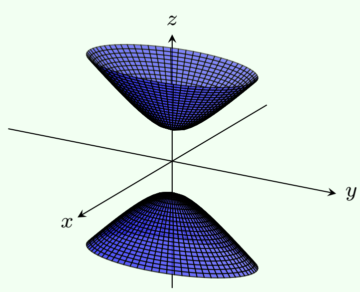
Equation: $\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1$

Horizontal traces are ellipses.

Vertical traces are hyperbolas.

The axis of symmetry corresponds to the variable whose coefficient is negative.

Hyperboloid of Two Sheet



Equation: $-\frac{x^2}{a^2} - \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$

Horizontal traces are ellipses if $k > c$ or $k < -c$.

Vertical traces are hyperbolas.

The two minus signs indicate two sheets.

Chapter 13

Vector Functions

Vector Function

A **vector-valued function**, or **vector function**, is simply a function whose domain is a set of real numbers and whose range is a set of vectors. In three-dimensional spaces, this means that for every number t in the domain of $\mathbf{r}$ there is a unique vector in V_3 denoted by $\mathbf{r}(t)$. If $f(t)$, $g(t)$, and $h(t)$ are the components of the vector $\mathbf{r}(t)$, then f, g, and h are real-valued functions called the component functions of $\mathbf{r}$ and we can write

$$\mathbf{r}(t) = \langle f(t), g(t), h(t) \rangle = f(t)\,\mathbf{i} + g(t)\,\mathbf{j} + h(t)\,\mathbf{k}.$$

We use the letter t to denote the independent variable because it represents time in most applications of vector functions.

Limit of a Vector Function

If $\mathbf{r}(t) = \langle f(t), g(t), h(t) \rangle$, then

$$\lim_{t \to a} \mathbf{r}(t) = \left\langle \lim_{t \to a} f(t), \lim_{t \to a} g(t), \lim_{t \to a} h(t) \right\rangle,$$

provided the limits of the component functions exist.

Continuous

A vector function $\mathbf{r}$ is **continuous at** a if

$$\lim_{t \to a} \mathbf{r}(t) = \mathbf{r}(a).$$

Parametric Equations - Plane and Space Curves

Suppose that f and g are continuous real-valued functions on an interval I. Then the set C of all points (x, y) in space, where

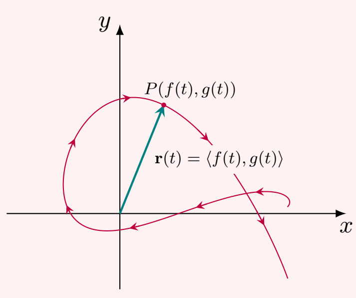
$$x = f(t), \qquad y = g(t)$$

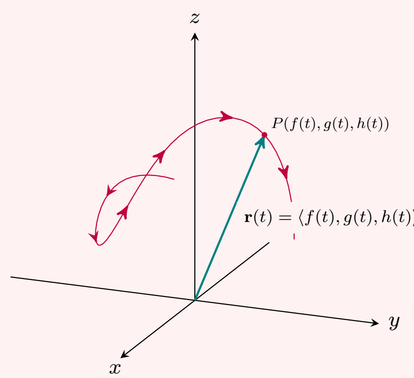
and t varies throughout the interval I, is called a **plane curve**. Theses equations are called **parametric equations** of C and t is called a **parameter**. The vector function $\mathbf{r}(t) = \langle f(t), g(t)\rangle$ is the position vector of the point $P(f(t), g(t))$ on C.

Similarly, suppose that f, g, and h are continuous real-valued functions on an interval I. Then the set C of all points (x, y, z) in space, where

$$x = f(t), \qquad y = g(t), \qquad z = h(t)$$

and t varies throughout the interval I, is called a **space curve**. The vector function $\mathbf{r}(t) = \langle f(t), g(t), h(t)\rangle$ is the position vector of the point $P(f(t), g(t), h(t))$ on C.





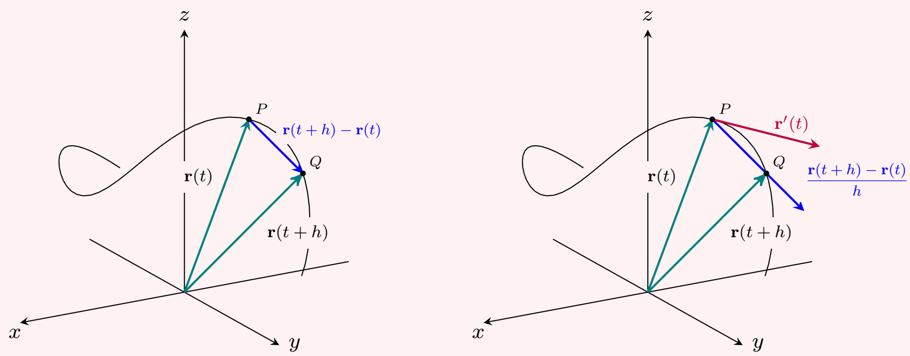
Derivative of a Vector Function

The **derivative $\mathbf{r}'$** of a vector function $\mathbf{r}$ is defined as

$$\frac{d\mathbf{r}}{dt} = \mathbf{r}'(t) = \lim_{h\to 0} \frac{\mathbf{r}(t+h) - \mathbf{r}(t)}{h}.$$

Geometrically, it is the **tangent vector** to the curve C defined by $\mathbf{r}$ at the point given by the position vector $\mathbf{r}(t)$, provided $\mathbf{r}'(t)$ exists and $\mathbf{r}'(t) \neq \mathbf{0}$. The **tangent line** to C at P is defined to be the line through P parallel to the tangent vector $\mathbf{r}'(t)$. The unit tangent vector is

$$\mathbf{T}(t) = \frac{\mathbf{r}'(t)}{|\,\mathbf{r}'(t)\,|}.$$



Component Differentiation

If $\mathbf{r}(t) = \langle f(t), g(t), h(t) \rangle = f(t)\,\mathbf{i} + g(t)\,\mathbf{j} + h(t)\,\mathbf{k}$, where f, g, and h are differentiable functions, then

$$\mathbf{r}'(t) = \langle f'(t), g'(t), h'(t) \rangle = f'(t)\,\mathbf{i} + g'(t)\,\mathbf{j} + h'(t)\,\mathbf{k}.$$

Justification

In fact,

$$\begin{aligned}
\mathbf{r}'(t) &= \lim_{\Delta t \to 0} \frac{1}{\Delta t}\left[\mathbf{r}(t+\Delta t) - \mathbf{r}(t)\right] \\
&= \lim_{\Delta t \to 0} \frac{1}{\Delta t}\left[\langle f(t+\Delta t), g(t+\Delta t), h(t+\Delta t)\rangle - \langle f(t), g(t), h(t)\rangle\right] \\
&= \lim_{\Delta t \to 0} \left\langle \frac{f(t+\Delta t) - f(t)}{\Delta t}, \frac{g(t+\Delta t) - g(t)}{\Delta t}, \frac{h(t+\Delta t) - h(t)}{\Delta t} \right\rangle \\
&= \left\langle \lim_{\Delta t \to 0} \frac{f(t+\Delta t) - f(t)}{\Delta t}, \lim_{\Delta t \to 0} \frac{g(t+\Delta t) - g(t)}{\Delta t}, \lim_{\Delta t \to 0} \frac{h(t+\Delta t) - h(t)}{\Delta t} \right\rangle \\
&= \langle f'(t), g'(t), h'(t) \rangle .
\end{aligned}$$

Properties of Differentiation for Vector Functions

Suppose $\mathbf{u}$ and $\mathbf{v}$ are differentiable vector functions, c is a scalar, and f is a real-valued function. Then

1. $\dfrac{d}{dt}\left[\mathbf{u}(t) + \mathbf{v}(t)\right] = \mathbf{u}'(t) + \mathbf{v}'(t)$.

2. $\dfrac{d}{dt}\left[c\,\mathbf{u}(t)\right] = c\,\mathbf{u}'(t)$.

3. $\dfrac{d}{dt}\left[f(t)\,\mathbf{u}(t)\right] = f'(t)\,\mathbf{u}(t) + f(t)\,\mathbf{u}'(t)$.

4. $\dfrac{d}{dt}[\mathbf{u}(t)\cdot\mathbf{v}(t)] = \mathbf{u}'(t)\cdot\mathbf{v}(t) + \mathbf{u}(\mathbf{t})\cdot\mathbf{v}'(t)$.

5. $\dfrac{d}{dt}[\mathbf{u}(t)\times\mathbf{v}(t)] = \mathbf{u}'(t)\times\mathbf{v}(t) + \mathbf{u}(\mathbf{t})\times\mathbf{v}'(t)$.

6. $\dfrac{d}{dt}[\mathbf{u}(f(t))] = f'(t)\,\mathbf{u}'(t)$. (Chain Rule)

Justification

All these properties can be proved by writing the vectors in terms of their components. For example, let

$$\mathbf{u}(t) = \langle f_1(t), f_2(t), f_3(t)\rangle, \qquad \mathbf{v}(t) = \langle g_1(t), g_2(t), g_3(t)\rangle.$$

Then

$$\begin{aligned}\frac{d}{dt}[\mathbf{u}(t)\cdot\mathbf{v}(t)] &= \frac{d}{dt}\sum_{i=1}^{3} f_i(t)g_i(t) = \sum_{i=1}^{3}\frac{d}{dt}[f_i(t)g_i(t)]\\ &= \sum_{i=1}^{3}[f_i'(t)g_i(t) + f_i(t)g_i'(t)]\\ &= \sum_{i=1}^{3} f_i'(t)g_i(t) + \sum_{i=1}^{3} f_i(t)g_i'(t)\\ &= \mathbf{u}'(t)\cdot\mathbf{v}(t) + \mathbf{u}(\mathbf{t})\cdot\mathbf{v}'(t).\end{aligned}$$

Definite Integral of Vector Function

If $\mathbf{r}(t) = \langle f(t), g(t), h(t)\rangle = f(t)\,\mathbf{i} + g(t)\,\mathbf{j} + h(t)\,\mathbf{k}$, where f, g, and h are integrable functions, then

$$\begin{aligned}\int_a^b \mathbf{r}(t)\,dt &= \left\langle \int_a^b f(t)\,dt, \int_a^b g(t)\,dt, \int_a^b h(t)\,dt \right\rangle\\ &= \left(\int_a^b f(t)\,dt\right)\mathbf{i} + \left(\int_a^b g(t)\,dt\right)\mathbf{j} + \left(\int_a^b h(t)\,dt\right)\mathbf{k}.\end{aligned}$$

Arc Length

Suppose a parametric curve is given by the parametric equations

$$x = f(t), \qquad y = g(t), \qquad z = h(t), \qquad \alpha \le t \le \beta$$

where f', g', and h' are continuous and the parametric curve is traversed exactly once as t increases from α to β, or equivalently, suppose the curve has the vector equation $\mathbf{r}(t) = \langle f(t), g(t), h(t)\rangle =$

$f(t)\,\mathbf{i} + g(t)\,\mathbf{j} + h(t)\,\mathbf{k}$, $\alpha \le t \le \beta$. Then the arc length is

$$\begin{aligned} L &= \int_{\alpha}^{\beta} |\,\mathbf{r}'(t)\,|\,dt = \int_{\alpha}^{\beta} \sqrt{[f'(t)]^2 + [g'(t)]^2 + [h'(t)]^2}\,dt \\ &= \int_{\alpha}^{\beta} \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}\,dt. \end{aligned}$$

Arc Length Function

Suppose the curve has the vector equation $\mathbf{r}(t) = \langle f(t), g(t), h(t)\rangle = f(t)\,\mathbf{i} + g(t)\,\mathbf{j} + h(t)\,\mathbf{k}$, $\alpha \le t \le \beta$. We define its **arc length function** s by

$$s(t) = \int_{\alpha}^{t} |\,\mathbf{r}'(u)\,|\,du = \int_{\alpha}^{t} \sqrt{[f'(u)]^2 + [g'(u)]^2 + [h'(u)]^2}\,du.$$

Derivative of the Arc Length Function

$$\frac{ds}{dt} = |\,\mathbf{r}'(t)\,|\,.$$

Smooth Curve

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. A curve is called **smooth** if it has a smooth parametrization.

Curvature

If C is a smooth curve defined by the vector function $\mathbf{r}$, the **curvature** of the curve is

$$\kappa = \left|\frac{d\mathbf{T}}{ds}\right|,$$

where $\mathbf{T}$ is the **unit tangent vector** defined by

$$\mathbf{T}(t) = \frac{\mathbf{r}'(t)}{|\,\mathbf{r}'(t)\,|}.$$

Curvature Formula

$$\kappa(t) = \frac{|\,\mathbf{T}'(t)\,|}{|\,\mathbf{r}'(t)\,|} = \frac{|\,\mathbf{r}'(t) \times \mathbf{r}''(t)\,|}{|\,\mathbf{r}'(t)\,|^3}.$$

Curvature Formula for Plane Curve

For a plane curve $y = f(x)$, its curvature is

$$\kappa(x) = \frac{|f''(x)|}{[1 + (f'(x))^2]^{3/2}}.$$

Justification

Since $\mathbf{T} = \mathbf{r}'/\,|\,\mathbf{r}'(t)\,|$ and $|\,\mathbf{r}'(t)\,| = ds/dt$, we have

$$\mathbf{r}' = |\,\mathbf{r}'(t)\,|\,\mathbf{T} = \frac{ds}{dt}\,\mathbf{T}.$$

Applying the product rule gives

$$\mathbf{r}'' = \frac{d^2s}{dt^2}\,\mathbf{T} + \frac{ds}{dt}\,\mathbf{T}'$$

so that

$$\begin{aligned}
\mathbf{r}' \times \mathbf{r}'' &= \left(\frac{ds}{dt}\,\mathbf{T}\right) \times \left(\frac{d^2s}{dt^2}\,\mathbf{T} + \frac{ds}{dt}\,\mathbf{T}'\right) \\
&= \left(\frac{ds}{dt}\right)\left(\frac{d^2s}{dt^2}\right)(\mathbf{T} \times \mathbf{T}) + \left(\frac{ds}{dt}\right)^2 (\mathbf{T} \times \mathbf{T}') = \left(\frac{ds}{dt}\right)^2 (\mathbf{T} \times \mathbf{T}')\,.
\end{aligned}$$

Because $\mathbf{T}$ is a unit vector, $\mathbf{T} \cdot \mathbf{T} = |\,\mathbf{T}|^2 = 1$. Thus,

$$0 = \frac{d}{dt}\,(\mathbf{T} \cdot \mathbf{T}) = \mathbf{T}' \cdot \mathbf{T} + \mathbf{T} \cdot \mathbf{T}' = 2\,\mathbf{T} \cdot \mathbf{T}',$$

which means that $\mathbf{T} \perp \mathbf{T}'$. This implies that

$$|\,\mathbf{T} \times \mathbf{T}'\,| = |\,\mathbf{T}\,|\,|\,\mathbf{T}'\,|\sin 90^\circ = 1 \cdot |\,\mathbf{T}'\,| \cdot 1 = |\,\mathbf{T}'\,|\,.$$

Therefore,

$$|\,\mathbf{r}' \times \mathbf{r}''\,| = \left(\frac{ds}{dt}\right)^2 |\,\mathbf{T}'\,|\,.$$

Thus,

$$|\,\mathbf{T}'\,| = \frac{|\,\mathbf{r}' \times \mathbf{r}''\,|}{(ds/dt)^2} = \frac{|\,\mathbf{r}' \times \mathbf{r}''\,|}{|\,\mathbf{r}'\,|^2}$$

and

$$\kappa = \frac{|\,\mathbf{T}'\,|}{|\,\mathbf{r}'\,|} = \frac{|\,\mathbf{r}' \times \mathbf{r}''\,|}{|\,\mathbf{r}'\,|^3}.$$

For a plane curve with equation $y = f(x)$, we choose x as the parameter and write $\mathbf{r}(x) = x\,\mathbf{i} + f(x)\,\mathbf{j}$. Then $\mathbf{r}' = \mathbf{i} + f'(x)\,\mathbf{j}$ and $\mathbf{r}'' = f''(x)\,\mathbf{j}$. Since

$$\begin{aligned}\mathbf{r}' \times \mathbf{r}'' &= (\mathbf{i} + f'(x)\,\mathbf{j}) \times (f''(x)\,\mathbf{j}) \\ &= f''(x)\,(\mathbf{i} \times \mathbf{j}) + f'(x)f''(x)\,(\mathbf{j} \times \mathbf{j}) = f''(x)\,\mathbf{k},\end{aligned}$$

we have

$$\kappa = \frac{|\,f''(x)\,\mathbf{k}\,|}{|\,\mathbf{i} + f'(x)\,\mathbf{j}\,|^3} = \frac{|f''(x)|}{[1 + (f'(x))^2]^{3/2}}.$$

The Normal and Binormal Vectors

If C is a smooth curve defined by the vector function $\mathbf{r}$, we use the unit tangent vector $\mathbf{T}(t) = \dfrac{\mathbf{r}'(t)}{|\,\mathbf{r}'(t)\,|}$ to define

$$\begin{aligned}&\text{the } \textbf{unit normal vector}: \quad && \mathbf{N}(t) = \frac{\mathbf{T}'(t)}{|\,\mathbf{T}'(t)\,|} \\ &\text{the } \textbf{binormal vector}: && \mathbf{B}(t) = \mathbf{T}(t) \times \mathbf{N}(t).\end{aligned}$$

The Normal Plane and Osculating Plane

The plane determined by the normal and binormal vectors $\mathbf{N}$ and $\mathbf{B}$ at a point P on a curve C is called the **normal plane** of C at P.
The plane determined by the vectors $\mathbf{T}$ and $\mathbf{N}$ is called the **osculating plane** of C at P.

The Osculating Circle

The circle that lies in the osculating plane of C at P, has the same tangent as C at P, lies on the concave side of C (toward which $\mathbf{N}$ points), and has radius $\rho = 1/\kappa$ (the reciprocal of the curvature) is called the **osculating circle** (or the **circle of curvature**) of C at P.

Properties of the Normal and Binormal Vectors

(1) $|\,\mathbf{T}(t)\,| = |\,\mathbf{N}(t)\,| = |\,\mathbf{B}(t)\,| = 1$.
(2) $\mathbf{T}(t) \perp \mathbf{N}(t)$, $\mathbf{T}(t) \perp \mathbf{B}(t)$, and $\mathbf{N}(t) \perp \mathbf{B}(t)$.

Justification

(1) By the definition,

$$\mathbf{T}(t) = \frac{\mathbf{r}'(t)}{|\,\mathbf{r}'(t)\,|}, \qquad \mathbf{N}(t) = \frac{\mathbf{T}'(t)}{|\,\mathbf{T}'(t)\,|}$$

we know that $\mathbf{T}$ and $\mathbf{N}$ are unit vectors, so that $|\,\mathbf{T}(t)\,| = |\,\mathbf{N}(t)\,| = 1$.
Since $\mathbf{B}(t) = \mathbf{T}(t) \times \mathbf{N}(t)$, we have

$$|\,\mathbf{B}(t)\,| = |\,\mathbf{T}(t)\,|\,|\,\mathbf{N}(t)\,| \sin 90^\circ = 1.$$

(2) Since $\mathbf{T} \cdot \mathbf{T} = |\,\mathbf{T}\,|^2 = 1$, we have

$$0 = \frac{d}{dt}(\mathbf{T} \cdot \mathbf{T}) = \mathbf{T}' \cdot \mathbf{T} + \mathbf{T} \cdot \mathbf{T}' = 2\,\mathbf{T} \cdot \mathbf{T}',$$

which implies that $\mathbf{T} \perp \mathbf{T}'$. Because $\mathbf{N} = \dfrac{\mathbf{T}'}{|\,\mathbf{T}'\,|}$, we have $\mathbf{T} \perp \mathbf{N}$.
Since $\mathbf{B} = \mathbf{T} \times \mathbf{N}$, we know that $\mathbf{T} \perp \mathbf{B}$, and $\mathbf{N} \perp \mathbf{B}$.

Velocity Vector - Application

Suppose a particle moves through space so that its position vector at time t is $\mathbf{r}(t)$. Then the average velocity vector over the period from t to $t + h$ is

$$\frac{\mathbf{r}(t+h) - \mathbf{r}(t)}{h}$$

and the **velocity vector** $\mathbf{v}(t)$ at time t is

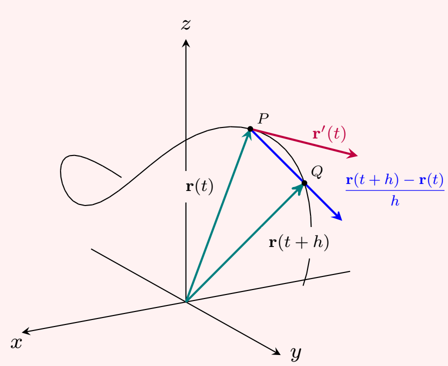
$$\mathbf{v}(t) = \mathbf{r}'(t) = \lim_{h \to 0} \frac{\mathbf{r}(t+h) - \mathbf{r}(t)}{h}.$$

The **acceleration** of the particle at time t is

$$\mathbf{a}(t) = \mathbf{v}'(t) = \mathbf{r}''(t).$$

The **speed** of the particle at time t is

$$v(t) = |\,\mathbf{v}(t)\,| = |\,\mathbf{r}'(t)\,|.$$



Tangential and Normal Components of Acceleration

Suppose a particle moves through space so that its position vector at time t is $\mathbf{r}(t)$. Then,

$$\mathbf{a} = a_T\mathbf{T} + a_N\mathbf{N},$$

where $\mathbf{T}$ is the unit tangent vector, $\mathbf{N}$ is the unit normal vector, and a_T and a_N for the tangential and normal components of acceleration. In terms of the speed of the particle and the curvature of the curve, we have

$$a_T = v', \qquad a_N = \kappa v^2,$$

or equivalently,

$$a_T = \frac{\mathbf{r}'(t)\cdot\mathbf{r}''(t)}{|\,\mathbf{r}'(t)\,|}, \qquad a_N = \frac{|\,\mathbf{r}'(t)\times\mathbf{r}''(t)\,|}{|\,\mathbf{r}'(t)\,|}.$$

Justification

By the definition,

$$\mathbf{T}(t) = \frac{\mathbf{r}'(t)}{|\,\mathbf{r}'(t)\,|} = \frac{\mathbf{v}(t)}{|\,\mathbf{v}(t)\,|} = \frac{\mathbf{v}}{v},$$

so that $\mathbf{v} = v\mathbf{T}$. Differentiating both sides of the equation with respect to t gives

$$\mathbf{a} = \mathbf{v}' = v'\mathbf{T} + v\mathbf{T}'.$$

Since

$$\kappa = \left|\frac{d\mathbf{T}}{ds}\right| = \left|\frac{d\mathbf{T}/dt}{ds/dt}\right| = \frac{|\,\mathbf{T}'\,|}{|\,\mathbf{r}'(t)\,|} = \frac{|\,\mathbf{T}'\,|}{v},$$

we have $|\,\mathbf{T}'\,| = \kappa v$. The unit normal vector was defined as $\mathbf{N} = \mathbf{T}'/\,|\,\mathbf{T}'\,|$, so that

$$\mathbf{T}' = |\,\mathbf{T}'\,|\,\mathbf{N} = \kappa v\mathbf{N}.$$

Hence, the acceleration $\mathbf{a}$ can be decomposed as

$$\mathbf{a} = v'\mathbf{T} + \kappa v^2\mathbf{N},$$

or

$$\mathbf{a} = a_T\mathbf{T} + a_N\mathbf{N},$$

where

$$a_T = v', \qquad a_N = \kappa v^2.$$

We can express the tangential and normal components of acceleration in terms of $\mathbf{r}$, $\mathbf{r}'$, and $\mathbf{r}''$ as follows. Since $\mathbf{v} = v\mathbf{T}$, we have

$$\begin{aligned}
\mathbf{v}\cdot\mathbf{a} &= v\mathbf{T}\cdot(v'\mathbf{T} + \kappa v^2\mathbf{N}) \\
&= vv'\mathbf{T}\cdot\mathbf{T} + \kappa v^3\mathbf{T}\cdot\mathbf{N} \\
&= vv' \qquad (\text{since } |\,\mathbf{T}(t)\,| = 1 \text{ and } \mathbf{T}\perp\mathbf{N}),
\end{aligned}$$

so that

$$a_T = v' = \frac{\mathbf{r}'(t) \cdot \mathbf{r}''(t)}{|\,\mathbf{r}'\,|}.$$

Since $\kappa(t) = \dfrac{|\,\mathbf{r}'(t) \times \mathbf{r}''(t)\,|}{|\,\mathbf{r}'(t)\,|^3}$, we have

$$a_N = \kappa v^2 = \frac{|\,\mathbf{r}'(t) \times \mathbf{r}''(t)\,|}{|\,\mathbf{r}'(t)\,|^3}\,|\,\mathbf{r}'(t)\,|^2 = \frac{|\,\mathbf{r}'(t) \times \mathbf{r}''(t)\,|}{|\,\mathbf{r}'(t)\,|}.$$

Position Vector of A projectile - Application

A projectile is fired with angle of elevation α and initial velocity $\mathbf{v}_0$. If we set up the axes so that the projectile starts at the origin and assume that air resistance is negligible and the only external force is due to gravity, then, the position vector of the projectile at time t is

$$\mathbf{r}(t) = (v_0 \cos\alpha)t\,\mathbf{i} + \left[(v_0 \sin\alpha)t - \tfrac{1}{2}gt^2\right]\mathbf{j},$$

where g is the gravitational constant $g \approx 9.8$ m/s^2. The parametric equations of the trajectory are

$$x = (v_0 \cos\alpha)t, \qquad y = (v_0 \sin\alpha)t - \tfrac{1}{2}gt^2.$$

The horizontal distance traveled is

$$d = \frac{v_0^2 \sin 2\alpha}{g},$$

which attains the maximum value when $\alpha = \pi/4$.

Justification

Since the force due to gravity acts downward, we have

$$\mathbf{F} = m\mathbf{a} = -mg\,\mathbf{j},$$

where $g = |\,\mathbf{a}\,| \approx 9.8$ m/s^2. Thus, $\mathbf{a} = -g\,\mathbf{j}$. Integrating this equation gives

$$\mathbf{v}(t) = -gt\,\mathbf{j} + \mathbf{C},$$

where $\mathbf{C} = \mathbf{v}(0) = \mathbf{v}_0$. Therefore

$$\mathbf{r}'(t) = \mathbf{v}(t) = -gt\,\mathbf{j} + \mathbf{v}_0.$$

Integrating again gives

$$\mathbf{r}(t) = -\tfrac{1}{2}gt^2\,\mathbf{j} + t\,\mathbf{v}_0 + \mathbf{D},$$

where $\mathbf{D} = \mathbf{r}(0) = \mathbf{0}$. Hence the position vector of the projectile is given by

$$\mathbf{r}(t) = -\tfrac{1}{2}gt^2\,\mathbf{j} + t\,\mathbf{v}_0.$$

If we write the initial speed of the projectile $|\,\mathbf{v}_0\,| = v_0$, then

$$\mathbf{v}_0 = v_0 \cos\alpha\,\mathbf{i} + v_0 \sin\alpha\,\mathbf{j}.$$

where α is the angle of elevation when the projectile is fired. Thus, the position vector of the projectile at time t is

$$\mathbf{r}(t) = (v_0\cos\alpha)t\,\mathbf{i} + \left[(v_0\sin\alpha)t - \tfrac{1}{2}gt^2\right]\mathbf{j}.$$

The parametric equations of the trajectory are therefore

$$x = (v_0\cos\alpha)t, \qquad y = (v_0\sin\alpha)t - \tfrac{1}{2}gt^2.$$

The maximal horizontal distance traveled is the value of x when $y = 0$. Setting $y = 0$, we obtain $t = 0$ or $t = (2v_0\sin\alpha)/g$. This second value of t then gives

$$d = x = (v_0\cos\alpha)\frac{2v_0\sin\alpha}{g} = \frac{v_0^2(2\sin\alpha\cos\alpha)}{g} = \frac{v_0^2\sin 2\alpha}{g}.$$

Clearly, d has its maximum value when $\sin 2\alpha = 1$, that is, $\alpha = \pi/4$.

Kepler's Laws - Application

1. A planet revolves around the sun in an elliptical orbit with the sun at one focus.
2. The line joining the sun to a planet sweeps out equal areas in equal times.
3. The square of the period of revolution of a planet is proportional to the cube of the length of the major axis of its orbit.

Justification - Kepler's First Law (Part 1)

We use a coordinate system with the sun at the origin and we let $\mathbf{r} = \mathbf{r}(t)$ be the position vector of the planet.
We first show that the planet moves in one plane. We use

Newton's Second Law of Motion:	$\mathbf{F} = m\mathbf{a}$,
Newton's Law of Universal Gravitation:	$\mathbf{F} = -\dfrac{GMm}{r^2}\dfrac{\mathbf{r}}{r} = -\dfrac{GMm}{r^2}\mathbf{u}$,

where $\mathbf{F}$ is the gravitational force on the planet, m and M are the masses of the planet and the sun, G is the gravitational constant, $r = |\,\mathbf{r}\,|$, and $\mathbf{u} = \mathbf{r}/r$ is the unit vector in the direction of $\mathbf{r}$. These equations gives

$$\mathbf{a} = -\frac{GM}{r^2}\mathbf{u},$$

so $\mathbf{a}$ is parallel to $\mathbf{r}$. Thus,

$$\begin{aligned}\frac{d}{dt}(\mathbf{r}\times\mathbf{v}) &= \mathbf{r}'\times\mathbf{v}+\mathbf{r}\times\mathbf{v}' = \mathbf{v}\times\mathbf{v}+\mathbf{r}\times\mathbf{a}\\ &= \mathbf{0}+\mathbf{r}\times\left(-\frac{GM}{r^3}\mathbf{r}\right) = \mathbf{0}.\end{aligned}$$

Hence, $\mathbf{r}\times\mathbf{v}=\mathbf{h}$, where $\mathbf{h}$ is a constant vector. This implies that $\mathbf{r}=\mathbf{r}(t)$ is perpendicular to a constant vector for all t. Therefore, the planet always lies in the plane through the origin perpendicular to a constant vecotor. Thus the orbit of the planet is a plane curve.

