

Calculus

Calculus

for scientists and engineers
with matlab

Jishan Hu, Weiping Li, and Yueping Wu

Contents

Preface	v
1 Limit and Continuity	1
1.1 Functions	1
Exponential Functions	4
Trigonometric Functions	5
1.2 Limit	9
Limit Rules	13
One-Sided Limits	16
Limits at Infinities	19
Infinite Limits	25
1.3 Continuous Functions	27
Composite Functions	29
Intermediate Value Theorem	30
1.4 Some Transcendental Functions	33
Exponential Functions	33
Inverse Function	37
Logarithmic Functions	40
Inverse Trigonometric Functions	42
Summary	50
2 Differentiation	55
2.1 Derivatives	55
Relationship Between Differentiability and Continuity	59
2.2 Rules of Differentiation	61
Differentiation of Trigonometric Functions	65
Chain Rule	68
Differentiation of Inverse Functions	71
2.3 Differentiation of Exponential and Logarithmic Functions	75
Logarithmic Differentiation	77

2.4	Implicit Differentiation	80
2.5	Related Rates	83
2.6	Linear Approximation	85
	Differential	87
2.7	Higher Derivatives	90
	Leibniz Formula	93
	Summary	95
3	Applications of Differentiation	97
3.1	Maximum and Minimum Values	97
3.2	Monotone Functions and Mean Value Theorem	107
3.3	Tests for Local Extrema and Concavity	114
	Local Extrema	114
	Concavity	121
3.4	l'Hospital's Rule	132
3.5	Curve Sketching	141
3.6	Parametric Curves	151
	Polar Coordinates and Polar Curves	155
3.7	Optimization	161
3.8	Newton's Method	168
	Summary	180
4	Integration	185
4.1	Areas under Curves	185
	Properties of Definite Integrals	190
	Average Value of a Function	193
4.2	Fundamental Theorem of Calculus	200
4.3	Indefinite Integrals	205
4.4	Substitution Method	210
	Trigonometric Substitutions	215
4.5	Integration by Parts	217
4.6	Some Common Integrals	223
	Trigonometric Integrals	223
	Rational Functions	228
4.7	Numerical Approximation of Integrals	235
4.8	Improper Integrals	247
	Summary	258
5	Applications of Integration	263
5.1	Areas of Regions Between Curves	263
	Area Under Parametric Curve	271
5.2	Volumes	273
	Method of Cylindrical Shells	279
	Solid Generated by Parametric Curve	283
5.3	Arc Lengths	286
	Arc Length of Parametric Curve	290
5.4	Surface Area of Revolution	294
	Surface Area of Revolution by Parametric Curve	300

5.5	Lengths and Areas in Polar Coordinates	305
	Arc Length of Polar Curve	305
	Area of Polar Region	307
5.6	Moments, Centers of Mass, and Centroids	311
	Discrete Systems	311
	Center of Mass Along a Curve	313
	Centroid for Planar Region	320
	Pappus's Theorem	324
5.7	Work	325
	Summary	330
6	Infinite Sequences and Series	335
6.1	Infinite Sequences	335
6.2	Series	348
	Convergence of Series	348
	Divergence Test	356
6.3	Integral Test	357
6.4	Convergence Tests for Nonnegative Series	367
	Comparison Test	367
	Ratio Test and Root Test	370
6.5	Convergence Tests for General Series	376
	Alternating Series	376
	Absolute Convergence and Conditional Convergence	380
	Summary	382
7	Series of Functions	389
7.1	Approximating Functions with Polynomials	389
	Taylor Polynomials	389
7.2	Represent Functions by Power Series	394
	Remainder Formulas	397
7.3	Properties of Power Series	401
	Radius of Convergence and Interval of Convergence	401
	Operations on Power Series	407
7.4	Applications of Taylor Series	418
	Limits by Taylor Series	418
	Approximation of Definite Integrals	419
	Solving Differential Equations	420
	Summary	422
Appendix: Introduction to MATLAB		427
Index		439

Preface

Limit and Continuity

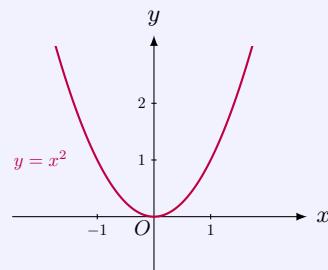
1.1 Functions

1.1 Definition: function

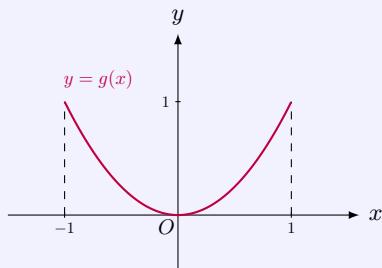
In the calculus of one variable, a **function** $y = f(x)$ is a rule assigning a unique real number y to a real number x in a subset $D \subset \mathbb{R}$. Here x is called the independent variable and y the dependent variable. The set D is called the **domain** of the function and the set R of all y values is called the **range** of the function. The function is denoted by f , while $f(x)$ represents the value y at x , i.e., $y = f(x)$.

1.2 Example: domain of function

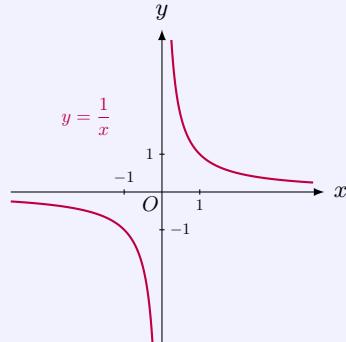
1. The function $f(x) = x^2$ assigns the square x^2 for any number x . The domain D is the whole real line, i.e., $D = (-\infty, +\infty)$. The range R is the interval $[0, +\infty)$.



2. The function $g(x) = x^2$ is defined on the interval $[-1, 1]$. The function g has the same rule as the function f in (i) except the domains are different. The function f has the whole real line as the domain, while g has the closed interval $[-1, 1]$ as its domain.



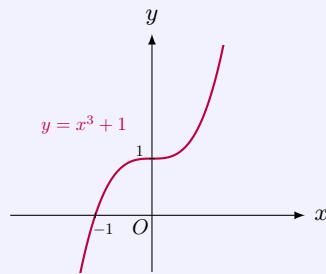
3. The function $h(x) = \frac{1}{x}$ is defined everywhere except at $x = 0$. So the domain D is the whole real line with 0 deleted, i.e., $D = (-\infty, 0) \cup (0, +\infty)$. The range R is the whole real line with 0 deleted, i.e., $R = (-\infty, 0) \cup (0, +\infty)$.



1.3 Example: natural domain of function

The domain of a function sometimes is not given explicitly. In this case, the domain can be found from the expression of the function or from the context. It is called the *natural domain*.

1. The natural domain of the polynomial $f(x) = x^3 + 1$ is the whole real line.

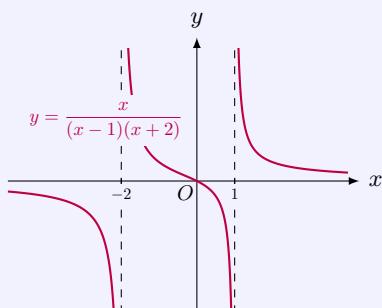


2. The natural domain of the rational function

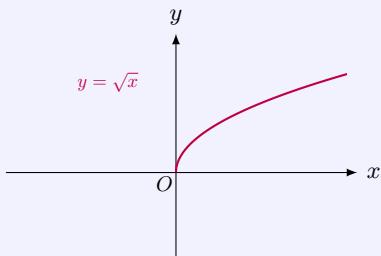
$$g(x) = \frac{x}{(x-1)(x+2)}$$

is the whole real line with -2 and 1 deleted, i.e.,

$$D = (-\infty, -2) \cup (-2, 1) \cup (1, +\infty).$$



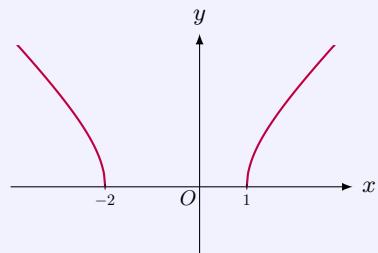
3. The function $h(x) = \sqrt{x}$ is defined for any $x \geq 0$, so the natural domain is $[0, +\infty)$.



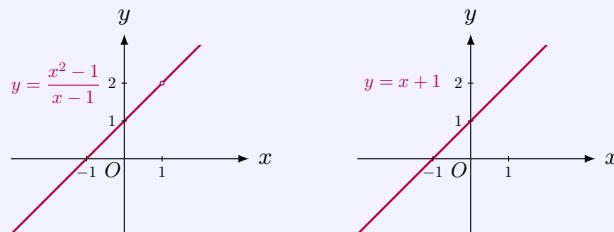
4. The function

$$k(x) = \sqrt{(x-1)(x+2)}$$

is defined for all x such that $(x-1)(x+2) \geq 0$. The inequality implies that $x \geq 1$ or $x \leq -2$. So the natural domain is $D = (-\infty, -2] \cup [1, +\infty)$.



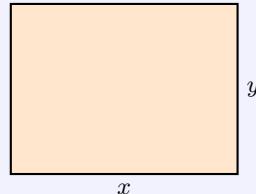
5. The natural domain of $p(x) = \frac{x^2 - 1}{x - 1}$ is $(-\infty, 1) \cup (1, +\infty)$. Note that $x = 1$ is not in the domain of p since $p(x)$ cannot be evaluated at $x = 1$. If we let $q(x) = x + 1$ whose natural domain is $(-\infty, +\infty)$, then $p(x) = q(x)$ when $x \neq 1$. Be alert that p and q are not the same functions, since they have different domains.



1.4 Example: application - natural domain

Consider a rectangle with one side with length x and the other side with length y . Suppose the sum of lengths of four sides is L . Then the area

$$A(x) = x \cdot y = x \cdot \frac{L - 2x}{2}.$$



Since x is the length of one side, we have $0 \leq x \leq \frac{1}{2}L$. So the natural domain of the function A is the closed interval $[0, \frac{1}{2}L]$. When $x = 0$ or $\frac{1}{2}L$, the rectangle is degenerate, i.e., the area of the rectangle is zero.

1.5 Example: transcendental functions

Typical transcendental functions are exponential functions, sine function, cosine function, logarithmic functions. The following are examples of transcendental functions:

$$x^\pi, \quad 2^x, \quad \log_2 x, \quad x^x, \quad x^{1/x}, \quad \sin x.$$

A general transcendental function is a function that does not satisfy a polynomial equation.

Exponential Functions

1.6 Definition: exponentiation and exponential function

For a positive number b (called the *base*) and a real number β (called the *exponent*), the exponentiation b^β defines a real number.

The exponential function is $f(x) = a^x$, where a is a positive number. It has domain \mathbb{R} and range $(0, +\infty)$.

1.7 Theorem: Laws of Exponents

Let a, b be positive numbers and α, β be real numbers. The following laws of exponents hold:

- | | |
|--|---|
| 1. $a^\alpha \cdot a^\beta = a^{\alpha+\beta}.$
2. $\frac{a^\alpha}{a^\beta} = a^{\alpha-\beta}.$
3. $(a^\alpha)^\beta = a^{\alpha \cdot \beta}.$ | 4. $(a \cdot b)^\alpha = a^\alpha \cdot b^\alpha.$
5. $a^{-\alpha} = \frac{1}{a^\alpha}.$
6. $a^1 = a, a^0 = 1.$ |
|--|---|

1.8 Example: application - exponential function

A hot object tends to cool down to the surrounding temperature. Newton expressed this as the following relationship (known as Newton's Law of Cooling):

$$T(t) = T_e + (T_0 - T_e)a^{-t},$$

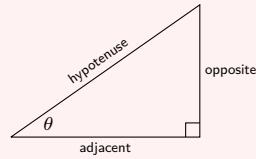
where $T(t)$ is the temperature at a time t , T_0 is the temperature at $t = 0$, T_e is the temperature of the environment around the cooling object, and a is some constant.

Trigonometric Functions

1.9 Definition: trigonometric functions

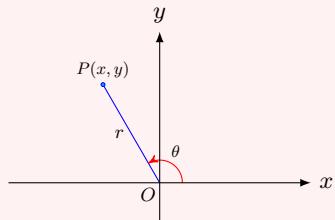
Trigonometric functions of an acute angle are ratios of different pairs of sides of a right-angled triangle, as shown in the figure:

$$\begin{aligned}\sin \theta &= \frac{\text{opposite}}{\text{hypotenuse}}, & \cos \theta &= \frac{\text{adjacent}}{\text{hypotenuse}}, \\ \tan \theta &= \frac{\text{opposite}}{\text{adjacent}}, & \cot \theta &= \frac{\text{adjacent}}{\text{opposite}}, \\ \csc \theta &= \frac{\text{hypotenuse}}{\text{opposite}}, & \sec \theta &= \frac{\text{hypotenuse}}{\text{adjacent}}.\end{aligned}$$



In general, 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

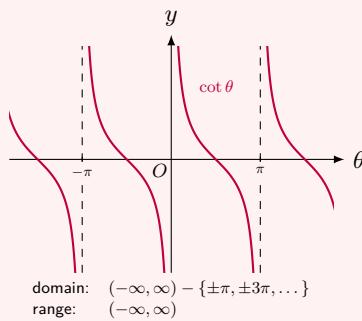
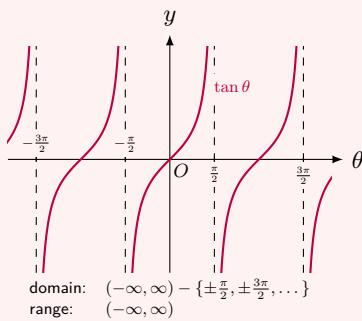
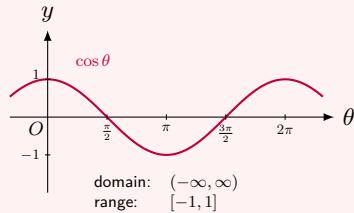
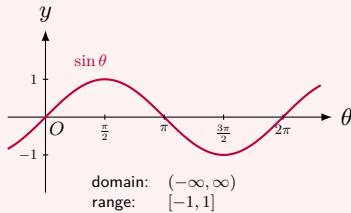
$$\sin \theta = \frac{y}{r}, \quad \cos \theta = \frac{x}{r}.$$