Justification - Kepler's First Law (Part 2)

To prove Kepler's First Law we rewrite the vector $\mathbf{h}$ as follows:

$$\begin{aligned}\mathbf{h} &= \mathbf{r}\times\mathbf{v} = \mathbf{r}\times\mathbf{r}' = r\mathbf{u}\times(r\mathbf{u})'\\ &= r\mathbf{u}\times(r\mathbf{u}'+r'\mathbf{u}) = r^2(\mathbf{u}\times\mathbf{u}')+rr'(\mathbf{u}\times\mathbf{u}) = r^2(\mathbf{u}\times\mathbf{u}').\end{aligned}$$

Thus,

$$\begin{aligned}\mathbf{a}\times\mathbf{h} &= \left(-\frac{GM}{r^2}\mathbf{u}\right)\times\left[r^2(\mathbf{u}\times\mathbf{u}')\right] = (-GM)\mathbf{u}\times(\mathbf{u}\times\mathbf{u}')\\ &= (-GM)\left[(\mathbf{u}\cdot\mathbf{u}')\mathbf{u}-(\mathbf{u}\cdot\mathbf{u})\mathbf{u}'\right] \qquad \text{(by a property of the cross product)}\end{aligned}$$

Since $|\,\mathbf{u}(t)\,|=1$, we have $\mathbf{u}\cdot\mathbf{u}=|\,\mathbf{u}(t)\,|^2=1$ so that

$$0=\frac{d}{dt}(\mathbf{u}\cdot\mathbf{u})=\mathbf{u}\cdot\mathbf{u}'+\mathbf{u}'\cdot\mathbf{u}=2\mathbf{u}\cdot\mathbf{u}'.$$

Hence

$$\mathbf{a}\times\mathbf{h}=(GM)\mathbf{u}'.$$

Therefore,

$$(\mathbf{v}\times\mathbf{h})'=\mathbf{v}'\times\mathbf{h}=\mathbf{a}\times\mathbf{h}=(GM)\mathbf{u}'.$$

Integrating both sides of this equation gives

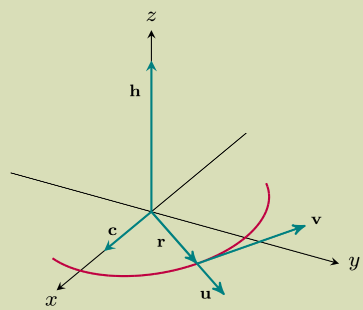
$$\mathbf{v}\times\mathbf{h}=(GM)\mathbf{u}+\mathbf{c},$$

where $\mathbf{c}$ is a constant vector.

Justification - Kepler's First Law (Part 3)

We choose the coordinate axes so that the standard basis vector $\mathbf{k}$ points in the direction of the vector $\mathbf{h}$. Then the planet moves in the xy-plane. Since both $\mathbf{v}\times\mathbf{h}$ and $\mathbf{u}$ are perpendicular to $\mathbf{h}$, we know that that $\mathbf{c}$ lies in the xy-plane. This means that we can choose the x- and y-axes so that the vector $\mathbf{i}$

lies in the direction of $\mathbf{c}$, as shown in the following figure.



Let θ be the angle between $\mathbf{c}$ and $\mathbf{r}$. Then (r, θ) are polar coordinates of the planet. From Part (2), we have

$$\begin{aligned}\mathbf{r}\cdot(\mathbf{v}\times\mathbf{h}) &= \mathbf{r}\cdot[(GM)\mathbf{u}+\mathbf{c}] = (GM)(r\mathbf{u}\cdot\mathbf{u}) + \mathbf{r}\cdot\mathbf{c}\\ &= GMr + |\,\mathbf{r}\,|\,|\,\mathbf{c}\,|\cos\theta = GMr + rc\cos\theta,\end{aligned}$$

where $c = |\,\mathbf{c}\,|$. On the other hand, the scalar triple product can be expressed as

$$\mathbf{r}\cdot(\mathbf{v}\times\mathbf{h}) = (\mathbf{r}\times\mathbf{v})\cdot\mathbf{h} = \mathbf{h}\cdot\mathbf{h} = |\,\mathbf{h}\,|^2 = h^2,$$

where $h = |\,\mathbf{h}\,|$. Thus, by solving from the equation $GMr + rc\cos\theta = h^2$, we get

$$r = \frac{h^2}{GM + c\cos\theta}?$$

or

$$r = \frac{ed}{1 + e\cos\theta},$$

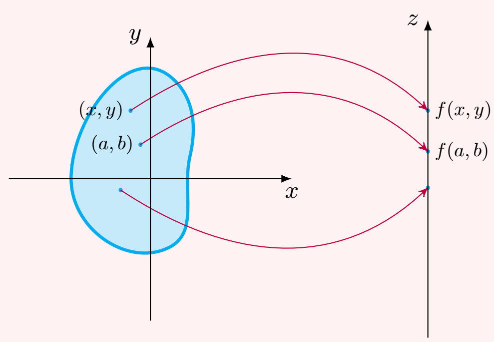
where $e = c/(GM)$ and $d = h^2/c$. This is the polar equation of a conic section with focus at the origin and eccentricity e. We know that the orbit of a planet is a closed curve and so the conic must be an ellipse.

Chapter 14

Partial Derivatives

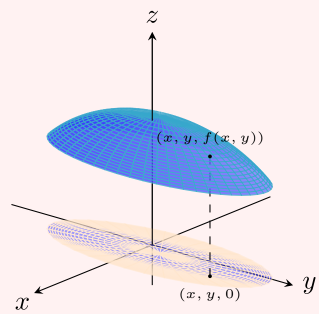
Function of Two Variables

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, $\{f(x,y) \,|\, (x,y) \in D\}$.
We often write $z = f(x, y)$ to make explicit the value taken on by f at the general point (x, y). The variables x and y are **independent variables** and z is the **dependent variable**.



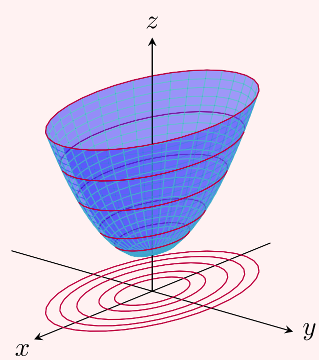
Graph of a Function of Two Variables

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.



Level Curves and Level Surfaces

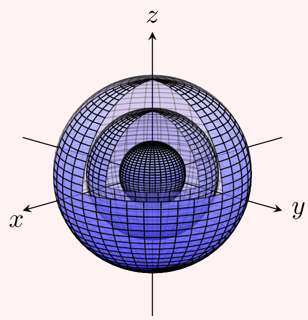
The **level curves**, or **contour lines**, of a function f of two variables are the curves with equations $f(x, y) = k$, where k is a constant (in the range of f).



The **level surfaces** of a function f of three variables are the surfaces with equations $f(x, y, z) = k$, where k is a constant (in the range of f).

Function of Three or More Variables

A **function of three variables**, f, is a rule that assigns to each ordered triple (x, y, z) in a domain $D \subset \mathbb{R}^3$ a unique real number denoted by $f(x, y, z)$.
It's very difficult to visualize a function f of three variables by its graph, since that would lie in a four-dimensional space. However, we do gain some insight into f by examining its level surfaces, which are the surfaces with equations $f(x, y, z) = k$, where k is a constant. If the point (x, y, z) moves along a level surface, the value of $f(x, y, z)$ remains fixed.



Level surfaces of the function
$f(x, y, z) = x^2 + y^2 + z^2$

Limit of Function of Two Variables

Let f be a function of two variables whose domain D includes points arbitrarily close to (a, b). Then we say that the **limit of** $f(x, y)$ **as** (x, y) **approaches** (a, b) is L and we write

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

if for every number $\varepsilon > 0$ there is a corresponding number $\delta > 0$ such that

$$\text{if } (x, y) \in D \text{ and } 0 \leq \sqrt{(x-a)^2 + (y-b)^2} < \delta \text{ then } |f(x, y) - L| < \varepsilon$$

Limits Along Different Paths

If $f(x, y) \to L_1$ as $(x, y) \to (a, b)$ along a path C_1 and $f(x, y) \to L_2$ as $(x, y) \to (a, b)$ along a path C_2, where $L_1 \neq L_2$, then $\lim_{(x,y)\to(a,b)} f(x, y)$ does not exist.

Continuous Function of Two Variables

A function f of two variables is called **continuous** at (a, b) if

$$\lim_{(x,y)\to(a,b)} f(x, y) = f(a, b).$$

We say f is **continuous on** D if f is continuous at every point (a, b) in D.

Continuous Function of n Variables

If f is defined on a subset D of $\mathbb{R}^n$, then $\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = L$ means that for every number $\varepsilon > 0$ there is a corresponding number $\delta > 0$ such that

$$\text{if } \mathbf{x} \in D \text{ and } 0 \leq |\mathbf{x} - \mathbf{a}| < \delta \text{ then } |f(\mathbf{x}) - L| < \varepsilon$$

A function f is **continuous** at $\mathbf{a}$ if $\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = f(\mathbf{a})$.

Properties of Continuous Functions

Arithmetic Operations of Continuous Functions

If f and g are continuous at $\mathbf{x}_0$, where $\mathbf{a} = (a_1, a_2)$ in 2D or $\mathbf{a} = (a_1, a_2, a_3)$ in 3D, and c is a constant, then the following functions are also continuous at a:

1. $f + g$ **2.** $f - g$ **3.** cf **4.** fg **5.** $\dfrac{f}{g}$ if $g(\mathbf{a}) \neq 0$

Continuity of Elementary Functions

The following types of functions are continuous at every number in their domains:

polynomials rational functions root functions
trigonometric functions inverse trigonometric functions
exponential functions logarithmic functions

Limit of Composite Function

If f is continuous at $\mathbf{b}$ and $\lim\limits_{\mathbf{x}\to\mathbf{a}} g(\mathbf{x}) = \mathbf{b}$, then $\lim\limits_{\mathbf{x}\to\mathbf{a}} f(g(\mathbf{x})) = f(\mathbf{b})$. In other words,

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

Continuity of Composite Function

If g is continuous at $\mathbf{a}$ and f is continuous at $g(\mathbf{a})$, then the composite function $f \circ g$ is continuous at $\mathbf{a}$.

Partial Derivatives at a Point

If f is a function of two variables, the **partial derivative of f with respect to x at** (a, b), denoted $f_x(a, b)$, is

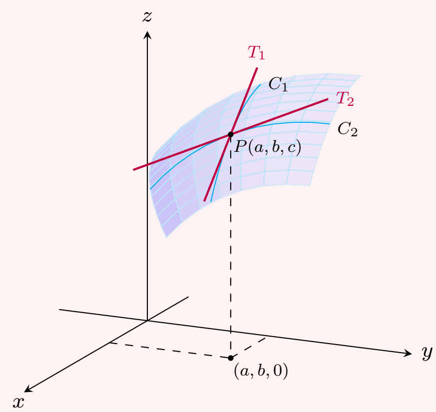
$$f_x(a,b) = \lim_{h\to 0} \frac{f(a+h,b) - f(a,b)}{h}.$$

Similarly, the **partial derivative of f with respect to y at** (a, b), denoted $f_y(a, b)$, is

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

Geometrically, the equation $z = f(x, y)$ represents a surface S. If $f(a, b) = c$, then the point $P(a, b, c)$

lies on S. The trace C_1 of S in the plane $y = b$ is the graph of the function $g(x) = f(x, b)$, so the slope of its tangent T_1 at P is $g'(a) = f_x(a, b)$. The trace C_2 of S in the plane $x = a$ is the graph of the function $h(y) = f(a, y)$, so the slope of its tangent T_2 at P is $h'(b) = f_y(a, b)$.



Partial derivatives can also be interpreted as *rates of change*. If $z = f(x, y)$, then $f_x(a, b)$ represents the rate of change of z with respect to x when y is fixed. Similarly, $f_y(a, b)$ represents the rate of change of z with respect to y when x is fixed.

Partial Derivatives as Functions

If f is a function of two variables, its **partial derivatives** are the functions f_x and f_y defined by

$$f_x(x, y) = \lim_{h \to 0} \frac{f(x + h, y) - f(x, y)}{h}$$
$$f_y(x, y) = \lim_{h \to 0} \frac{f(x, y + h) - f(x, y)}{h}$$

Notations for Partial Derivatives

If $z = f(x, y)$, we write

$$f_x(x, y) = f_x = \frac{\partial f}{\partial x} = \frac{\partial}{\partial x} f(x, y) = \frac{\partial z}{\partial x} = f_1 = D_1 f = D_x f$$
$$f_y(x, y) = f_y = \frac{\partial f}{\partial y} = \frac{\partial}{\partial y} f(x, y) = \frac{\partial z}{\partial y} = f_2 = D_2 f = D_y f$$

Rule for Finding Partial Derivatives of $z = f(x, y)$

1. To find f_x, regard y as a constant and differentiate $f(x, y)$ with respect to x.
2. To find f_y, regard x as a constant and differentiate $f(x, y)$ with respect to y.

Higher Partial Derivatives

If $z = f(x, y)$, the **second partial derivatives** are

$$(f_x)_x = f_{xx} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial x^2} = \frac{\partial^2 z}{\partial x^2} = f_{11} = D_{11}f = D_{xx}f$$
$$(f_x)_y = f_{xy} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial y\,\partial x} = \frac{\partial^2 z}{\partial y\,\partial x} = f_{12} = D_{12}f = D_{xy}f$$
$$(f_y)_x = f_{yx} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial x\,\partial y} = \frac{\partial^2 z}{\partial x\,\partial y} = f_{21} = D_{21}f = D_{yx}f$$
$$(f_y)_y = f_{yy} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial y^2} = \frac{\partial^2 z}{\partial y^2} = f_{22} = D_{22}f = D_{yy}f$$

Partial derivatives of order 3 or higher can also be defined. For instance,

$$f_{xyy} = (f_{xy})_y = \frac{\partial}{\partial y}\left(\frac{\partial^2 f}{\partial y\,\partial x}\right) = \frac{\partial^3 f}{\partial y^2\,\partial x}.$$

Clairaut's Theorem

Suppose f is defined on a disk D that contains the point (a, b). If the functions f_{xy} and f_{yx} are both continuous on D, then

$$f_{xy}(a, b) = f_{yx}(a, b).$$

Justification

For small values of h, $h \neq 0$, consider the difference

$$\Delta(h) == \left[f(a+h, b+h) - f(a+h, b)\right] - \left[f(a, b+h) - f(a, b)\right].$$

If we denote $g(x) = f(x, b+h) - f(x, b)$, then

$$\Delta(h) = g(a+h) - g(a).$$

By the Mean Value Theorem, there is a number c between a and $a+h$ such that

$$g(a+h) - g(a) = g'(c)h.$$

Since $g'(c) = f_x(c, b+h) - f_x(c, b)$, we get

$$\Delta(h) = h\left[f_x(c, b+h) - f_x(c, b)\right].$$

Applying the Mean Value Theorem again, this time to f_x, we get a number d between b and $b+h$ such that

$$f_x(c, b+h) - f_x(c, b) = f_{xy}(c, d)h.$$

Thus, we obtain

$$\Delta(h) = h^2 f_{xy}(c, d).$$

If $h \to 0$, then $(c, d) \to (a, b)$, so the continuity of f_{xy} at (a, b) gives

$$\lim_{h\to 0} \frac{\Delta(h)}{h^2} = \lim_{h\to 0} f_{xy}(c, d) = f_{xy}(a, b).$$

Similarly, by writing

$$\Delta(h) = [f(a+h, b+h) - f(a, b+h)] - [f(a+h, b) - f(a, b)]$$

and using the Mean Value Theorem twice and the continuity of f_{yx} at (a, b), we obtain

$$\lim_{h\to 0} \frac{\Delta(h)}{h^2} = f_{yx}(a, b).$$

Therefore, $f_{xy}(a, b) = f_{yx}(a, b)$.

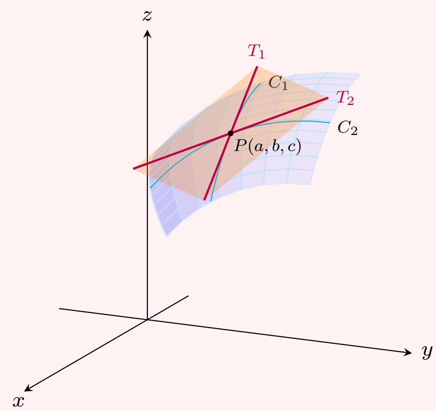
Marginal Productivity - Application

If the production function is denoted by $P = P(L, K)$, where P is the total production, L is the amount of labor, and K is the amount of capital invested, then the partial derivative $\partial P/\partial L$ is called the marginal production with respect to labor or the **marginal productivity of labor**. Likewise, the partial derivative $\partial P/\partial K$ is called the **marginal productivity of capital**.

Tangent Plane

Suppose f has continuous partial derivatives. An equation of the tangent plane to the surface $z = f(x, y)$ at the point $P(a, b, c)$ is

$$z - c = f_x(a, b)(x - a) + f_y(a, b)(y - b).$$



Linear Approximation

Suppose f has continuous partial derivatives. Then

$$L(x, y) = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b)$$

is called the **linearization** of f at (a, b) and the approximation

$$f(x, y) \approx f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b)$$

is called the **linear approximation** or the **tangent plane approximation** of f at (a, b).

Differentiable of a Function of Two Variables

If $z = f(x, y)$, the **increment** of z from (a, b) to $(a + \Delta x, b + \Delta y)$ is

$$\Delta z = f(a + \Delta x, b + \Delta y) - f(a, b).$$

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

$$\Delta z = f_x(a, b)\Delta x + f_y(a, b)\Delta y + \varepsilon_1 \Delta x + \varepsilon_2 \Delta y,$$

where ε_1 and $\varepsilon_2 \to 0$ as $(\Delta x, \Delta y) \to (0, 0)$.

Sufficient Condition for Being Differentiable

If the partial derivatives f_x and f_y exist near (a, b) and are continuous at (a, b), then f is differentiable at (a, b).

Total Differential of a Function of Two Variables

If $z = f(x, y)$, then the **total differential** at (a, b) is defined by

$$dz = f_x(a, b)\, dx + f_y(a, b)\, dy = \frac{\partial z}{\partial x}\, dx + \frac{\partial z}{\partial y}\, dy,$$

where the differentials dx and dy are independent variables.

Linear Approximation and Total Differential of a Function of Three Variables

If $w = f(x, y, z)$, then the **linearization** of f at (a, b, c) is

$$L(x, y, z) = f(a, b, c) + f_x(a, b, c)(x - a) + f_y(a, b, c)(y - b) + f_z(a, b, c)(z - c).$$

The **linear approximation** of f at (a, b, c) is

$$f(x, y, z) \approx f(a, b, c) + f_x(a, b, c)(x - a) + f_y(a, b, c)(y - b) + f_z(a, b, c)(z - c).$$

The **increment** of w from (a, b, c) to $(a + \Delta x, b + \Delta y, c + \Delta z)$ is

$$\Delta w = f(a + \Delta x, b + \Delta y, c + \Delta z) - f(a, b, c).$$

The **total differential** at (a, b, c) is

$$dw = f_x(a, b, c)\, dx + f_y(a, b, c)\, dy + f_z(a, b, c)\, dz = \frac{\partial w}{\partial x}\, dx + \frac{\partial w}{\partial y}\, dy + \frac{\partial w}{\partial z}\, dz.$$

The Chain Rule

Case 1: Suppose that $z = f(x, y)$ is a differentiable function of x and y, where $x = g(t)$ and $y = h(t)$ are both differentiable functions of t. Then z is a differentiable function of t and

$$\begin{aligned} \frac{dz}{dt} &= \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} \\ &= \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt} \end{aligned}$$

Case 2: Suppose that $z = f(x, y)$ is a differentiable function of x and y, where $x = g(s, t)$ and $y = h(s, t)$ are both differentiable functions of s and t. Then

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s}, \qquad \frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t}.$$

General Version of the Chain Rule

Suppose that u is a differentiable function of the n variables $x_1, x_2, \ldots, x_n$ and each x_j is a differentiable function of the m variables $t_1, t_2, \ldots, t_m$. Then u is a function of $t_1, t_2, \ldots, t_m$ 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 $i = 1, 2, \ldots, m$.

Justification - Case 1

A change of Δt in t produces changes of Δx in x and Δy in y. These, in turn, produce a change of Δz in z. Since $z = f(x, y)$ is differentiable, we have

$$\Delta z = \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y + \varepsilon_1 \Delta x + \varepsilon_2 \Delta y,$$

where ε_1 and $\varepsilon_2 \to 0$ as $(\Delta x, \Delta y) \to (0, 0)$. It gives

$$\frac{\Delta z}{\Delta t} = \frac{\partial f}{\partial x}\frac{\Delta x}{\Delta t} + \frac{\partial f}{\partial y}\frac{\Delta y}{\Delta t} + \varepsilon_1 \frac{\Delta x}{\Delta t} + \varepsilon_2 \frac{\Delta y}{\Delta t}.$$

If we now let $\Delta t \to 0$, then $\Delta x = g(t + \Delta t) - g(t) \to 0$ because g is differentiable and therefore continuous. Similarly, $\Delta y \to 0$. Thus, if $\Delta t \to 0$, then ε_1 and $\varepsilon_2 \to 0$. Hence,

$$\begin{aligned}
\frac{dz}{dt} &= \lim_{\Delta t \to 0} \frac{\Delta z}{\Delta t} \\
&= \frac{\partial f}{\partial x} \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} + \frac{\partial f}{\partial y} \lim_{\Delta t \to 0} \frac{\Delta y}{\Delta t} + \left(\lim_{\Delta t \to 0} \varepsilon_1\right) \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} + \left(\lim_{\Delta t \to 0} \varepsilon_2\right) \lim_{\Delta t \to 0} \frac{\Delta y}{\Delta t} \\
&= \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + 0 \cdot \frac{dx}{dt} + 0 \cdot \frac{dy}{dt} \\
&= \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}.
\end{aligned}$$

Since we often write $\partial z/\partial x$ and $\partial z/\partial y$ in place of $\partial f/\partial x$ and $\partial f/\partial y$, we can rewrite the Chain Rule in the form

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}.$$

Justification - Case 2

When $z = f(x, y)$ but each of x and y is a function of two variables s and t: $x = g(s, t)$, $y = h(s, t)$. Then z is indirectly a function of s and t and we wish to find $\partial z/\partial s$ and $\partial z/\partial t$. Recall that in computing $\partial z/\partial t$ we hold s fixed and compute the ordinary derivative of z with respect to t. Therefore we can apply the result in Case 1 to obtain

$$\frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t}.$$

Similarly, we have

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s}.$$

Implicit Differentiation

Implicit Differentiation for Function of Single Variable

Suppose that an equation of the form $F(x, y) = 0$ defines y implicitly as a differentiable function of x. If $\partial F/\partial y \neq 0$, then

$$\frac{dy}{dx} = -\frac{\dfrac{\partial F}{\partial x}}{\dfrac{\partial F}{\partial y}} = -\frac{F_x}{F_y}.$$

Implicit Differentiation for Function of Two Variables

Suppose that an equation of the form $F(x, y, z) = 0$ defines z implicitly as a differentiable function of x and y. If $\partial F/\partial z \neq 0$, then

$$\frac{\partial z}{\partial x} = -\frac{\dfrac{\partial F}{\partial x}}{\dfrac{\partial F}{\partial z}} = -\frac{F_x}{F_z}, \qquad \frac{\partial z}{\partial y} = -\frac{\dfrac{\partial F}{\partial y}}{\dfrac{\partial F}{\partial z}} = -\frac{F_y}{F_z}.$$

Justification - for Function of Single Variable

The equation $F(x, y) = 0$ defines a differentiable function $y = f(x)$. Since F is differentiable, we apply the Chain Rule to differentiate both sides of the equation $F(x, y) = 0$ with respect to x and obtain

$$\frac{\partial F}{\partial x}\frac{dx}{dx} + \frac{\partial F}{\partial y}\frac{dy}{dx} = 0.$$

But $dx/dx = 1$, so if $\partial F/\partial y \neq 0$ we solve for dy/dx and obtain

$$\frac{dy}{dx} = -\frac{\dfrac{\partial F}{\partial x}}{\dfrac{\partial F}{\partial y}} = -\frac{F_x}{F_y}.$$

Justification - for Function of Two Variables

The equation $F(x, y, z) = 0$ defines a differentiable function $z = f(x, y)$. Since F is differentiable, we apply the Chain Rule to differentiate both sides of the equation $F(x, y, z) = 0$ with respect to x and obtain

$$\frac{\partial F}{\partial x}\frac{\partial x}{\partial x} + \frac{\partial F}{\partial y}\frac{\partial y}{\partial x} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial x} = 0.$$

But

$$\frac{\partial}{\partial x}(x) = 1, \qquad \frac{\partial}{\partial x}(y) = 0,$$

so this equation becomes

$$\frac{\partial F}{\partial x} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial x} = 0.$$

If $\partial F/\partial z \neq 0$ we solve for $\dfrac{\partial z}{\partial x}$ and obtain

$$\frac{\partial z}{\partial x} = -\frac{\dfrac{\partial F}{\partial x}}{\dfrac{\partial F}{\partial z}} = -\frac{F_x}{F_z}.$$

Similarly, we can get

$$\frac{\partial z}{\partial y} = -\frac{\dfrac{\partial F}{\partial y}}{\dfrac{\partial F}{\partial z}} = -\frac{F_y}{F_z}.$$

Rate of Change in the Direction of $\mathbf{u}$

Suppose $z = f(x, y)$ is a differentiable function. Then the rate of change of z at (x_0, y_0) in the direction of an arbitrary unit vector $\mathbf{u} = \langle a, b\rangle$ is

$$D_{\mathbf{u}}f(x_0, y_0) = \lim_{h\to 0}\frac{f(x_0 + ha, y_0 + hb) - f(x_0, y_0)}{h}.$$

Directional Derivative

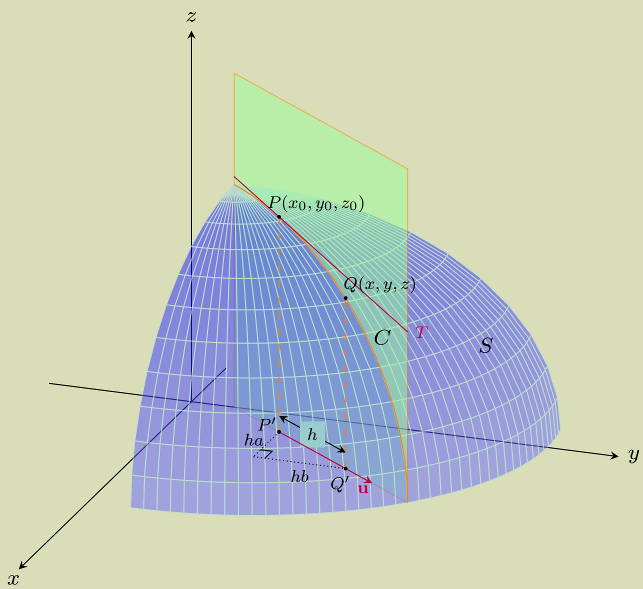
The limit is also called the **directional derivative** of f at (x_0, y_0) in the direction of $\mathbf{u}$.

In particular,

$$D_{\mathbf{i}}f = f_x, \qquad D_{\mathbf{j}}f = f_y.$$

Justification

To find the rate of change of z at (x_0, y_0) in the direction of an arbitrary unit vector $\mathbf{u} = \langle a, b\rangle$, we consider the surface S with the equation $z = f(x, y)$ and let $z_0 = f(x_0, y_0)$. Then the point $P(x_0, y_0, z_0)$ lies on 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 of $\mathbf{u}$.



Suppose $Q(x, y, z)$ is another point on C. Let P', Q' be the projections of P, Q onto the xy-plane. Then the vector $\overrightarrow{P'Q'}$ is parallel to $\mathbf{u}$, so

$$\overrightarrow{P'Q'} = \langle x - x_0, y - y_0\rangle = h\mathbf{u} = \langle ha, hb\rangle$$

for some scalar h. Thus,

$$\frac{\Delta z}{h} = \frac{z - z_0}{h} = \frac{f(x, y) - f(x_0, y_0)}{h} = \frac{f(x_0 + ha, y_0 + hb) - f(x_0, y_0)}{h}.$$

Therefore, by taking the limit as $h \to 0$, we have the rate of change of z (with respect to distance) in the direction of $\mathbf{u}$.
If $\mathbf{u} = \mathbf{i} = \langle 1, 0\rangle$, then

$$D_{\mathbf{i}} f(x_0, y_0) = \lim_{h\to 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h} = f_x(x_0, y_0).$$

Similarly, $D_{\mathbf{j}} f(x_0, y_0) = f_y(x_0, y_0)$.

Directional Derivative and Partial Derivatives

If f is a differentiable function of x and y, then f has a directional derivative in the direction of any unit vector $\mathbf{u} = \langle a, b\rangle$ and

$$D_{\mathbf{u}} f(x, y) = f_x(x, y)\, a + f_y(x, y)\, b.$$

Justification

We define a function g of the single variable h by

$$g(h) = f(x_0 + ha, y_0 + hb).$$

By the definition of a derivative,

$$g'(0) = \lim_{h\to 0} \frac{g(h) - g(0)}{h} = \lim_{h\to 0} \frac{f(x_0 + ha, y_0 + hb) - f(x_0, y_0)}{h} = D_{\mathbf{u}} f(x_0, y_0).$$

On the other hand, for $g(h) = f(x_0 + ha, y_0 + hb)$, by applying the Chain Rule,

$$g'(h) = f_x(x_0 + ha, y_0 + hb) \cdot a + f_y(x_0 + ha, y_0 + hb) \cdot b.$$

Putting $h = 0$ gives

$$g'(0) = f_x(x_0, y_0) \cdot a + f_y(x_0, y_0) \cdot b.$$

Therefore,

$$D_{\mathbf{u}} f(x_0, y_0) = f_x(x_0, y_0) \cdot a + f_y(x_0, y_0) \cdot b.$$

2D Gradient

If f is a function of two variables x and y, then the **gradient** of f is the vector function ∇f defined by

$$\nabla f(x, y) = \langle f_x(x, y), f_y(x, y)\rangle = \frac{\partial f}{\partial x}\,\mathbf{i} + \frac{\partial f}{\partial y}\,\mathbf{j}.$$

Gradient and Directional Derivative

$$D_{\mathbf{u}} f(x, y) = \nabla f(x, y) \cdot \mathbf{u}.$$

Directional Derivative and Gradient for Function of Three Variables

Directional Derivative

The **directional derivative** of f at (x_0, y_0, z_0) in the direction of a unit vector $\mathbf{u} = \langle a, b, c\rangle$ is

$$D_{\mathbf{u}} f(x_0, y_0, z_0) = \lim_{h\to 0} \frac{f(x_0 + ha, y_0 + hb, z_0 + hc) - f(x_0, y_0, z_0)}{h},$$

provided that the limit exists. Or, in vector form,

$$D_{\mathbf{u}} f(\mathbf{x}_0) = \lim_{h\to 0} \frac{f(\mathbf{x}_0 + h\mathbf{u}) - f(\mathbf{x}_0)}{h}.$$

Directional Derivative and Partial Derivatives

If f is a differentiable function of x, y, and z, then f has a directional derivative in the direction of any unit vector $\mathbf{u} = \langle a, b, c\rangle$ and

$$D_{\mathbf{u}} f(x, y, z) = f_x(x, y, z)\, a + f_y(x, y, z)\, b + f_y(x, y, z)\, c.$$

3D Gradient

If f is a function of three variables x, y, and z, then the **gradient** of f is the vector function ∇f defined by

$$\nabla f(x, y, z) = \langle f_x(x, y, z), f_y(x, y, z), f_z(x, y, z)\rangle = \frac{\partial f}{\partial x}\,\mathbf{i} + \frac{\partial f}{\partial y}\,\mathbf{j} + \frac{\partial f}{\partial z}\,\mathbf{k}.$$

Gradient and Directional Derivative

$$D_{\mathbf{u}} f(x, y, z) = \nabla f(x, y, z) \cdot \mathbf{u}.$$

Maximizing the Directional Derivative

Suppose f is a differentiable function of two or three variables. The maximum value of the directional derivative $D_{\mathbf{u}} f(\mathbf{x})$ is $|\nabla f(\mathbf{x})|$ and it occurs when $\mathbf{u}$ has the same direction as the gradient vector $\nabla f(\mathbf{x})$.

Justification

Since $|\,\mathbf{u}\,| = 1$,

$$D_{\mathbf{u}} f(\mathbf{x}) = \nabla f(\mathbf{x}) \cdot \mathbf{u} = |\nabla f(\mathbf{x})|\,|\,\mathbf{u}\,|\cos\theta = |\nabla f(\mathbf{x})| \cos\theta,$$

where θ is the angle between $\nabla f(\mathbf{x})$ and $\mathbf{u}$. The maximum value of $\cos\theta$ is 1 and this occurs when $\theta = 0$. Therefore the maximum value of $D_{\mathbf{u}} f(\mathbf{x})$ is $|\nabla f(\mathbf{x})|$ and it occurs when $\theta = 0$, that is, when $\mathbf{u}$ has the same direction as $\nabla f(\mathbf{x})$.

Tangent Plane and Normal Line to a Level Surface

Suppose S with equation $F(x, y, z) = k$ is a level surface of a function F of three variables and let $P(x_0, y_0, z_0)$ be a point on S. If $\nabla F(x_0, y_0, z_0) \neq \mathbf{0}$, then the **tangent plane to the level surface at** P is given by

$$F_x(x_0, y_0, z_0)(x - x_0) + F_y(x_0, y_0, z_0)(y - y_0) + F_z(x_0, y_0, z_0)(z - z_0) = 0.$$

The **normal line** to S at P is given by

$$\frac{x - x_0}{F_x(x_0, y_0, z_0)} = \frac{y - y_0}{F_y(x_0, y_0, z_0)} = \frac{z - z_0}{F_z(x_0, y_0, z_0)}.$$

Justification

Let C be any curve that lies on the surface S and passes through the point P. Suppose C is given by a vector function $\mathbf{r}(t) = \langle x(t), y(t), z(t)\rangle$ and t_0 is the parameter value corresponding to P: $\mathbf{r}(t_0) = \langle x_0, y_0, z_0\rangle$. Since C lies on S,

$$F(x(t), y(t), z(t)) = k.$$

From the Chain Rule, we have

$$\frac{\partial F}{\partial x} x'(t) + \frac{\partial F}{\partial y} y'(t) + \frac{\partial F}{\partial z} z'(t) = 0.$$

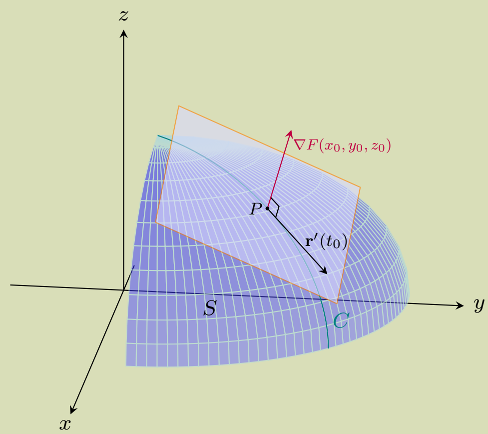
Because $\nabla F = \langle F_x, F_y, F_z\rangle$ and $\mathbf{r}'(t) = \langle x'(t), y'(t), z'(t)\rangle$, we get

$$\nabla F \cdot \mathbf{r}'(t) = 0.$$

In particular, when t_0, we get

$$\nabla F(x_0, y_0, z_0) \cdot \mathbf{r}'(t_0) = 0.$$

The last equation means that the gradient vector at P, $\nabla F(x_0, y_0, z_0)$, is perpendicular to the tangent vector $\mathbf{r}'(t_0)$ to any curve C on S that passes through P. Hence, it makes sense to define the tangent plane to the surface S at P and the normal line to S at P, as stated.



Local Extreme of a Function of Two Variables

A function f of two variables has a **local maximum** at (a, b) if $f(x, y) \leq f(a, b)$ when (x, y) is near (a, b). The number $f(a, b)$ is called a **local maximum value**. If $f(x, y) \geq f(a, b)$ when (x, y) is near (a, b), then f has a **local minimum** at (a, b) and $f(a, b)$ is a **local minimum value**. A point (a, b) is called a **saddle point** of f if $f(a, b)$ is a local maximum in one direction but a local minimum in another.

Necessary Condition

If f has a local maximum or minimum at (a, b) and the first-order partial derivatives of f exist there, then $f_x(a, b) = 0$ and $f_y(a, b) = 0$.

Critical Point

A point (a, b) is called a **critical point** of f if $f_x(a, b) = 0$ and $f_y(a, b) = 0$, or if one of these partial derivatives does not exist.

Justification

Let $g(x) = f(x, b)$. If f has a local maximum (or minimum) at (a, b), then g has a local maximum (or minimum) at a, so $g'(a) = 0$ by Fermat's Theorem. But $g'(a) = f_x(a, b)$ and so $f_x(a, b) = 0$. Similarly, by applying Fermat's Theorem to the function $h(y) = f(a, y)$, we obtain $f_y(a, b) = 0$.

Second Derivatives Test

Suppose the second partial derivatives of f are continuous on a disk with center (a, b), and suppose that $f_x(a,b) = 0$ and $f_y(a,b) = 0$. 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 $D < 0$, then (a,b) is a saddle point and $f(a,b)$ is not a local maximum or minimum.

Close Set and Bounded Set in $\mathbb{R}^2$

Suppose D is a subset in $\mathbb{R}^2$. A boundary point of D is a point (a,b) such that every disk with center (a,b) contains points in D and also points not in D. D is said to be a **closed set** in $\mathbb{R}^2$ if it contains all its boundary points. D is said to be a **bounded set** in $\mathbb{R}^2$ is if it is contained within some disk.

Extreme Value Theorem for Functions of Two Variables

If f is continuous on a closed, 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 points (x_1, y_1) and (x_2, y_2) in D.

Extension of the Closed Interval Method

To find the absolute maximum and minimum values of a continuous function f on a closed, bounded set D:

1. Find the values of f at the critical points of f in D.

2. Find the extreme values of f on the boundary of D.

3. The largest of the values from steps 1 and 2 is the absolute maximum value; the smallest of these values is the absolute minimum value.

Method of Lagrange Multipliers

Assume that the maximum and minimum values of $f(x,y,z)$ subject to the constraint $g(x,y,z) = k$ exist. Suppose $\nabla g \neq \mathbf{0}$ on the surface $g(x,y,z) = k$. To find the these extreme values:

(a) Find all values of x, y, z, and λ such that

$$\nabla f(x,y,z) = \lambda \nabla g(x,y,z)$$

and

$$g(x,y,z) = k.$$

(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.

Method of Lagrange Multipliers with Two Constraints

Assume that the maximum and minimum values of $f(x, y, z)$ subject to two constraints $g(x, y, z) = k$ and $h(x, y, z) = c$ exist. Suppose $\nabla g \neq \mathbf{0}$ on the surface $g(x, y, z) = k$ and $\nabla h \neq \mathbf{0}$ on the surface $h(x, y, z) = c$. To find the these extreme values:

(a) Find all values of x, y, z, λ, and μ such that

$$\nabla f(x, y, z) = \lambda \nabla g(x, y, z) + \mu \nabla h(x, y, z)$$

and

$$g(x, y, z) = k, \qquad h(x, y, z) = c.$$

(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.

Chapter 15

Multiple Integrals

Double Integral over a Rectangle

Suppose f is a function of two variables defined on a closed rectangle

$$R = [a, b] \times [c, d] = \left\{ (x, y) \in \mathbb{R}^2 \,\middle|\, a \le x \le b,\, c \le y \le d \right\}.$$

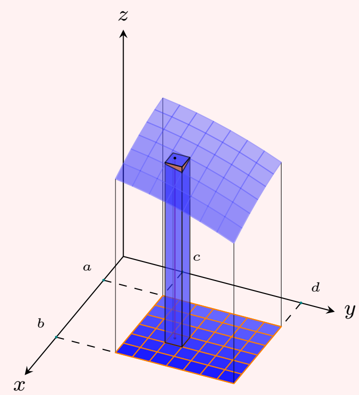
We divide the rectangle R into subrectangles, by dividing the interval $[a, b]$ into m subintervals $[x_{i-1}, x_i]$, $i = 1, 2, \ldots, m$, of equal width $\Delta x = (b - a)/m$ and dividing the interval $[c, d]$ into n subintervals $[y_{j-1}, y_j]$, $j = 1, 2, \ldots, n$, of equal width $\Delta y = (d - c)/n$. We choose a sample point (x_{ij}^*, y_{ij}^*) in each subrectangle $R_{ij} = [x_{i-1}, x_i] \times [y_{j-1}, y_j]$, which has area $\Delta A = \Delta x\, \Delta y$, and form the **double Riemann sum** $\displaystyle\sum_{i=1}^{m} \sum_{j=1}^{n} f\left(x_{ij}^*, y_{ij}^*\right) \Delta A$. If the limit

$$\lim_{m,n \to \infty} \sum_{i=1}^{m} \sum_{j=1}^{n} f\left(x_{ij}^*, y_{ij}^*\right) \Delta A$$

exists for any choice of sample points (x_{ij}^*, y_{ij}^*) in R_{ij}, we call the limit as the **double integral** of f over the rectangle R and denote the value as

$$\iint_R f(x, y)\, dA.$$

In this case, f is called **integrable**.



Volume

If $f(x,y) \geq 0$, then the volume V of the solid that lies above the rectangle R and below the surface$z = f(x,y)$ is

$$V = \iint_R f(x,y)\,dA.$$

The Midpoint Rule

$$\iint_R f(x,y)\,dA \approx \sum_{i=1}^{m}\sum_{j=1}^{n} f\left(\bar{x}_i, \bar{y}_j\right)\Delta A$$

where $\bar{x}_i$ is the midpoint of $[x_{i-1}, x_i]$ and $\bar{y}_j$ is the midpoint of $[y_{j-1}, y_j]$.

Average Value

The **average value** of a function f of two variables defined on a rectangle R is

$$f_{\text{ave}} = \frac{1}{A(R)} \iint_R f(x,y)\,dA,$$

where $A(R)$ is the area of R.

Properties of Double Integrals

Suppose f and g are integrable functions of of two variables on R, then the following properties hold:

1. $\displaystyle\iint_R [f(x,y) + g(x,y)]\,dA = \iint_R f(x,y)\,dA + \iint_R g(x,y)\,dA$

2. $\displaystyle\iint_R cf(x,y)\,dA = c\iint_R f(x,y)\,dA$, where c is a constant.

3. If $f(x,y) \geq g(x,y)$ for all (x,y) in R, then

$$\iint_R f(x,y)\,dA \geq \iint_R g(x,y)\,dA.$$

Fubini's Theorem

If f is continuous on the rectangle

$$R = [a,b]\times[c,d] = \left\{(x,y)\in\mathbb{R}^2 \,\middle|\, a\leq x\leq b,\, c\leq y\leq d\right\},$$

then

$$\iint_R f(x,y)\,dA = \int_a^b\int_c^d f(x,y)\,dy\,dx = \int_c^d\int_a^b f(x,y)\,dx\,dy,$$

where

$$\int_a^b\int_c^d f(x,y)\,dy\,dx = \int_a^b\left[\int_c^d f(x,y)\,dy\right]dx, \qquad \int_c^d\int_a^b f(x,y)\,dx\,dy = \int_c^d\left[\int_a^b f(x,y)\,dx\right]dy$$

are called **iterated integrals**.
More generally, this is true if we assume that f is bounded on R, f is discontinuous only on a finite number of smooth curves, and the iterated integrals exist.
In particular,

$$\iint_R g(x)\,h(y)\,dA = \int_a^b g(x)\,dx\int_c^d h(y)\,dy.$$

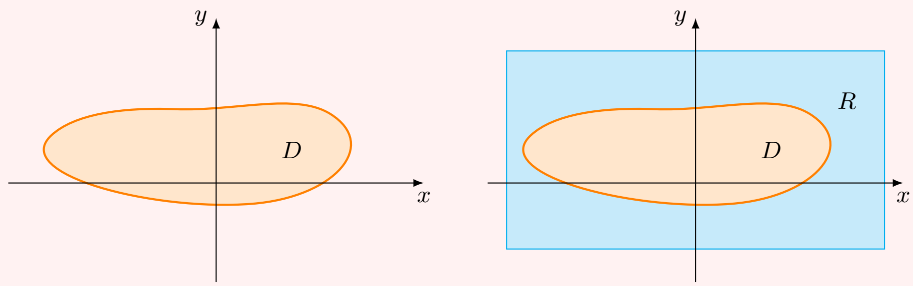
Double Integral over General Regions

Suppose f is a function of two variables defined on a bounded region D. Let R be a rectangular region that encloses D and denote

$$F(x,y) = \begin{cases} f(x,y), & \text{if } (x,y) \text{ is in } D,\\ 0, & \text{if } (x,y) \text{ is in } R \text{ but not in } D.\end{cases}$$

The **double integral of f over D** by

$$\iint_D f(x,y)\,dA = \iint_R F(x,y)\,dA.$$



Double Integral over types I and II Regions

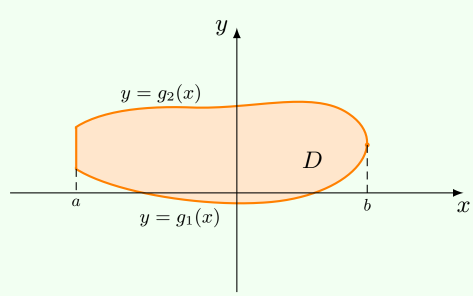
Double Integral over a type I Region

If f is continuous on a type I region D such that

$$D = \left\{ (x, y) \in \mathbb{R}^2 \,\middle|\, a \le x \le b,\ g_1(x) \le y \le g_2(x) \right\},$$

then

$$\iint_D f(x, y)\, dA = \int_a^b \int_{g_1(x)}^{g_2(x)} f(x, y)\, dy\, dx.$$



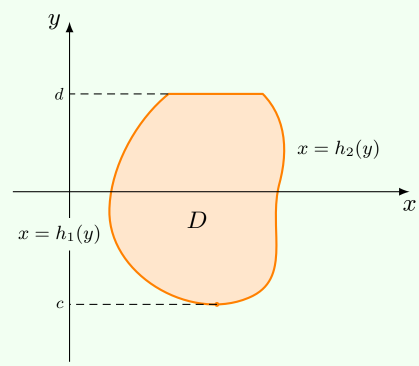
Double Integral over a type II Region

If f is continuous on a type II region D such that

$$D = \left\{(x, y) \in \mathbb{R}^2 \,\middle|\, c \le y \le d,\, h_1(y) \le x \le h_2(y)\right\},$$

then

$$\iint_D f(x, y)\, dA = \int_c^d \int_{h_1(y)}^{h_2(y)} f(x, y)\, dx\, dy.$$



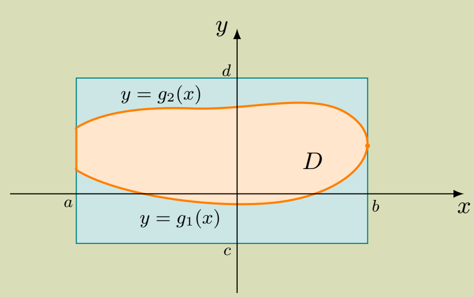
Justification

Let D be of type I that lies between the graphs of two continuous functions of x:

$$D = \left\{(x, y) \in \mathbb{R}^2 \,\middle|\, a \le x \le b,\, g_1(x) \le y \le g_2(x)\right\}.$$

To evaluate the double integral $\iint_D f(x, y)\, dA$, we choose a rectangle $R = [a, b] \times [c, d]$ that contains D, as in shown in the figure, and we let F be the function given by

$$F(x, y) = \begin{cases} f(x, y), & \text{if } (x, y) \text{ is in } D, \\ 0, & \text{if } (x, y) \text{ is in } R \text{ but not in } D. \end{cases}$$



Then, by Fubini's Theorem,

$$\iint_D f(x,y)\,dA = \iint_R F(x,y)\,dA = \int_a^b \int_c^d F(x,y)\,dy\,dx.$$

For fixed x, we have

$$\begin{aligned}
\int_c^d F(x,y)\,dy &= \int_c^{g_1(x)} F(x,y)\,dy + \int_{g_1(x)}^{g_2(x)} F(x,y)\,dy + \int_{g_2(x)}^d F(x,y)\,dy \\
&= \int_c^{g_1(x)} 0\,dy + \int_{g_1(x)}^{g_2(x)} f(x,y)\,dy + \int_{g_2(x)}^d 0\,dy \\
&= 0 + \int_{g_1(x)}^{g_2(x)} f(x,y)\,dy + 0 = \int_{g_1(x)}^{g_2(x)} f(x,y)\,dy.
\end{aligned}$$

Hence, for a type I region D,

$$\iint_D f(x,y)\,dA = \int_a^b \int_{g_1(x)}^{g_2(x)} f(x,y)\,dy\,dx.$$

In a similar manner, we can obtain the equality in the case of type II region.

Properties of Double Integrals

Suppose f and g are integrable functions of two variables over a region D.

1. $\displaystyle\iint_D [f(x,y)+g(x,y)]\,dA = \iint_D f(x,y)\,dA + \iint_D g(x,y)\,dA.$

2. $\displaystyle\iint_D cf(x,y)\,dA = c\iint_D f(x,y)\,dA$, where c is any constant.

3. If $f(x,y) \geq g(x,y)$ for all (x,y) in D, then $\displaystyle\iint_D f(x,y)\,dA \geq \iint_D g(x,y)\,dA.$

4. $\displaystyle\iint_D f(x,y)\,dA = \iint_{D_1} f(x,y)\,dA + \iint_{D_2} f(x,y)\,dA$, provided that the two double integrals on the right-hand side exist.

5. $\displaystyle\iint_D 1\,dA =$ area of D.

6. If $m \leq f(x,y) \leq M$ for all (x,y) in D, then

$$m\,A(D) \leq \iint_D f(x,y)\,dA \leq M\,A(D),$$

where $A(D)$ is the area of D.

Change to Polar Coordinates in a Double Integral over a Polar Rectangle

If f is continuous on a **polar rectangle** R given by $0 \le a \le r \le b$, $\alpha \le \theta \le \beta$, where $0 \le \beta - \alpha \le 2\pi$, then

$$\iint_R f(x,y)\,dA = \int_\alpha^\beta \int_a^b f(r\cos\theta, r\sin\theta)\,r\,dr\,d\theta.$$

Justification

In case of a polar rectangle

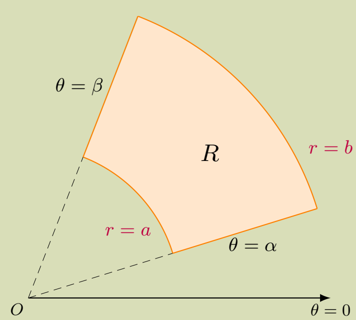
$$R = \left\{(r,\theta) \,\middle|\, a \le r \le b,\ \alpha \le \theta \le \beta\right\},$$

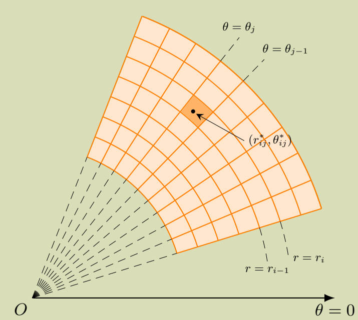
we divide the interval $[a,b]$ into m subintervals $[r_{i-1}, r_i]$ of equal width $\Delta r = (b-a)/m$ and we divide the interval $[\alpha,\beta]$ into n subintervals $[\theta_{j-1},\theta_j]$ of equal width $\Delta\theta = (\beta-\alpha)/n$. So, we have a collection of polar subrectangles $\{R_{ij}\}$:

$$R_{ij} = \left\{(r,\theta) \,\middle|\, r_{i-1} \le r \le r_i,\ \theta_{j-1} \le \theta \le \theta_j\right\}.$$

For fixed i, j, we take sample point $(r_{ij}^*, \theta_{ij}^*) \in R_{ij}$. Then, the area of R_{ij} is

$$\begin{aligned}\Delta A_{ij} &= \tfrac{1}{2} r_i^2\,\Delta\theta - \tfrac{1}{2} r_{i-1}^2\,\Delta\theta = \tfrac{1}{2}(r_i^2 - r_{i-1}^2)\,\Delta\theta \\ &= \tfrac{1}{2}(r_i + r_{i-1})(r_i - r_{i-1})\,\Delta\theta \approx r_{ij}^*\,\Delta r\,\Delta\theta.\end{aligned}$$





Thus, the double integral $\iint_R f(x,y)\,dA$ in terms of ordinary rectangles has a typical Riemann sum

$$\sum_{i=1}^m \sum_{j=1}^n f(r_{ij}^*\cos\theta_{ij}, r_{ij}^*\sin\theta_{ij})\,\Delta A_{ij} \approx \sum_{i=1}^m \sum_{j=1}^n f(r_{ij}^*\cos\theta_{ij}, r_{ij}^*\sin\theta_{ij}) r_{ij}^*\,\Delta r\,\Delta\theta.$$

If we write $g(r,\theta) = rf(r\cos\theta, r\sin\theta)$, then the last Riemann sum can be written as

$$\sum_{i=1}^m \sum_{j=1}^n g(r_{ij}^*, \theta_{ij}^*)\,\Delta r\,\Delta\theta,$$

which is a Riemann sum for the double integral $\int_\alpha^\beta \int_a^b g(r,\theta)\,dr\,d\theta$. Hence,

$$\begin{aligned}\iint_R f(x,y)\,dA &= \lim_{m,n\to\infty} \sum_{i=1}^m \sum_{j=1}^n f(r_{ij}^* \cos\theta_{ij}, r_{ij}^* \sin\theta_{ij})\,\Delta A_{ij} \\ &= \lim_{m,n\to\infty} \sum_{i=1}^m \sum_{j=1}^n g(r_{ij}^*, \theta_{ij}^*)\,\Delta r\,\Delta\theta = \int_\alpha^\beta \int_a^b g(r,\theta)\,dr\,d\theta = \int_\alpha^\beta \int_a^b f(r\cos\theta, r\sin\theta) r\,dr\,d\theta.\end{aligned}$$

Change to Polar Coordinates in a Double Integral over a Polar Region

If f is continuous on a polar region of the form

$$D = \left\{(r,\theta) \,\middle|\, \alpha \le \theta \le \beta,\ h_1(\theta) \le r \le h_2(\theta)\right\},$$

where $0 \le \beta - \alpha \le 2\pi$, then

$$\iint_D f(x,y)\,dA = \int_\alpha^\beta \int_{h_1(\theta)}^{h_2(\theta)} f(r\cos\theta, r\sin\theta)\,r\,dr\,d\theta.$$

Justification

In case of a polar region D bounded by two polar curves:

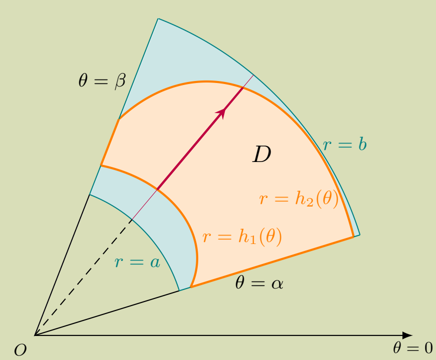
$$D = \left\{(r,\theta) \,\middle|\, \alpha \le \theta \le \beta,\ h_1(\theta) \le r \le h_2(\theta)\right\},$$

we choose a polar rectangle

$$R = \left\{(r,\theta) \,\middle|\, a \le r \le b,\ \alpha \le \theta \le \beta\right\},$$

that contains D, as in shown in the figure, and we let F be the function given by

$$F(x,y) = \begin{cases} f(x,y), & \text{if } (r,\theta) \text{ is in } D, \\ 0, & \text{if } (r,\theta) \text{ is in } R \text{ but not in } D. \end{cases}$$



Since R is a polar rectangle, we have

$$\iint_R F(x,y)\,dA = \int_\alpha^\beta \int_a^b F(r\cos\theta, r\sin\theta) r\,dr\,d\theta.$$

For fixed θ, we have

$$\begin{aligned}
&\int_a^b F(r\cos\theta, r\sin\theta) r\,dr \\
&= \int_a^{h_1(\theta)} F(r\cos\theta, r\sin\theta) r\,dr + \int_{h_1(\theta)}^{h_2(\theta)} F(r\cos\theta, r\sin\theta) r\,dr + \int_{h_2(\theta)}^b F(r\cos\theta, r\sin\theta) r\,dr \\
&= \int_a^{h_1(\theta)} 0\,dr + \int_{h_1(\theta)}^{h_2(\theta)} f(r\cos\theta, r\sin\theta) r\,dr + \int_{h_2(\theta)}^b 0\,dr \\
&= \int_{h_1(\theta)}^{h_2(\theta)} f(r\cos\theta, r\sin\theta) r\,dr.
\end{aligned}$$

Hence,

$$\iint_D f(x,y)\,dA = \iint_R F(x,y)\,dA = \int_\alpha^\beta \int_{h_1(\theta)}^{h_2(\theta)} f(r\cos\theta, r\sin\theta) r\,dr.$$

Center of Mass - Application

The coordinates $(\overline{x}, \overline{y})$ of the **center of mass** of a lamina occupying the region D and having density function $\rho(x,y)$ are

$$\overline{x} = \frac{M_y}{m}, \qquad \overline{y} = \frac{M_x}{m},$$

where m is the **mass**

$$m = \iint_D \rho(x,y)\,dA$$

and M_x and M_y are the **moments** about the x- and y-axis

$$M_x = \iint_D y\rho(x,y)\,dA, \qquad M_y = \iint_D x\rho(x,y)\,dA.$$

Moment of Inertia - Application

Suppose a lamina occupies the region D and has density function $\rho(x,y)$. Then the **moment of inertia** of the lamina about the x-axis is

$$I_x = \iint_D y^2\rho(x,y)\,dA.$$

Similarly, the moment of inertia about the y-axis is

$$I_x = \iint_D x^2 \rho(x,y)\, dA.$$

The moment of inertia about the origin is

$$I_0 = \iint_D (x^2+y^2)\rho(x,y)\, dA.$$

Radius of Gyration - Application

The **radius of gyration of a lamina about an axis** is the number R such that $mR^2 = I$, where m is the mass of the lamina and I is the moment of inertia about the given axis.
In particular, the radius of gyration $\bar{\bar{y}}$ with respect to the x-axis and the radius of gyration $\bar{\bar{x}}$ with respect to the y-axis are given by the equations

$$m\bar{\bar{y}}^2 = I_x, \qquad m\bar{\bar{x}}^2 = I_y.$$

Joint Density Function

Consider a pair of continuous random variables X and Y. The **joint density function** of X and Y is a function f of two variables such that the probability that (X,Y) lies in a region D is

$$P\left((X,Y)\in D\right) = \iint_D f(x,y)\, dA.$$

In particular, if the region is a rectangle, the probability that X lies between a and b and Y lies between c and d is

$$P(a \le X \le b,\, c \le Y \le d) = \int_a^b \int_c^d f(x,y)\, dy\, dx.$$

The joint density function f satisfies the following properties:

$$f(x,y) \ge 0, \qquad \iint_{\mathbb{R}^2} f(x,y)\, dA = 1.$$

Suppose X is a random variable with probability density function $f_1(x)$ and Y is a random variable with density function $f_2(y)$. Then X and Y are called **independent random variables** if their joint density function is the product of their individual density functions:

$$f(x,y) = f_1(x) f_2(y).$$

The X**-mean** and Y**-mean**, also called the **expected values** of X and Y are

$$\mu_1 = \iint_{\mathbb{R}^2} x f(x,y)\, dA, \qquad \mu_2 = \iint_{\mathbb{R}^2} y f(x,y)\, dA.$$