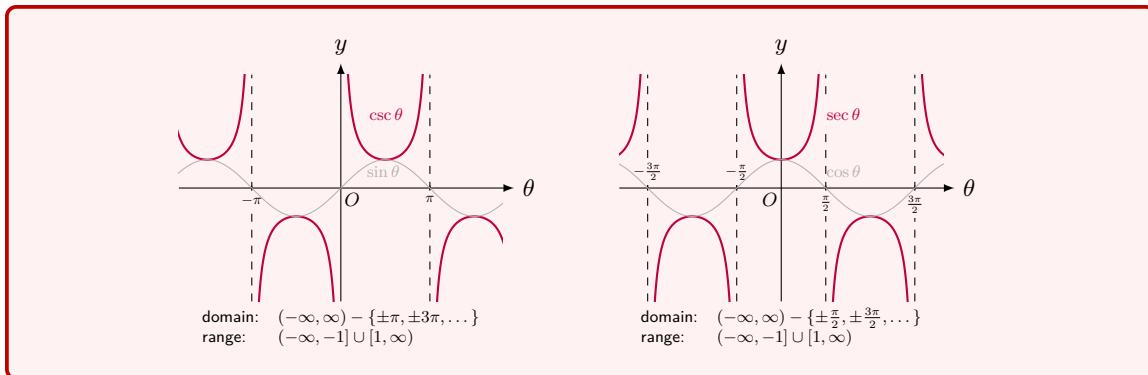


For θ outside $[0, 2\pi)$, the values of the sine and the cosine are defined in terms of the following periodic property:

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

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





1.10 Theorem: exact values on special angles

For some special angles between 0 and $\frac{\pi}{2}$, we have the following table of exact values:

degree	radian	sine	cosine
0°	0	0	1
30°	$\frac{\pi}{6}$	$\frac{1}{2}$	$\frac{\sqrt{3}}{2}$
45°	$\frac{\pi}{4}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{2}}{2}$
60°	$\frac{\pi}{3}$	$\frac{\sqrt{3}}{2}$	$\frac{1}{2}$
90°	$\frac{\pi}{2}$	1	0

For other trigonometric functions, the exact values at these special angles can be obtained by using

$$\tan \theta = \frac{\sin \theta}{\cos \theta}, \quad \cot \theta = \frac{\cos \theta}{\sin \theta}, \quad \sec \theta = \frac{1}{\cos \theta}, \quad \csc \theta = \frac{1}{\sin \theta}.$$

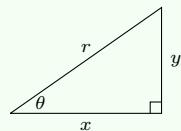
For angles in other ranges, one may use the following equalities:

$$\begin{aligned} \sin(x + \pi) &= -\sin x, & \cos(x + \pi) &= -\cos x, & \tan(x + \pi) &= \tan x, \\ \csc(x + \pi) &= -\csc x, & \cos(x + \pi) &= -\sec x, & \cot(x + \pi) &= \cot x. \end{aligned}$$

1.11 Theorem: Two Key Identities of Trigonometric Functions

- The basic relationship between the sine and the cosine is given by the Pythagorean identity:

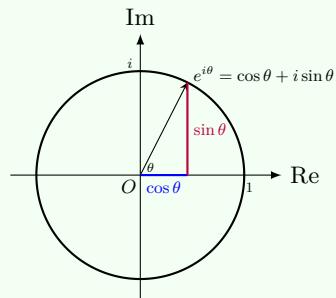
$$\sin^2 \theta + \cos^2 \theta = 1.$$



- *Euler's formula*, the fundamental relationship between the trigonometric functions and the complex exponential function: for any real number θ :

$$e^{i\theta} = \cos \theta + i \sin \theta,$$

where i is the imaginary unit satisfying $i^2 = -1$.



Remark: complex functions

- Any complex number can be expressed in the form $a + bi$, where a and b are real numbers. The number a is called the real part of $a + bi$, and b is its imaginary part.
- The complex exponential function $e^{i\theta}$ can be treated similarly as other real exponential functions.
- Many trigonometric identities can be proved by using the Pythagorean identity $\sin^2 \theta + \cos^2 \theta = 1$ and Euler's formula $e^{i\theta} = \cos \theta + i \sin \theta$. For instance, from the following:

$$\begin{aligned} & \cos(\alpha + \beta) + i \sin(\alpha + \beta) \\ &= e^{i(\alpha+\beta)} = e^{i\alpha} \cdot e^{i\beta} \\ &= (\cos \alpha + i \sin \alpha) \cdot (\cos \beta + i \sin \beta) \\ &= (\cos \alpha \cos \beta - \sin \alpha \sin \beta) + i(\sin \alpha \cos \beta + \cos \alpha \sin \beta), \end{aligned}$$

by identifying the real and imaginary parts on both sides, we have the angle sum identities:

$$\begin{aligned} \cos(\alpha + \beta) &= \cos \alpha \cos \beta - \sin \alpha \sin \beta, \\ \sin(\alpha + \beta) &= \sin \alpha \cos \beta + \cos \alpha \sin \beta. \end{aligned}$$

- The proof of Euler's formula requires advanced knowledge of mathematics, so is omitted here.

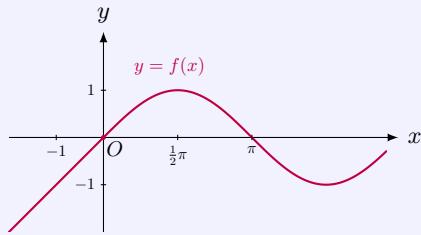
1.12 Example: piecewise function

There are various ways to express functions, numerically, graphically, or using mathematical expressions.

Some functions may have different mathematical expressions over different parts of its domain. Here are some examples.

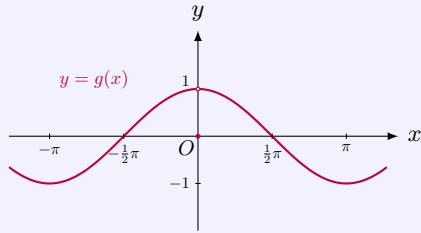
1.

$$f(x) = \begin{cases} x, & \text{if } x \leq 0; \\ \sin x, & \text{if } x > 0. \end{cases}$$



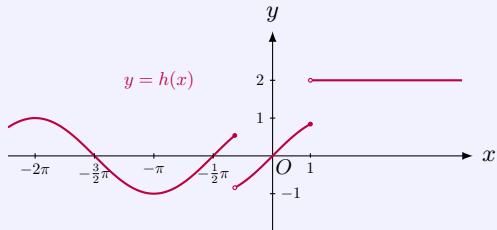
2.

$$g(x) = \begin{cases} \cos x, & \text{if } x \neq 0; \\ 0, & \text{if } x = 0. \end{cases}$$



3.

$$h(x) = \begin{cases} \cos x, & \text{if } x \leq -1; \\ \sin x, & \text{if } -1 < x \leq 1; \\ 2, & \text{if } x > 1. \end{cases}$$



1.13 MATLAB: symbolic function

MATLAB supports creating the symbolic function with the help of the command `syms`. There are two ways to create a symbolic function in MATLAB. The first way creates the symbolic function by first creating the symbolic variable. The following example shows how to create the function $y = f(x) = x^2$ through creating the symbolic variable `x`.

```
>> syms x;
>> y = x^2;
```

where `syms x` creates the symbolic variable `x`.

The second way is to create the symbolic function and the symbolic variable simultaneously, which is shown in the following:

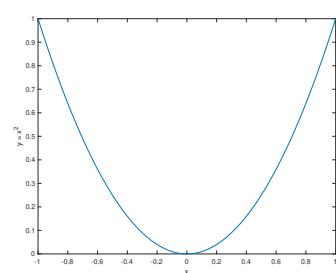
```
>> syms f(x);
>> f(x) = x^2;
```

where `syms f(x)` creates the symbolic function `f(x)` as well as the symbolic variable `x`.

Moreover, we can plot the figure of $y = f(x) = x^2$ using the `plot` command in MATLAB. The MATLAB code below shows how to plot the function $y = f(x) = x^2$ in the domain $D = [-1, 1]$.

```
>> x = -1:0.01:1;
>> figure
>> plot(x,x.^2)
>> xlabel('x')
>> ylabel('y = x^2')
```

With the above MATLAB code, we can generate the figure of $y = f(x) = x^2$ in the domain $D = [-1, 1]$, shown in the following:



1.14 Engineering Example: signal

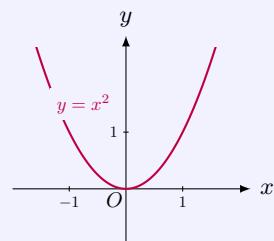
The signal is a basic concept in signal processing, communication systems and electrical engineering. Signals in nature are mathematical functions. For example, the independent variable can be the time; while the dependent variable can be electric current, flow rate, sound, motion of an object, and energy, etc.

1.2 Limit

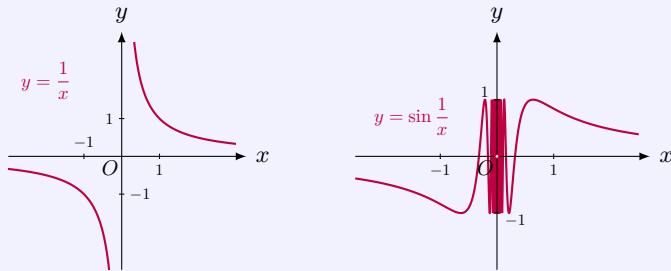
1.15 Example: behaviors near a point

Consider two types of behaviors of functions near a point.

1. Consider $f(x) = x^2$. When x approaches 0, $f(x)$ approaches 0.



2. Consider $g(x) = \frac{1}{x}$ and $h(x) = \sin \frac{1}{x}$. When x approaches 0, both $g(x)$ and $h(x)$ do not approach any finite number.



1.16 Definition: limit of function

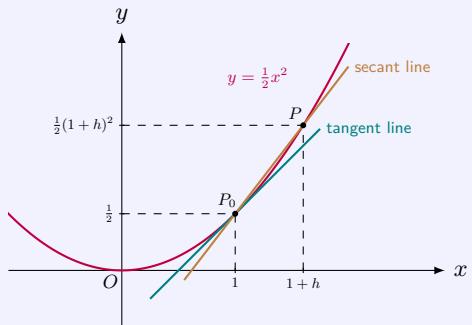
We say a function f with domain D has a finite **limit** L at a if $f(x)$ approaches L when $x \in D$ approaches a . In this case, we write $\lim_{x \rightarrow a} f(x) = L$.

1.17 Example: tangent line

Consider the function $f(x) = \frac{1}{2}x^2$ and the tangent line of the graph of $y = f(x)$ at $x = 1$. We first consider the secant line of the graph near $x = 1$.

The secant line passes through two points $P_0 = (1, \frac{1}{2})$ and $P = (1 + h, \frac{1}{2}(1 + h)^2)$. The slope of the secant line is

$$S(h) = \frac{\frac{1}{2}(1 + h)^2 - \frac{1}{2}}{(1 + h) - 1} = \frac{h + \frac{1}{2}h^2}{h} = 1 + \frac{1}{2}h.$$



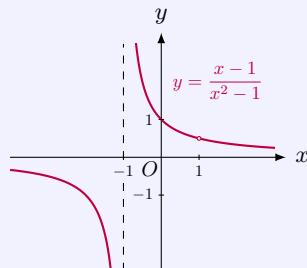
As the point P approaches P_0 , i.e., as h approaches 0, the secant line approaches the tangent line. In the language of limits, the tangent line is the limit of the secant line. Hence the *limit* of the slope of the secant line is the slope S of the tangent line, i.e.,

$$S = \lim_{h \rightarrow 0} S(h) = \lim_{h \rightarrow 0} (1 + \frac{1}{2}h) = 1.$$

1.18 Example: eliminating the zero factor

Consider the function $f(x) = \frac{x-1}{x^2-1}$. The function is not defined at 1. In fact, if we substitute x by 1 in $\frac{x-1}{x^2-1}$, we will get $\frac{0}{0}$, which is meaningless.

However, the limit of the function, $\lim_{x \rightarrow 1} \frac{x-1}{x^2-1}$, only involves values of the function *near* $x = 1$, but not the value at $x = 1$.



Indeed, the value of the limit is $\frac{1}{2}$, that can be obtained by eliminating the same “zero” factor $x - 1$ in the numerator and the denominator:

$$\begin{aligned} \lim_{x \rightarrow 1} \frac{x-1}{x^2-1} &= \lim_{x \rightarrow 1} \frac{x-1}{(x+1)(x-1)} && \text{factoring} \\ &= \lim_{x \rightarrow 1} \frac{1}{x+1} && \text{cancelling 0-factor} \\ &= \frac{1}{2}, \end{aligned}$$

This example shows that the limit of a function at $x = 1$ is irrelevant to the actual value of the function at $x = 1$.

Remark:

Note that in the definition of limit, we do not require a in the domain D . In particular, the actual value of $f(a)$ even when $a \in D$ plays no role in the definition of limit as shown in the current example.

1.19 Example: search for rigorous definition of limits (optional)

The intuitive definition of the limit of a function $f(x)$ at $x = a$ is that $f(x)$ approaches a number L when x approaches a . To phrase in another way, $\lim_{x \rightarrow a} f(x) = L$ if, when x is close to a , $f(x)$ is close to L .

Consider two functions

$$f(x) = x^2, \quad g(x) = \sqrt{x}.$$

We know that

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

When x is close to 0, say $|x| < 0.01$, $f(x) = x^2$ is close to $L = 0$, i.e. $|f(x) - 0| = |x^2| < 0.0001$. For the function $g(x) = \sqrt{x}$, if we choose $|x| < 0.01$, then we can only obtain $|g(x)| < 0.1$. Even though 0.1 is a small number, the number 0.0001 is much smaller than 0.1. If we require $|g(x)| < 0.0001 = 10^{-4}$, we need to choose $|x| < 10^{-8}$.

Now we can rephrase $\lim_{x \rightarrow a} f(x) = L$ as follows: we can make $f(x)$ arbitrarily close to L by choose x sufficiently close to a .

In the above example, if we make $f(x) = x^2$ close to 0 within 0.01, we choose x close to 0 within 0.1; if we make $g(x) = \sqrt{x}$ close to 0 within 0.01, we choose x close to 0 within 10^{-8} .

The statement “we can make $f(x)$ arbitrarily close to L ” can be phrased as “Given a positive ϵ , which might be very small, we require $|f(x) - L| < \epsilon$ ”; the statement “choosing x sufficiently close to a ” can be rephrased as “there exists a positive number δ such that $|x - a| < \delta$.”

Now we can write the rigorous definition of limit $\lim_{x \rightarrow a} f(x) = L$ as follows:

Given an $\epsilon > 0$, there exists $\delta > 0$ such that, whenever $|x - a| < \delta$, we have $|f(x) - L| < \epsilon$.

For the function $f(x) = x^2$, given $\epsilon = 0.01$, there exists $\delta = 0.1$ such that, whenever $|x| < \delta = 0.1$, we have $|f(x) - 0| = |x^2| < \epsilon = 0.01$; given $\epsilon = 0.0001$, there exists $\delta = 0.01$ such that, whenever $|x| < \delta = 0.01$, we have $|f(x) - 0| = |x^2| < \epsilon = 0.0001$.

1.20 Definition: rigorous definition of limit (optional)

The limit of $f(x)$ equals L as x approaches a , denoted by

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

if, given an $\epsilon > 0$, there exists $\delta > 0$ such that, whenever $|x - a| < \delta$, we have $|f(x) - L| < \epsilon$.