Surface Area

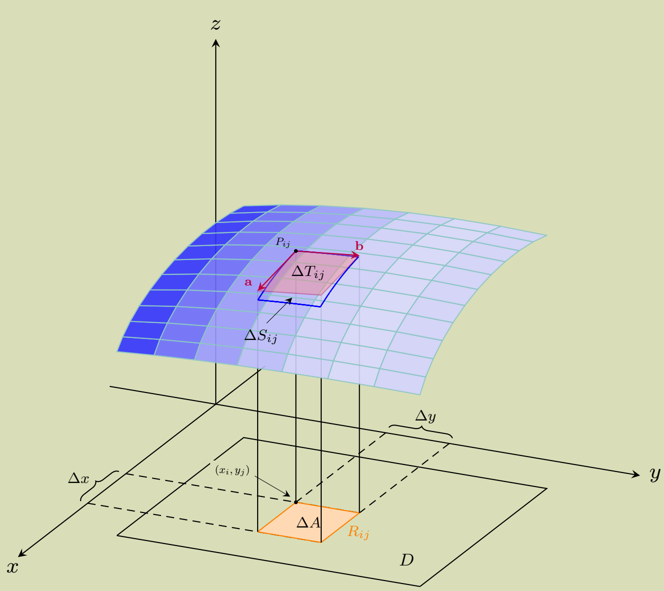
The area of the surface with equation $z = f(x, y)$, $(x, y) \in D$, where f_x an f_y are continuous, is

$$\begin{aligned} A(S) &= \iint_D \sqrt{\left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2 + 1}\, dA \\ &= \iint_D \sqrt{[f_x(x, y)]^2 + [f_y(x, y)]^2 + 1}\, dA. \end{aligned}$$

Justification - Part 1

Let S be a surface with equation $z = f(x, y)$, where f has continuous partial derivatives. For simplicity, we assume that $f(x, y) \geq 0$ and the domain D of f is a rectangle. We divide D into small rectangles R_{ij} with area $\Delta A = \Delta x \Delta y$. If (x_i, y_j) is the corner of R_{ij} closest to the origin, let $P_{ij}(x_i, y_j, f(x_i, y_j))$ be the point on S directly above it. The tangent plane to S at P_{ij} is an approximation to S near P_{ij}. So the area ΔT_{ij} of the part of this tangent plane (a parallelogram) that lies directly above R_{ij} is an approximation to the area ΔS_{ij} of the part of S that lies directly above R_{ij}. Thus the sum $\sum\sum \Delta T_{ij}$ is an approximation to the total area of S, and this approximation appears to improve as the number of rectangles increases. Therefore we define the surface area of S to be

$$A(S) = \lim_{m,n\to\infty} \sum_{i=1}^{m} \sum_{j=1}^{n} \Delta T_{ij}.$$



Justification - Part 2

To derive the formula for the surface area, we let $\mathbf{a}$ and $\mathbf{b}$ be the vectors that start at P_{ij} and lie along the sides of the parallelogram with area ΔT_{ij}. We know that $\Delta T_{ij} = |\,\mathbf{a}\times\mathbf{b}\,|$. Since $f_x(x_i, y_j)$ and $f_y(x_i, y_j)$ are the slopes of the tangent lines through P_{ij} in the directions of $\mathbf{a}$ and $\mathbf{b}$, we have

$$\begin{aligned}\mathbf{a} &= \Delta x\,\mathbf{i} + f_x(x_i, y_j)\,\Delta x\,\mathbf{k},\\ \mathbf{b} &= \Delta y\,\mathbf{i} + f_y(x_i, y_j)\,\Delta y\,\mathbf{k},\end{aligned}$$

so that

$$\begin{aligned}\mathbf{a}\times\mathbf{b} &= \begin{vmatrix}\mathbf{i} & \mathbf{j} & \mathbf{k}\\ \Delta x & 0 & f_x(x_i, y_j)\,\Delta x\\ 0 & \Delta y & f_y(x_i, y_j)\,\Delta y\end{vmatrix}\\ &= -f_x(x_i, y_j)\,\Delta x\,\Delta y\,\mathbf{i} - f_y(x_i, y_j)\,\Delta x\,\Delta y\,\mathbf{j} + \Delta x\,\Delta y\,\mathbf{k}\\ &= \left[-f_x(x_i, y_j)\,\mathbf{i} - f_y(x_i, y_j)\,\mathbf{j} + \mathbf{k}\right]\,\Delta A.\end{aligned}$$

Thus,

$$\Delta T_{ij} = |\,\mathbf{a}\times\mathbf{b}\,| = \sqrt{[f_x(x_i, y_j)]^2 + [f_y(x_i, y_j)]^2 + 1}\,\Delta A.$$

Hence, the desired surface area is

$$\begin{aligned}A(S) &= \lim_{m,n\to\infty}\sum_{i=1}^{m}\sum_{j=1}^{n}\Delta T_{ij}\\ &= \lim_{m,n\to\infty}\sum_{i=1}^{m}\sum_{j=1}^{n}\sqrt{[f_x(x_i, y_j)]^2 + [f_y(x_i, y_j)]^2 + 1}\,\Delta A\\ &= \iint_D \sqrt{[f_x(x,y)]^2 + [f_y(x,y)]^2 + 1}\,dA.\end{aligned}$$

Triple Integral over a Rectangular Box

Suppose f is a function of three variables defined on a rectangular box:

$$B = [a,b]\times[c,d]\times[r,s] = \left\{(x,y,z)\in\mathbb{R}^4 \,\middle|\, a\le x\le b,\ c\le y\le d,\ r\le z\le s\right\}.$$

We divide B into sub-boxes, by dividing the interval $[a,b]$ into l subintervals $[x_{i-1}, x_i]$, $i = 1, 2, \ldots, l$, of equal width $\Delta x = (b-a)/l$, dividing the interval $[c,d]$ into m subintervals $[y_{j-1}, y_j]$, $j = 1, 2, \ldots, m$, of equal width $\Delta y = (d-c)/m$, and dividing the interval $[r,s]$ into n subintervals $[z_{k-1}, z_k]$, $k = 1, 2, \ldots, n$, of equal width $\Delta z = (s-r)/n$. If we choose a sample point $(x^*_{ijk}, y^*_{ijk}, z^*_{ijk})$ in each sub-box $R_{ijk} = [x_{i-1}, x_i]\times[y_{j-1}, y_j]\times[z_{k-1}, z_k]$, which has volume $\Delta V = \Delta x\,\Delta y\,\Delta z$, and if the limit $\lim_{l,m,n\to\infty}\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} f\left(x^*_{ijk}, y^*_{ijk}, z^*_{ijk}\right)\Delta V$ exists for any choice of sample points (x^*_{ij}, y^*_{ij}) in R_{ij}, we call the limit as the **triple integral** of f over the box B and denote the value as

$$\iiint_B f(x,y,z)\,dV.$$

Volume as a Triple Integral

$$V(E) = \iiint_E dV$$

Fubini's Theorem for Triple Integrals

If f is continuous on the rectangular box

$$B = [a,b] \times [c,d] \times [r,s] = \left\{(x,y) \in \mathbb{R}^2 \,\middle|\, a \le x \le b,\, c \le y \le d,\, r \le z \le s\right\},$$

then

$$\iiint_B f(x,y,z)\,dV = \int_r^s \int_c^d \int_a^b f(x,y,z)\,dx\,dy\,dz.$$

Triple Integral over General Regions

Suppose f is a function of three variables defined on a bounded region E in three-dimensional space. Let B be a rectangular box that encloses E and denote

$$F(x,y,z) = \begin{cases} f(x,y,z), & \text{if } (x,y,z) \text{ is in } E, \\ 0, & \text{if } (x,y,z) \text{ is in } B \text{ but not in } E. \end{cases}$$

The **triple integral of f over E** by

$$\iiint_E f(x,y,z)\,dV = \iiint_B F(x,y,z)\,dV.$$

Triple Integral over a General Bounded Region

If f is defined in a bounded region E of **type 1** that lies between the graphs of two continuous functions of x and y:

$$E = \left\{(x,y,z) \in \mathbb{R}^3 \,\middle|\, (x,y) \in D,\, u_1(x,y) \le z \le u_2(x,y)\right\},$$

where D is the projection of E onto the xy-plane, then

$$\iiint_E f(x,y,z)\,dV = \iint_D \left[\int_{u_1(x,y)}^{u_2(x,y)} f(x,y,z)\,dz\right] dA.$$

In particular, if the projection D of E onto the xy-plane is a type I plane region:

$$E = \left\{(x,y,z) \in \mathbb{R}^3 \,\middle|\, a \le x \le b,\ g_1(x) \le y \le g_2(x),\ u_1(x,y) \le z \le u_2(x,y)\right\},$$

then

$$\iiint_E f(x,y,z)\,dV = \int_a^b \int_{g_1(x)}^{g_2(x)} \int_{u_1(x,y)}^{u_2(x,y)} f(x,y,z)\,dz\,dy\,dx.$$

If the projection D of E onto the xy-plane is a type II plane region:

$$E = \left\{(x,y,z) \in \mathbb{R}^3 \,\middle|\, c \le y \le d,\ h_1(y) \le x \le h_2(y),\ u_1(x,y) \le z \le u_2(x,y)\right\},$$

then

$$\iiint_E f(x,y,z)\,dV = \int_c^d \int_{h_1(y)}^{h_2(y)} \int_{u_1(x,y)}^{u_2(x,y)} f(x,y,z)\,dz\,dx\,dy.$$

Similarly, if E is a bounded region of **type 2**:

$$E = \left\{(x,y,z) \in \mathbb{R}^3 \,\middle|\, (y,z) \in D,\ u_1(y,z) \le x \le u_2(y,z)\right\},$$

where D is the projection of E onto the yz-plane, then

$$\iiint_E f(x,y,z)\,dV = \iint_D \left[\int_{u_1(y,z)}^{u_2(y,z)} f(x,y,z)\,dx\right] dA.$$

If E is a bounded region of **type 3**:

$$E = \left\{(x,y,z) \in \mathbb{R}^3 \,\middle|\, (x,z) \in D,\ u_1(x,z) \le y \le u_2(x,z)\right\},$$

where D is the projection of E onto the xz-plane, then

$$\iiint_E f(x,y,z)\,dV = \iint_D \left[\int_{u_1(x,z)}^{u_2(x,z)} f(x,y,z)\,dy\right] dA.$$

Center of Mass - Application

The coordinates $(\overline{x}, \overline{y}, \overline{z})$ of the **center of mass** of a solid object occupying the region E and having density function $\rho(x,y,z)$ are

$$\overline{x} = \frac{M_{yz}}{m}, \qquad \overline{y} = \frac{M_{xz}}{m}, \qquad \overline{y} = \frac{M_{xy}}{m},$$

where m is the **mass**

$$m = \iiint_E \rho(x,y)\,dV$$

and M_{yz}, M_{xz}, and M_{xy} are the **moments** about the three coordinate planes:

$$M_{yz} = \iiint_E x\rho(x,y,z)\,dV,$$
$$M_{xz} = \iiint_E y\rho(x,y,z)\,dV,$$
$$M_{xy} = \iiint_E z\rho(x,y,z)\,dV.$$

Moment of Inertia - Application

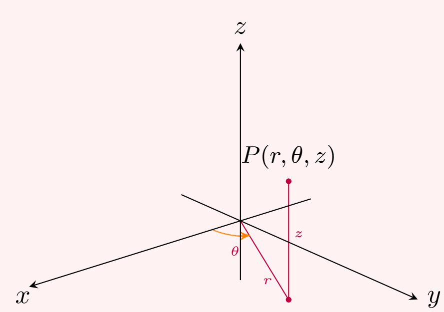
Suppose a solid occupies the region E and has density function $\rho(x,y,z)$. Then the **moment of inertia** of the solid about the three coordinate axes are

$$I_x = \iiint_E (y^2+z^2)\rho(x,y,z)\,dV,$$
$$I_y = \iiint_E (x^2+z^2)\rho(x,y,z)\,dV,$$
$$I_z = \iiint_E (x^2+y^2)\rho(x,y,z)\,dV.$$

Cylindrical Coordinates

To convert from cylindrical to rectangular coordinates, use the equations

$$x = r\cos\theta, \qquad y = r\sin\theta, \qquad z = z.$$



From Rectangular to Cylindrical coordinates

To convert rectangular to cylindrical coordinates, use the equations

$$r^2 = x^2 + y^2, \qquad \tan\theta = \frac{y}{x}, \qquad z = z.$$

Evaluating Triple Integrals with Cylindrical Coordinates

Suppose that E is a type 1 region whose projection D onto the xy-plane is conveniently described in polar coordinates:

$$E = \left\{(x, y, z) \in \mathbb{R}^3 \,\middle|\, (x, y) \in D,\, u_1(x, y) \le z \le u_2(x, y)\right\},$$

where D is given in polar coordinates by

$$D = \left\{(r, \theta) \,\middle|\, \alpha \le \theta \le \beta,\, h_1(\theta) \le r \le h_2(\theta)\right\}$$

Suppose that f is continuous. Then

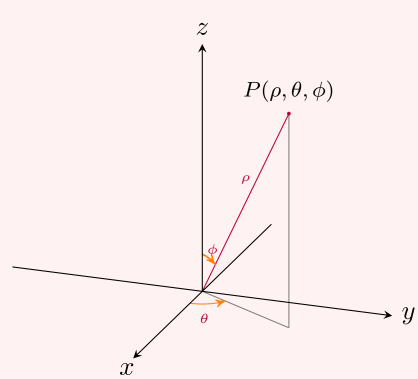
$$\begin{aligned}\iiint_E f(x, y, z)\, dV &= \iint_D \left[\int_{u_1(x,y)}^{u_2(x,y)} f(x, y, z)\, dz\right] dA \\ &= \int_\alpha^\beta \int_{h_1(\theta)}^{h_2(\theta)} \int_{u_1(r\cos\theta, r\sin\theta)}^{u_2(r\cos\theta, r\sin\theta)} f(r\cos\theta, r\sin\theta, z)\, r\, dz\, dr\, d\theta.\end{aligned}$$

Spherical Coordinates

To convert from spherical to rectangular coordinates, use the equations

$$x = \rho \sin\phi \, \cos\theta, \qquad y = \rho \sin\phi \, \sin\theta, \qquad z = \rho\cos\phi,$$

where $\rho \ge 0$ and $0 \le \phi \le \pi$.



Distance Formula

$$\rho^2 = x^2 + y^2 + z^2.$$

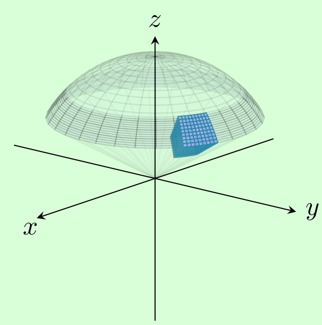
Evaluating Triple Integrals with Spherical Coordinates

Suppose that E is a spherical wedge given by

$$E = \left\{ (\rho, \theta, \phi) \,\middle|\, a \le \rho \le b,\, \alpha \le \theta \le \beta,\, c \le \phi \le d \right\}.$$

Suppose that f is continuous. Then

$$\iiint_E f(x,y,z)\,dV = \int_c^d \int_\alpha^\beta \int_a^b f(\rho \sin\phi\, \cos\theta, \rho \sin\phi\, \sin\theta, \rho\cos\phi)\,\rho^2\, \sin\phi\, d\rho\, d\theta\, d\phi.$$



In general, if E is given by

$$E = \left\{ (\rho, \theta, \phi) \,\middle|\, \alpha \le \theta \le \beta,\, c \le \phi \le d,\, g_1(\theta,\phi) \le \rho \le g_2(\theta,\phi) \right\},$$

then the iterated integral becomes

$$\int_c^d \int_\alpha^\beta \int_{g_1(\theta,\phi)}^{g_2(\theta,\phi)} f(\rho \sin\phi\, \cos\theta, \rho \sin\phi\, \sin\theta, \rho\cos\phi)\,\rho^2\, \sin\phi\, d\rho\, d\theta\, d\phi.$$

Jacobian

In 2-dimensional spaces, the **Jacobian** of the transformation T given by $x = g(u,v)$ and $y = h(u,v)$ is

$$\frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}.$$

In 3-dimensional spaces, the Jacobian of the transformation T given by

$$x = g(u,v,w), \qquad y = h(u,v,w), \qquad z = k(u,v,w)$$

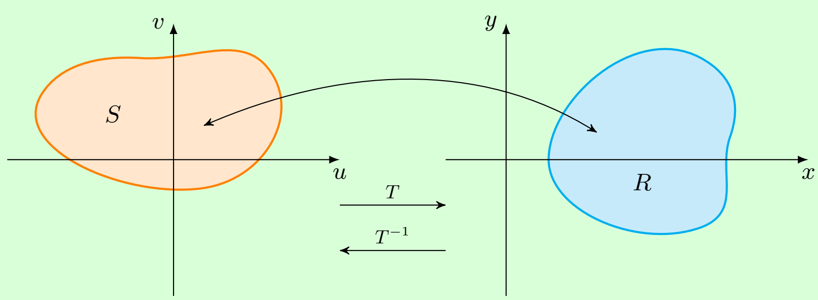
is

$$\frac{\partial(x,y,z)}{\partial(u,v,w)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} & \frac{\partial z}{\partial w} \end{vmatrix}.$$

Change of Variables in a Double Integral

Suppose that T is a C^1 transformation whose Jacobian is nonzero and that maps a region S in the uv-plane onto a region R in the xy-plane. Suppose that f is continuous on R and that R and S are type I or type II plane regions. Suppose also that T is one-to-one, except perhaps on the boundary of S. Then

$$\iint_R f(x,y)\,dA = \iint_S f(x(u,v),y(u,v)) \left|\frac{\partial(x,y)}{\partial(u,v)}\right| du\,dv$$



Change of Variables in a Triple Integral

$$\iiint_R f(x,y)\,dA = \iiint_S f(x(u,v,w),y(u,v,w),z(u,v,w)) \left|\frac{\partial(x,y,z)}{\partial(u,v,w)}\right| du\,dv\,dw$$

Justification - Part 1

We start with a small rectangle S in the uv-plane whose lower left corner is the point (u_0, v_0) and whose dimensions are Δu and Δv. The image of S is a region R in the xy-plane, with $(x_0, y_0) = T(u_0, v_0)$. The vector

$$\mathbf{r}(u,v) = g(u,v)\,\mathbf{i} + h(u,v)\,\mathbf{j}$$

is the position vector of the image of the point (u,v). The vector functions of the boundary of S

intersecting at (x_0, y_0) are $\mathbf{r}(u, v_0)$ and $\mathbf{r}(u_0, v)$, respectively. The tangent vectors at (x_0, y_0) are

$$\mathbf{r}_u = g_u(u_0, v_0)\,\mathbf{i} + h_u(u_0, v_0)\,\mathbf{j} = \frac{\partial x}{\partial u}\,\mathbf{i} + \frac{\partial y}{\partial u}\,\mathbf{j},$$
$$\mathbf{r}_v = g_v(u_0, v_0)\,\mathbf{i} + h_v(u_0, v_0)\,\mathbf{j} = \frac{\partial x}{\partial v}\,\mathbf{i} + \frac{\partial y}{\partial v}\,\mathbf{j}.$$

The image region $R = T(S)$ can be approximated by a parallelogram determined by the secant vectors

$$\mathbf{a} = \mathbf{r}(u_0 + \Delta u, v_0) - \mathbf{r}(u_0, v_0), \qquad \mathbf{b} = \mathbf{r}(u_0, v_0 + \Delta v) - \mathbf{r}(u_0, v_0).$$

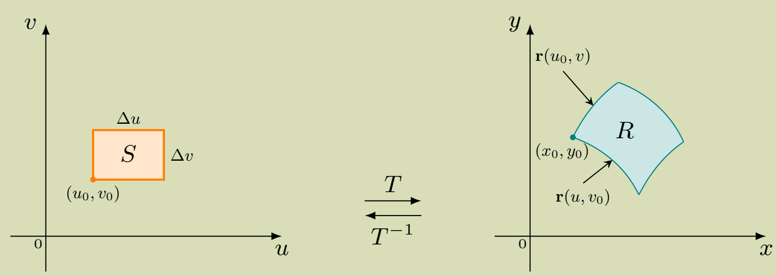
Since

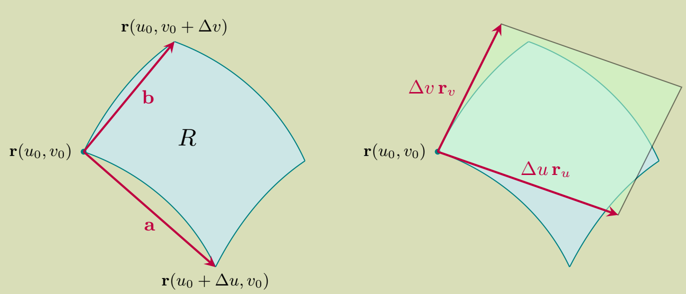
$$\mathbf{r}_u = \lim_{\Delta u \to 0} \frac{\mathbf{r}(u_0 + \Delta u, v_0) - \mathbf{r}(u_0, v_0)}{\Delta u}, \qquad \mathbf{r}_v = \lim_{\Delta v \to 0} \frac{\mathbf{r}(u_0, v_0 + \Delta v) - \mathbf{r}(u_0, v_0)}{\Delta v},$$

we have

$$\mathbf{a} = \mathbf{r}(u_0 + \Delta u, v_0) - \mathbf{r}(u_0, v_0) \approx \Delta u\,\mathbf{r}_u, \qquad \mathbf{b} = \mathbf{r}(u_0, v_0 + \Delta v) - \mathbf{r}(u_0, v_0) \approx \Delta v\,\mathbf{r}_v.$$

Thus, we can approximate R by a parallelogram determined by the vectors $\Delta u\,\mathbf{r}_u$ and $\Delta v\,\mathbf{r}_v$.





Justification - Part 2

The area of the parallelogram determined by the vectors $\Delta u\,\mathbf{r}_u$ and $\Delta v\,\mathbf{r}_v$ is given by the cross product

$$|(\Delta u\,\mathbf{r}_u) \times (\Delta v\,\mathbf{r}_v)| = |\,\mathbf{r}_u \times \mathbf{r}_v\,|\,\Delta u\,\Delta v.$$

By the definition of the cross product,

$$\mathbf{r}_u \times \mathbf{r}_v = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial x}{\partial u} & \dfrac{\partial y}{\partial u} & 0 \\ \dfrac{\partial x}{\partial v} & \dfrac{\partial y}{\partial v} & 0 \end{vmatrix} = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial y}{\partial u} \\ \dfrac{\partial x}{\partial v} & \dfrac{\partial y}{\partial v} \end{vmatrix} \mathbf{k} = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix} \mathbf{k}.$$

It follows that, in terms of the Jacobian,

$$\Delta A \approx \left| \frac{\partial(x,y)}{\partial(u,v)} \right| \Delta u \, \Delta v,$$

where the Jacobian is evaluated at (u_0, v_0).

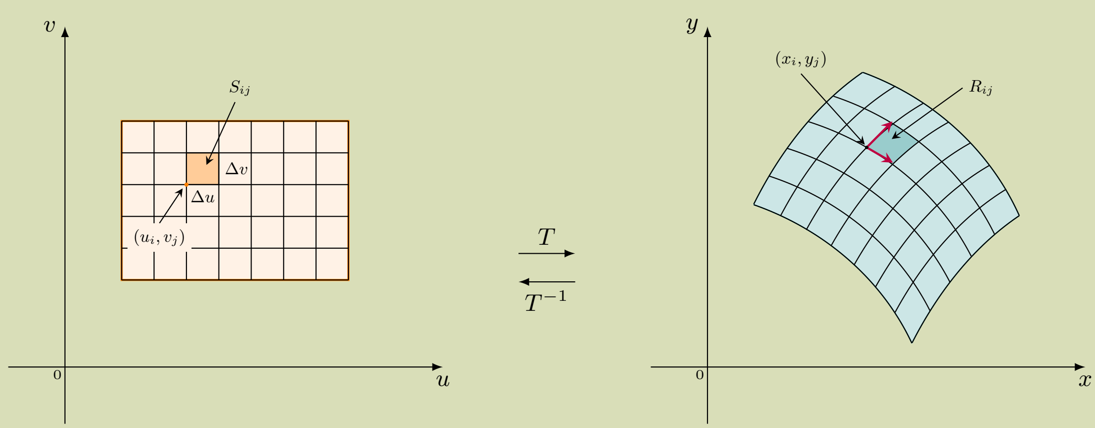
Justification - Part 3

Now are can approximate the double integral of f over R as follows:

$$\begin{aligned} \iint_R f(x,y)\,dA &\approx \sum_{i=1}^{m} \sum_{j=1}^{n} f(x_i, y_j)\,\Delta A \\ &\approx \sum_{i=1}^{m} \sum_{j=1}^{n} f(g(u_i, v_j), h(u_i, v_j)) \left| \frac{\partial(x,y)}{\partial(u,v)} \right| \Delta u \, \Delta v, \end{aligned}$$

where the Jacobian is evaluated at (u_i, v_j). The last double sum is a Riemann sum for the integral

$$\iint_S f(g(u,v), h(u,v)) \left| \frac{\partial(x,y)}{\partial(u,v)} \right| du\,dv.$$

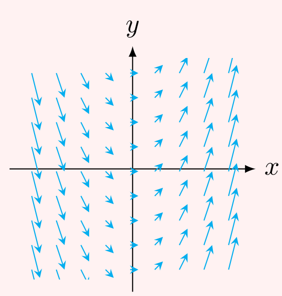


Chapter 16

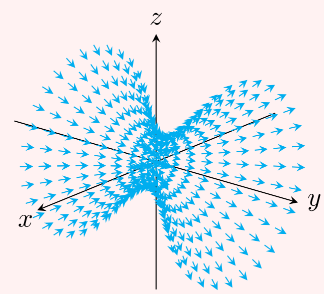
Vector Calculus

Vector Field

Let D be a subset in $\mathbb{R}^2$ (a plane region). A **vector field on** $\mathbb{R}^2$ is a function $\mathbf{F}$ that assigns to each point (x, y) in D a two-dimensional vector $\mathbf{F}(x, y)$.



Let E be a subset in $\mathbb{R}^3$. A **vector field on** $\mathbb{R}^3$ is a function $\mathbf{F}$ that assigns to each point (x, y, z) in D a three-dimensional vector $\mathbf{F}(x, y, z)$.



Gradient

If f is a scalar function of two variables, the **gradient** ∇f (or grad f) is defined by

$$\nabla f(x,y) = f_x(x,y)\,\mathbf{i} + f_y(x,y)\,\mathbf{j}.$$

If f is a scalar function of three variables, its **gradient** is a vector field on $\mathbb{R}^3$ given by

$$\nabla f(x,y,z) = f_x(x,y,z)\,\mathbf{i} + f_y(x,y,z)\,\mathbf{j} + f_z(x,y,z)\,\mathbf{k}.$$

Conservative Vector Field

A vector field $\mathbf{F}$ is called a **conservative vector field** if it is the gradient of some scalar function, that is, if there exists a function f such that $\mathbf{F} = \nabla f$. In this situation f is called a **potential function** for $\mathbf{F}$.

Line Integral

Suppose a plane curve C is given by the parametric equations

$$x = x(t), \qquad y = y(t), \qquad a \le t \le b$$

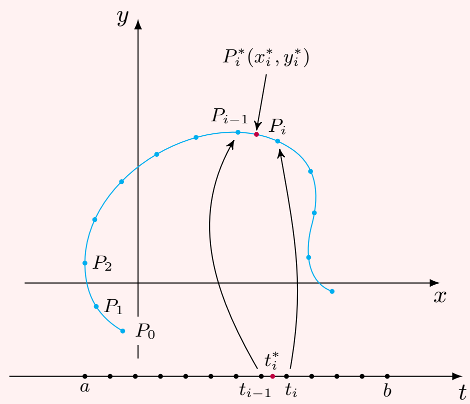
or, equivalently, by the vector equation $\mathbf{r}(t) = x(t)\,\mathbf{i} + y(t)\,\mathbf{j}$, and we assume that C is a smooth curve, that is, $\mathbf{r}'$ is continuous and $\mathbf{r}'(t) \neq 0$. If we divide the parameter interval $[a,b]$ into n subintervals $[t_{i-1}, t_i]$ of equal width and we let $x_i = x(t_i)$ and $y_i = y(t_i)$, then the corresponding points $P_i(x_i, y_i)$ divide C into n subarcs with lengths $\Delta s_1, \Delta s_2, \dots, \Delta s_n$. By choosing any point $P_i^*(x_i^*, y_i^*)$ in the ith subarc, we form the sum

$$\sum_{i=1}^{n} f(x_i^*, y_i^*)\,\Delta s_i,$$

where f is a function of two variables whose domain includes the curve C. The **line integral of f along C** is defined by

$$\int_C f(x,y)\,ds = \lim_{n\to\infty}\sum_{i=1}^{n} f(x_i^*, y_i^*)\,\Delta s_i,$$

provided the last limit exists.



Line Integrals With Respect to x and y

If replacing Δs_i by either $\Delta x_i = x_i - x_{i-1}$ or $\Delta y_i = y_i - y_{i-1}$, we get the **line integrals of f along C with respect to x and y**:

$$\int_C f(x,y)\,dx = \lim_{n\to\infty}\sum_{i=1}^{n} f(x_i^*, y_i^*)\,\Delta x_i,$$

$$\int_C f(x,y)\,dy = \lim_{n\to\infty}\sum_{i=1}^{n} f(x_i^*, y_i^*)\,\Delta y_i.$$

Evaluation of Line Integral

$$\int_C f(x,y)\,ds = \int_a^b f(x(t),y(t))\sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\,dt$$

Line Integral Along a Union of a Finite Number of Smooth Curves

Suppose that C is a piecewise-smooth curve; that is, C is a union of a finite number of smooth curves $C_1, C_2, \ldots, C_n$, where the initial point of C_{i+1} is the terminal point of C_i. Then

$$\int_C f(x,y)\,ds = \int_{C_1} f(x,y)\,ds + \int_{C_2} f(x,y)\,ds + \cdots + \int_{C_n} f(x,y)\,ds$$

Evaluation of Line Integrals With Respect to x and y

$$\int_C f(x,y)\,dx = \int_a^b f(x(t),y(t))\,x'(t)\,dt,$$
$$\int_C f(x,y)\,dy = \int_a^b f(x(t),y(t))\,y'(t)\,dt.$$

Justification

If $s(t)$ is the length of C between $\mathbf{r}(a)$ and $\mathbf{r}(t)$, then

$$\frac{ds}{dt} = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}.$$

So,

$$ds = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\,dt.$$

Thus, in terms of the variable t, we have

$$\int_C f(x,y)\,ds = \int_a^b f(x(t),y(t))\sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\,dt$$

For the line integral with Respect to x, we have $dx = x'(t)\,dt$, so that

$$\int_C f(x,y)\,dx = \int_a^b f(x(t),y(t))\,x'(t)\,dt.$$

Similarly, we have

$$\int_C f(x,y)\,dy = \int_a^b f(x(t),y(t))\,y'(t)\,dt$$

Line Integrals in Space

Suppose that C us a smooth space curve given by the parametric equations

$$x = x(t), \qquad y = y(t), \qquad z = z(t), \qquad a \le t \le b$$

If f is a function of three variables that is continuous on some region containing C, then we define the line integral of f along C (with respect to arc length) in a manner similar to that for plane curves:

$$\int_C f(x, y, z)\, ds = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*, y_i^*, z_i^*)\, \Delta s_i,$$

Line Integrals With Respect to x, y and z

By replacing Δs_i by either $\Delta x_i = x_i - x_{i-1}$, $\Delta y_i = y_i - y_{i-1}$, or $\Delta z_i = z_i - z_{i-1}$, we get the **line integrals of f along C with respect to x, y and z:**

$$\int_C f(x, y, z)\, dx = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*, y_i^*, z_i^*)\, \Delta x_i,$$
$$\int_C f(x, y, z)\, dy = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*, y_i^*, z_i^*)\, \Delta y_i,$$
$$\int_C f(x, y, z)\, dz = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*, y_i^*, z_i^*)\, \Delta z_i.$$

Evaluation of Line Integral

$$\int_C f(x, y, z)\, ds = \int_a^b f(x(t), y(t), z(t)) \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}\, dt$$

Evaluation of Line Integrals With Respect to x, y and z

$$\int_C f(x, y, z)\, dx = \int_a^b f(x(t), y(t), z(t))\, x'(t)\, dt,$$
$$\int_C f(x, y, z)\, dy = \int_a^b f(x(t), y(t), z(t))\, y'(t)\, dt,$$
$$\int_C f(x, y, z)\, dz = \int_a^b f(x(t), y(t), z(t))\, z'(t)\, dt.$$

Line Integrals of Vector Fields

Suppose that $\mathbf{F}$ is a continuous force field on $\mathbb{R}^3$. To compute the work done by this force in moving a particle along a smooth curve C given by the parametric equations

$$x = x(t), \qquad y = y(t), \qquad z = z(t), \qquad a \le t \le b$$

we divide the parameter interval $[a, b]$ into n subintervals $[t_{i-1}, t_i]$ of equal width and we let $x_i = x(t_i)$, $y_i = y(t_i)$, and $z_i = z(t_i)$, then the corresponding points $P_i(x_i, y_i, z_i)$ divide C into n subarcs with lengths $\Delta s_1, \Delta s_2, \ldots, \Delta s_n$. Choose a point $P_i^*(x_i^*, y_i^*, z_i^*)$ on the ith subarc corresponding to the parameter value t_i^*. For small Δs_i, as the particle moves from P_{i-1} to P_i along the curve, it proceeds approximately in the direction of $\mathbf{T}(x_i^*, y_i^*, z_i^*)$, the unit tangent vector at P_i^*. Thus the work done by the force F in moving the particle from P_{i-1} to P_i is approximately

$$\mathbf{F}(x_i^*, y_i^*, z_i^*) \cdot [\Delta s_i \mathbf{T}(x_i^*, y_i^*, z_i^*)] = [\mathbf{F}(x_i^*, y_i^*, z_i^*) \cdot \mathbf{T}(x_i^*, y_i^*, z_i^*)]\ \Delta s_i,$$

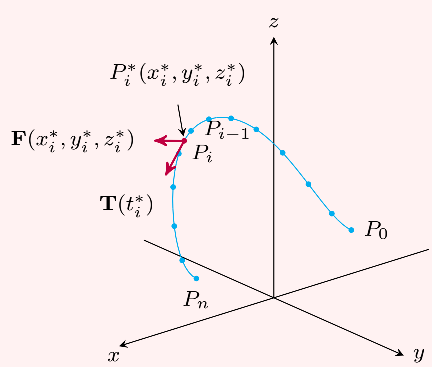
and the total work done in moving the particle along C is approximately

$$\sum_{i=1}^{n} [\mathbf{F}(x_i^*, y_i^*, z_i^*) \cdot \mathbf{T}(x_i^*, y_i^*, z_i^*)]\ \Delta s_i,$$

Thus, we define the **work** W done by the force field $\mathbf{F}$ as

$$W = \int_C \mathbf{F}(x, y, z) \cdot \mathbf{T}(x, y, z)\, ds = \lim_{n\to\infty} [\mathbf{F}(x_i^*, y_i^*, z_i^*) \cdot \mathbf{T}(x_i^*, y_i^*, z_i^*)]\ \Delta s_i,$$

provided the last limit exists.



If the curve C is given by the vector equation $\mathbf{r}(t) = x(t)\,\mathbf{i} + y(t)\,\mathbf{j} + z(t)\,\mathbf{k}$, then $\mathbf{T}(t) = \mathbf{r}'(t)/\,|\mathbf{r}'(t)|$. Thus,

$$W = \int_a^b \left[\mathbf{F}(\mathbf{r}(t)) \cdot \frac{\mathbf{r}'(t)}{|\mathbf{r}'(t)|}\right] |\mathbf{r}'(t)|\ dt = \int_a^b \mathbf{F}(\mathbf{r}(t)) \cdot \mathbf{r}'(t)\, dt.$$

Hence, if $\mathbf{F}$ is a continuous vector field defined on a smooth curve C given by a vector function $\mathbf{r}(t)$, $a \le t \le b$, then we can define the **line integral of F along** C as

$$\int_C \mathbf{F} \cdot d\mathbf{r} = \int_a^b \mathbf{F}(\mathbf{r}(t)) \cdot \mathbf{r}'(t)\, dt = \int_C \mathbf{F} \cdot \mathbf{T}\, ds.$$

Line Integrals of Vector Fields in Component Form

Let $\mathbf{F} = P\,\mathbf{i} + Q\,\mathbf{j} + R\,\mathbf{k}$ be a continuous vector field defined on a smooth curve C given by a vector function $\mathbf{r}(t)$, $a \le t \le b$. Then the **line integral of F along** C is

$$\int_C \mathbf{F}\cdot\mathbf{T}\,ds = \int_C \mathbf{F}\cdot d\mathbf{r} = \int_C P\,dx + Q\,dy + R\,dz.$$

Justification

Using the position vector of the curve C:

$$\mathbf{r}(t) = x(t)\,\mathbf{i} + y(t)\,\mathbf{j} + z(t)\,\mathbf{k},$$

we express the line integral of the vector field $\mathbf{F} = P\,\mathbf{i} + Q\,\mathbf{j} + R\,\mathbf{k}$ along C as:

$$\begin{aligned}
\int_C \mathbf{F}\cdot\mathbf{T}\,ds &= \int_a^b \mathbf{F}(\mathbf{r}(t))\cdot\mathbf{r}'(t)\,dt \\
&= \int_a^b (P\,\mathbf{i} + Q\,\mathbf{j} + R\,\mathbf{k})\cdot(x'(t)\,\mathbf{i} + y'(t)\,\mathbf{j} + z'(t)\,\mathbf{k})\,dt \\
&= \int_a^b \left[P(x(t),y(t),z(t))\,x'(t) + Q(x(t),y(t),z(t))\,y'(t) + R(x(t),y(t),z(t))\,z'(t)\right]\,dt.
\end{aligned}$$

By using the line integrals with respect to x, y, and z:

$$\begin{aligned}
\int_a^b P(x(t),y(t),z(t))\,x'(t)\,dt &= \int_C P\,dx, \\
\int_a^b Q(x(t),y(t),z(t))\,y'(t)\,dt &= \int_C Q\,dy, \\
\int_a^b R(x(t),y(t),z(t))\,z'(t)\,dt &= \int_C R\,dz,
\end{aligned}$$

we have

$$\int_C \mathbf{F}\cdot\mathbf{T}\,ds = \int_C P\,dx + \int_C Q\,dy + \int_C R\,dz = \int_C P\,dx + Q\,dy + R\,dz.$$

The Mass and the Center of Mass of a Wire

Suppose C is a smooth plane wire and $\rho(x,y)$ represents the linear density at a point (x,y) along C. Then the mass of the wire is

$$m = \int_C \rho(x,y)\,ds$$

and the center of mass $(\overline{x}, \overline{y})$ of the wire is given by

$$\overline{x} = \frac{1}{m}\int_C x\rho(x,y)\,ds, \qquad \overline{y} = \frac{1}{m}\int_C y\rho(x,y)\,ds.$$

Justification

The mass of the part of the wire from P_{i-1} to P_i is approximately $\rho(x_i^*, y_i^*)\Delta s_i$ and so the total mass of the wire is approximately $\sum_{i=1}^n \rho(x_i^*, y_i^*)\Delta s_i$. By taking more and more points on the curve, we obtain the mass m of the wire as the limiting value of these approximations:

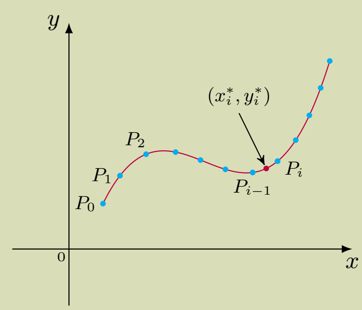
$$m = \lim_{n\to\infty}\sum_{i=1}^n \rho(x_i^*, y_i^*)\Delta s_i = \int_C \rho(x,y)\,ds.$$

Similarly, we can computer the moments about the x- and y-axis

$$M_x = \int_C y\rho(x,y)\,ds, \qquad M_y = \int_C x\rho(x,y)\,ds$$

and obtain the center of mass of the wire, $(\overline{x}, \overline{y})$, where

$$\overline{x} = \frac{M_x}{m} = \frac{1}{m}\int_C x\rho(x,y)\,ds, \qquad \overline{x} = \frac{M_y}{m} = \frac{1}{m}\int_C y\rho(x,y)\,ds.$$



Topological Concepts

(1) A region D in $\mathbb{R}^2$ is **open** if for every point P in D there is a disk with center P that lies entirely in D.
(2) A region D is **connected** if any two points in D can be joined by a path that lies in D.
(3) A curve is a **simple curve** if it doesn't intersect itself anywhere between its endpoints.
(4) A **simply-connected** region in the plane is a connected region D such that every simple closed curve in D encloses only points that are in D.

Fundamental Theorem for Line Integrals

Let C be a smooth curve given by the vector function $\mathbf{r}(t)$, $a \le t \le b$. Let f be a differentiable function of two or three variables whose gradient vector ∇f is continuous on C. Then

$$\int_C \nabla f \cdot d\mathbf{r} = f(\mathbf{r}(b)) - f(\mathbf{r}(a)).$$

Justification

For the given function f,

$$\nabla f = \frac{\partial f}{\partial x}(x,y,z)\,\mathbf{i} + \frac{\partial f}{\partial y}(x,y,z)\,\mathbf{j} + \frac{\partial f}{\partial z}(x,y,z)\,\mathbf{k}.$$

Along the curve C given by the vector function $\mathbf{r}(t)$,

$$\mathbf{r}(t) = x(t)\,\mathbf{i} + y(t)\,\mathbf{j} + z(t)\,\mathbf{k},$$

we have

$$\nabla f(\mathbf{r}(t)) = \frac{\partial f}{\partial x}(x(t),y(t),z(t))\,\mathbf{i} + \frac{\partial f}{\partial y}(x(t),y(t),z(t))\,\mathbf{j} + \frac{\partial f}{\partial z}(x(t),y(t),z(t))\,\mathbf{k},$$

so that

$$\begin{aligned}
&\nabla f(\mathbf{r}(t)) \cdot \mathbf{r}'(t)\,dt \\
&= \left[\frac{\partial f}{\partial x}(x(t),y(t),z(t))\,\mathbf{i} + \frac{\partial f}{\partial y}(x(t),y(t),z(t))\,\mathbf{j} + \frac{\partial f}{\partial z}(x(t),y(t),z(t))\,\mathbf{k}\right] \cdot [x'(t)\,\mathbf{i} + y'(t)\,\mathbf{j} + z'(t)\,\mathbf{k}] \\
&= \frac{\partial f}{\partial x}(x(t),y(t),z(t))\,x'(t) + \frac{\partial f}{\partial y}(x(t),y(t),z(t))\,y'(t) + \frac{\partial f}{\partial z}(x(t),y(t),z(t))\,z'(t) \\
&= \frac{d}{dt}\left[f(x(t),y(t),z(t))\right].
\end{aligned}$$

Thus, by the Fundamental Theorem of Calculus,

$$\begin{aligned}
\int_C \nabla f \cdot d\mathbf{r} &= \int_a^b \nabla f(\mathbf{r}(t)) \cdot \mathbf{r}'(t)\,dt = \int_a^b \frac{d}{dt}\left[f(x(t),y(t),z(t))\right]\,dt \\
&= \left[f(x(t),y(t),z(t))\right]_{t=a}^{t=b} = f(x(b),y(b),z(b)) - f(x(a),y(a),z(a)) \\
&= f(\mathbf{r}(b)) - f(\mathbf{r}(a)).
\end{aligned}$$

Independence of Path

Suppose that $\mathbf{F}$ is a continuous vector field with domain D, we say that the line integral $\int_C \mathbf{F} \cdot d\mathbf{r}$ is

independent of path if

$$\int_{C_1} \mathbf{F} \cdot d\mathbf{r} = \int_{C_2} \mathbf{F} \cdot d\mathbf{r}$$

for any two paths C_1 and C_2 in D that have the same initial and terminal points.

Path Independence and Line Integral Along a Closed Path

The line integral $\int_C \mathbf{F} \cdot d\mathbf{r}$ is independent of path in D if and only if $\int_C \mathbf{F} \cdot d\mathbf{r} = 0$ for every closed path C in D.

Justification

Suppose $\int_C \mathbf{F} \cdot d\mathbf{r} = 0$ for every closed path C in D. Take any two paths C_1 and C_2 from A to B in D and define C to be the curve consisting of C_1 followed by $-C_2$. Then

$$0 = \int_C \mathbf{F} \cdot d\mathbf{r} = \int_{C_1} \mathbf{F} \cdot d\mathbf{r} + \int_{-C_2} \mathbf{F} \cdot d\mathbf{r} = \int_{C_1} \mathbf{F} \cdot d\mathbf{r} - \int_{C_2} \mathbf{F} \cdot d\mathbf{r}.$$

So,

$$\int_{C_1} \mathbf{F} \cdot d\mathbf{r} = \int_{C_2} \mathbf{F} \cdot d\mathbf{r}.$$

Conversely, suppose the line integral $\int_C \mathbf{F} \cdot d\mathbf{r}$ is independent of path in D. For any closed path C in D, we choose any two points A and B on C and regard C as being composed of the path C_1 from A to B followed by the path C_2 from B to A. Then

$$\int_C \mathbf{F} \cdot d\mathbf{r} = \int_{C_1} \mathbf{F} \cdot d\mathbf{r} + \int_{C_2} \mathbf{F} \cdot d\mathbf{r} = \int_{C_1} \mathbf{F} \cdot d\mathbf{r} - \int_{-C_2} \mathbf{F} \cdot d\mathbf{r} = 0,$$

since C_1 and $-C_2$ have the same initial and terminal points.

Path Independence and Conservative Vector Field

If $\mathbf{F}$ is a vector field that is continuous on an open connected region D and if $\int_C \mathbf{F} \cdot d\mathbf{r}$ is independent of path in D, then $\mathbf{F}$ is a conservative vector field on D; that is, there exists a function f such that $\nabla f = F$.

Justification

Assume (a, b) is a point in D. Since $\int_C \mathbf{F} \cdot d\mathbf{r}$ is independent of path, the integral

$$\int_{(a,b)}^{(x,y)} \mathbf{F} \cdot d\mathbf{r}$$

gives the same value no matter which path it takes from (a, b) to (x, y). Thus, the integral defines a function $f(x,y)$ in D. Since D is open, there exists a disk contained in D with center (x,y). Choose any point (x_1,y) in the disk with $x_1 < x$ and let C consist of any path C_1 from (a,b) to (x_1,y) followed by the horizontal line segment C_2 from (x_1,y) to (x,y). Then

$$f(x,y) = \int_{C_1} \mathbf{F}\cdot d\mathbf{r} + \int_{C_2} \mathbf{F}\cdot d\mathbf{r} = \int_{(a,b)}^{(x_1,y)} \mathbf{F}\cdot d\mathbf{r} + \int_{C_2} \mathbf{F}\cdot d\mathbf{r}.$$

Because the first of these integrals does not depend on x, we get

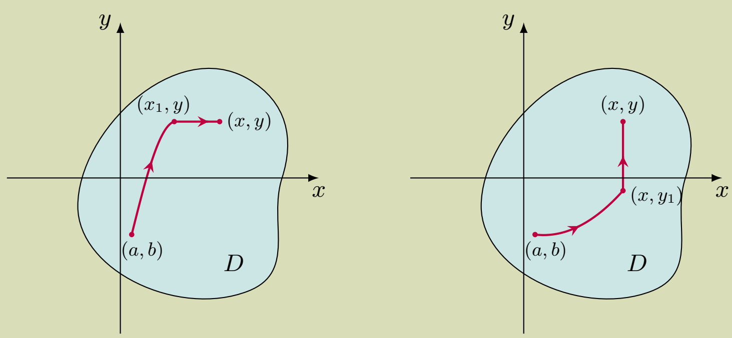
$$f_x(x,y) = 0 + \frac{\partial}{\partial x}\int_{C_2} \mathbf{F}\cdot d\mathbf{r}.$$

By writing $\mathbf{F} = P\,\mathbf{i} + Q\,\mathbf{j}$ and $\mathbf{r} = x\,\mathbf{i} + y\,\mathbf{j}$, we have

$$\int_{C_2} \mathbf{F}\cdot d\mathbf{r} = \int_{C_2} P\,dx + Q\,dy.$$

On C_2, y is constant, so $dy = 0$. Using t as the parameter, where $x_1 \le t \le x$, we have

$$f_x(x,y) = \frac{\partial}{\partial x}\int_{C_2} P\,dx + Q\,dy = \frac{\partial}{\partial x}\int_{x_1}^{x} P(t,y)\,dt = P(x,y).$$



Similarly, by using a path C_1 from (a,b) to (x,y_1) followed by the vertical line segment C_2 from (x,y_1) to (x,y), we can have

$$f_y(x,y) = \frac{\partial}{\partial y}\int_{C_2} P\,dx + Q\,dy = \frac{\partial}{\partial y}\int_{y_1}^{y} Q(x,t)\,dt = Q(x,y).$$

Thus,

$$\mathbf{F} = P\,\mathbf{i} + Q\,\mathbf{j} = f_x(x,y)\,\mathbf{i} + f_y(x,y)\,\mathbf{j} = \nabla f(x,y),$$

which says that $\mathbf{F}$ is conservative.

Plane Conservative Vector Fields

Necessary Condition

If $\mathbf{F}(x,y) = P(x,y)\,\mathbf{i} + Q(x,y)\,\mathbf{j}$ is a conservative vector field, where P and Q have continuous first-order partial derivatives on a domain D, then

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x} \qquad \text{throughout } D.$$

Sufficient Condition

Let $\mathbf{F}(x,y) = P(x,y)\,\mathbf{i} + Q(x,y)\,\mathbf{j}$ be a vector field on an open simply-connected region D. Suppose that P and Q have continuous first-order derivatives and

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x} \qquad \text{throughout } D.$$

Then $\mathbf{F}$ is conservative.

Justification

Necessary Condition: If $\mathbf{F}(x,y) = P(x,y)\,\mathbf{i} + Q(x,y)\,\mathbf{j}$ is a conservative vector field, there is a function f such that $\nabla f(x,y) = \mathbf{F}(x,y)$, that is,

$$\frac{\partial f}{\partial x} = P, \qquad \frac{\partial f}{\partial y} = Q.$$

Since P and Q have continuous first-order derivatives, by Clairaut's Theorem,

$$\frac{\partial P}{\partial y} = \frac{\partial^2 f}{\partial y\,\partial x} = \frac{\partial^2 f}{\partial x\,\partial y} = \frac{\partial Q}{\partial x}.$$

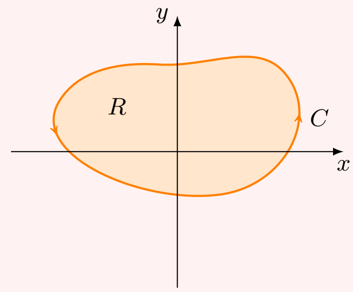
Sufficient Condition: If C is any simple closed path in D and R is the region that C encloses, then Green's Theorem gives

$$\oint_C \mathbf{F}\cdot d\mathbf{r} = \oint_C P\,dx + Q\,dy = \iint_R \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dA.$$

Since $\dfrac{\partial Q}{\partial x} = \dfrac{\partial P}{\partial y}$ throughout D, we get $\oint_C \mathbf{F}\cdot d\mathbf{r} = 0$. We see that $\int_C \mathbf{F}\cdot d\mathbf{r} = 0$ for any closed curve C. Therefore $\int_C \mathbf{F}\cdot d\mathbf{r}$ is independent of path in D. It follows that $\mathbf{F}$ is a conservative vector field.

Positive Orientation

A simple closed curve has **positive orientation** if the region R enclosed by C is on the left when traveling along C.



Green's Theorem

Let C be a positively oriented, piecewise-smooth, simple closed curve in the plane and let D be the region bounded by C. If P and Q have continuous partial derivatives on an open region that contains D, then

$$\int_C P\,dx + Q\,dy = \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dA.$$

Justification - When D Is a Simple Region

We shall prove that

$$\int_C P\,dx = -\iint_D \frac{\partial P}{\partial y}\,dA, \qquad \int_C Q\,dy = \iint_D \frac{\partial Q}{\partial x}\,dA.$$

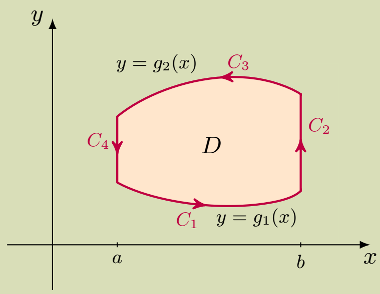
Suppose D can be expressed as a type I region:

$$D = \left\{(x, y) \in \mathbb{R}^2 \,\middle|\, a \le x \le b,\ g_1(x) \le y \le g_2(x)\right\},$$

where g_1 and g_2 are continuous functions. Thus, by applying the Fundamental Theorem of Calculus,

$$\iint_D \frac{\partial P}{\partial y}\,dA = \int_a^b \int_{g_1(x)}^{g_2(x)} \frac{\partial P}{\partial y}(x, y)\,dy\,dx = \int_a^b \left[P(x, g_2(x)) - P(x, g_1(x))\right] dx.$$

As shown in the figure, the closed path $C = C_1 \cup C_2 \cup C_3 \cup C_4$.



C_1 : parametrization $x = x,\ y = g_1(x),\ a \le x \le b$

$$\Longrightarrow \int_{C_1} P(x,y)\,dx = \int_a^b P(x, g_1(x))\,dx$$

C_2 : parametrization $x = b \Longrightarrow dx = 0$

$$\Longrightarrow \int_{C_2} P(x,y)\,dx = 0$$

C_3 : parametrization $x = x,\ y = g_2(x),\ a \le x \le b$

$$\Longrightarrow \int_{C_3} P(x,y)\,dx = \int_b^a P(x, g_2(x))\,dx = -\int_a^b P(x, g_2(x))\,dx$$

C_4 : parametrization $x = a \Longrightarrow dx = 0$

$$\Longrightarrow \int_{C_4} P(x,y)\,dx = 0$$

So, we get

$$\begin{aligned}\int_C P\,dx = \int_{C_1 \cup C_2 \cup C_3 \cup C_4} P\,dx &= \int_{C_1} P\,dx + \int_{C_2} P\,dx + \int_{C_3} P\,dx + \int_{C_4} P\,dx \\ &= \int_a^b P(x, g_1(x))\,dx - \int_a^b P(x, g_2(x))\,dx.\end{aligned}$$

Hence,

$$\int_C P\,dx = -\iint_D \frac{\partial P}{\partial y}\,dA.$$

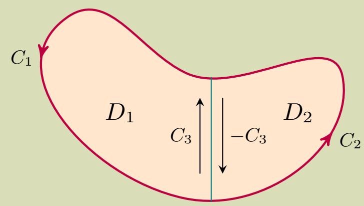
In much the same way by expressing D as a type II region, we can get $\int_C Q\,dy = \iint_D \frac{\partial Q}{\partial x}\,dA$. Adding these two equalities, we obtain Green's Theorem.

Justification - When D Is a Finite Union of Simple Regions

For example, if D is the region shown in the Figure, then we can write $D = D_1 \cup D_2$, where D_1 and D_2 are both simple. The boundary of D_1 is $C_1 \cup C_3$ and the boundary of D_2 is $C_2 \cup (-C_3)$. By applying the Green's Theorem to D_1 and D_2 separately, we get

$$\int_{C_1 \cup C_3} P\,dx + Q\,dy = \iint_{D_1} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dA,$$

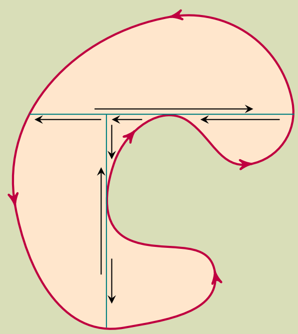
$$\int_{C_2 \cup (-C_3)} P\,dx + Q\,dy = \iint_{D_2} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dA.$$



If we add these two equations, the line integrals along C_3 and $-C_3$ cancel, so we get

$$\int_C P\,dx + Q\,dy = \int_{C_1 \cup C_2} P\,dx + Q\,dy = \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dA.$$

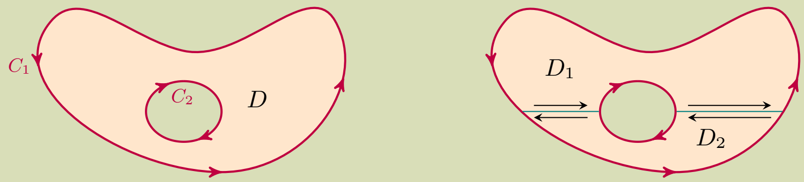
The same sort of argument allows us to establish Green's Theorem for any finite union of nonoverlapping simple regions, as in the following figure.



Justification - When D Contains Holes

Green's Theorem can be extended to apply to regions with holes, that is, regions that are not simply-connected. For example, the boundary C of the region D in the figure consists of two simple closed curves C_1 and C_2. We assume that these boundary curves are oriented so that the region D is always on the left as the curve C is traversed. Thus the positive direction is counterclockwise for the outer curve C_1 but clockwise for the inner curve C_2. If we divide D into two regions D_1 and D_2 by means of the lines shown in the figure and then apply Green's Theorem to each of D_1 and D_2, we get

$$\begin{aligned}\iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dA &= \iint_{D_1} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dA + \iint_{D_2} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dA \\ &= \int_{\partial D_1} P\,dx + Q\,dy + \int_{\partial D_2} P\,dx + Q\,dy.\end{aligned}$$



Since the line integrals along the common boundary lines are in opposite directions, they cancel and we get

$$\iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dA = \int_{C_1} P\,dx + Q\,dy + \int_{C_2} P\,dx + Q\,dy = \int_C P\,dx + Q\,dy,$$

which is Green's Theorem for the region D.

Use Line Integral to Compute Area

Green's Theorem gives the following formulas for the area of D:

$$A = \oint_C x\,dy = -\oint_C y\,dx = \tfrac{1}{2}\oint_C x\,dy - y\,dx.$$

Justification

Green's Theorem gives

$$\iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dA = \int_C P\,dx + Q\,dy.$$

If $\dfrac{\partial Q}{\partial x} - \dfrac{\partial P}{\partial y} = 1$, the double integral, therefore the line integral, gives the area of D. There are several possibilities:

$$\begin{cases} P(x,y) = 0, \\ Q(x,y) = x; \end{cases} \qquad \begin{cases} P(x,y) = -y, \\ Q(x,y) = 0; \end{cases} \qquad \begin{cases} P(x,y) = -\frac{1}{2}y, \\ Q(x,y) = \frac{1}{2}x. \end{cases}$$

These give the formulas for the area of D:

$$A = \oint_C x\,dy = -\oint_C y\,dx = \tfrac{1}{2}\oint_C x\,dy - y\,dx.$$

Curl

If $\mathbf{F} = P\,\mathbf{i} + Q\,\mathbf{j} + R\,\mathbf{k}$ is a vector field on $\mathbb{R}^3$ and the partial derivatives of P, Q, and R all exist, then the **curl of F** is the vector field on $\mathbb{R}^3$ defined by

$$\operatorname{curl}\mathbf{F} = \nabla\times\mathbf{F} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ P & Q & R \end{vmatrix}$$
$$= \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right)\mathbf{i} + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right)\mathbf{j} + \left(\frac{\partial Q}{\partial X} - \frac{\partial P}{\partial Y}\right)\mathbf{k}.$$

When $\mathbf{F}$ represents the velocity field in fluid flow, particles near (x, y, z) in the fluid tend to rotate about the axis that points in the direction of $\operatorname{curl}\mathbf{F}(x, y, z)$, and the length of this curl vector is a measure of how quickly the particles move around the axis. If $\operatorname{curl}\mathbf{F} = \mathbf{0}$ at a point P, then the fluid is free from rotations at P and $\mathbf{F}$ is called **irrotational** at P.