1.21 MATLAB: limit

MATLAB can calculate the limit of a function with command `limit`. The following example shows how to calculate the limit of $f(x) = x^2$ when x approaches 2, i.e., $\lim_{x \rightarrow 2} x^2$, in MATLAB.

```
>> syms x;
>> limit(x^2,x,2)

ans =
4
```

MATLAB returns 4 as the result of the limit.

The following example shows the calculation of $\lim_{x \rightarrow 1} \frac{x-1}{x^2-1}$ in MATLAB.

```
>> syms x;
>> f = (x - 1)/(x^2 - 1);
>> limit(f,x,1)

ans =
1/2
```

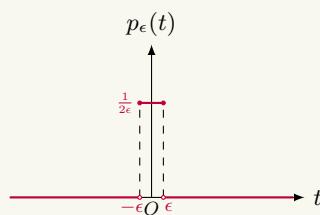
MATLAB returns $\frac{1}{2}$ as the result of the limit.

1.22 Engineering Example: unit-impulse signal

Consider the function

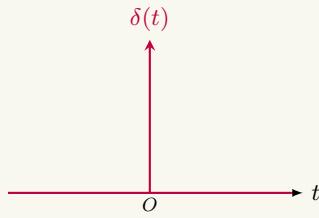
$$p_\epsilon(t) = \begin{cases} \frac{1}{2\epsilon}, & \text{if } -\epsilon \leq t \leq \epsilon; \\ 0, & \text{otherwise.} \end{cases}$$

The graph of $p_\epsilon(t)$ is shown.



Clearly, as ϵ gets smaller, the graph of $p_\epsilon(t)$ gets narrower and taller. In particular, as $\epsilon \rightarrow 0$, the function $p_\epsilon(t)$ approaches the Dirac delta function $\delta(t)$, and can be mathematically written as

$$\delta(t) = \lim_{\epsilon \rightarrow 0} p_\epsilon(t).$$



Note that the Dirac delta function equals to 0 everywhere except for 0, and the “area” under the function equals 1. The Dirac delta function $\delta(t)$ is depicted as shown.

In engineering fields, especially signal processing, the Dirac delta function is also known as the unit-impulse signal. The output of a system with the unit-impulse signal input is called the unit-impulse response. The unit-impulse response can completely describe a linear time-invariant (LTI) system, such as communication channels.

Limit Rules

1.23 Theorem: arithmetic rules of limits

Suppose $\lim_{x \rightarrow a} f(x)$ and $\lim_{x \rightarrow a} g(x)$ exist. Then the following rules hold.

1. Sum Rule:

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

2. Difference Rule:

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

3. Product Rule:

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

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

4. Quotient Rule:

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

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

1.24 Example: arithmetic rules

Using the product rule of limits, we get $\lim_{x \rightarrow a} x^2 = (\lim_{x \rightarrow a} x)^2 = a^2$. By induction, it is easy to have $\lim_{x \rightarrow a} x^n = a^n$ for any positive integer n .

Then using the sum rule of limits, we get

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

Thus for *any polynomial* $f(x)$, we have $\lim_{x \rightarrow a} f(x) = f(a)$.

If $f(x)$ and $g(x)$ are polynomials with $g(a) \neq 0$, then for the *rational function* $r(x) = \frac{f(x)}{g(x)}$, by using the quotient rule, we have

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

1.25 Pitfall: arithmetic rules - conditions not met

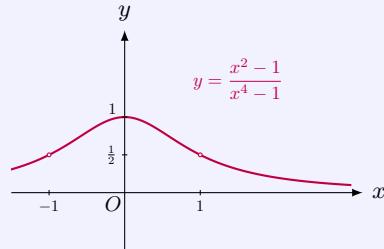
For every arithmetic rule, there are associated conditions under which the rule holds. If these conditions are not all satisfied, then the corresponding rule may fail.

For example, if we hope to calculate $\lim_{x \rightarrow 1} \frac{x^2 - 1}{x^4 - 1}$, we cannot directly apply the quotient rule since $\lim_{x \rightarrow 1} (x^4 - 1) = 0$.

The correct way to calculate the limit is first to simplify the expression by eliminating the “zero” factor:

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

Then we can apply the quotient rule to get



$$\lim_{x \rightarrow 1} \frac{x^2 - 1}{x^4 - 1} = \lim_{x \rightarrow 1} \frac{1}{x^2 + 1} = \frac{1}{2}.$$

1.26 Theorem: Squeeze Theorem

Suppose

1. $f(x) \leq g(x) \leq h(x)$ near a (not necessarily at a).
2. $\lim_{x \rightarrow a} f(x) = \lim_{x \rightarrow a} h(x) = L$.

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

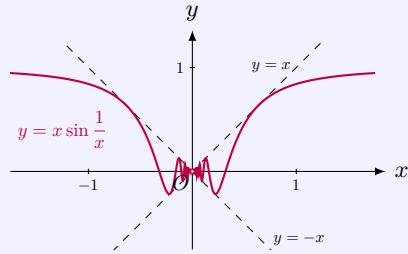
1.27 Example: Squeeze Theorem

Calculate $\lim_{x \rightarrow 0} x \sin \frac{1}{x}$.

We cannot use the product rule since $\lim_{x \rightarrow 0} \frac{\sin \frac{1}{x}}{x}$ does not exist. However, we can apply the Squeeze Theorem. In fact, since $0 \leq \left| x \sin \frac{1}{x} \right| \leq |x|$, we have

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

Because $\lim_{x \rightarrow 0} |x| = \lim_{x \rightarrow 0} (-|x|) = 0$, by the Squeeze Theorem, we get $\lim_{x \rightarrow 0} x \sin \frac{1}{x} = 0$.

**1.28 Example:** Squeeze Theorem and arithmetic rules

Calculate $\lim_{x \rightarrow 1} (x-1)^2 \left(\frac{1}{x^2-1} + \sin \frac{1}{x-1} \right)$.

Since the values of $\frac{1}{x^2-1}$ can be made arbitrarily large when x approaches 1 and $\sin \frac{1}{x-1}$ is bounded by 1, $\left(\frac{1}{x^2-1} + \sin \frac{1}{x-1} \right)$ does not have limit (or approaches ∞) and is not bounded either. Thus we can apply neither the product rule nor the Squeeze Theorem. However, we can do the following changes before we take the limit:

$$\begin{aligned} & (x-1)^2 \left(\frac{1}{x^2-1} + \sin \frac{1}{x-1} \right) \\ &= (x-1)^2 \cdot \frac{1}{(x-1)(x+1)} + (x-1)^2 \sin \frac{1}{x-1} && \text{expanding} \\ &= \frac{x-1}{x+1} + (x-1)^2 \sin \frac{1}{x-1}. && \text{cancelling 0-factor} \end{aligned}$$

For the first term above, we have

$$\lim_{x \rightarrow 1} \frac{x-1}{x+1} = 0. \quad \text{applying the quotient rule}$$

For the second term, since

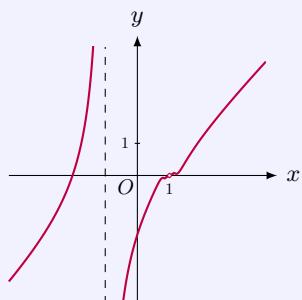
$$-(x-1)^2 \leq (x-1)^2 \sin \frac{1}{x-1} \leq (x-1)^2$$

and both $(x-1)^2$ and $-(x-1)^2$ approaches zero as $x \rightarrow 1$, we apply the Squeeze Theorem and get

$$\lim_{x \rightarrow 1} (x-1)^2 \sin \frac{1}{x-1} = 0.$$

Thus,

$$\begin{aligned} & \lim_{x \rightarrow 1} (x-1)^2 \left(\frac{1}{x^2-1} + \sin \frac{1}{x-1} \right) \\ &= \lim_{x \rightarrow 1} \frac{x-1}{x+1} + \lim_{x \rightarrow 1} (x-1)^2 \sin \frac{1}{x-1} \\ &= 0 + 0 = 0. \end{aligned}$$



One-Sided Limits

1.29 Definition: one-sided limits

We write

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

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 $f(x)$ is arbitrarily close to L for all x sufficiently close to a with $x < a$.

Similarly, we write

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

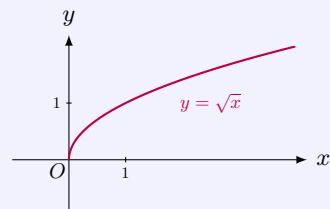
and say the **right-hand limit of $f(x)$ as x approaches a** [or the limit of $f(x)$ as x approaches a from the right] is equal to L , if $f(x)$ is arbitrarily close to L for all x sufficiently close to a with $x > a$.

1.30 Theorem: relationship - the limit and the one-sided limits

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

1.31 Example: one-sided limits

- When we calculate the limit $\lim_{x \rightarrow 0} \sqrt{x}$, since x must be non-negative, we in fact are calculating the right-hand limit $\lim_{x \rightarrow 0^+} \sqrt{x}$.



2. Consider the function

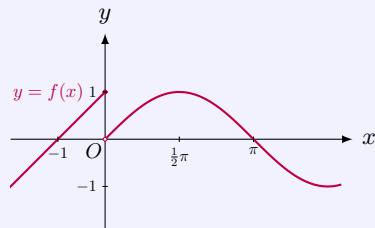
$$f(x) = \begin{cases} x+1, & \text{if } x \leq 0; \\ \sin x, & \text{if } x > 0. \end{cases}$$

- When $x < 0$, $f(x) = x + 1$. So,

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

- When $x > 0$, $f(x) = \sin x$. So,

$$\lim_{x \rightarrow 0^+} f(x) = \lim_{x \rightarrow 0^+} \sin x = 0.$$



Because $\lim_{x \rightarrow 0^-} f(x) \neq \lim_{x \rightarrow 0^+} f(x)$, we conclude that the limit $\lim_{x \rightarrow 0} f(x)$ does not exist.

3. Consider the function

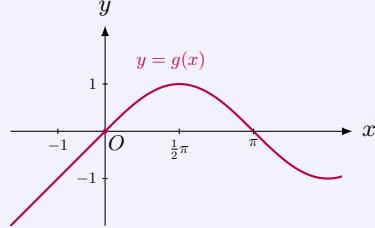
$$g(x) = \begin{cases} x, & \text{if } x \leq 0; \\ \sin x, & \text{if } x > 0. \end{cases}$$

- When $x < 0$, $g(x) = x$. So,

$$\lim_{x \rightarrow 0^-} g(x) = \lim_{x \rightarrow 0^-} x = 0.$$

- When $x > 0$, $g(x) = \sin x$. So,

$$\lim_{x \rightarrow 0^+} g(x) = \lim_{x \rightarrow 0^+} \sin x = 0.$$



Because $\lim_{x \rightarrow 0^-} g(x) = 0 = \lim_{x \rightarrow 0^+} g(x)$, we conclude that $\lim_{x \rightarrow 0} g(x) = 0$.

1.32 MATLAB: one-sided limits

The MATLAB command `limit` can also calculate the one-sided limit of a function. The following example shows how to calculate the left-hand limit of $f(x) = x^2$ as x approaches 2, i.e., $\lim_{x \rightarrow 2^-} x^2$, in MATLAB.

```
>> syms x;
>> limit(x^2,x,2,'left')

ans =
4
```

MATLAB returns 4 as the result of the left-hand limit.

The right-hand limit of $f(x) = x^2$ as x approaches 2, i.e., $\lim_{x \rightarrow 2^+} x^2$, can also be calculated using similar MATLAB code, shown as follows:

```
>> syms x;
```

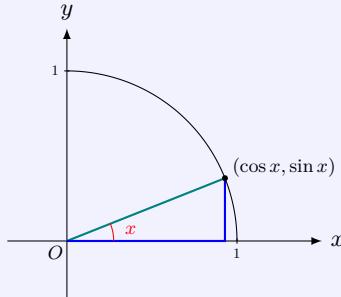
```
>> limit(x^2,x,2,'right')
ans =
4
```

MATLAB returns 4 as the result of the right-hand limit.

1.33 Example: $\lim_{x \rightarrow a} \sin x$ and $\lim_{x \rightarrow a} \cos x$

Firstly, let us calculate $\lim_{x \rightarrow 0} \sin x$.

From the diagram, we see that the arc-length from the point $(1, 0)$ to the point $(\cos x, \sin x)$ is equal to x which is bigger than the blue side of the triangle with length equal to $\sin x$. Thus when $0 < x < \pi/2$, we have $0 < \sin x < x$.



By the Squeeze Theorem, we get $\lim_{x \rightarrow 0^+} \sin x = 0$. Because $\sin x$ is an odd function, i.e., $\sin(-x) = -\sin x$, we further have $\lim_{x \rightarrow 0^-} \sin x = \lim_{-x \rightarrow 0^+} [-\sin(-x)] = 0$. Thus we get $\lim_{x \rightarrow 0} \sin x = 0$.

Furthermore, since

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

by the Squeeze Theorem, we get $\lim_{x \rightarrow a} \sin x = \sin a$.

Since $\cos x = \sin(\frac{\pi}{2} - x) = -\sin(x - \frac{\pi}{2})$, we have

$$\begin{aligned} \lim_{x \rightarrow a} \cos x &= \lim_{x \rightarrow a} [-\sin(x - \frac{\pi}{2})] \\ &= -\lim_{x \rightarrow a - \frac{\pi}{2}} \sin x = -\sin(a - \frac{\pi}{2}) = \cos a. \end{aligned}$$

1.34 Example: $\lim_{\theta \rightarrow 0} \frac{\sin \theta}{\theta} = 1$

We cannot apply the quotient rule to calculate the limit $\lim_{\theta \rightarrow 0} \frac{\sin \theta}{\theta}$, since the limit of the denominator is zero.

We use geometry and the Squeeze Theorem to calculate the limit. In fact, from the diagram shown, for $0 < \theta < \frac{\pi}{2}$, we have

$$\begin{aligned}\text{Area } \triangle OSP &< \text{Area sector } OSP \\ &< \text{Area } \triangle OST,\end{aligned}$$

or

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

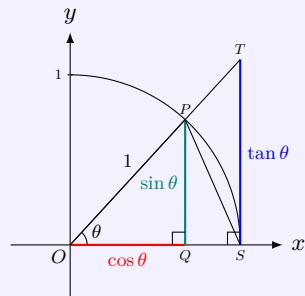
These inequalities give

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

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

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

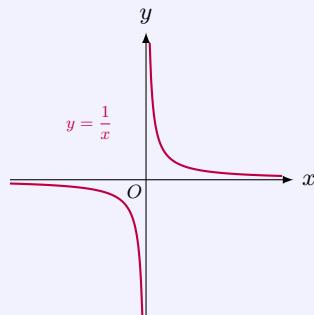
$$\text{Hence, } \lim_{\theta \rightarrow 0} \frac{\sin \theta}{\theta} = 1.$$



Limits at Infinities

1.35 Example: limits at infinities

For the function $y = \frac{1}{x}$, we see that when x takes arbitrarily large values, $f(x)$ will approach 0. The line $y = 0$ is called the *asymptote* of the function. This is one of the key components in curve sketching.