Conservative Vector Field in $\mathbb{R}^3$

If f is a function of three variables that has continuous second-order partial derivatives, then

$$\operatorname{curl}(\nabla f) = \mathbf{0}.$$

Conversely, if $\mathbf{F}$ is a vector field defined on all of $\mathbb{R}^3$ whose component functions have continuous partial derivatives and $\operatorname{curl}\mathbf{F} = \mathbf{0}$, then $\mathbf{F}$ is a conservative vector field.

Justification

If f is a function of three variables that has continuous second-order partial derivatives, then

$$\operatorname{curl}(\nabla f) = \nabla\times(\nabla f) = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} & \dfrac{\partial f}{\partial z} \end{vmatrix}$$
$$= \left(\frac{\partial^2 f}{\partial y\,\partial z} - \frac{\partial^2 f}{\partial z\,\partial y}\right)\mathbf{i} + \left(\frac{\partial^2 f}{\partial z\,\partial x} - \frac{\partial^2 f}{\partial x\,\partial z}\right)\mathbf{j} + \left(\frac{\partial^2 f}{\partial x\,\partial y} - \frac{\partial^2 f}{\partial y\,\partial x}\right)\mathbf{k}$$
$$= 0\,\mathbf{i} + 0\,\mathbf{j} + 0\,\mathbf{k} = \mathbf{0},$$

by Clairaut's Theorem.
Suppose $\operatorname{curl}\mathbf{F} = \mathbf{0}$ on $\mathbb{R}^3$. For every closed path C, let S be an orientable surface whose boundary is C. Then Stokes' Theorem gives

$$\int_C \mathbf{F}\cdot d\mathbf{r} = \iint_S \operatorname{curl}\mathbf{F}\cdot d\mathbf{S} = \iint_S \mathbf{0}\cdot d\mathbf{S} = 0.$$

This implies that $\mathbf{F}$ is conservative.

Divergence

If $\mathbf{F} = P\,\mathbf{i} + Q\,\mathbf{j} + R\,\mathbf{k}$ is a vector field on $\mathbb{R}^3$ and $\partial P/\partial x$, $\partial Q/\partial y$, and $\partial R/\partial z$ exist, then the **divergence of F** is the function of three variables defined by

$$\operatorname{div}\mathbf{F} = \nabla \cdot \mathbf{F} = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}.$$

If $\mathbf{F}$ is the velocity of a fluid (or gas), then $\operatorname{div}\mathbf{F}(x, y, z)$ represents the net rate of change (with respect to time) of the mass of fluid (or gas) flowing from the point (x, y, z) per unit volume. In other words, $\operatorname{div}\mathbf{F}(x, y, z)$ measures the tendency of the fluid to diverge from the point (x, y, z). If $\operatorname{div}\mathbf{F} = 0$, then $\mathbf{F}$ is said to be **incompressible**.

Divergence of the Curl

If $\mathbf{F} = P\,\mathbf{i} + Q\,\mathbf{j} + R\,\mathbf{k}$ is a vector field on $\mathbb{R}^3$ and P, Q, and R have continuous second-order partial derivatives, then

$$\operatorname{div}\operatorname{curl}\mathbf{F} = 0.$$

Justification

Using the definitions of divergence and curl, we have

$$\begin{aligned}
\operatorname{div}\operatorname{curl}\mathbf{F} &= \nabla \cdot (\nabla \times \mathbf{F}) \\
&= \frac{\partial}{\partial x}\left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right) + \frac{\partial}{\partial y}\left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right) + \frac{\partial}{\partial z}\left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) \\
&= \frac{\partial^2 R}{\partial x\,\partial y} - \frac{\partial^2 Q}{\partial x\,\partial z} + \frac{\partial^2 P}{\partial y\,\partial z} - \frac{\partial^2 R}{\partial y\,\partial x} + \frac{\partial^2 Q}{\partial z\,\partial x} - \frac{\partial^2 P}{\partial z\,\partial y} \\
&= 0,
\end{aligned}$$

by Clairaut's Theorem.

Vector Forms of Green's Theorem

Suppose that the plane region D, its boundary curve C, and the functions P and Q satisfy the hypotheses of Green's Theorem. Then we can rewrite the equation in Green's Theorem in the vector form

$$\oint_C \mathbf{F} \cdot d\mathbf{r} = \iint_D (\operatorname{curl}\mathbf{F}) \cdot \mathbf{k}\, dA,$$

where $\mathbf{r} = x\,\mathbf{i} + y\,\mathbf{j}$ is the position vector. Or equivalently,

$$\oint_C \mathbf{F} \cdot \mathbf{n}\, ds = \iint_D \operatorname{div}\mathbf{F}\, dA,$$

where $\mathbf{n}$ is the outward unit normal vector to C.

Justification - Part 1

Suppose that the plane region D, its boundary curve C, and the functions P and Q satisfy the hypotheses of Green's Theorem. Denote $\mathbf{F} = P\,\mathbf{i} + Q\,\mathbf{j}$. Then the line integral along C can be expressed in its components:

$$\oint_C \mathbf{F} \cdot d\mathbf{r} = \oint_C P\,dx + Q\,dy.$$

We regard $\mathbf{F}$ as a vector field in $\mathbb{R}^3$ with third component 0 and have

$$\operatorname{curl} \mathbf{F} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ P(x,y) & Q(x,y) & 0 \end{vmatrix} = \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) \mathbf{k}.$$

Therefore

$$(\operatorname{curl} \mathbf{F}) \cdot \mathbf{k} = \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) \mathbf{k} \cdot \mathbf{k} = \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}$$

and we can now rewrite the equation in Green's Theorem in the vector form

$$\oint_C \mathbf{F} \cdot d\mathbf{r} = \iint_D (\operatorname{curl} \mathbf{F}) \cdot \mathbf{k}\,dA.$$

This equality means that the line integral of the tangential component of $\mathbf{F}$ along C equals the double integral of the vertical component of $\operatorname{curl} \mathbf{F}$ over the region D enclosed by C.

Justification - Part 2

We now derive the formula involving the normal component of $\mathbf{F}$. If C is given by the vector equation

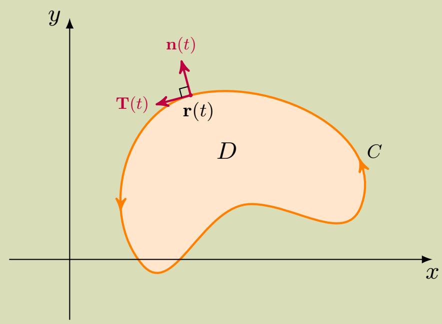
$$\mathbf{r}(t) = x(t)\,\mathbf{i} + y(t)\,\mathbf{j}, \qquad a \le t \le b,$$

then the unit tangent vector is

$$\mathbf{T}(t) = \frac{\mathbf{r}'(t)}{|\mathbf{r}'(t)|} = \frac{x'(t)}{|\mathbf{r}'(t)|}\,\mathbf{i} + \frac{y'(t)}{|\mathbf{r}'(t)|}\,\mathbf{j}.$$

and the outward unit normal vector is

$$\mathbf{n}(t) = \frac{y'(t)}{|\mathbf{r}'(t)|}\,\mathbf{i} - \frac{x'(t)}{|\mathbf{r}'(t)|}\,\mathbf{j}.$$



Thus, by applying Green's Theorem, we have

$$\begin{aligned}\oint_C \mathbf{F}\cdot\mathbf{n}\,ds &= \int_a^b (\mathbf{F}\cdot\mathbf{n})\,(t)\;|\mathbf{r}'(t)|\;dt\\ &= \int_a^b \left[\frac{P(x(t),y(t))\;y'(t)}{|\mathbf{r}'(t)|} - \frac{Q(x(t),y(t))\;x'(t)}{|\mathbf{r}'(t)|}\right]|\mathbf{r}'(t)|\;dt\\ &= \int_a^b P(x(t),y(t))\;y'(t)\,dt - Q(x(t),y(t))\;x'(t)\,dt\\ &= \int_C P\,dy - Q\,dx = \iint_D \left(\frac{\partial P}{\partial x}+\frac{\partial Q}{\partial y}\right)dA.\end{aligned}$$

Since

$$\operatorname{div}\mathbf{F}(x,y) = \frac{\partial P}{\partial x}+\frac{\partial Q}{\partial y}+\frac{\partial 0}{\partial z} = \frac{\partial P}{\partial x}+\frac{\partial Q}{\partial y},$$

we have a second vector form of Green's Theorem:

$$\oint_C \mathbf{F}\cdot\mathbf{n}\,ds = \iint_D \operatorname{div}\mathbf{F}(x,y)\,dA.$$

It says that the line integral of the normal component of $\mathbf{F}$ along C is equal to the double integral of the divergence of $\mathbf{F}$ over the region D enclosed by C.

Parametric Surfaces

Suppose that

$$\mathbf{r}(u,v) = x(u,v)\,\mathbf{i} + y(u,v)\,\mathbf{j} + z(u,v)\,\mathbf{k}$$

is a vector-valued function defined on a region D in the uv-plane. So x, y, and z, the component functions of $\mathbf{r}$, are functions of the two variables u and v with domain D. The set of all points (x,y,z) in $\mathbb{R}^3$ such that

$$x = x(u,v),\quad y = y(u,v),\quad z = z(u,v)$$

and (u,v) varies throughout D, is called a **parametric surface** S and these equations are called **parametric equations** of S. Each choice of u and v gives a point on S; by making all choices, we get

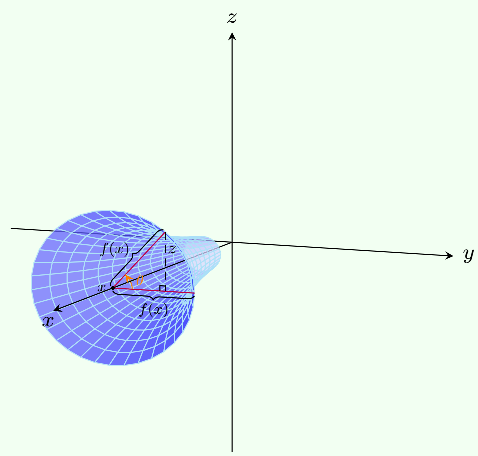
all of S. In other words, the surface S is traced out by the tip of the position vector $\mathbf{r}(u, v)$ as (u, v) moves throughout the region D.
If a parametric surface S is given by a vector function $\mathbf{r}(u, v)$, then there are two useful families of curves that lie on S, one family with u constant and the other with v constant. These families correspond to vertical and horizontal lines in the uv-plane. If we keep u constant by putting $u = u_0$, then $\mathbf{r}(u_0, v)$ becomes a vector function of the single parameter v and defines a curve lying on S.
Similarly, if we keep v constant by putting $v = v_0$, we get a curve given by $\mathbf{r}(u, v_0)$ that lies on S. We call these curves **grid curves**.

Parametric Surface of Revolution

Suppose the surface S is obtained by rotating the curve $y = f(x)$, $a \leq x \leq b$, about the x-axis, where $f(x) \geq 0$. Let θ be the angle of rotation as shown in the figure. If (x, y, z) is a point on S, then the parametric equations of S are

$$x = x, \quad y = f(x)\cos\theta, \quad z = f(x)\sin\theta,$$

where $a \leq x \leq b$ and $0 \leq \theta \leq 2\pi$.



Tangent Plane of Parametric Surfaces

Suppose a parametric surface S is given by a vector function

$$\mathbf{r}(u, v) = x(u, v)\,\mathbf{i} + y(u, v)\,\mathbf{j} + z(u, v)\,\mathbf{k}.$$

To find the tangent plane to the surface S at a point P_0 with position vector $\mathbf{r}(u_0, v_0)$, we consider the grid curves C_1 and C_2, given by $\mathbf{r}(u_0, v)$ and $\mathbf{r}(u, v_0)$ respectively. The tangent vector to C_1 at P_0 is

$$\mathbf{r}_v = \frac{\partial x}{\partial v}(u_0, v_0)\,\mathbf{i} + \frac{\partial y}{\partial v}(u_0, v_0)\,\mathbf{j} + \frac{\partial z}{\partial v}(u_0, v_0)\,\mathbf{k}$$

and the tangent vector to C_2 at P_0 is

$$\mathbf{r}_u = \frac{\partial x}{\partial u}(u_0, v_0)\,\mathbf{i} + \frac{\partial y}{\partial u}(u_0, v_0)\,\mathbf{j} + \frac{\partial z}{\partial u}(u_0, v_0)\,\mathbf{k}.$$

If $\mathbf{r}_u \times \mathbf{r}_v$ is not $\mathbf{0}$, then the surface S is called smooth (it has no "corners"). For a smooth surface, the **tangent plane** is the plane that contains the tangent vectors $\mathbf{r}_u$ and $\mathbf{r}_v$, and the vector $\mathbf{r}_u \times \mathbf{r}_v$ is a normal vector to the tangent plane.

Surface Area

Suppose a parametric surface S is given by a smooth vector function

$$\mathbf{r}(u, v) = x(u, v)\,\mathbf{i} + y(u, v)\,\mathbf{j} + z(u, v)\,\mathbf{k},$$

whose parameter domain D is a rectangle, that is, $\mathbf{r}_u \times \mathbf{r}_v \neq \mathbf{0}$ on D. We divide D into subrectangles R_{ij}. Let (u_i^*, v_j^*) be the lower left corner of R_{ij}. For the patch S_{ij} of the surface S that corresponds to R_{ij}, the point P_{ij} with position vector $\mathbf{r}(u_i^*, v_j^*)$ as one of its corners. Let

$$\mathbf{r}_u^* = \mathbf{r}_u(u_i^*, v_j^*), \qquad \mathbf{r}_v^* = \mathbf{r}_v(u_i^*, v_j^*)$$

be the tangent vectors at P_{ij}, where

$$\mathbf{r}_u = \frac{\partial x}{\partial u}\,\mathbf{i} + \frac{\partial y}{\partial u}\,\mathbf{j} + \frac{\partial z}{\partial u}\,\mathbf{k}, \qquad \mathbf{r}_v = \frac{\partial x}{\partial v}\,\mathbf{i} + \frac{\partial y}{\partial v}\,\mathbf{j} + \frac{\partial z}{\partial v}\,\mathbf{k}.$$

Then the two edges of the patch that meet at P_{ij} can be approximated by the vectors $\Delta u\,\mathbf{r}_u^*$ and $\Delta v\,\mathbf{r}_v^*$. Thus we approximate S_{ij} by the parallelogram determined by the vectors $\mathbf{r}_u(u_i^*, v_j^*)$ and $\mathbf{r}_v(u_i^*, v_j^*)$. The area of this parallelogram is

$$|(\Delta u\,\mathbf{r}_u^*) \times (\Delta v\,\mathbf{r}_v^*)| = |\mathbf{r}_u^* \times \mathbf{r}_v^*|\;\Delta u\,\Delta v,$$

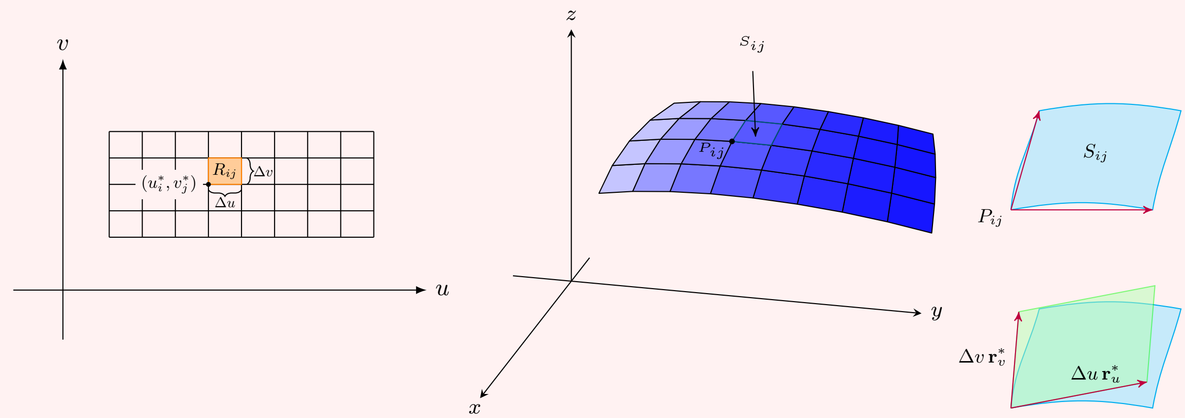
and so an approximation to the area of S is

$$\sum_{i=1}^{m}\sum_{j=1}^{n} \Delta S_{ij} \approx \sum_{i=1}^{m}\sum_{j=1}^{n} |\mathbf{r}_u^* \times \mathbf{r}_v^*|\;\Delta u\,\Delta v.$$

The **surface area** of S is defined by

$$A(S) = \iint_D |\mathbf{r}_u \times \mathbf{r}_v|\,dA = \lim_{m,n\to\infty} \sum_{i=1}^{m}\sum_{j=1}^{n} |\mathbf{r}_u^* \times \mathbf{r}_v^*|\;\Delta u\,\Delta v,$$

provided the last limit exists.



Surface Integral of f over the Surface S

Suppose a parametric surface S is given by a smooth vector function

$$\mathbf{r}(u, v) = x(u, v)\,\mathbf{i} + y(u, v)\,\mathbf{j} + z(u, v)\,\mathbf{k},$$

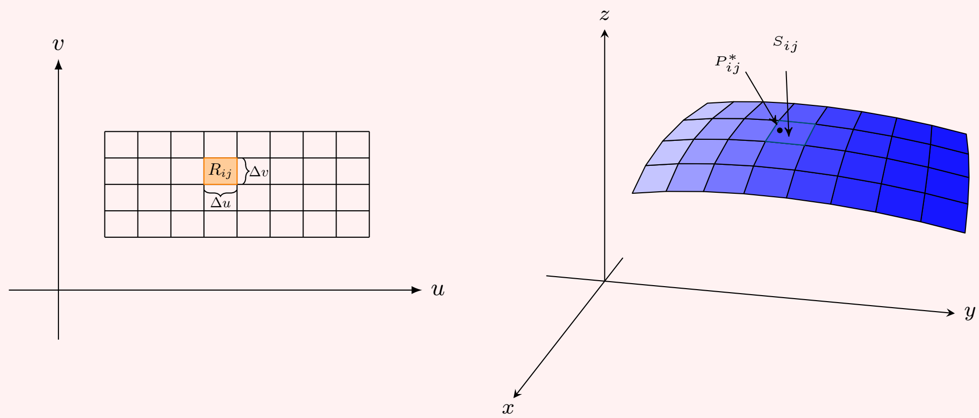
whose parameter domain D is a rectangle, that is, $\mathbf{r}_u \times \mathbf{r}_v \neq \mathbf{0}$ on D. Let f be a function defined on S. As in the definition of the surface area, we form the sum

$$\sum_{i=1}^{m} \sum_{j=1}^{n} f(P_{ij}^*)\Delta S_{ij},$$

where P_{ij}^* is a sample point in the patch S_{ij}. Then the **surface integral f over the surface S** is defined by

$$\iint_S f(x, y, z)\, dS = \lim_{m,n\to\infty} \sum_{i=1}^{m} \sum_{j=1}^{n} f(P_{ij}^*)\Delta S_{ij},$$

provided the last limit exists.



Evaluation of Surface Integrals

$$\iint_S f(x,y,z)\,dS = \iint_D f(\mathbf{r}(u,v))|\mathbf{r}_u \times \mathbf{r}_v|\,dA$$

Justification

For the parametric surface S given by the smooth vector function

$$\mathbf{r}(u,v) = x(u,v)\,\mathbf{i} + y(u,v)\,\mathbf{j} + z(u,v)\,\mathbf{k},$$

we approximate the patch area S_{ij} by the area of an approximating parallelogram in the tangent plane. We know that

$$\Delta S_{ij} \approx |(\Delta u\,\mathbf{r}_u) \times (\Delta v\,\mathbf{r}_v)| = |\mathbf{r}_u \times \mathbf{r}_v|\ \Delta u\,\Delta v,$$

where

$$\mathbf{r}_u = \frac{\partial x}{\partial u}\,\mathbf{i} + \frac{\partial y}{\partial u}\,\mathbf{j} + \frac{\partial z}{\partial u}\,\mathbf{k}, \qquad \mathbf{r}_v = \frac{\partial x}{\partial v}\,\mathbf{i} + \frac{\partial y}{\partial v}\,\mathbf{j} + \frac{\partial z}{\partial v}\,\mathbf{k}$$

are the tangent vectors at a corner of S_{ij}. If the components are continuous and $\mathbf{r}_u$ and $\mathbf{r}_v$ are nonzero and nonparallel in the interior of D, we have

$$\iint_S f(x,y,z)\,dS = \iint_D f(\mathbf{r}(u,v))|\mathbf{r}_u \times \mathbf{r}_v|\,dA.$$

Surface Integral over the Graph of a Function

Suppose that a surface S is given by the equation $z = g(x,y)$, or equivalently, by a vector function

$$\mathbf{r}(x,y) = x\,\mathbf{i} + y\,\mathbf{j} + g(x,y)\,\mathbf{k},$$

where (x,y) lies in D and g has continuous partial derivatives. Then the surface integral of f over the surface S is

$$\iint_S f(x,y,z)\,dS = \iint_D f(x,y,g(x,y))\sqrt{1+\left(\frac{\partial g}{\partial x}\right)^2+\left(\frac{\partial g}{\partial y}\right)^2}\,dA.$$

Similarly, if S is given by the equation $y = h(x,z)$ and D is its projection onto the xz-plane, then

$$\iint_S f(x,y,z)\,dS = \iint_D f(x,h(x,z),z)\sqrt{1+\left(\frac{\partial h}{\partial x}\right)^2+\left(\frac{\partial h}{\partial z}\right)^2}\,dA.$$

In particular, the surface area of the graph given by $z = g(x,y)$, $(x,y) \in D$ is

$$A(S) = \iint_S 1\,dS = \iint_D \sqrt{1+\left(\frac{\partial g}{\partial x}\right)^2+\left(\frac{\partial g}{\partial y}\right)^2}\,dA.$$

Justification

For the vector function

$$\mathbf{r}(x,y) = x\,\mathbf{i} + y\,\mathbf{j} + g(x,y)\,\mathbf{k},$$

we have

$$\begin{aligned}
\mathbf{r}_x &= \mathbf{i} + g_x\,\mathbf{k},\\
\mathbf{r}_y &= \mathbf{j} + g_y\,\mathbf{k},\\
\mathbf{r}_x \times \mathbf{r}_y &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 0 & g_x \\ 0 & 1 & g_y \end{vmatrix} = -g_x\,\mathbf{i} - g_y\,\mathbf{j} + \mathbf{k}.
\end{aligned}$$

Thus,

$$|\mathbf{r}_x \times \mathbf{r}_y| = \sqrt{g_x^2 + g_y^2 + 1}.$$

Thus, for the surface S given by the equation $z = g(x,y)$, we have

$$\begin{aligned}
\iint_S f(x,y,z)\,dS &= \iint_D f(\mathbf{r}(u,v))|\mathbf{r}_u \times \mathbf{r}_v|\,dA\\
&= \iint_D f(x,y,g(x,y))\sqrt{1+\left(\frac{\partial g}{\partial x}\right)^2+\left(\frac{\partial g}{\partial y}\right)^2}\,dA.
\end{aligned}$$

In the same manner, for the surface S given by the equation $y = h(x,z)$ and D is its projection onto the xz-plane, we have

$$\iint_S f(x,y,z)\,dS = \iint_D f(x,h(x,z),z)\sqrt{1+\left(\frac{\partial h}{\partial x}\right)^2+\left(\frac{\partial h}{\partial z}\right)^2}\,dA.$$

For the graph given by $z = g(x, y)$, $(x, y) \in D$, we have $f(x, y, z) = 1$ and get

$$A(S) = \iint_S 1\, dS = \iint_D \sqrt{1 + \left(\frac{\partial g}{\partial x}\right)^2 + \left(\frac{\partial g}{\partial y}\right)^2}\, dA.$$

Mass and the Center of Mass of a Thin Sheet

If a thin sheet (say, of aluminum foil) has the shape of a surface S and the density (mass per unit area) at the point (x, y, z) is $\rho(x, y, z)$, then the total **mass** of the sheet is

$$m = \iint_S \rho(x, y, z)\, dS$$

and the **center of mass** is $(\overline{x}, \overline{y}, \overline{z})$, where

$$\overline{x} = \frac{1}{m}\iint_S x\rho(x, y, z)\, dS, \qquad \overline{y} = \frac{1}{m}\iint_S y\rho(x, y, z)\, dS, \qquad \overline{y} = \frac{1}{m}\iint_S z\rho(x, y, z)\, dS.$$

Oriented Surface

If it is possible to choose a unit normal vector $\mathbf{n}$ at every such point (x, y, z) so that $\mathbf{n}$ varies continuously over S, then S is called an **oriented surface** and the given choice of $\mathbf{n}$ provides S with an **orientation**.

Orientation of a Parametric Surface

If S is a smooth orientable surface given in parametric form by a vector function $\mathbf{r}(u, v)$, then it is automatically supplied with the orientation of one of the unit normal vectors

$$\mathbf{n} = \pm\frac{\mathbf{r}_u \times \mathbf{r}_v}{|\mathbf{r}_u \times \mathbf{r}_v|}.$$

Orientation of a Surface Given by a Graph of a Function

For a surface $z = g(x, y)$, the upward orientation of the surface is given by

$$\mathbf{n} = \frac{-g_x\,\mathbf{i} - g_y\,\mathbf{j} + \mathbf{k}}{\sqrt{1 + g_x^2 + g_y^2}};$$

the downward orientation of the surface is given by

$$\mathbf{n} = \frac{g_x\,\mathbf{i} + g_y\,\mathbf{j} - \mathbf{k}}{\sqrt{1 + g_x^2 + g_y^2}}.$$

Justification

The surface $z = g(x, y)$ is given by the vector function

$$\mathbf{r}(x, y) = x\,\mathbf{i} + y\,\mathbf{j} + g(x, y)\,\mathbf{k},$$

so we have

$$\begin{aligned} \mathbf{r}_x &= \mathbf{i} + g_x\,\mathbf{k}, \\ \mathbf{r}_y &= \mathbf{j} + g_y\,\mathbf{k}, \\ \mathbf{r}_x \times \mathbf{r}_y &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 0 & g_x \\ 0 & 1 & g_y \end{vmatrix} = -g_x\,\mathbf{i} - g_y\,\mathbf{j} + \mathbf{k}. \end{aligned}$$

Since the vector $\mathbf{r}_x \times \mathbf{r}_y$ has positive the component in $\mathbf{k}$, the upward orientation of the surface is given by

$$\mathbf{n} = \frac{\mathbf{r}_x \times \mathbf{r}_y}{|\mathbf{r}_x \times \mathbf{r}_y|} = \frac{-g_x\,\mathbf{i} - g_y\,\mathbf{j} + \mathbf{k}}{\sqrt{g_x^2 + g_y^2 + 1}}.$$

The downward orientation of the surface is given by

$$\mathbf{n} = -\frac{\mathbf{r}_x \times \mathbf{r}_y}{|\mathbf{r}_x \times \mathbf{r}_y|} = \frac{g_x\,\mathbf{i} + g_y\,\mathbf{j} - \mathbf{k}}{\sqrt{g_x^2 + g_y^2 + 1}}.$$

Surface Integral of a Vector Field over an Oriented Surface

If $\mathbf{F}$ is a continuous vector field defined on an oriented surface S with unit normal vector $\mathbf{n}$, then the **surface integral of F over** S is

$$\iint_S \mathbf{F} \cdot d\mathbf{S} = \iint_S \mathbf{F} \cdot \mathbf{n}\, dS.$$

This integral is also called the **flux** of $\mathbf{F}$ across S.

Evaluation of Surface Integrals

If S is given by a vector function $\mathbf{r}(u, v)$, with its orientation given by the unit normal vector $\mathbf{n} = \dfrac{\mathbf{r}_u \times \mathbf{r}_v}{|\mathbf{r}_u \times \mathbf{r}_v|}$, then

$$\iint_S \mathbf{F} \cdot d\mathbf{S} = \iint_D \mathbf{F}(\mathbf{r}(u, v)) \cdot (\mathbf{r}_u \times \mathbf{r}_v)\, dA.$$

In particular, if $\mathbf{F} = P\,\mathbf{i} + Q\,\mathbf{j} + R\,\mathbf{k}$ be a continuous vector field and if S is given by a graph $z = g(x, y)$, then the upward flux is

$$\iint_S \mathbf{F} \cdot d\mathbf{S} = \iint_D (-Pg_x - Qg_y + R)\, dA.$$

Justification

If S is given by a vector function $\mathbf{r}(u,v)$, with its orientation given by the unit normal vector $\mathbf{n} = \dfrac{\mathbf{r}_u \times \mathbf{r}_v}{|\mathbf{r}_u \times \mathbf{r}_v|}$, then

$$\begin{aligned}\iint_S \mathbf{F}\cdot d\mathbf{S} &= \iint_S \mathbf{F}\cdot \frac{\mathbf{r}_u \times \mathbf{r}_v}{|\mathbf{r}_u \times \mathbf{r}_v|}\, dS = \iint_D \mathbf{F}\cdot \frac{\mathbf{r}_u \times \mathbf{r}_v}{|\mathbf{r}_u \times \mathbf{r}_v|}|\mathbf{r}_u \times \mathbf{r}_v|\, dA \\ &= \iint_D \mathbf{F}(\mathbf{r}(u,v))\cdot(\mathbf{r}_u \times \mathbf{r}_v)\, dA.\end{aligned}$$

The surface $z = g(x,y)$ is given by the vector function

$$\mathbf{r}(x,y) = x\,\mathbf{i} + y\,\mathbf{j} + g(x,y)\,\mathbf{k},$$

so we have

$$\begin{aligned}\mathbf{r}_x &= \mathbf{i} + g_x\,\mathbf{k}, \\ \mathbf{r}_y &= \mathbf{j} + g_y\,\mathbf{k}, \\ \mathbf{r}_x \times \mathbf{r}_y &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 0 & g_x \\ 0 & 1 & g_y \end{vmatrix} = -g_x\,\mathbf{i} - g_y\,\mathbf{j} + \mathbf{k}.\end{aligned}$$

Thus, for $\mathbf{F} = P\,\mathbf{i} + Q\,\mathbf{j} + R\,\mathbf{k}$, we have

$$\begin{aligned}\iint_S \mathbf{F}\cdot d\mathbf{S} &= \iint_D \mathbf{F}(\mathbf{r}(x,y))\cdot(\mathbf{r}_x \times \mathbf{r}_y)\, dA \\ &= \iint_D (P\,\mathbf{i} + Q\,\mathbf{j} + R\,\mathbf{k})\cdot(-g_x\,\mathbf{i} - g_y\,\mathbf{j} + \mathbf{k})\, dA \\ &= \iint_D (-Pg_x - Qg_y + R)\, dA.\end{aligned}$$

Electric Flux - application

If $\mathbf{E}$ is an electric field, then the surface integral

$$\iint_S \mathbf{E}\cdot d\mathbf{S}$$

is called the **electric flux of E** through the surface S. Gauss's Law says that the net charge enclosed by a closed surface S is

$$Q = \varepsilon_0 \iint_S \mathbf{E}\cdot d\mathbf{S},$$

where ε_0 is a constant, called the permittivity of free space.

Heat Flux - application

Suppose the temperature at a point (x, y, z) in a body is $u(x, y, z)$. Then the **heat flow** is defined as the vector field

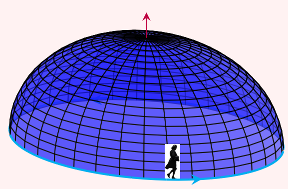
$$\mathbf{F} = -K\nabla u,$$

where K is an experimentally determined constant called the conductivity of the substance. The rate of heat flow across the surface S in the body is then given by the surface integral

$$\iint_S \mathbf{F} \cdot d\mathbf{S} = -K \iint_S \nabla u \cdot d\mathbf{S}.$$

Induced Positive Orientation of Boundary Curve

For an oriented surface with unit normal vector $\mathbf{n}$. The orientation of S induces the positive orientation of the boundary curve C by requiring that if you walk in the positive direction around C with your head pointing in the direction of $\mathbf{n}$, then the surface will always be on your left.



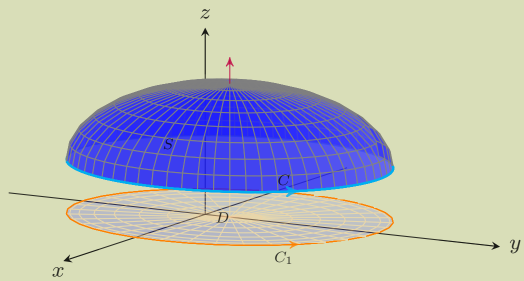
Stokes' Theorem

Let S be an oriented piecewise-smooth surface that is bounded by a simple, closed, piecewise-smooth boundary curve C with positive orientation. Let $\mathbf{F}$ be a vector field whose components have continuous partial derivatives on an open region in $\mathbb{R}^3$ that contains S. Then

$$\int_C \mathbf{F} \cdot d\mathbf{r} = \iint_S \operatorname{curl} \mathbf{F} \cdot d\mathbf{S}.$$

Justification - Part 1

We will show that Stokes' Theorem holds in the case that the surface S is given by $z = g(x, y)$, $(x, y) \in D$, where g has continuous second-order partial derivatives and D is a simple plane region whose boundary curve C_1 corresponds to C. As shown in the figure, suppose the orientation of S is upward, then the positive orientation of C corresponds to the positive orientation of C_1.



The surface S is given by the vector function

$$\mathbf{r}(x,y) = x\,\mathbf{i} + y\,\mathbf{j} + g(x,y)\,\mathbf{k},$$

so we have

$$\begin{aligned}
\mathbf{r}_x &= \mathbf{i} + g_x\,\mathbf{k},\\
\mathbf{r}_y &= \mathbf{j} + g_y\,\mathbf{k},\\
\mathbf{r}_x \times \mathbf{r}_y &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 0 & g_x \\ 0 & 1 & g_y \end{vmatrix} = -g_x\,\mathbf{i} - g_y\,\mathbf{j} + \mathbf{k}.
\end{aligned}$$

Thus, for $\mathbf{F} = P\,\mathbf{i} + Q\,\mathbf{j} + R\,\mathbf{k}$, we have

$$\operatorname{curl}\mathbf{F} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ P & Q & R \end{vmatrix} = (R_y - Q_z)\,\mathbf{i} - (R_x - P_z)\,\mathbf{j} + (Q_x - P_y)\,\mathbf{k}.$$

So,

$$\begin{aligned}
\iint_S \operatorname{curl}\mathbf{F}\cdot d\mathbf{S} &= \iint_D \operatorname{curl}\mathbf{F}(\mathbf{r}(x,y))\cdot(\mathbf{r}_x \times \mathbf{r}_y)\,dA\\
&= \iint_D \left[(R_y - Q_z)(-g_x) - (R_x - P_z)(-g_y) + (Q_x - P_y)\right]\,dA,
\end{aligned}$$

where the partial derivatives of P, Q, and R are evaluated at $(x, y, g(x,y))$.

Justification - Part 2

On the other hand, if $x = x(t)$, $y = y(t)$, $a \le t \le b$, is a parametric representation of C_1, then a parametric representation of C is

$$x = x(t), \quad y = y(t), \quad z = g(x(t), y(t)), \quad a \le t \le b.$$

Thus,

$$\begin{aligned}
\int_C \mathbf{F}\cdot d\mathbf{r} &= \int_a^b \left(P\,\frac{dx}{dt} + Q\,\frac{dy}{dt} + R\,\frac{dz}{dt}\right) dt \\
&= \int_a^b \left[P\,\frac{dx}{dt} + Q\,\frac{dy}{dt} + R\left(\frac{\partial g}{\partial x}\frac{dx}{dt} + \frac{\partial g}{\partial y}\frac{dy}{dt}\right)\right] dt \qquad \text{(chain rule)} \\
&= \int_a^b \left[\left(P + R\,\frac{\partial g}{\partial x}\right)\frac{dx}{dt} + \left(Q + R\,\frac{\partial g}{\partial y}\right)\frac{dy}{dt}\right] dt \\
&= \int_{C_1} \left(P + R\,\frac{\partial g}{\partial x}\right) dx + \left(Q + R\,\frac{\partial g}{\partial y}\right) dy \\
&= \iint_D \left[\frac{\partial}{\partial x}\left(Q + R\,\frac{\partial g}{\partial y}\right) - \frac{\partial}{\partial y}\left(P + R\,\frac{\partial g}{\partial x}\right)\right] dA \qquad \text{(Green's Theorem)}
\end{aligned}$$

We note that P, Q, and R are functions of x, y, and z and that z is itself a function of x and y. By the Chain Rule,

$$\begin{aligned}
&\frac{\partial}{\partial x}\left(Q + R\,\frac{\partial g}{\partial y}\right) - \frac{\partial}{\partial y}\left(P + R\,\frac{\partial g}{\partial x}\right) \\
&= Q_x + Q_z g_x + (R_x + R_z g_x)g_y + Rg_{xy} - [P_y + P_z g_y + (R_y + R_z g_y)g_x - Rg_{xy}] \\
&= Q_x + Q_z g_x + R_x g_y - P_y - P_z g_y - R_y g_x \\
&= (R_y - Q_z)(-g_x) - (R_x - P_z)(-g_y) + (Q_x - P_y).
\end{aligned}$$

Therefore, we have

$$\int_C \mathbf{F}\cdot d\mathbf{r} = \iint_S \operatorname{curl}\mathbf{F}\cdot d\mathbf{S}.$$

Surface Independence

If S_1 and S_2 are oriented surfaces with the same oriented boundary curve C and both satisfy the hypotheses of Stokes' Theorem, then

$$\iint_{S_1} \operatorname{curl}\mathbf{F}\cdot d\mathbf{S} = \int_C \mathbf{F}\cdot d\mathbf{r} = \iint_{S_2} \operatorname{curl}\mathbf{F}\cdot d\mathbf{S}.$$

The Divergence Theorem

Let E be a simple solid region and let S be the boundary surface of E, given with positive (outward) orientation. Let $\mathbf{F}$ be a vector field whose component functions have continuous partial derivatives on an open region that contains E. Then

$$\iint_S \mathbf{F}\cdot d\mathbf{S} = \iiint_E \operatorname{div}\mathbf{F}\cdot dV.$$

Justification - Part 1

Let $\mathbf{F} = P\,\mathbf{i} + Q\,\mathbf{j} + R\,\mathbf{k}$. Then

$$\operatorname{div}\mathbf{F} = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}$$

so

$$\iiint_E \operatorname{div}\mathbf{F}\,dV = \iiint_E \frac{\partial P}{\partial x}\,dV + \iiint_E \frac{\partial Q}{\partial y}\,dV + \iiint_E \frac{\partial R}{\partial z}\,dV.$$

If $\mathbf{n}$ is the unit outward normal of S, then the surface integral on the left side of the Divergence Theorem is

$$\begin{aligned}\iint_S \mathbf{F}\cdot\,d\mathbf{S} &= \iint_S \mathbf{F}\cdot\mathbf{n}\,dS = \iint_S (P\,\mathbf{i} + Q\,\mathbf{j} + R\,\mathbf{k})\cdot\mathbf{n}\,dS \\ &= \iint_S P\,\mathbf{i}\cdot\mathbf{n}\,dS + \iint_S Q\,\mathbf{j}\cdot\mathbf{n}\,dS + \iint_S R\,\mathbf{k}\cdot\mathbf{n}\,dS\end{aligned}$$

Therefore, to prove the Divergence Theorem, it suffices to prove the following three equations:

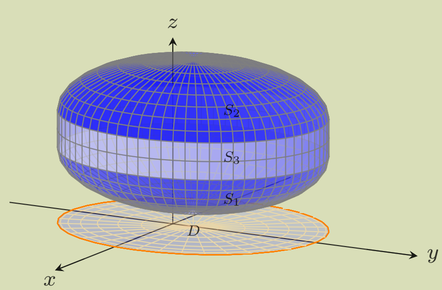
$$\begin{aligned}\iint_S P\,\mathbf{i}\cdot\mathbf{n}\,dS &= \iiint_E \frac{\partial P}{\partial x}\,dV, \\ \iint_S Q\,\mathbf{j}\cdot\mathbf{n}\,dS &= \iiint_E \frac{\partial Q}{\partial y}\,dV, \\ \iint_S R\,\mathbf{k}\cdot\mathbf{n}\,dS &= \iiint_E \frac{\partial R}{\partial z}\,dV.\end{aligned}$$

Justification - Part 2

We shall prove the equation

$$\iint_S R\,\mathbf{k}\cdot\mathbf{n}\,dS = \iiint_E \frac{\partial R}{\partial z}\,dV.$$

The other two equations can be proved similarly.



Suppose E is a region given by

$$E = \left\{(x,y,z) \,\middle|\, (x,y)\in D,\; u_1(x,y) \le z \le u_2(x,y)\right\},$$

where D is the projection of E onto the xy-plane. Then, by the Fundamental Theorem of Calculus,

$$\iiint_E \frac{\partial R}{\partial z}\,dV = \iint_D \left[\int_{u_1(x,y)}^{u_2(x,y)} \frac{\partial R}{\partial z}(x,y,z)\,dz\right] dA$$
$$= \iint_D \left[R(x,y,u_2(x,y)) - R(x,y,u_1(x,y))\right] dA.$$

The boundary surface S consists of three pieces: the bottom surface S_1, the top surface S_2, and possibly a vertical surface S_3, which lies above the boundary curve of D. Notice that on S_3 we have $\mathbf{k}\cdot\mathbf{n} = 0$, because $\mathbf{k}$ is vertical and $\mathbf{n}$ is horizontal, and so

$$\iint_S R\,\mathbf{k}\cdot\mathbf{n}\,dS = \iint_{S_1} R\,\mathbf{k}\cdot\mathbf{n}\,dS + \iint_{S_2} R\,\mathbf{k}\cdot\mathbf{n}\,dS + \iint_{S_3} R\,\mathbf{k}\cdot\mathbf{n}\,dS$$
$$= \iint_{S_1} R\,\mathbf{k}\cdot\mathbf{n}\,dS + \iint_{S_2} R\,\mathbf{k}\cdot\mathbf{n}\,dS.$$

On S_2, $z = u_2(x,y)$, $(x,y) \in D$, and the outward normal $\mathbf{n}$ points upward, so we have

$$\iint_{S_2} R\,\mathbf{k}\cdot\mathbf{n}\,dS = \iint_D R(x,y,u_2(x,y))\,dA.$$

On S_1, $z = u_1(x,y)$, $(x,y) \in D$, and the outward normal $\mathbf{n}$ points downward, so we have

$$\iint_{S_1} R\,\mathbf{k}\cdot\mathbf{n}\,dS = -\iint_D R(x,y,u_1(x,y))\,dA.$$

Hence,

$$\iint_S R\,\mathbf{k}\cdot\mathbf{n}\,dS = \iint_D R(x,y,u_2(x,y))\,dA - \iint_D R(x,y,u_1(x,y))\,dA = \iiint_E \frac{\partial R}{\partial z}\,dV.$$

Divergence Viewed as Flux Density

$$\operatorname{div}\mathbf{F}(P_0) = \lim_{a\to 0} \frac{1}{V(B_a)} \iint_{S_a} \mathbf{F}\cdot d\mathbf{S},$$

where B_a is a ball with center P_0 and very small radius a.
Thus, If $\operatorname{div}\mathbf{F}(P) > 0$, the net flow is outward near P and P is called a **source**. If $\operatorname{div}\mathbf{F}(P) < 0$, the net flow is inward near P and P is called a **sink**.

Justification

Let $\mathbf{v}(x,y,z)$ be the velocity field of a fluid with constant density. Then $\mathbf{F} = \rho\mathbf{v}$ is the rate of flow per unit area. Since B_a is a ball with center P_0 and very small radius a $\operatorname{div}\mathbf{F}(P) \approx \operatorname{div}\mathbf{F}(P_0)$ for all points in B_a since $\operatorname{div}\mathbf{F}$ is continuous. We approximate the flux over the boundary sphere S_a as follows:

$$\iint_{S_a} \mathbf{F}\cdot d\mathbf{S} = \iiint_{B_a} \operatorname{div}\mathbf{F}\, dV \approx \iiint_{B_a} \operatorname{div}\mathbf{F}(P_0)\, dV = \operatorname{div}\mathbf{F}(P_0)\, V(B_a),$$

which yields the result:

$$\operatorname{div}\mathbf{F}(P_0) = \lim_{a\to 0} \frac{1}{V(B_a)} \iint_{S_a} \mathbf{F}\cdot d\mathbf{S}.$$

Chapter 17

Second-Order Differential Equations

Second-Order Linear Differential Equation

A **second-order linear differential equation** has the form

$$P(x)\frac{d^2y}{dx^2} + Q(x)\frac{dy}{dx} + R(x)y = G(x),$$

where P, Q, R, and G are continuous functions. When $G = 0$, such equations are said to be **homogeneous**. Otherwise, they are **nonhomogeneous**.

Linear Combination of Solutions for Homogeneous Equations

If $y_1(x)$ and $y_2(x)$ are both solutions of the linear homogeneous equation

$$P(x)\frac{d^2y}{dx^2} + Q(x)\frac{dy}{dx} + R(x)y = 0,$$

and c_1 and c_2 are any constants, then the function

$$y(x) = c_1\, y_1(x) + c_2\, y_2(x)$$

is also a solution of the equation.

Linear Independence

Suppose that $y_1(x)$ and $y_2(x)$ are functions. If neither y_1 nor y_2 is a constant multiple of the other, then we say that y_1 and y_2 are **linearly independent**.

Justification

Since y_1 and y_2 are solutions, we have

$$P(x)y_1'' + Q(x)y_1' + R(x)y_1 = 0,$$
$$P(x)y_2'' + Q(x)y_2' + R(x)y_2 = 0.$$

Thus, for $y = c_1y_1 + c_2y_2$, we have

$$\begin{aligned} P(x)y'' + Q(x)y' + R(x)y &= P(x)(c_1y_1 + c_2y_2)'' + Q(x)(c_1y_1 + c_2y_2)' + R(x)(c_1y_1 + c_2y_2) \\ &= P(x)(c_1y_1'' + c_2y_2'') + Q(x)(c_1y_1' + c_2y_2') + R(x)(c_1y_1 + c_2y_2) \\ &= c_1\left[P(x)y_1'' + Q(x)y_1' + R(x)y_1\right] + c_2\left[P(x)y_2'' + Q(x)y_2' + R(x)y_2\right] \\ &= c_1(0) + c_2(0) = 0. \end{aligned}$$

General Solution

If y_1 and y_2 are linearly independent solutions of the linear homogeneous equation

$$P(x)\frac{d^2y}{dx^2} + Q(x)\frac{dy}{dx} + R(x)y = 0$$

on an interval, and $P(x)$ is never 0, then the **general solution** is given by

$$y(x) = c_1\, y_1(x) + c_2\, y_2(x),$$

where c_1 and c_2 are arbitrary constants.

Initial-Value and Boundary-Value Problems

An **initial-value problem** for the second-order equation

$$P(x)\frac{d^2y}{dx^2} + Q(x)\frac{dy}{dx} + R(x)y = G(x),$$

or

$$P(x)\frac{d^2y}{dx^2} + Q(x)\frac{dy}{dx} + R(x)y = 0,$$

consists of finding a solution y of the differential equation that also satisfies initial conditions of the form

$$y(x_0) = y_0, \qquad y'(x_0) = y_1,$$

where y_0 and y_1 are given constants.
An **boundary-value problem** consists of finding a solution y of the differential equation that also satisfies boundary conditions of the form

$$y(x_0) = y_0, \qquad y(x_1) = y_1.$$

Second-Order Homogeneous Linear Equation with Constant Coefficients

A **second-order homogeneous linear differential equation with constant coefficients** has the form

$$ay'' + by' + cy = 0,$$

where a, b, and c are constants and $a \neq 0$. The equation induces the following **auxiliary equation** (or **characteristic equation**):

$$ar^2 + br + c = 0.$$

General Solution: when the auxiliary equation has two real distinct roots

If the roots r_1 and r_2 of the auxiliary equation $ar^2 + br + c = 0$ are real and unequal, then the general solution of $ay'' + by' + cy = 0$ is

$$y = c_1 e^{r_1 x} + c_2 e^{r_2 x}.$$

General Solution: when the auxiliary equation has one repeated real root

If the auxiliary equation $ar^2 + br + c = 0$ has only one real root r, then the general solution of $ay'' + by' + cy = 0$ is

$$y = c_1 e^{rx} + c_2 x e^{rx}.$$

General Solution: when the auxiliary equation has one pair of conjugate complex roots

If the roots of the auxiliary equation $ar^2 + br + c = 0$ are the complex numbers $r = \alpha \pm i\,\beta$, then the general solution of $ay'' + by' + cy = 0$ is

$$y = e^{\alpha x}\left(c_1 \cos \beta x + c_2 \sin \beta x\right).$$

Justification

Case 1: when the auxiliary equation has two real distinct roots. This corresponds to that $b^2 - 4ac > 0$. If r is a real root of the auxiliary equation, i.e., $ar^2 + br + c = 0$, then

$$a(e^{rx})'' + b(e^{rx})' + c(e^{rx}) = ar^2 e^{rx} + bre^{rx} + ce^{rx} = e^{rx}(ar^2 + br + c) = e^{rx}(0) = 0.$$

We see that e^{rx} is a solution of the differentia equation $ay'' + by' + cy = 0$. Thus, if r_1 and r_2 are two distinct real roots of the auxiliary equation, we have two solutions $y_1 = e^{r_1 x}$ and $y_2 = e^{r_2 x}$ of the differentia equation. Because neither y_1 nor y_2 is a constant multiple of the other, we know that the general solution is given by

$$y(x) = c_1\, y_1(x) + c_2\, y_2(x) = c_1 e^{r_1 x} + c_2 e^{r_2 x}.$$

Case 2: when the auxiliary equation has one repeated real root. This corresponds to that $b^2 - 4ac = 0$.

Let's denote by r the common value of r_1 and r_2. Since the roots are repeated, we have

$$r = -\frac{b}{2a} \qquad \text{or} \qquad 2ar + b = 0.$$

From Case 1, we know that $y_1 = e^{rx}$ is one solution of the differential equation. We now verify that $y_2 = xe^{rx}$ is also a solution:

$$\begin{aligned} ay_2'' + by_2' + cy_2 &= a(xe^{rx})'' + b(xe^{rx})' + c(xe^{rx}) \\ &= a(2re^{rx} + r^2xe^{rx}) + b(e^{rx} + rxe^{rx}) + c(xe^{rx}) \\ &= (2ar + b)e^{rx} + (ar^2 + br + c)xe^{rx} = 0(e^{rx}) + 0(xe^{rx}) = 0. \end{aligned}$$

Since neither $y_1 = e^{rx}$ nor $y_2 = xe^{rx}$ is a constant multiple of the other, we know that they are linearly independent solutions. Hence, the general general solution is given by

$$y(x) = c_1\, y_1(x) + c_2\, y_2(x) = c_1 e^{r_1 x} + c_2 x e^{r_2 x}.$$

Case 3: when the auxiliary equation has one pair of conjugate complex roots. This corresponds to that $b^2 - 4ac < 0$.
We write

$$r_1 = \alpha + i\beta, \qquad r_2 = \alpha - i\beta,$$

where α and β are real numbers, with $\beta \neq 0$. Using Euler's equation $e^{i\theta} = \cos\theta + i\sin\theta$, we write the solution of the differential equation as

$$\begin{aligned} y &= C_1 e^{r_1 x} + C_2 e^{r_2 x} = C_1 e^{(\alpha + i\beta)x} + C_2 e^{(\alpha - i\beta)x} \\ &= C_1 e^{\alpha x}(\cos\beta x + i\sin\beta x) + C_2 e^{\alpha x}(\cos\beta x - i\sin\beta x) \\ &= e^{\alpha x}\left[(C_1 + C_2)\cos\beta x + i(C_1 - C_2)\sin\beta x\right] \\ &= e^{\alpha x}\left(c_1\cos\beta x + c_2\sin\beta x\right), \end{aligned}$$

where $c_1 = C_1 + C_2$, $c_2 = i(C_1 - C_2)$. This gives all solutions (real or complex) of the differential equation. The solutions are real when the constants c_1 and c_2 are real.

Second-Order Nonhomogeneous Linear Equation with Constant Coefficients

A **second-order nonhomogeneous linear differential equation with constant coefficients** has the form

$$ay'' + by' + cy = G(x),$$

where a, b, and c are constants, $a \neq 0$, and G is a continuous function. The related homogeneous equation

$$ay'' + by' + cy = 0$$

is called the **complementary equation**.

General Solution

The general solution of the nonhomogeneous differential equation $ay'' + by' + cy = G(x)$ can be written as

$$y(x) = y_p(x) + y_c(x),$$

where y_p is a particular solution of the nonhomogeneous equation and y_c is the general solution of the complementary equation.

Justification

If y is any solution of the nonhomogeneous equation, then $y - y_p$ is a solution of the complementary equation. Indeed

$$\begin{aligned} a(y-y_p)'' + b(y-y_p)' + c(y-y_p) &= ay'' - ay_p'' + by' - by_p' + cy - cy_p \\ &= (ay'' + by' + cy) - (ay_p'' + by_p' + cy_p) \\ &= G(x) - G(x) = 0. \end{aligned}$$

This shows that every solution is of the form $y(x) = y_p(x) + y_c(x)$.
By substituting $y_p(x)+y_c(x)$ into the nonhomogeneous equation, we can easily verify that every function of this form is a solution.

Method of Undetermined Coefficients

1. If $G(x) = e^{kx}P(x)$, where P is a polynomial of degree n, then try $y_p(x) = e^{kx}Q(x)$, where $Q(x)$ is an nth-degree polynomial (whose coefficients are determined by substituting in the differential equation).

2. If $G(x) = e^{kx}P(x)\cos mx$ or $G(x) = e^{kx}P(x)\sin mx$, where P is an nth-degree polynomial, then try

$$y_p(x) = e^{kx}Q(x)\cos mx + e^{kx}R(x)\sin mx,$$

where Q and R are nth-degree polynomials.

Modification: If any term of y_p is a solution of the complementary equation, multiply y_p by x (or by x^2 if necessary).

Method of Variation of Parameters

1. If $y_1(x)$ and $y_2(x)$ are two linearly independent solutions of the homogeneous equation $ay'' + by' + c = 0$, we may look for a particular solution of the nonhomogeneous equation $ay'' + by' + c = G(x)$ of the form

$$y_p(x) = u_1(x)\,y_1(x) + u_2(x)\,y_2(x).$$

2. If we choose $u_1(x)$ and $u_2(x)$ so that

$$u_1'(x)\,y_1(x) + u_2'(x)\,y_2(x) = 0,$$

then

$$u_1'(x)\,y_1'(x) + u_2'(x)\,y_2'(x) = G(x)/a.$$

3. Solve the linear system in Step 2 for the functions u_1' and u_2' and use them for have a particular solution of the nonhomogeneous equation.

Power Series Solution for Second-Order Linear Equation

Consider a second-order linear differential equation of the form

$$y'' + p(x)y' + q(x)y = r(x).$$

Suppose p, q, and r can be expanded in power series. We may seek the general solution y of the equation in a power series

$$y(x) = \sum_{n=0}^{\infty} c_n x^n.$$

1. Write

$$p(x) = \sum_{n=0}^{\infty} p_n x^n, \qquad q(x) = \sum_{n=0}^{\infty} q_n x^n, \qquad r(x) = \sum_{n=0}^{\infty} r_n x^n,$$

   and substitute them and $y(x) = \sum_{n=0}^{\infty} c_n x^n$ into the differential equation.

2. Differentiate the power series term-by-term and compute the products of power series. Express both sides of the equation as power series.

3. By equating coefficients of like terms on both sides of the equation to determine c_n for $n \geq 2$ in terms of c_0 and c_1. Sometimes this can be done by solving a recursion relation.

Note: The power series method can also be applied to other equations, like $y' + p(x)y = q(x)$.

Chapter 18

Appendixes

Rules for Inequalities

1. If $a < b$, then $a + c < b + c$.
2. If $a < b$ and $c < d$, then $a + c < b + d$.
3. If $a < b$ and $c > 0$, then $ac < bc$.
4. If $a < b$ and $c < 0$, then $ac > bc$.
5. If $0 < a < b$, then $1/a > 1/b$.

Absolute Value

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

Square Root of a Number Squared Equals to the Absolute Value of That Number

For any real number a,

$$\sqrt{a^2} = |a|.$$

Properties of Absolute Values

Suppose a and b are any real numbers and n is an integer. Then

1. $|ab| = |a||b|$.
2. $\left|\dfrac{a}{b}\right| = \dfrac{|a|}{|b|}$ $(b \neq 0)$.
3. $|a^n| = |a|^n$.

4. Suppose $a > 0$. Then $|x| = a \iff x = \pm a$.

5. Suppose $a > 0$. Then $|x| < a \iff -a < x < a$.

6. Suppose $a > 0$. Then $|x| > a \iff x > a$ or $x < -a$.

7. The Triangle Inequality If a and b are any real numbers, then

$$|a + b| \leq |a| + |b|.$$

Distance Formula

The distance between the points $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$ is

$$|P_1 P_2| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

Slope of a Line

The **slope** of a nonvertical line that passes through the points $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$ is

$$m = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}.$$

The slope of a vertical line is not defined.

Straight Line

The equation of every line can be written in the form

$$Ax + By + C = 0.$$

Point-Slope Form of the Equation of a Line

An equation of the line passing through the point $P_1(x_1, y_1)$ and having slope m is

$$y - y_1 = m(x - x_1).$$

Slope-Intercept Form of the Equation of a Line

An equation of the line with slope m and y-intercept b is

$$y = mx + b.$$

Parallel and Perpendicular Lines

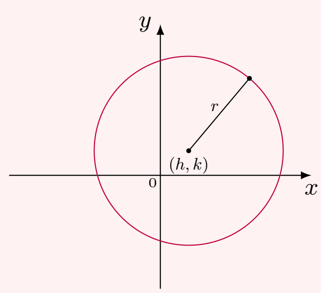
1. Two nonvertical lines are parallel if and only if they have the same slope.
2. Two lines with slopes m_1 and m_2 are perpendicular if and only if $m_1 m_2 = -1$; that is, their slopes are negative reciprocals:

$$m_2 = -\frac{1}{m_1}.$$

Equation of a Circle

An equation of the circle with center (h, k) and radius r is

$$(x-h)^2 + (y-k)^2 = r^2.$$



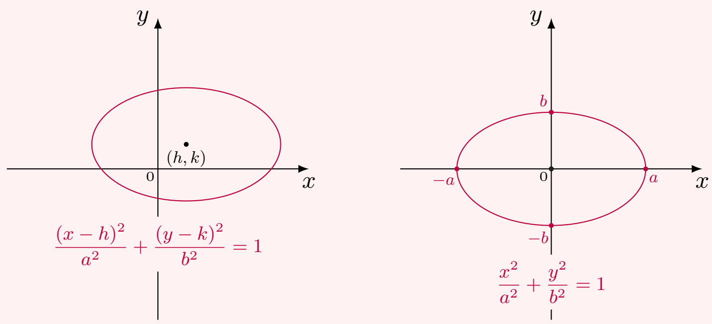
In particular, if the center is the origin $(0, 0)$, the equation is

$$x^2 + y^2 = r^2.$$

Equation of an Ellipse

An equation of the ellipse with center (h, k) is

$$\frac{(x-h)^2}{a^2} + \frac{(y-k)^2}{b^2} = 1.$$



In particular, if the center is the origin $(0,0)$, the ellipse has x-intercepts $\pm a$ and the y-intercepts $\pm b$.