1.36 Definition: limits at infinities

- If the values of $f(x)$ can be made arbitrarily close to L by taking x sufficiently large positive values, then we write

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

- If the values of $f(x)$ can be made arbitrarily close to L by taking x sufficiently large

negative values, then we write

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

- If the values of $f(x)$ can be made arbitrarily close to L by taking $|x|$ sufficiently large values, then we write

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

1.37 Definition: horizontal asymptote

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

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

or

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

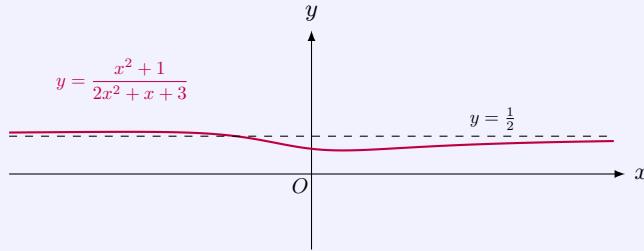
1.38 Example: horizontal asymptote

The arithmetic rules and the Squeeze Theorem still hold for limits at infinities.

For the rational function $\frac{x^2 + 1}{2x^2 + x + 3}$, we can apply some arithmetic rules of limits to calculate the limit at infinity:

$$\begin{aligned} & \lim_{x \rightarrow \infty} \frac{x^2 + 1}{2x^2 + x + 3} && \text{comparing the leading powers} \\ &= \lim_{x \rightarrow \infty} \frac{1 + \frac{1}{x^2}}{2 + \frac{1}{x} + \frac{3}{x^2}} && \text{dividing each term by } x^2 \\ &= \frac{\lim_{x \rightarrow \infty} \left(1 + \frac{1}{x^2}\right)}{\lim_{x \rightarrow \infty} \left(2 + \frac{1}{x} + \frac{3}{x^2}\right)} = \frac{1}{2}. && \text{using the arithmetic rules of limits} \end{aligned}$$

Thus, this function has a horizontal asymptote $x = \frac{1}{2}$.

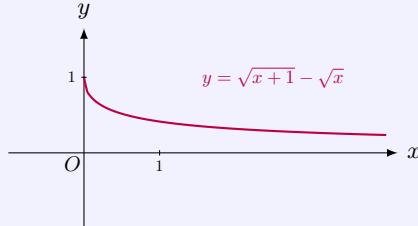


1.39 Example: limit at infinity - difference of two radicals

Rationalizing radicals is a usual approach to deal with limits of this type before applying arithmetic rules:

$$\begin{aligned}
 & \lim_{x \rightarrow +\infty} (\sqrt{x+1} - \sqrt{x}) \\
 &= \lim_{x \rightarrow +\infty} \frac{(\sqrt{x+1} - \sqrt{x})(\sqrt{x+1} + \sqrt{x})}{\sqrt{x+1} + \sqrt{x}} && \text{multiplying } \frac{\sqrt{x+1} + \sqrt{x}}{\sqrt{x+1} + \sqrt{x}} \\
 &= \lim_{x \rightarrow +\infty} \frac{(\sqrt{x+1})^2 - (\sqrt{x})^2}{\sqrt{x+1} + \sqrt{x}} && \text{simplifying} \\
 &= \lim_{x \rightarrow +\infty} \frac{1}{\sqrt{x+1} + \sqrt{x}} && \text{simplifying} \\
 &= 0.
 \end{aligned}$$

This shows that $y = 0$ is a horizontal asymptote of the function.

**1.40 MATLAB:** limits at infinities

With the help of positive infinity `inf`, MATLAB can calculate the limit of a function at infinities. The MATLAB code in the following shows how to calculate $\lim_{x \rightarrow +\infty} (\sqrt{x+1} - \sqrt{x})$.

```

>> syms x;
>> f = sqrt(x+1)-sqrt(x);
>> limit(f,x,inf)

ans =
0

```

From the MATLAB code above, we see that

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

1.41 Example: limit at infinity - polynomials

Let $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$ be a polynomial. We can rewrite

$$f(x) = x^n \left(1 + \frac{a_{n-1}}{x} + \cdots + \frac{a_0}{x^n}\right).$$

When x approaches $\pm\infty$, the term $1 + \frac{a_{n-1}}{x} + \cdots + \frac{a_0}{x^n}$ approaches 1, and the term x^n approaches ∞ . In fact, if n is an even integer, x^n approaches $+\infty$; if n is an odd integer, x^n approaches $\pm\infty$. Thus we get the following useful result:

- When n is an even positive integer, a degree- n polynomial $f(x) = x^n + \cdots + a_1x + a_0$ approaches positive infinity when x approaches infinities.
- When n is an odd positive integer, $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$ approaches positive infinity (respectively, negative infinity) when x approaches positive infinity (respectively, negative infinity).

1.42 Example: limits at infinity - slant asymptote

Let us study the behaviors of the function $f(x) = \frac{x^3 - 2x^2 + 1}{2x^2 + 1}$ near the infinities.

We factor out the highest power in numerator and denominator and have

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

By applying the arithmetic rules, we have

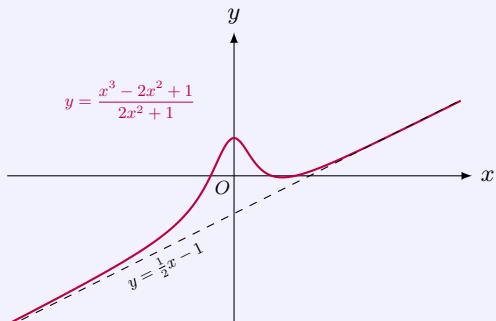
$$\frac{\lim_{x \rightarrow +\infty} \left(1 - \frac{2}{x} + \frac{1}{x^3}\right)}{\lim_{x \rightarrow +\infty} \left(2 + \frac{1}{x^2}\right)} = \frac{1}{2},$$

so that

$$\lim_{x \rightarrow +\infty} \frac{f(x)}{x} = \frac{1}{2}.$$

Furthermore, we have

$$\begin{aligned} & \lim_{x \rightarrow +\infty} [f(x) - \frac{1}{2}x] \\ &= \lim_{x \rightarrow +\infty} \left(\frac{x^3 - 2x^2 + 1}{2x^2 + 1} - \frac{1}{2}x \right) \\ &= \lim_{x \rightarrow +\infty} \left(-\frac{4x^2 + x - 2}{4x^2 + 2} \right) = -1. \end{aligned}$$



We see that as $x \rightarrow +\infty$, the difference between $f(x)$ and $\frac{1}{2}x - 1$ approaches zero. This means that, near the positive infinity, the function $y = f(x)$ behaves like the function $y = \frac{1}{2}x - 1$. Similarly, the function also behaves like the function $y = \frac{1}{2}x - 1$ near the negative infinity. The line $y = \frac{1}{2}x - 1$ is called a **slant asymptote** of the function $f(x)$.

Remark:

An alternating approach of finding a slant asymptote for a rational function is to use polynomial long division. For the above example, the long division gives

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

so that

$$\frac{x^3 - 2x^2 + 1}{2x^2 + 1} = \frac{1}{2}x - 1 + \frac{-x + 4}{4x^2 + 2},$$

The quotient $\frac{1}{2}x - 1$ gives the slant asymptote of the function f .

1.43 Engineering Example: initial value theorem

In electronic engineering and control system, the time-domain behaviour of a system can be characterized by the step response, which is the output of a system in the time domain when the input is a unit-step function. The unit-step input means the input changes from 0 to 1 in a very short time. In other words, there is a sudden input. Note that a sudden input may cause a fast and possibly large system performance deviation from its long-term steady state, which may have severe effects on a certain system component or even the overall system. Moreover, if a sudden input occurs, it may take time for the system output to settle down to some vicinity of its final state before the system can act properly. Hence, it is important to understand how the system responds to a sudden input, i.e., understand the step response.

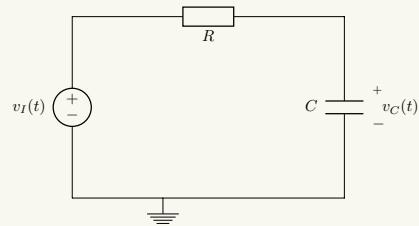
The initial slope of the step response is one of the important quantities to describe the step response. One way to calculate the initial slope of the step response is to use the **Initial Value Theorem**. Let the initial slope of the step response of a system be denoted by $\lim_{t \rightarrow 0} f(t)$, the Initial Value Theorem says that

$$\lim_{t \rightarrow 0} f(t) = \lim_{s \rightarrow +\infty} sF(s),$$

where $F(s)$ is the (one-sided) Laplace transform of the step response $f(t)$ and s is a complex number frequency parameter. Note that the Laplace transform provides an alternative way of analyzing the system in the frequency domain.

Now, we use the Initial Value Theorem to calculate the initial slope of the step response for a resistor-capacitor (RC) circuit.

Given that the step response of the RC circuit is $v_C(t) = \left(1 - e^{-\frac{t}{RC}}\right) u(t)$, where R is the resistance of the resistor, C is the capacitance of the capacitor and $u(t)$ is the unit-step function, then the Laplace transform of this step response is $V_C(s) = \frac{1}{s(RC \cdot s + 1)}$.



According to the property of Laplace transform, the Laplace transform of the slope of the step response can be calculated as

$$F(s) = sV_C(s) = \frac{1}{RC \cdot s + 1}.$$

Then, we use the Initial Value Theorem to calculate the initial slope of the step response and have

$$\lim_{t \rightarrow 0} f(t) = \lim_{s \rightarrow +\infty} sF(s) = \lim_{s \rightarrow +\infty} \frac{s}{RC \cdot s + 1} = \frac{1}{RC}.$$

This shows that the initial slope of the step response is $\frac{1}{RC}$.

1.44 Engineering Example: final value theorem

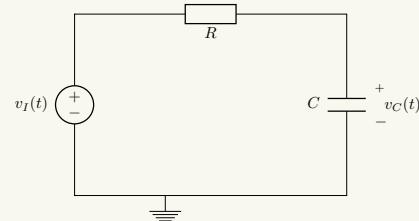
A system or process is in a steady state when the behaviour of this system or process is unchanging in time. The steady state is an important concept in many fields, e.g., economics, electronics, chemical engineering, etc. For example, many design specifications of the electronic systems are given in terms of the steady-state characteristics. Given that the step response of a system is $f(t)$, then the steady state of a system is defined as $\lim_{t \rightarrow +\infty} f(t)$. According to the definition, the steady state can be obtained by calculating the limit of the step response when time t approaches infinity. Alternatively, we can use the **Final Value Theorem** to calculate the steady state. The Final Value Theorem is provided in the following.

$$\lim_{t \rightarrow +\infty} f(t) = \lim_{s \rightarrow 0} sF(s),$$

where $F(s)$ is the (one-sided) Laplace transform of $f(t)$ and s is a complex number frequency parameter. Note that the Laplace transform provides an alternative way of analyzing the system in the frequency domain.

Now, we use the Final Value Theorem to calculate the steady state of a resistor-capacitor (RC) circuit.

Given that the Laplace transform of the step response is $V_C(s) = \frac{1}{s(RC \cdot s + 1)}$, where R is the resistance of the resistor and C is the capacitance of the capacitor.



By applying the Final Value Theorem, we have

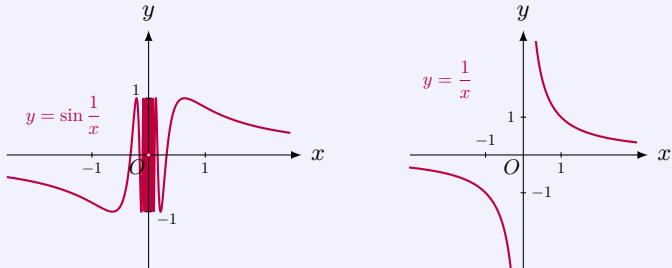
$$\lim_{t \rightarrow +\infty} v_C(t) = \lim_{s \rightarrow 0} sV_C(s) = \lim_{s \rightarrow 0} \frac{s}{s(RC \cdot s + 1)} = 1.$$

This shows that the steady state of the RC circuit is 1.

Infinite Limits

1.45 Example: infinite limits

We know both functions $f(x) = \sin \frac{1}{x}$ and $g(x) = \frac{1}{x}$ do not have limit at $x = 0$. However, we do know the behavior of $g(x)$ near 0. In fact, when x approaches 0 from the right ($x > 0$), the values of $g(x)$ can be made arbitrarily large positive, and when x approaches 0 from the left ($x < 0$), the values of $g(x)$ can be made arbitrarily large negative. The infinite limits are related to vertical asymptotes of the graph of a function.



1.46 Definition: infinite limits

- If the values of $f(x)$ can be made arbitrarily large by taking x sufficiently close to a (but not equal to a), then we write

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

- If the values of $f(x)$ can be made arbitrarily large positive by taking x sufficiently close to a (but not equal to a), then we write

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

- If the values of $f(x)$ can be made arbitrarily large negative by taking x sufficiently close to a (but not equal to a), then we write

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

Similarly, if we take x sufficiently close to from the left-hand side or the right-hand side of a , then we can define

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

and

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

1.47 Definition: vertical asymptote

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

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

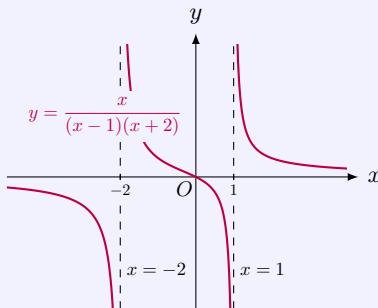
1.48 Example: vertical asymptotes

Consider the function $f(x) = \frac{x}{(x-1)(x+2)}$.

When x approaches 1 from the right, $x-1$ approaches 0 and the values of $f(x)$ can be made arbitrarily large. To determine whether it is the positive infinity or negative infinity, let us check the sign of $f(x)$. In fact, when x approaches 1 from the right, $(x-1) > 0$, $x+2$ approaches 3 and hence, for x near 1 and $x > 1$, $f(x)$ has the positive sign. So $\lim_{x \rightarrow 1^+} f(x) = +\infty$. Hence $x = 1$ is a vertical asymptote for the graph of the function. One can calculate similarly to get

$$\begin{aligned} \lim_{x \rightarrow 1^-} f(x) &= -\infty, \\ \lim_{x \rightarrow -2^+} f(x) &= +\infty, \\ \lim_{x \rightarrow -2^-} f(x) &= -\infty. \end{aligned}$$

It also has a horizontal asymptote line $y = 0$ since when $x \rightarrow \infty$, if we ignore -1 in $(x-1)$ and 2 in $(x+2)$, we get $\lim_{x \rightarrow \infty} f(x) = \lim_{x \rightarrow \infty} \frac{x}{x^2} = 0$.



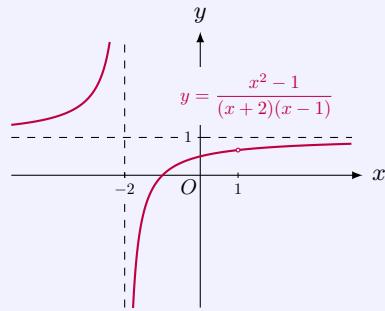
1.49 Pitfall: zero of the denominator $\not\Rightarrow$ vertical asymptote

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

The denominator equals $(x+2)(x-1)$ which have two roots $x = -2$ and $x = 1$. It seems $f(x)$ should have vertical asymptotes $x = -2$ and $x = 1$. However,

$$\begin{aligned} &\lim_{x \rightarrow 1} \frac{x^2 - 1}{(x+2)(x-1)} \\ &= \lim_{x \rightarrow 1} \frac{(x-1)(x+1)}{(x+2)(x-1)} && \text{factoring the numerator} \\ &= \lim_{x \rightarrow 1} \frac{x+1}{x+2} && \text{cancelling 0-factor} \\ &= \frac{2}{3}. \end{aligned}$$

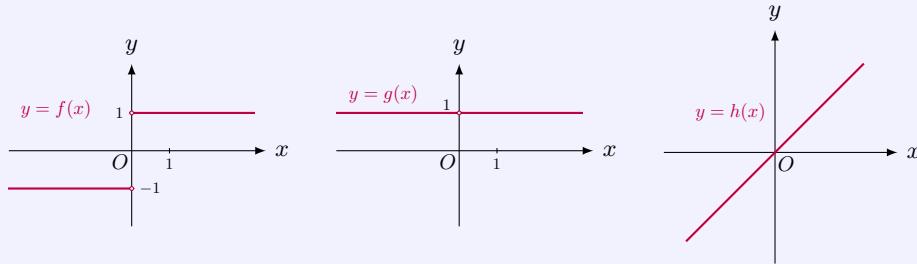
Thus $x = 1$ is *not* a vertical asymptote of the function. This example shows that a zero of the denominator is not necessarily a vertical asymptote.



1.3 Continuous Functions

1.50 Example: continuous function - motivation

Consider the following functions: $f(x) = \frac{|x|}{x}$, $g(x) = \frac{x}{x}$, and $h(x) = x$. The graphs are given below.



We can see that the graph of $h(x)$ moves continuously as x moves along the x -axis while the graphs of both $f(x)$ and $g(x)$ have a break at $x = 0$. In order to distinguish these two different behaviors of functions, we introduce the concept of continuous function. Let us check the difference between these three functions in order to get a correct definition. For the function $f(x)$, the limit $\lim_{x \rightarrow 0} f(x)$ does not exist. For the function $g(x)$, the limit $\lim_{x \rightarrow 0} g(x)$ exists but $g(x)$ is not defined at $x = 0$. For the function $h(x)$, the limit $\lim_{x \rightarrow 0} h(x)$ exists and equals the function value at $x = 0$, i.e., $\lim_{x \rightarrow 0} h(x) = h(0)$.

1.51 Definition: continuous function

- A function f is **continuous** at $x = a$ if
 1. $\lim_{x \rightarrow a} f(x)$ exists;
 2. $f(x)$ is defined at $x = a$;

$$3. \lim_{x \rightarrow a} f(x) = f(a).$$

- A function f is **continuous in an interval** (α, β) if it is continuous at every number in the interval.
- A function f is **continuous on an closed interval** $[\alpha, \beta]$ if it is continuous at every number in the interval (α, β) , and also continuous at the endpoints. The latter means that

$$\lim_{x \rightarrow \alpha^+} f(x) = f(\alpha) \quad \text{and} \quad \lim_{x \rightarrow \beta^-} f(x) = f(\beta).$$

- If f is not continuous at a point a in its domain, we say that f is **discontinuous** at a .

1.52 Theorem: continuity with arithmetic operations

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

$$1. f + g;$$

$$4. f \cdot g;$$

$$2. f - g;$$

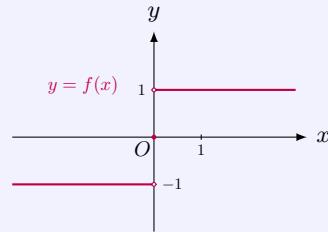
$$5. \frac{f}{g}, \text{ provided that } g(a) \neq 0.$$

1.53 Example: discontinuous functions

1. Consider the function

$$f(x) = \begin{cases} \frac{|x|}{x}, & \text{if } x \neq 0; \\ 0, & \text{if } x = 0. \end{cases}$$

It is easy to calculate the one-sided limits:



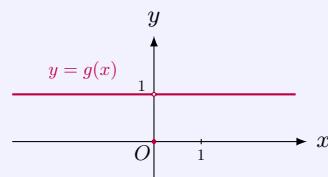
$$\lim_{x \rightarrow 0^-} f(x) = \lim_{x \rightarrow 0^-} \frac{|x|}{x} = \lim_{x \rightarrow 0^-} \frac{-x}{x} = -1,$$

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

Since the left and right-hand limits are not equal, the limit $\lim_{x \rightarrow 0} f(x)$ does not exist. Hence, the function f is not continuous at $x = 0$.

2. Consider the function

$$g(x) = \begin{cases} \frac{x}{x}, & \text{if } x \neq 0; \\ 0, & \text{if } x = 0. \end{cases}$$



It is clear that $g(x) = 1$ if $x \neq 0$, so that

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

Since the left and right-hand limits are both equal to 1, the limit $\lim_{x \rightarrow 0} g(x)$ exists and equals to 1. However, since $\lim_{x \rightarrow 0} g(x) = 1 \neq g(0) = 0$, the function is not continuous at $x = 0$.

1.54 Example: continuous elementary functions

Besides the facts on continuity under arithmetic operations, we can prove the following elementary functions are continuous in their domains.

1. Polynomials are continuous in \mathbb{R} .
2. The functions $\sin x$ and $\cos x$ are continuous in \mathbb{R} .
3. If n is an odd positive integer, the function $f(x) = \sqrt[n]{x}$ is continuous for all x . If n is an even positive integer, the function $f(x) = \sqrt[n]{x}$ is continuous for all $x \geq 0$.

Composite Functions

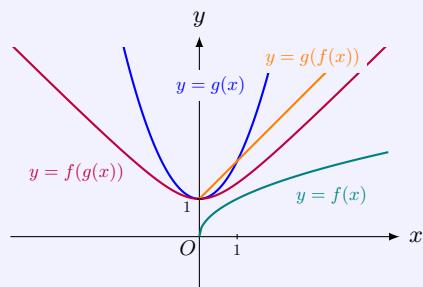
1.55 Definition: composite function

Let f and g be two functions such that the range of g is contained in the domain of f , we can get a new function, called the **composite function** h , defined by $h(x) = f(g(x))$.

1.56 Example: composite functions

Let $f(x) = \sqrt{x}$ and $g(x) = x^2 + 1$. Then the composite function $f(g(x)) = \sqrt{x^2 + 1}$. Note that the composition of two functions depends on the order of the composition. For example, $g(f(x)) = (\sqrt{x})^2 + 1 = x + 1$. Thus $f(g(x)) \neq g(f(x))$.

In fact, $f(g(x))$ and $g(f(x))$ have not only different expressions, but also different domains: the function $f(g(x))$ is defined for all real number x ; the function $g(f(x))$ is defined for $x \geq 0$.



1.57 MATLAB: composite function

MATLAB can show the composite function with command `compose`. The following exam-

ple shows the composite function $h(x) = f(g(x))$, where $f(x) = \frac{1}{x}$ and $g(x) = 1 + x^2$.

```
>> syms x;
>> f = 1/x;
>> g = 1+x^2;
>> h = compose(f,g)

h =
1/(x^2 + 1)
```

1.58 Theorem: continuity of composition of continuous functions

Given two *continuous* functions f and g such that the composition $f(g(x))$ exists. Then the composite function $f(g(x))$ is continuous.

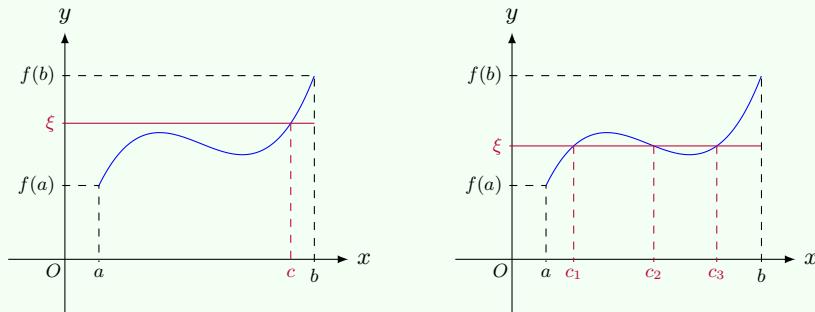
1.59 Example: continuous composite functions

Let $f(x) = \sin x$ and $g(x) = x^2 + 1$. Then the composite functions $f(g(x)) = \sin(x^2 + 1)$ and $g(f(x)) = (\sin x)^2 + 1$ are continuous.

Intermediate Value Theorem

1.60 Theorem: Intermediate Value Theorem

Let f be a *continuous* function over a *closed* interval $[a, b]$. For any number ξ between $f(a)$ and $f(b)$, there exists at least one c in the interval $[a, b]$ such that $f(c) = \xi$.



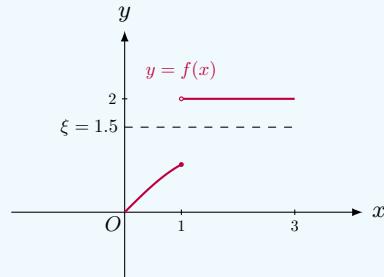
1.61 Remark: key hypothesis in the Intermediate Value Theorem

Continuity of f is the key hypothesis in the Intermediate Value Theorem. If f is not continuous, then the conclusion of the theorem may not hold.

For example, consider the function f defined on $[-1, 3]$:

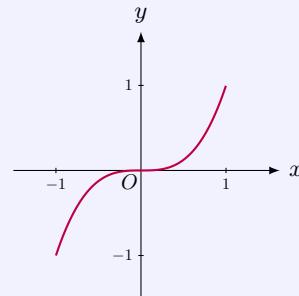
$$f(x) = \begin{cases} \sin x, & \text{if } 0 \leq x \leq 1; \\ 2, & \text{if } 1 < x \leq 3. \end{cases}$$

Clearly, $f(0) = 0$ and $f(3) = 2$. For the intermediate value $\xi = 1.5$, there is no values c in $[0, 3]$ such that $f(c) = 1.5$.

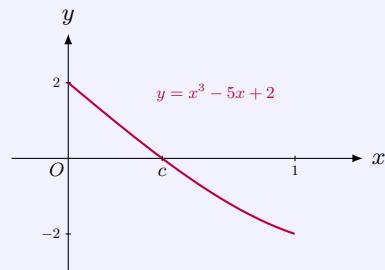


1.62 Example: Intermediate Value Theorem - existence of roots

- Consider the function $f(x) = x^3$. It is clear that f is continuous on $[-1, 1]$, with $f(-1) = -1$ and $f(1) = 1$. To see whether there exists c in the interval $[-1, 1]$ such that $f(c) = 0$, we examine the equation $f(x) = 0$. Such a solution is a root of f . It is easy to see that f indeed has a root $c = 0$, for which we do not need to apply the Intermediate Value Theorem.



- Consider another function $g(x) = x^3 - 5x + 2$. It is not obvious that g has a root in $[0, 1]$. However, it is clear g is continuous, with $g(0) = 2$ and $g(1) = -2$. Since $-2 < 0 < 2$ (i.e., 0 is an intermediate value), from the Intermediate Value Theorem, there exists c in the interval $[0, 1]$ such that $g(c) = 0$, i.e., $g(x) = 0$ has a root in the interval $[0, 1]$.



1.63 Example: Intermediate Value Theorem - third-order polynomial

The Intermediate Value Theorem can be used to show that $f(x) = x^3 + ax^2 + bx + c$, which is continuous everywhere, always has a real root for any numbers a, b and c .

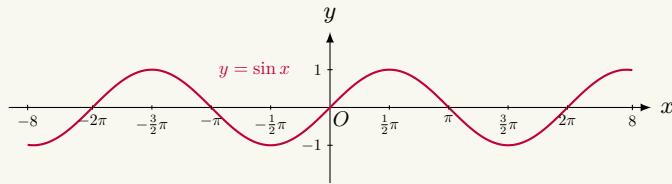
In fact, since $\lim_{x \rightarrow +\infty} f(x) = +\infty$, there exists $b > 0$ such that $f(b) > 0$. Similarly since $\lim_{x \rightarrow -\infty} f(x) = -\infty$, there exists $a < 0$ such that $f(a) < 0$. Because 0 is an intermediate value of $f(a)$ and $f(b)$, from the Intermediate Value Theorem, on the closed interval $[a, b]$, there exists c in $[a, b]$ such that $f(c) = 0$.

1.64 Engineering Example: amplitude modulation

Communication technology plays an important role in modern society by connecting people together. For example, radio broadcast program has been very popular for providing information and entertainment to people. There are two popular transmission technologies used for radio broadcasting, namely, AM (amplitude modulation) and FM (frequency modulation). For both schemes, the programs (music, news, stories, etc.) are brought to people by radio waves, which serve as the carrier signal. In particular, in AM, the carrier signal is a **sinusoidal function**, and the sinusoidal waveform is modulated by an information-bearing audio waveform before transmission. In this way, the amplitude (signal strength) of the carrier signal is varied in proportion to the information-bearing signal to be transmitted.

1.65 Engineering Example: sinusoidal function - signal processing

An engineering system usually can be described by its response to a certain input function. The sinusoidal function is one of such input functions. The sinusoidal function, denoted as $\sin x$, is one of the trigonometric functions, which is thus a periodical function. The sinusoidal function $\sin x$ (i.e., with period 2π) is plotted in the following.

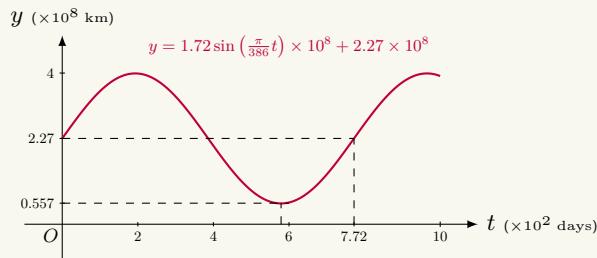


Many systems, e.g., audio system, can be described by their responses to the sinusoidal function. For linear time-invariant (LTI) systems, the sinusoidal input is especially important, because the output is also a sinusoidal function with the same frequency as the input, but possibly different amplitudes and phase angles. Then, the frequency response, which is an important way of characterizing a system, will be utilized to measure the amplitude and the phase angle of the sinusoidal output.