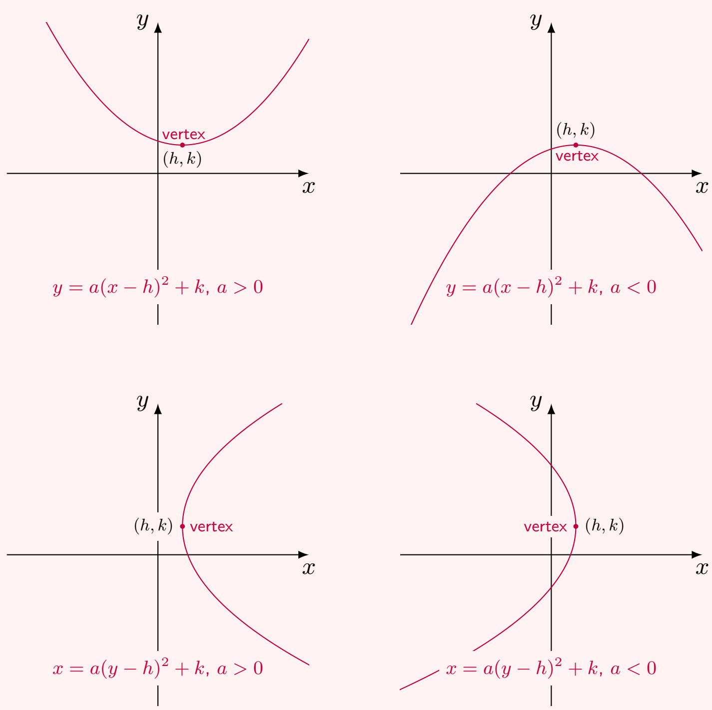
Equation of a Parabola

An equation of the parabola with vertex (h,k) is either

$$y = a(x-h)^2 + k$$

or

$$x = a(y-k)^2 + h.$$



Equation of a Hyperbola

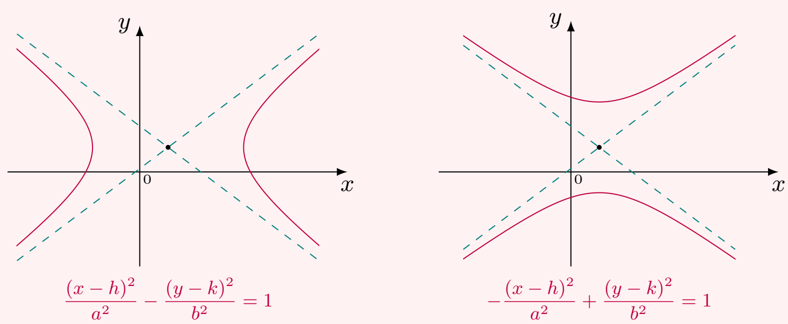
An equation of the hyperbola with center (h, k) is either

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

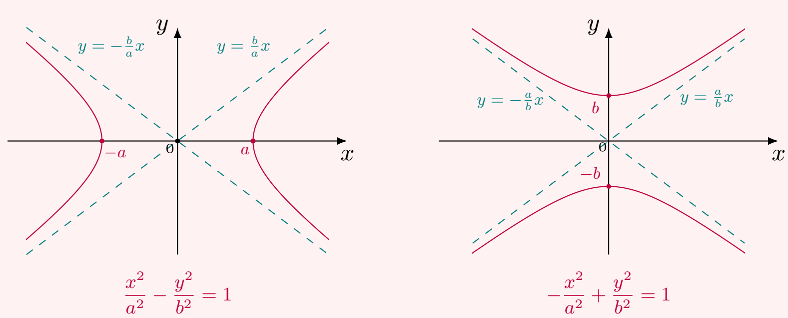
or

$$-\frac{(x-h)^2}{a^2} + \frac{(y-k)^2}{b^2} = 1.$$

It consists of two branches as shown in the following figure.



In particular, if the center is the origin $(0, 0)$, the hyperbola has x-intercepts $\pm a$ or the y-intercepts $\pm b$. The lines $y = (b/a)x$ and $y = -(b/a)x$ are two asymptotes.



Relation Between Degrees and Radians

Angles can be measured in degrees or in radians (abbreviated as rad). The angle given by a complete revolution contains 360°, which is the same as 2π rad. So,

$$\pi \text{ rad} = 180^\circ.$$

Relation Between Angle and the Length of the Arc

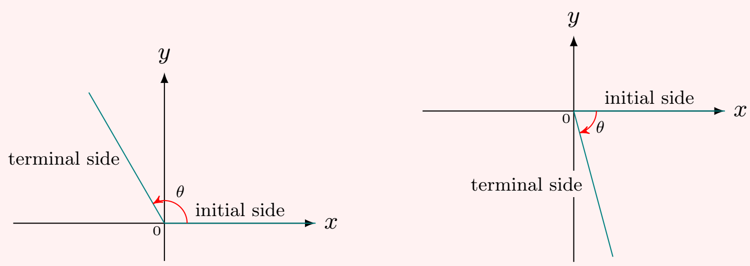
Suppose a sector of a circle with central angle θ and radius r subtending an arc with length a. Since the length of the arc is proportional to the size of the angle, and since the entire circle has circumference $2\pi r$ and central angle 2π, we have

$$a = r\theta,$$

where θ is measured in radians.

Standard Position of an Angle

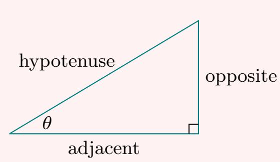
The **standard position** of an angle occurs when we place its vertex at the origin of a coordinate system and its initial side on the positive x-axis as in the left figure. A **positive** angle is obtained by rotating the initial side counterclockwise until it coincides with the terminal side. Likewise, **negative** angles are obtained by clockwise rotation as in the right figure.



$\theta \geq 0$ $\qquad$ $\theta < 0$

Trigonometric Functions

For an acute angle θ the six trigonometric functions are defined as

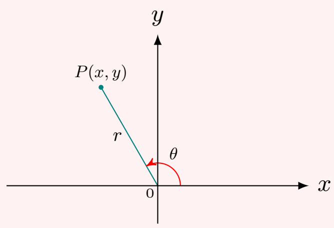


$$\sin\theta = \frac{\text{opp}}{\text{hyp}} \qquad \csc\theta = \frac{\text{hyp}}{\text{opp}}$$

$$\cos\theta = \frac{\text{adj}}{\text{hyp}} \qquad \sec\theta = \frac{\text{hyp}}{\text{adj}}$$

$$\tan\theta = \frac{\text{opp}}{\text{adj}} \qquad \cot\theta = \frac{\text{adj}}{\text{opp}}$$

For a general angle θ in standard position we let $P(x, y)$ be any point on the terminal side of θ and we let r be the distence $|OP|$. Then we define



$$\sin\theta = \frac{y}{r} \qquad \csc\theta = \frac{r}{y}$$

$$\cos\theta = \frac{x}{r} \qquad \sec\theta = \frac{r}{x}$$

$$\tan\theta = \frac{y}{x} \qquad \cot\theta = \frac{x}{y}$$

Since division by 0 is not defined, $\tan\theta$ and $\sec\theta$ are undefined when $x = 0$ and $\csc\theta$ and $\cot\theta$ are undefined when $y = 0$.

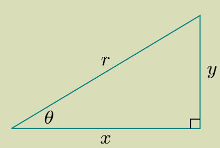
Trigonometric Values of Special Angles

θ	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	$\frac{2\pi}{3}$	$\frac{3\pi}{4}$	$\frac{5\pi}{6}$	π	$\frac{3\pi}{2}$	2π
$\sin\theta$	0	$\frac{1}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{\sqrt{3}}{2}$	1	$\frac{\sqrt{3}}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{2}$	0	-1	0
$\cos\theta$	1	$\frac{\sqrt{3}}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{2}$	0	$-\frac{1}{\sqrt{2}}$	$-\frac{1}{\sqrt{2}}$	$-\frac{\sqrt{3}}{2}$	-1	0	1

Pythagorean Trigonometric Identity

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

Justification



The Pythagorean Theorem says that $x^2 + y^2 = r^2$. By the definitions of the sine and cosine functions, we have

$$\sin\theta = \frac{y}{r}, \qquad \cos\theta = \frac{x}{r}.$$

Thus, we get the identity:

$$\sin^2\theta + \cos^2\theta = \frac{y^2}{r^2} + \frac{x^2}{r^2} = \frac{x^2 + y^2}{r^2} = \frac{r^2}{r^2} = 1.$$

Pythagorean Related Trigonometric Identities

$$\tan^2\theta + 1 = \sec^2\theta$$
$$1 + \cot^2\theta = \csc^2\theta$$

Justification

The Pythagorean Trigonometric Identity gives

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

The first identity is obtained by dividing $\cos^2\theta$, while the second identity by dividing $\sin^2\theta$.

Addition and Subtraction Formulas

The **addition formulas**:

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

The **subtraction formulas**:

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

Addition and Subtraction Formulas for the Tangent Function

$$\tan(x+y) = \frac{\tan x + \tan y}{1 - \tan x\,\tan y}$$
$$\tan(x-y) = \frac{\tan x - \tan y}{1 + \tan x\,\tan y}$$

Justification

By using the addition formulas for the sine and cosine functions, we have

$$\begin{aligned}\tan(x+y) = \frac{\sin(x+y)}{\cos(x+y)} &= \frac{\sin x\,\cos y + \cos x\,\sin y}{\cos x\,\cos y - \sin x\,\sin y}\\ &= \frac{\dfrac{\sin x\,\cos y}{\cos x\,\cos y} + \dfrac{\cos x\,\sin y}{\cos x\,\cos y}}{\dfrac{\cos x\,\cos y}{\cos x\,\cos y} - \dfrac{\sin x\,\sin y}{\cos x\,\cos y}} = \frac{\tan x + \tan y}{1 - \tan x\,\tan y}.\end{aligned}$$

Similarly, by using the subtraction formulas for the sine and cosine functions, we have

$$\begin{aligned}\tan(x-y) &= \frac{\sin(x-y)}{\cos(x-y)} = \frac{\sin x\,\cos y - \cos x\,\sin y}{\cos x\,\cos y + \sin x\,\sin y} \\ &= \frac{\dfrac{\sin x\,\cos y}{\cos x\,\cos y} - \dfrac{\cos x\,\sin y}{\cos x\,\cos y}}{\dfrac{\cos x\,\cos y}{\cos x\,\cos y} + \dfrac{\sin x\,\sin y}{\cos x\,\cos y}} = \frac{\tan x - \tan y}{1 + \tan x\,\tan y}.\end{aligned}$$

Double-Angle Formulas

$$\begin{aligned}\sin 2x &= 2\sin x\,\cos x \\ \cos 2x &= \cos^2 x - \sin^2 x\end{aligned}$$

Justification

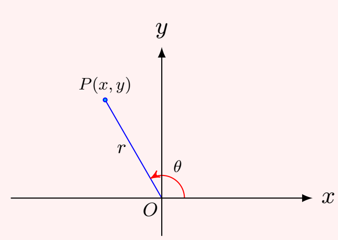
By the addition formulas,

$$\begin{aligned}\sin 2x &= \sin(x+x) = \sin x\,\cos x + \cos x\,\sin x = 2\sin x\,\cos x, \\ \cos 2x &= \cos(x+x) = \cos x\,\cos x - \sin x\,\sin x = \cos^2 x - \sin^2 x.\end{aligned}$$

Trigonometric Functions

For an angle θ, with $\theta \in [0, 2\pi)$, if we let $P(x, y)$ be the point on the terminal side of θ and if let r be the distance $|OP|$, the values of the sine and the cosine are defined as

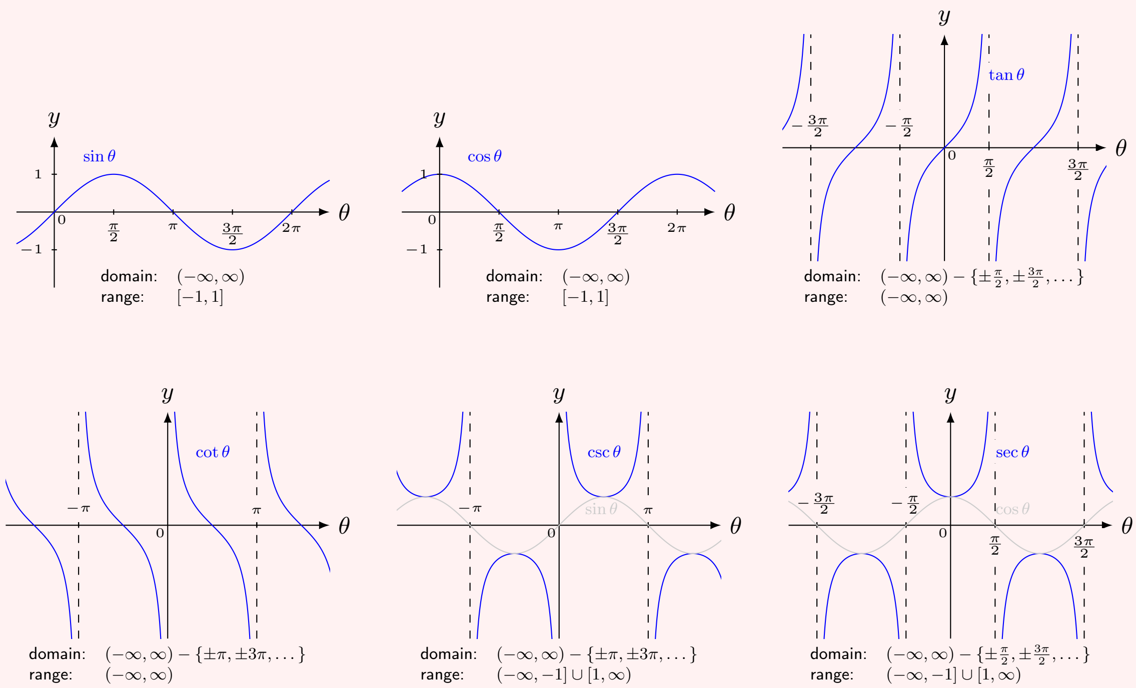
$$\sin\theta = \frac{y}{r}, \qquad \cos\theta = \frac{x}{r}.$$



For θ outside $[0, 2\pi)$, the values of the sine and the cosine are defined in terms of the following periodic property:

$$\sin(\theta + 2\pi) = \sin\theta, \qquad \cos(\theta + 2\pi) = \cos\theta.$$

The tangent function is defined by $\tan\theta = \dfrac{\sin\theta}{\cos\theta}$. The cosecant, secant, and cotangent function are the reciprocals of the sine, cosine, and tangent functions.



Finite Sum

If $a_m, a_{m+1}, \dots, a_n$ are real numbers and m and n are integers such that $m \le n$, then we use the **sigma notation** to denote that

$$\sum_{i=m}^{n} a_i = a_m + a_{m+1} + a_{m+2} + \cdots + a_{n-1} + a_n$$

Properties of Sigma Notation

If c is any constant (that is, it does not depend on i), then

(a) $\displaystyle\sum_{i=m}^{n} ca_i = c\sum_{i=m}^{n} a_i$.

(b) $\displaystyle\sum_{i=m}^{n} (a_i + b_i) = \sum_{i=m}^{n} a_i + \sum_{i=m}^{n} b_i$.

(c) $\displaystyle\sum_{i=m}^{n}(a_i - b_i) = \sum_{i=m}^{n} a_i - \sum_{i=m}^{n} b_i$.

Justification

By the definition of sigma notation,

$$\sum_{i=m}^{n} ca_i = ca_m + ca_{m+1} + \cdots + ca_n = c(a_m + a_{m+1} + \cdots + a_n) = c\sum_{i=m}^{n} a_i,$$

$$\begin{aligned}\sum_{i=m}^{n}(a_i + b_i) &= (a_m + b_m) + (a_{m+1} + b_{m+1}) + \cdots + (a_n + b_n)\\ &= (a_m + a_{m+1} + \cdots + a_n) + (b_m + b_{m+1} + \cdots + b_n) = \sum_{i=m}^{n} a_i + \sum_{i=m}^{n} b_i,\\ \sum_{i=m}^{n}(a_i - b_i) &= (a_m - b_m) + (a_{m+1} - b_{m+1}) + \cdots + (a_n - b_n)\\ &= (a_m + a_{m+1} + \cdots + a_n) - (b_m + b_{m+1} + \cdots + b_n) = \sum_{i=m}^{n} a_i - \sum_{i=m}^{n} b_i.\end{aligned}$$

Basic Sums

Let c be a constant and n a positive integer. Then

(a) $\displaystyle\sum_{i=1}^{n} 1 = n$.

(b) $\displaystyle\sum_{i=1}^{n} c = nc$.

(c) $\displaystyle\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$.

(d) $\displaystyle\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$.

(e) $\displaystyle\sum_{i=1}^{n} i^3 = \left[\frac{n(n+1)}{2}\right]^2$.

Justification

Formulas (a) and (b) can be obtained by using the definition of sigma notation:

$$\sum_{i=1}^{n} 1 = \underbrace{1 + 1 + \cdots + 1}_{n \text{ summands}} = n,$$

$$\sum_{i=1}^{n} c = \underbrace{c + c + \cdots + c}_{n \text{ summands}} = nc.$$

To prove Formula (c), we write the sum $S = \sum_{i=1}^{n} i$ twice, once in the usual order and once in reverse order:

$$\begin{array}{lccccccccc} S = 1 & + & 2 & + & 3 & + & \cdots & + & n, \\ S = n & + & (n-1) & + & (n-2) & + & \cdots & + & 1. \end{array}$$

Adding all columns vertically, we get

$$2S = \underbrace{(n+1) + (n+1) + (n+1) + \cdots + (n+1)}_{n \text{ summands}} = n(n+1)$$

so that

$$S = \frac{n(n+1)}{2}.$$

Formulas (d) and (e) can be proved by mathematical induction. Here we only show how to prove Formula (d), since Formula (e) can be handled similarly.

Let S_n be the given formula: $S_n = \sum_{i=1}^{n} i^2 = \dfrac{n(n+1)(2n+1)}{6}$.

1. S_1 is true because

$$1^2 = \frac{1(1+1)(2 \cdot 1 + 1)}{6}.$$

2. Assume that S_k is true; that is,

$$1^2 + 2^2 + \cdots + k^2 = \frac{k(k+1)(2k+1)}{6}.$$

Then

$$\begin{aligned} 1^2 + 2^2 + \cdots + k^2 + (k+1)^2 &= \left(1^2 + 2^2 + \cdots + k^2\right) + (k+1)^2 \\ &= \frac{k(k+1)(2k+1)}{6} + (k+1)^2 = (k+1)\frac{k(2k+1) + (k+1)}{6} \\ &= (k+1)\frac{(k+2)(2k+3)}{6} = \frac{(k+1)[(k+1)+1][2(k+1)+1]}{6}. \end{aligned}$$

So S_{k+1} is true.

By the Principle of Mathematical Induction, S_n is true for all n.

Arithmetic Operations of Complex Numbers

The **sum** and **difference** of two complex numbers are defined by adding or subtracting their real parts and their imaginary parts:

$$(a + b\,i) + (c + d\,i) = (a + c) + (b + d)\,i$$
$$(a + b\,i) - (c + d\,i) = (a - c) + (b - d)\,i.$$

The **product** of complex numbers is defined so that the usual commutative and distributive laws hold:

$$\begin{aligned}(a + b\,i)(c + d\,i) &= a(c + d\,i) + (b\,i)(c + d\,i)\\ &= ac + ad\,i + bc\,i + bd\,i^2.\end{aligned}$$

The equality can be implied by using $i^2 = -1$:

$$(a + b\,i)(c + d\,i) = (ac - bd) + (ad + bc)\,i.$$

Complex Conjugate

For the complex number $z = a + b\,i$, its **complex conjugate** is $\overline{z} = a - b\,i$.

(a) $\overline{z + w} = \overline{z} + \overline{w}$
(b) $\overline{zw} = \overline{z}\overline{w}$
(c) $\overline{z^n} = \overline{z}^n$

Justification

Let $z = a + b\,i$ and $w = c + d\,i$.
(a) By the definition,

$$\overline{z + w} = \overline{(a + b\,i) + (c + d\,i)} = \overline{(a + c) + (b + d)\,i} = (a + c) - (b + d)\,i,$$
$$\overline{z} + \overline{w} = \overline{a + b\,i} + \overline{c + d\,i} = (a - b\,i) + (c - d\,i) = (a + c) - (b + d)\,i.$$

So, $\overline{z + w} = \overline{z} + \overline{w}$.
(b) By the definition,

$$\begin{aligned}\overline{zw} &= \overline{(a + b\,i)(c + d\,i)} = \overline{ac + ad\,i + bc\,i + bd(-1)}\\ &= \overline{(ac - bd) + (ad + bc)\,i} = (ac - bd) - (ad + bc)\,i,\\ \overline{z}\,\overline{w} &= \overline{(a + b\,i)} \cdot \overline{(c + d\,i)} = (a - b\,i)(c - d\,i)\\ &= ac - ad\,i - bc\,i + bd(-1) = (ac - bd) - (ad + bc)\,i.\end{aligned}$$

So, $\overline{zw} = \overline{z}\,\overline{w}$.
(c) We prove the equality $\overline{z^n} = \overline{z}^n$ by mathematical induction.

When $n = 1$, we have
$$\overline{z^1} = \overline{z} = \overline{z}^1.$$

So the equality holds when $n = 1$.
Suppose the equality holds when $n = k$, that is, $\overline{z^k} = \overline{z}^k$. Then, by part (b),

$$\begin{aligned}\overline{z^{k+1}} &= \overline{z^k \cdot z} = \overline{z^k}\,\overline{z} \\ &= \overline{z}^k\,\overline{z} = \overline{z}^{k+1}.\end{aligned}$$

Therefore, by mathematical induction, $\overline{z^n} = \overline{z}^n$ for every positive integer n.

Modulus of a Complex Number

The **modulus**, or **absolute value**, $|z|$ of a complex number $z = a + b\,i$ is its distance from the origin:

$$|z| = \sqrt{a^2 + b^2}.$$

Division of Complex Numbers

For any complex number z,
$$|z|^2 = z\overline{z}.$$

This leads to the following way to compute the **division** of complex numbers:

$$\frac{z}{w} = \frac{z\overline{w}}{|w|^2}.$$

Justification

Write $z = a + b\,i$. Then

$$z\overline{z} = (a + b\,i)(a - b\,i) = a^2 - ab\,i + ab\,i - b^2\,i^2 = a^2 + b^2 = |z|^2.$$

Thus,
$$\frac{z}{w} = \frac{z\overline{w}}{w\overline{w}} = \frac{z\overline{w}}{|w|^2}.$$

Square Root of a Complex Number

If c is any positive number, we write
$$\sqrt{-c} = \sqrt{c}\,i.$$

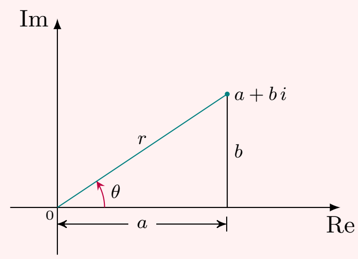
Polar Form of a Complex Number

Any complex number $z = a + b\,i$ can be considered as a point (a, b) and that any such point can be represented by polar coordinates (r, θ) with $r \geq 0$. In fact, as shown in the figure, we have

$$z = a + b\,i = (r\cos\theta) + (r\sin\theta)\,i = r(\cos\theta + i\sin\theta),$$

where

$$r = |z| = \sqrt{a^2 + b^2}, \quad \tan\theta = \frac{b}{a}.$$



The angle θ is called the **argument** of z and we write $\theta = \arg z$. Note that $\arg z$ is not unique; any two arguments of z differ by an integer multiple of 2π.

Multiplication and Division of Complex Numbers in Polar Form

Let

$$z_1 = r_1(\cos\theta_1 + i\sin\theta_1), \quad z_2 = r_2(\cos\theta_2 + i\sin\theta_2).$$

Then

$$z_1 z_2 = r_1 r_2 \left[\cos(\theta_1 + \theta_2) + i\sin(\theta_1 + \theta_2)\right],$$
$$\frac{z_1}{z_2} = \frac{r_1}{r_2}\left[\cos(\theta_1 - \theta_2) + i\sin(\theta_1 - \theta_2)\right].$$

In particular, if $z_1 = 1$ and $z_2 = z = r(\cos\theta + i\sin\theta)$, then

$$\frac{1}{z} = \frac{1}{r}\left[\cos(-\theta) + i\sin(-\theta)\right].$$

Justification

For

$$z_1 = r_1(\cos\theta_1 + i\sin\theta_1), \quad z_2 = r_2(\cos\theta_2 + i\sin\theta_2),$$

we have

$$\begin{aligned} z_1 z_2 &= r_1 r_2 (\cos\theta_1 + i\sin\theta_1)(\cos\theta_2 + i\sin\theta_2) \\ &= r_1 r_2 \left[(\cos\theta_1\,\cos\theta_2 - \sin\theta_1\,\sin\theta_2) + i(\sin\theta_1\,\cos\theta_2 + \cos\theta_1\,\sin\theta_2)\right]. \end{aligned}$$

thus, by the addition formulas for the cosine and sine functions, we have

$$z_1 z_2 = r_1 r_2 \left[\cos(\theta_1 + \theta_2) + i\sin(\theta_1 + \theta_2)\right].$$

If $z_2 \neq 0$, then, by using the subtraction formulas for the sine and cosine functions,

$$\begin{aligned} \frac{z_1}{z_2} = \frac{r_1(\cos\theta_1 + i\sin\theta_1)}{r_2(\cos\theta_2 + i\sin\theta_2)} &= \frac{r_1(\cos\theta_1 + i\sin\theta_1)}{r_2(\cos\theta_2 + i\sin\theta_2)} \cdot \frac{\cos\theta_2 - i\sin\theta_2}{\cos\theta_2 - i\sin\theta_2} \\ &= \frac{r_1}{r_2} \frac{(\cos\theta_1\,\cos\theta_2 + \sin\theta_1\,\sin\theta_2) + i(\sin\theta_1\,\cos\theta_2 - \cos\theta_1\,\sin\theta_2)}{\cos^2\theta_2 + \sin^2\theta_2} \\ &= \frac{r_1}{r_2} \frac{\cos(\theta_1 - \theta_2) + i\sin(\theta_1 - \theta_2)}{1} \\ &= \frac{r_1}{r_2} \left[\cos(\theta_1 - \theta_2) + i\sin(\theta_1 - \theta_2)\right]. \end{aligned}$$

In particular, if $z_1 = 1 = 1(\cos 0 + i\sin 0)$ and $z_2 = z = r(\cos\theta + i\sin\theta)$, then

$$\frac{1}{z} = \frac{1}{r}\left[\cos(-\theta) + i\sin(-\theta)\right].$$

De Moivre's Theorem

If $z = r(\cos\theta + i\sin\theta)$ and n is a positive integer, then

$$z^n = r^n(\cos n\theta + i\sin n\theta).$$

Justification

When $n = 1$, the equality holds trivially.
Suppose the equality holds when $n = k$, that is, $z^k = r^k(\cos k\theta + i\sin k\theta)$. Then, by the addition formulas for the trigonometric functions, we have

$$\begin{aligned} z^{k+1} = z^k \cdot z &= r^k(\cos k\theta + i\sin k\theta) \cdot r(\cos\theta + i\sin\theta) \\ &= r^{k+1}(\cos k\theta\,\cos\theta + \cos k\theta\,\sin\theta\, i + \sin k\theta\,\cos\theta\, i - \sin k\theta\,\sin\theta) \\ &= r^{k+1}\left[(\cos k\theta\,\cos\theta - \sin k\theta\,\sin\theta) + (\cos k\theta\,\sin\theta + \sin k\theta\,\cos\theta)\, i\right] \\ &= r^{k+1}\left[\cos(k+1)\theta + \sin(k+1)\theta\, i\right] \end{aligned}$$

That is, the equality also hold for $n = k + 1$. Therefore, by mathematical induction,

$$z^n = r^n(\cos n\theta + i\sin n\theta)$$

for every positive integer n.

Roots of a Complex Number in Polar Form

If $z = r(\cos\theta + i\sin\theta)$ and n is a positive integer, then z has the n distinct nth roots

$$w_k = r^{1/n}\left[\cos\left(\frac{\theta + 2k\pi}{n}\right) + i\sin\left(\frac{\theta + 2k\pi}{n}\right)\right],$$

where $k = 0, 1, 2, \ldots, n-1$.

Justification

An nth root of the complex number z is a complex number w such that

$$w^n = z.$$

For $z = r(\cos\theta + i\sin\theta)$, we write w in polar form:

$$w = s(\cos\phi + i\sin\phi).$$

By De Moivre's Theorem, we get

$$s^n(\cos n\phi + i\sin n\phi) = r(\cos\theta + i\sin\theta).$$

The equality of these two complex numbers shows that

$$s^n = r \Longrightarrow s = r^{1/n}$$

and

$$\cos n\phi = \cos\theta \qquad \text{and} \qquad \sin n\phi = \sin\theta.$$

From the fact that sine and cosine have period 2π it follows that

$$n\phi = \theta + 2k\pi \qquad \text{or} \qquad \phi = \frac{\theta + 2k\pi}{n}.$$

Thus,

$$w = r^{1/n}\left[\cos\left(\frac{\theta + 2k\pi}{n}\right) + i\sin\left(\frac{\theta + 2k\pi}{n}\right)\right].$$

This expression gives a different value of w for $k = 0, 1, 2, \ldots, n-1$.

Euler's Formula

$$e^{iy} = \cos y + i\sin y.$$

Justification

We extend the Taylor series to complex numbers to have

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!} = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots .$$

Thus,

$$e^{iy} = 1 + iy + \frac{(iy)^2}{2!} + \frac{(iy)^3}{3!} + \frac{(iy)^4}{4!} + \frac{(iy)^5}{5!} + \cdots .$$

Using

$$i^2 = -1, \qquad i^3 = -i, \qquad i^4 = 1, \qquad \ldots$$

we get

$$\begin{aligned} e^{iy} &= 1 + iy - \frac{y^2}{2!} - \frac{iy^3}{3!} + \frac{y^4}{4!} + \frac{iy^5}{5!} + \cdots \\ &= \left(1 - \frac{y^2}{2!} + \frac{y^4}{4!} + \cdots\right) + i\left(y - \frac{y^3}{3} + \frac{y^5}{5} - \cdots\right) \\ &= \cos y + i \sin y. \end{aligned}$$

Here we have used the Taylor series for $\cos y$ and $\sin y$.