1.66 Engineering Example: sinusoidal function - aeronautics and astronautics

As one of the four “earth-like” terrestrial planets in the solar system, Mars has been explored by multiple organizations and space agencies. For example, the National Aeronautics and Space Administration (NASA) has been exploring Mars for more than 40 years. In January 2004, NASA landed on the surface of Mars with the Mars exploration rovers, Spirit and Opportunity. However, humans have not landed on Mars yet. To successfully complete the “human landing on Mars” mission, the communication between Mars and Earth is vital. But, there is normally a significant communication delay due to the long distance between these two planets (on average, over 100 million kilometers). Since Mars and Earth orbit Sun at different rates, the distance between them is constantly varying. A natural question thus arises as to when is Mars closest to Earth (so that the communication delay is the smallest)? According to some data collected by NASA scientists and equipments, it turns out that the distance between Mars and Earth can be modelled as a **sinusoidal function** (with shift),

which is plotted in the following.



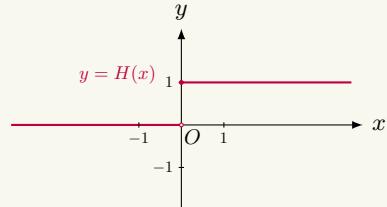
We can see from the above figure that the distance between Mars and Earth is in a 772-day (i.e., around 26 months) cycle and the closest distance between these two planets is around 5.57×10^7 kilometers. Given that the radio waves which can carry information travel in space at the speed of light, the one-way communication time is around 3.1 minutes when Mars is closest to Earth.

1.67 Engineering Example: unit-step function

In mathematics, the unit-step function is defined as follows:

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

It is a discontinuous function and not continuous at $x = 0$.



The unit-step function is a fundamental function in engineering and can be widely applied to many fields, e.g., circuit analysis and communication systems. In particular, for circuit analysis, the unit-step function can model the behaviour of a switch (on/off). Moreover, it can also describe sudden changes of the voltage or the current in a circuit. In communication systems, the sum of unit-step functions can be used to represent waveforms.

1.4 Some Transcendental Functions

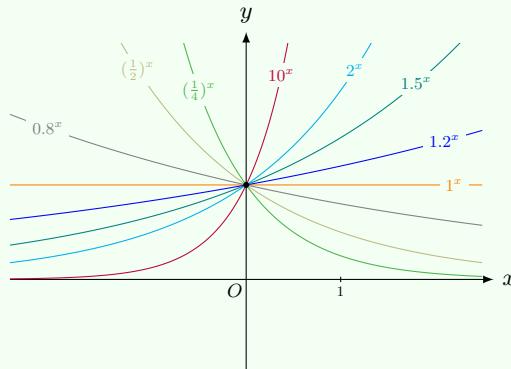
Exponential Functions

1.68 Theorem: properties of exponential functions

The exponential function is $f(x) = a^x$, where a is a positive number.

1. The exponential function $f(x) = a^x$ has domain \mathbb{R} and range $(0, +\infty)$.
2. The exponential function $f(x) = a^x$ is a continuous function.

3. The graph of the exponential function $f(x) = a^x$ depends the value of a .
4. If $a > 1$, then $\lim_{x \rightarrow +\infty} a^x = +\infty$ and $\lim_{x \rightarrow -\infty} a^x = 0$; if $0 < a < 1$, then $\lim_{x \rightarrow +\infty} a^x = 0$ and $\lim_{x \rightarrow -\infty} a^x = +\infty$.
5. The following are the graphs of a^x for various bases.



1.69 Definition: the number e

Among all exponential functions a^x with different base a , there is a base which will play an important role in calculus. This number is denoted by e , defined by

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

This limit exists and e is an irrational number, approximately 2.71828.

1.70 MATLAB: exponential functions

In MATLAB, the exponential functions can be simply written in the following form.

```
>> syms x;
>> f = 3^x

f =
3^x
```

where 3 is the base.

Specifically, the command `exp` can be used for the exponential functions with base e . The following examples show the exponential functions e and e^2 , respectively.

```
>> exp(1)
```

```

ans =
2.7183
>> exp(2)
ans =
7.3891

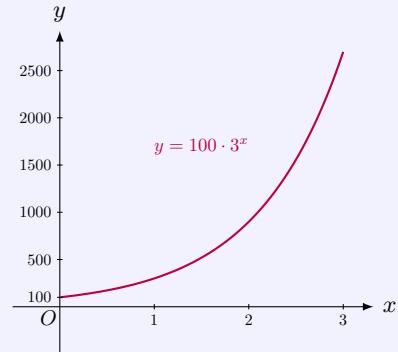
```

1.71 Example: exponential function - population growth

Suppose a bacteria culture with 100 bacteria triples in every hour. To write the function describing the bacteria population growth, let x be the variable representing number of hours, and $y = f(x)$ be the function of bacteria population by x hours. We observe

$$\begin{aligned}f(1) &= 100 \times 3, \\f(2) &= 100 \times 3^2, \\f(3) &= 100 \times 3^3.\end{aligned}$$

So, in general, we get $y = 100 \cdot 3^x$.



1.72 Example: exponential function - radioactive decay

Carbon-14 is a radioactive substance with half-life 5730 years. Half-life means that 1-pound of Carbon-14 will be half-pound after 5730 years.

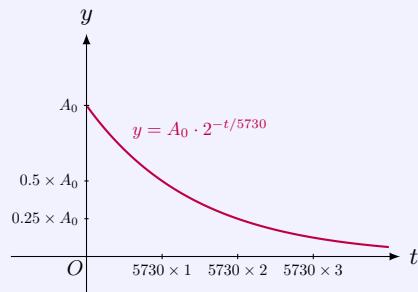
Suppose the initial amount of Carbon-14 is A_0 and $f(t)$ represents the amount after t years. Then

$$\begin{aligned}f(0) &= A_0, \\f(5730) &= A_0/2, \\f(2 \times 5730) &= A_0/2^2.\end{aligned}$$

So, $f(k \times 5730) = A_0/2^k$. Let $k \times 5730 = t$.
Then we get $k = \frac{t}{5730}$, so that the amount
of Carbon-14 after t -years is

$$f(t) = \frac{A_0}{2^{\frac{t}{5730}}} = A_0 \cdot 2^{-t/5730}.$$

If we further write $a = 2^{-\frac{1}{5730}}$, then we get
 $y = f(t) = A_0 a^t$, which is an exponential
function.



1.73 Example: exponential function - bank interest

A bank has a savings account with the annual interest rate r . Let P be the initial deposit. If the bank provides interest m times per year, then after t years, the amount A is given by

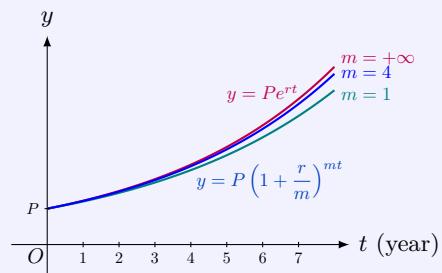
$$A = P \left(1 + \frac{r}{m}\right)^{mt}.$$

If the interest is calculated continuously, i.e. let m go to infinity, the amount A is given by

$$\begin{aligned} A &= \lim_{m \rightarrow +\infty} P \left(1 + \frac{r}{m}\right)^{mt} \\ &= P \lim_{m \rightarrow +\infty} \left[\left(1 + \frac{r}{m}\right)^{\frac{m}{r}} \right]^{rt} \\ &= P \left[\lim_{m \rightarrow +\infty} \left(1 + \frac{r}{m}\right)^{\frac{m}{r}} \right]^{rt} \\ &= P \left[\lim_{x \rightarrow +\infty} \left(1 + \frac{1}{x}\right)^x \right]^{rt} = Pe^{rt}. \end{aligned}$$

Here the number e appears naturally.

The graph compares different models for fixed P and r . It shows that the yield increases as m increases. In other words, the bank will pay you more money if the interest is compounded more often. The amount will reach its maximum if the bank calculates the compound interest (for which they will avoid).



Inverse Function

1.74 Definition: one-to-one

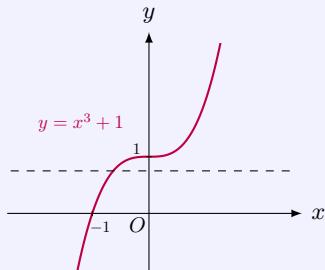
We say $f(x)$ is **one-to-one** if it never takes on the same value twice, i.e., if $x_1 \neq x_2$, then $f(x_1) \neq f(x_2)$, or equivalently, if $f(x_1) = f(x_2)$, then $x_1 = x_2$.

1.75 Theorem: horizontal line test

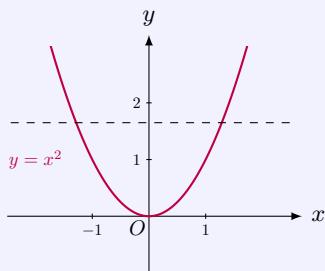
A function is one-to-one if and only if no *horizontal line* intersects its graph more than once.

1.76 Example: one-to-one functions

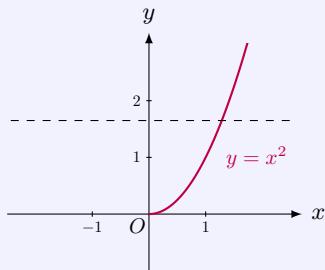
- The function $f(x) = x^3$ is a one-to-one function. It is easy to check that if $f(x_1) = f(x_2)$, then $x_1 = x_2$. It is also easy to verify that the function f passes the horizontal line test.



- The function $g(x) = x^2$ defined over $(-\infty, +\infty)$ is not one-to-one because $g(1) = 1 = g(-1)$. It is easy to see that the function fails the horizontal line test.



- The function $h(x) = x^2$ with the domain $[0, +\infty)$ is one-to-one. In fact, if $x_1^2 = x_2^2$ and both x_1 and x_2 are non-negative, we must have $x_1 = x_2$. It is also easy to verify that the function h passes the horizontal line test.



1.77 Definition: inverse function

One can regard a function f with domain D and range R as a map f from D to R . If f is one-to-one, then we can construct a map from R to D whose composition with f is the

identity map. The corresponding function of the new map is called the **inverse function**.

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

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

for any y in R . Thus,

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

Theorem: cancellation equations

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

$$\begin{aligned} f^{-1}(f(x)) &= x \quad \text{for every } x \text{ in } D; \\ f(f^{-1}(x)) &= x \quad \text{for every } x \text{ in } R. \end{aligned}$$

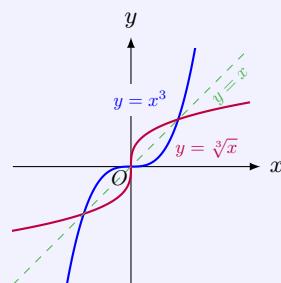
1.78 Method: find the inverse function

1. Write $y = f(x)$.
2. Solve this equation for x in terms of y (if possible).
3. To express f^{-1} as a function of x , interchange x and y . The resulting equation is $y = f^{-1}(x)$.

1.79 Example: inverse functions

1. The function $f(x) = x^3$ is a one-to-one function. The inverse function of f is $f^{-1}(x) = \sqrt[3]{x}$ since $f(f^{-1}(x)) = (\sqrt[3]{x})^3 = x$.

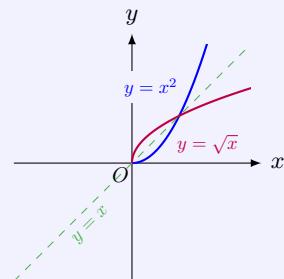
The inverse function f^{-1} can be found by the following steps: solving $y = x^3$ to get $x = \sqrt[3]{y}$, then interchanging x and y to have $y = \sqrt[3]{x}$.



One can see the graph of the inverse function $f^{-1}(x) = \sqrt[3]{x}$ is the mirror image of that of the function $f(x) = x^3$ with respect to the diagonal line $y = x$ as the mirror.

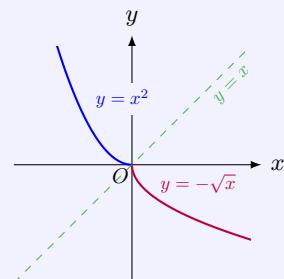
2. The function $h(x) = x^2$ defined on the domain $[0, +\infty)$ is one-to-one. Its inverse function is $h^{-1}(x) = \sqrt{x}$.

One also sees the graph of the inverse function $h^{-1}(x) = \sqrt{x}$ is the mirror image of that of the function $h(x) = x^2$ with respect to the diagonal line $y = x$ as the mirror.



3. The function $p(x) = x^2$ defined on the domain $[-\infty, 0)$ is one-to-one. Its inverse function is $p^{-1}(x) = -\sqrt{x}$.

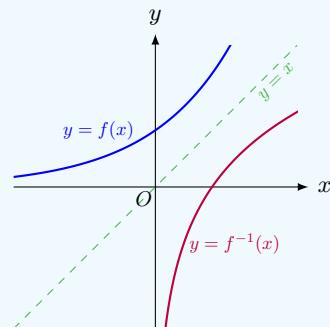
One again sees the graph of the inverse function $p^{-1}(x) = -\sqrt{x}$ is the mirror image of that of the function $p(x) = x^2$ with respect to the diagonal line $y = x$ as the mirror.



1.80 Method:

obtain 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 $y = x$.



1.81 Theorem:

continuity of the inverse function

The inverse function of any *continuous* one-to-one function is also continuous.

1.82 MATLAB:

inverse function

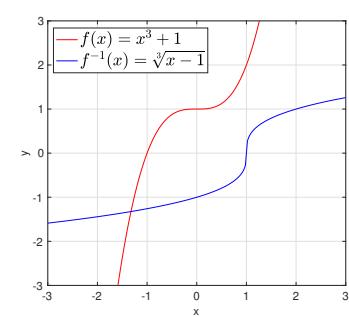
In MATLAB, the inverse function can be calculated using command `finverse`. The following example shows how to calculate the inverse function of $f(x) = x^3 + 1$ in MATLAB.

```
>> syms x;
>> f = x^3 + 1;
>> finverse(f)
```

```
ans =
(x - 1)^(1/3)
```

The following figure plots both the function $f(x) = x^3 + 1$ and its inverse function $f^{-1}(x) = \sqrt[3]{x - 1}$.

```
>> x = -3:0.02:3;
>> y = x.^3 + 1;
>> y_i = nthroot(x - 1,3);
>> figure
>> plot(x,y, 'r')
>> hold on
>> plot(x,y_i, 'b')
>> axis([-3 3 -3 3]) % Change the axis limits so that both ...
    the x-axis and the y-axis range from -3 to 3
>> xlabel('x')
>> ylabel('y')
>> grid
>> legend({'$f(x) = x^3 + 1$', '$f^{-1}(x) = ...',
    '\sqrt[3]{x-1}$'}, 'Interpreter','latex')
```



1.83 Definition: logarithmic function

Since the exponential function a^x is one-to-one, it has an inverse. The inverse function of $f(x) = a^x$ is called **logarithmic function with base a** , denoted by $\log_a x$. We use $\ln x$ for $\log_e x$.

1.84 Theorem: Laws of Logarithms

Let a, b be positive numbers not equal to 1, A, B be positive numbers, and α be a real numbers.

$$1. \log_a(A \cdot B) = \log_a A + \log_a B.$$

$$2. \log_a\left(\frac{A}{B}\right) = \log_a A - \log_a B.$$

$$3. \log_a A^\alpha = \alpha \log_a A.$$

$$4. a^{\log_a A} = A.$$

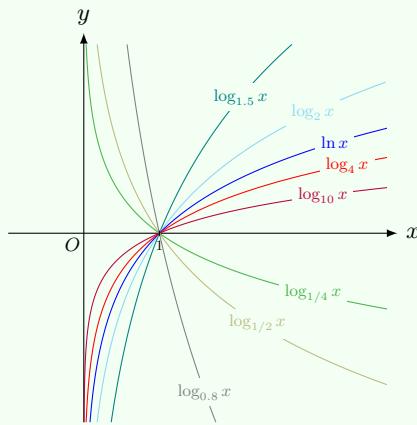
$$5. \log_a A = \frac{\log_b A}{\log_b a}.$$

$$6. \log_a a = 1, \log_a 1 = 0.$$

1.85 Theorem: properties of logarithmic functions

Assume a is positive number not equal to 1.

1. The logarithmic function $f(x) = \log_a x$ has domain $(0, +\infty)$ and range \mathbb{R} .
2. The logarithmic function $f(x) = \log_a x$ is a continuous function.
3. The graph of the logarithmic function $f(x) = \log_a x$ depends on the value of a .
4. If $a > 1$, then $\lim_{x \rightarrow +\infty} \log_a x = +\infty$ and $\lim_{x \rightarrow 0^+} \log_a x = -\infty$; if $0 < a < 1$, then $\lim_{x \rightarrow +\infty} \log_a x = -\infty$ and $\lim_{x \rightarrow 0^+} \log_a x = +\infty$.
5. The following are the graphs of $\log_a x$ for various bases.

**1.86 Example:** logarithmic functions

Let us look at some examples using the fact that exponential functions and logarithmic functions are inverse to each other.

1. To solve the equation $\ln(x - 1) = 0$, we take exponential on both sides of the equation simultaneously to have $e^{\ln(x-1)} = e^0$. Using the identity $e^{\ln a} = a$, we get $x - 1 = 1$, so that $x = 2$.

2. To solve the equation $2^{x-1} = 5$, we take logarithm on both sides of the equation simultaneously to have $\log_2 2^{x-1} = \log_2 5$. Using the identity $\log_a a^\alpha = \alpha$, we get $x - 1 = \log_2 5$, so that

$$x = 1 + \log_2 5 = \log_2 2 + \log_2 5 = \log_2 10.$$

1.87 MATLAB: logarithmic functions

MATLAB provides commands for logarithmic functions with different bases, e.g., `log2` for base 2, `log` for base e , and `log10` for base 10. The following examples show functions $f(x) = \log_2(x)$, $g(x) = \ln(x)$ and $h(x) = \log_{10}(x)$ in MATLAB.

```
>> syms x;
>> f = log2(x)

f =

log(x)/log(2)

>> g = log(x)

g =

log(x)

>> h = log10(x)

h =

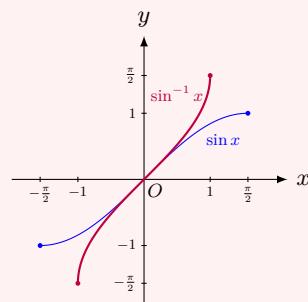
log(x)/log(10)
```

Inverse Trigonometric Functions

1.88 Definition: inverse trigonometric functions

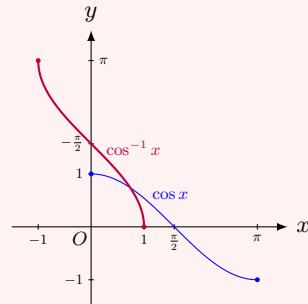
- The function $\sin x$ restricted on $[-\frac{\pi}{2}, \frac{\pi}{2}]$ has unique inverse function, denoted by $\sin^{-1} x$ or $\arcsin x$, that has domain $[-1, 1]$ and range $[-\frac{\pi}{2}, \frac{\pi}{2}]$. That is,

$$\begin{aligned}\sin^{-1} x &= y \quad (|x| \leq 1) \\ \iff \sin y &= x, \quad -\frac{\pi}{2} \leq y \leq \frac{\pi}{2}.\end{aligned}$$



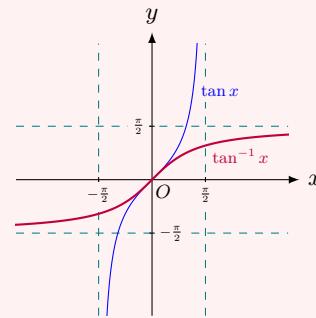
- The function of $\cos x$ restricted on $[0, \pi]$ has unique inverse function, denoted by $\cos^{-1} x$ or $\arccos x$, that has domain $[-1, 1]$ and range $[0, \pi]$. That is,

$$\begin{aligned}\cos^{-1} x &= y \quad (|x| \leq 1) \\ \iff \cos y &= x, \quad 0 \leq y \leq \pi.\end{aligned}$$



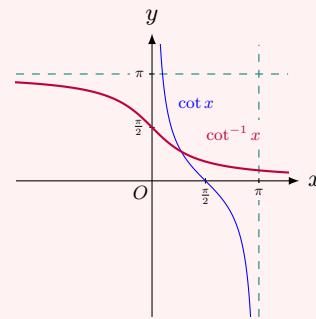
- The function of $\tan x$ restricted on $(-\frac{\pi}{2}, \frac{\pi}{2})$ has unique inverse function, denoted by $\tan^{-1} x$ or $\arctan x$, that has domain $(-\infty, +\infty)$ and range $(-\frac{\pi}{2}, \frac{\pi}{2})$. That is,

$$\begin{aligned}\tan^{-1} x &= y \quad (x \in \mathbb{R}) \\ \iff \tan y &= x, \quad -\frac{\pi}{2} < y < \frac{\pi}{2}.\end{aligned}$$



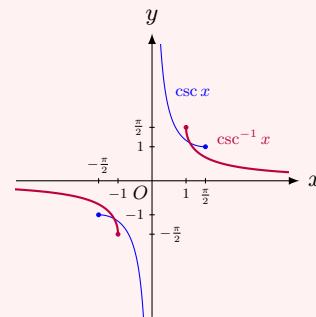
- The function of $\cot x$ restricted on $(0, \pi)$ has unique inverse function, denoted by $\cot^{-1} x$ or $\text{arccot } x$, that has domain $(-\infty, \infty)$ and range $(0, \pi)$. That is,

$$\begin{aligned}\cot^{-1} x &= y \quad (x \in \mathbb{R}) \\ \iff \cot y &= x, \quad y \in (0, \pi).\end{aligned}$$



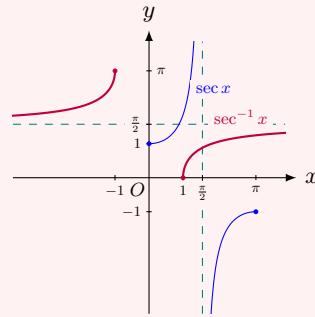
- The function of $\csc x$ restricted on $[-\frac{\pi}{2}, 0) \cup (0, \frac{\pi}{2}]$, has unique inverse function, denoted by $\csc^{-1} x$ or $\text{arccsc } x$, that has domain $|x| \geq 1$ and range $[-\frac{\pi}{2}, 0) \cup (0, \frac{\pi}{2}]$. That is,

$$\begin{aligned}\csc^{-1} x &= y \quad (|x| \geq 1) \\ \iff \csc y &= x, \quad y \in [-\frac{\pi}{2}, 0) \cup (0, \frac{\pi}{2}].\end{aligned}$$



- The function of $\sec x$ restricted on $[0, \frac{\pi}{2}) \cup (\frac{\pi}{2}, \pi]$, has unique inverse function, denoted by $\sec^{-1} x$ or $\text{arcsec } x$, that has domain $|x| \geq 1$ and range $[0, \frac{\pi}{2}) \cup (\frac{\pi}{2}, \pi]$. That is,

$$\begin{aligned}\sec^{-1} x &= y \quad (|x| \geq 1) \\ \iff \sec y &= x, \quad y \in [0, \frac{\pi}{2}) \cup (\frac{\pi}{2}, \pi].\end{aligned}$$



1.89 Example: inverse trigonometric functions

- Calculate $\sin^{-1} \left(\frac{\sqrt{2}}{2} \right)$.

Solution Let $\theta = \sin^{-1} \left(\frac{\sqrt{2}}{2} \right)$. Then, by the definition of \sin^{-1} , θ is an angle in $[-\frac{\pi}{2}, \frac{\pi}{2}]$ and $\sin \theta = \frac{\sqrt{2}}{2}$. So $\theta = \frac{\pi}{4}$.

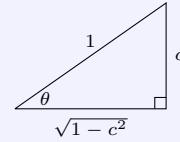
- Calculate $\cos^{-1} \left(-\frac{\sqrt{2}}{2} \right)$.

Solution Let $\theta = \cos^{-1} \left(-\frac{\sqrt{2}}{2} \right)$. Then, by the definition of \cos^{-1} , θ is an angle in $[0, \pi]$ and $\cos \theta = -\frac{\sqrt{2}}{2}$. So $\theta = \frac{3\pi}{4}$.

- Find $\cos(\sin^{-1} c)$ for c in $[0, 1]$.

Solution Let $\theta = \sin^{-1} c$, then $\sin \theta = c$. We can draw a triangle with the opposite side of length c and hypotenuse of length 1. Then the adjacent side has length $\sqrt{1 - c^2}$. Hence,

$$\cos(\sin^{-1} c) = \cos \theta = \sqrt{1 - c^2}/1 = \sqrt{1 - c^2}.$$



1.90 MATLAB: trigonometric and inverse trigonometric functions

In MATLAB, there are several commands for calculating the trigonometric and inverse trigonometric functions, which are provided in the following table.

Command	Function
<code>sin</code>	Sine function
<code>cos</code>	Cosine function
<code>tan</code>	Tangent function
<code>cot</code>	Cotangent function
<code>sec</code>	Secant function
<code>csc</code>	Cosecant function
<code>asin</code>	Inverse sine function
<code>acos</code>	Inverse cosine function
<code>atan</code>	Inverse tangent function
<code>acot</code>	Inverse cotangent function
<code>asec</code>	Inverse secant function
<code>acsc</code>	Inverse cosecant function

The following examples show the calculations of $\sin\left(\frac{\pi}{2}\right)$ and $\tan^{-1}(1)$, respectively, in MATLAB.

```
>> sin(pi/2)

ans =
1

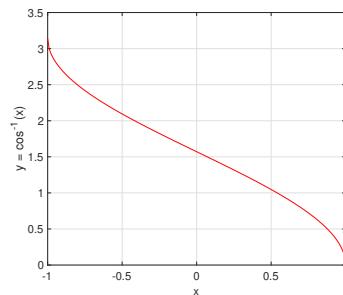
>> atan(1)

ans =
0.7854
```

Besides calculating the values of the trigonometric and inverse trigonometric functions, graphs of the trigonometric and inverse trigonometric functions can also be easily plotted by MATLAB. The following code shows the plot of $\cos^{-1}(x)$ within the interval $[-1, 1]$:

```
>> x = -1:0.01:1;
>> y = acos(x);
>> figure
>> plot(x,y, 'r')
>> xlabel('x')
>> ylabel('y = cos^{-1}(x)')
>> grid
```

The plotted figure is shown below.



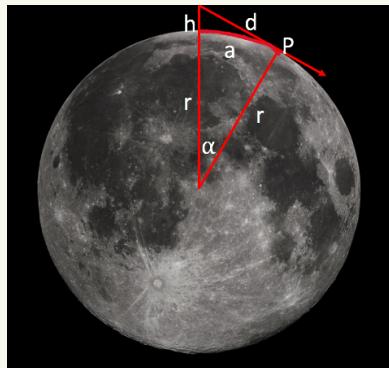
1.91 Theorem: continuity of basic elementary functions

The following types of basic elementary functions are continuous at every point in their *domains*:

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

1.92 Engineering Example: inverse tangent - aeronautics and astronautics

It is said that the National Aeronautics and Space Administration (NASA) astronauts will explore the surface of Moon again by 2020. This time, the astronauts are going to stay, build outposts, and pave a way for the journeys to Mars and beyond. It is thus critical to have a reliable and robust surface-to-surface communication between astronauts on Moon, and also keep constant communications with Earth. This will likely be accomplished by a combined use of communication satellites in orbit around Moon and communication equipments on the lunar surface. In the case of no satellite coverage over an outpost, the astronauts there will purely rely on the communication equipments on the lunar surface, e.g., communication towers. Since Moon is a sphere, the surface-to-surface communication is limited by the height of the communication tower. In particular, the communication signals cannot be received beyond the point of tangency P in the following figure.



Parameter	Description
r	radius of Moon (1738.14 km)
h	tower height
d	line of sight distance
P	point of tangency
a	arc length
α	central angle measure

In the case of no satellite coverage over a certain area on the lunar surface, if there is a communication tower with height 60 meters, then calculate the maximum distance that the astronaut can explore in this area.

First, we note that the area which the astronaut can explore should be within the communication range of the communication tower, i.e., not beyond the point of tangency P. Moreover, the distance that the astronaut can explore refers to the one on the lunar surface. Hence, the maximum distance is the arc length a in the above figure.

In order to calculate the arc length a , we first need to calculate the central angle measure α which can be obtained by using the **inverse tangent function**. Specifically, we have

$$\alpha = \tan^{-1} \left(\frac{d}{r} \right).$$

We have known the Moon radius $r = 1738.14$ km. Now, we need to calculate the line of sight distance d . We note that $r^2 + d^2 = (r + h)^2$, so we have

$$\begin{aligned} d &= \sqrt{(r + h)^2 - r^2} \\ &= \sqrt{(1738.14 + 0.06)^2 - 1738.14^2} \approx 14.44 \text{ (km)}. \end{aligned}$$

Substituting the value of d to the first equation, we obtain

$$\alpha = \tan^{-1} \left(\frac{14.44}{1738.14} \right) \approx 0.0083.$$

Finally, we can calculate the arc length a , i.e., the maximum distance that the astronaut can explore, as follows:

$$a = \frac{\theta}{2\pi} \times 2\pi r = 0.0083 \times 1738.14 \approx 14.43 \text{ (km)}.$$

1.93 Engineering Example: decibel

The decibel (dB) is used to measure the ratio of two quantities with the same unit. For example, in terms of decibel, the ratio of two quantities A and B , which have the same unit, is defined as

$$10 \log_{10} \left(\frac{A}{B} \right) \text{ (dB)}.$$

The decibel has been widely used in a variety of science and engineering fields, such as electronics and control theory. For example, in electronics, the gains of amplifiers, the signal-to-noise ratio (SNR), and the attenuations of signals are often represented in decibel.

Note that the decibel can be used to express the ratio of A with respect to any absolute value B or a fixed reference value B . In the latter case, a suffix which indicates the reference value is usually appended to the decibel symbol. For example, if the reference value B is 1 volt, then the suffix is "V", e.g., 10 dBV; if the reference value B is 1 milliwatt, then the suffix is "m", e.g., 10 dBm.

The use of the decibel has several advantages. First, since the decibel is on a logarithmic

scale, it means that a very big ratio can be represented in a convenient way. For example, if A is a *trillion* times of B , then in terms of decibel, we have

$$10 \log_{10} \left(\frac{A}{B} \right) = 10 \log_{10} 10^{12} = 120 \text{ (dB)},$$

which is a more compact representation than 1,000,000,000,000.

Second, due to the logarithmic scale property, the decibel can be used for simplifying the calculations of multiplied effects of a system. For example, a multi-component system has three components with gains $g_{1,\text{dB}}$ (dB), $g_{2,\text{dB}}$ (dB) and $g_{3,\text{dB}}$ (dB), respectively. Then the overall gain $g_{o,\text{dB}}$ (dB) of the system can be calculated as

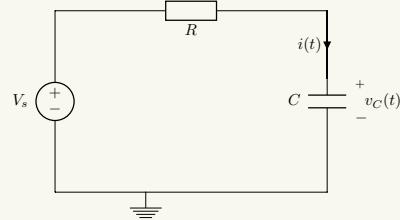
$$\begin{aligned} g_{o,\text{dB}} &= 10 \log_{10} (g_o) \\ &= 10 \log_{10} (g_1 g_2 g_3) \\ &= 10 \log_{10} \left(10^{\frac{g_{1,\text{dB}}}{10}} 10^{\frac{g_{2,\text{dB}}}{10}} 10^{\frac{g_{3,\text{dB}}}{10}} \right) \\ &= 10 \log_{10} \left(10^{\frac{g_{1,\text{dB}}}{10}} \right) + 10 \log_{10} \left(10^{\frac{g_{2,\text{dB}}}{10}} \right) + 10 \log_{10} \left(10^{\frac{g_{3,\text{dB}}}{10}} \right) \\ &= g_{1,\text{dB}} + g_{2,\text{dB}} + g_{3,\text{dB}}, \end{aligned}$$

where g_o , g_1 , g_2 and g_3 are the overall gain, the gain of the 1st component, the gain of the 2nd component and the gain of the 3rd component in the linear scale, respectively. The above additive operation is more convenient than the multiplication of multiple gains.

1.94 Engineering Example: RC charging circuit

The capacitor is an electronic component which can store electrical energy. If a resistor is connected in series with the capacitor to form a resistor-capacitor (RC) circuit, then the capacitor will charge up gradually through the resistor until the voltage across this capacitor reaches the value of the supply voltage.

An example of the RC circuit is shown, where R represents the resistance of the resistor, C represents the capacitance of the capacitor, V_s represents the supply voltage, and $v_C(t)$ represents the voltage across the capacitor.



Assume that $V_s = 10$ V, $R = 4.7 \times 10^4$ Ω, $C = 10^{-3}$ F, and the initial voltage (i.e., the voltage at $t = 0$) across the capacitor is 0, it can be shown that the voltage across the capacitor is given by

$$v_C(t) = V_s (1 - e^{-\frac{t}{RC}}) \quad (t \geq 0).$$

We can find the time taken for the capacitor to achieve $v_C(t) = 9.8$ V.

In fact, to achieve $v_C(t) = 9.8$ V, we need to solve the following equation:

$$V_s (1 - e^{-\frac{t}{RC}}) = 9.8.$$

Substituting the values of V_s , R and C into the equation above, we have

$$10(1 - e^{-\frac{t}{4.7 \times 10^4 \times 10^{-3}}}) = 9.8.$$

After some algebraic manipulations, we find that the time taken can be represented in terms of a *logarithmic function*, and it is shown as follows:

$$t = -47 \ln(1 - 0.98) \approx 183.86 \text{ (s).}$$

Note that the maximum voltage the capacitor can achieve is the supply voltage V_s . The result above thus shows that it takes 183.86 seconds which is around 3 minutes for the capacitor to virtually fully charged (i.e., achieve 98% of its maximum value).

Moreover, in RC circuit, the term $\tau = RC$ is defined as the *time constant*. In this example, $\tau = RC = 47$ (s). We notice that $\frac{183.86}{47} \approx 4$, which means the time taken for the capacitor to achieve 98% of its maximum voltage is 4 times of the time constant. The time period taken for the capacitor to reach this 4τ point (i.e., achieve 98% of the maximum voltage) is known as the *transient period*.

Summary

Definition: basic concepts

- **function :** $f: D \subset \mathbb{R} \rightarrow R$
(D = domain, R = range)
- **function continuous at a :** $\lim_{x \rightarrow a} f(x) = f(a)$
- **function continuous in I :** $\lim_{x \rightarrow a} f(x) = f(a)$ for every $a \in I$
- **composite function :** $f(g(x))$
- **one-to-one function :** if $x_1 \neq x_2$, then $f(x_1) \neq f(x_2)$

$\iff f$ passes the horizontal line test.
- **inverse function :** $f^{-1}(y) = x \iff f(x) = y$
(f has to be one-to-one)

Definition: limit

limit : $\lim_{x \rightarrow a} f(x) = L$
 $\iff f(x)$ approaches L when $x \in D$ approaches a

Definition: asymptotes

- **horizontal asymptote $y = b$** if at least one of the following is true:

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

- **vertical asymptote $x = a$** if at least one of the following is true:

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

Theorem: limit rules and theorems

Suppose $\lim_{x \rightarrow a} f(x)$ and $\lim_{x \rightarrow a} g(x)$ exist. Then the following rules hold.

1. Sum Rule:

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

2. Difference Rule:

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

3. Product Rule:

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

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

4. Quotient Rule: If $\lim_{x \rightarrow a} g(x) \neq 0$, then

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

5. Squeeze Theorem: If $f(x) \leq g(x) \leq h(x)$ and $\lim_{x \rightarrow a} f(x) = \lim_{x \rightarrow a} h(x) = L$, then $\lim_{x \rightarrow a} g(x) = L$.

6. Limit and One-Sided Limit: $\lim_{x \rightarrow a} f(x) = L$ if and only if $\lim_{x \rightarrow a^-} f(x) = \lim_{x \rightarrow a^+} f(x) = L$.

Theorem: continuity of basic elementary functions

The following types of basic elementary functions are continuous at every point in their domains:

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

Theorem: composite function and inverse function

The composite function of two continuous functions is a continuous function.

The inverse function of a one-to-one continuous function is a continuous function.

Theorem: continuity of arithmetic operations

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. $f \cdot g$;

5. $\frac{f}{g}$, provided that $g(a) \neq 0$.

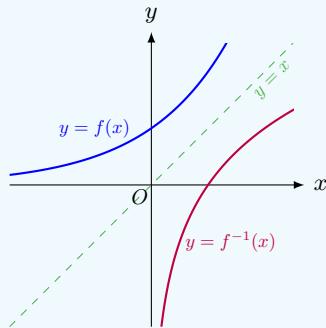
Theorem: two important limits

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

$$2. \lim_{x \rightarrow +\infty} \left(1 + \frac{1}{x}\right)^x = e.$$

Method: find and graph the inverse function

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 $y = x$.



Theorem: Laws of Exponents

Let a, b be positive numbers and α, β be real numbers.

$$1. a^\alpha \cdot a^\beta = a^{\alpha+\beta}.$$

$$4. (a \cdot b)^\alpha = a^\alpha \cdot b^\alpha.$$

$$2. \frac{a^\alpha}{a^\beta} = a^{\alpha-\beta}.$$

$$5. a^{-\alpha} = \frac{1}{a^\alpha}.$$

$$3. (a^\alpha)^\beta = a^{\alpha \cdot \beta}.$$

$$6. a^1 = a, a^0 = 1.$$

Theorem: Laws of Logarithms

Let a, b be positive numbers not equal to 1, A, B be positive numbers, and α be a real numbers.

$$1. \log_a(A \cdot B) = \log_a A + \log_a B.$$

$$4. a^{\log_a A} = A.$$

$$2. \log_a \left(\frac{A}{B}\right) = \log_a A - \log_a B.$$

$$5. \log_a A = \frac{\log_b A}{\log_b a}.$$

$$3. \log_a A^\alpha = \alpha \log_a A.$$

$$6. \log_a a = 1, \log_a 1 = 0.$$

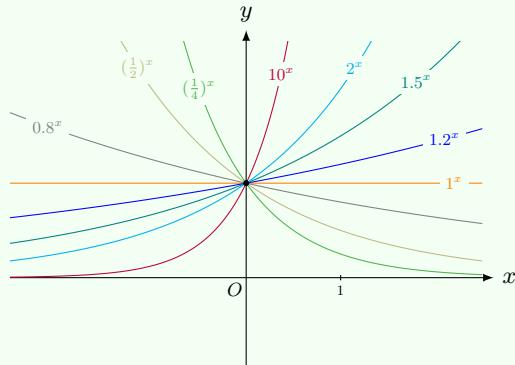
Theorem: properties of exponential functions

The exponential function is $f(x) = a^x$, where a is a positive number.

1. The exponential function $f(x) = a^x$ has domain \mathbb{R} and range $(0, +\infty)$.
2. The exponential function $f(x) = a^x$ is a continuous function.
3. The graph of the exponential function $f(x) = a^x$ depends the value of a .

4. If $a > 1$, then $\lim_{x \rightarrow +\infty} a^x = +\infty$ and $\lim_{x \rightarrow -\infty} a^x = 0$; if $0 < a < 1$, then $\lim_{x \rightarrow +\infty} a^x = 0$ and $\lim_{x \rightarrow -\infty} a^x = +\infty$.

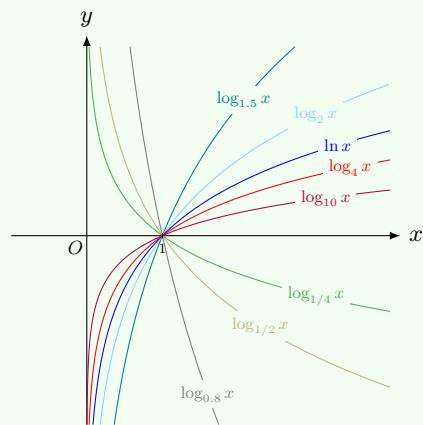
5. The following are the graphs of a^x for various bases.



Theorem: properties of logarithmic functions

Assume a is positive number not equal to 1.

1. The logarithmic function $f(x) = \log_a x$ has domain $(0, +\infty)$ and range \mathbb{R} .
2. The logarithmic function $f(x) = \log_a x$ is a continuous function.
3. The graph of the logarithmic function $f(x) = \log_a x$ depends the value of a .
4. If $a > 1$, then $\lim_{x \rightarrow +\infty} \log_a x = +\infty$ and $\lim_{x \rightarrow 0^+} \log_a x = -\infty$; if $0 < a < 1$, then $\lim_{x \rightarrow +\infty} \log_a x = -\infty$ and $\lim_{x \rightarrow 0^+} \log_a x = +\infty$.
5. The following are the graphs of $\log_a x$ for various bases.

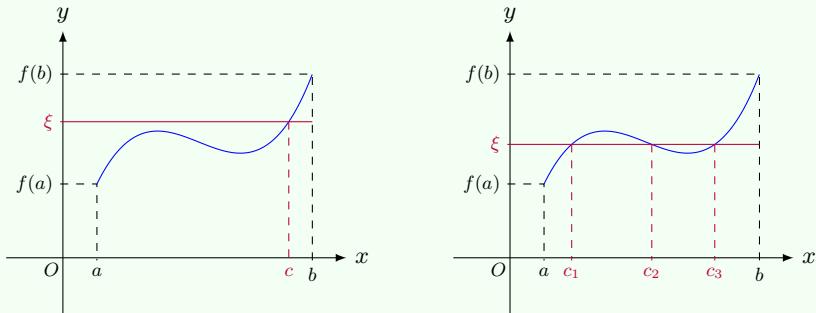


Definition: some inverse functions

- $y = \log_a x \iff x = a^y;$
- $y = \sin^{-1} x \ (|x| \leq 1) \iff \sin y = x, -\frac{\pi}{2} \leq y \leq \frac{\pi}{2};$
- $y = \cos^{-1} x \ (|x| \leq 1) \iff \cos y = x, 0 \leq y \leq \pi;$
- $y = \tan^{-1} x \ (x \in \mathbb{R}) \iff \tan y = x, -\frac{\pi}{2} < y < \frac{\pi}{2};$
- $y = \cot^{-1} x \ (x \in \mathbb{R}) \iff \cot y = x, y \in (0, \pi);$
- $y = \csc^{-1} x \ (|x| \geq 1) \iff \csc y = x, y \in [-\frac{\pi}{2}, 0) \cup (0, \frac{\pi}{2}];$
- $y = \sec^{-1} x \ (|x| \geq 1) \iff \sec y = x, y \in [0, \frac{\pi}{2}) \cup (\frac{\pi}{2}, \pi].$

Theorem: Intermediate Value Theorem

Let f be a *continuous* function over a *closed* interval $[a, b]$, for any number ξ between $f(a)$ and $f(b)$, there exists at least one c in the interval $[a, b]$ such that $f(c) = \xi$.



Differentiation

2.1 Derivatives

2.1 Example: instantaneous velocity

Let $y = f(t) = t^2$ be the distant function describing a movement of a particle on the x -axis and t is the time variable. The velocity of the particle is not a constant and depends upon the time t . So we call the velocity at time t the **instantaneous velocity** $v(t)$. How to calculate the instantaneous velocity? We know how to find the average velocity. So let's calculate the average velocity from time $t = 1$ to time t . The distance traveled by the particle within the time is $f(t) - f(1) = t^2 - 1^2$. The time spent is $t - 1$. Thus the average velocity is

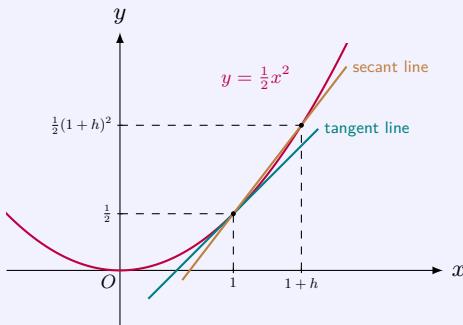
$$\frac{f(t) - f(1)}{t - 1} = \frac{t^2 - 1}{t - 1} = t + 1.$$

Clearly this average velocity doesn't equal the instantaneous velocity at $t = 1$. However, if we use the average velocity to approximate the instantaneous velocity at $t = 1$, the approximation is better if t is closer to 1. Using the idea of limits, it is reasonable to say that the instantaneous velocity $v(1)$ is the limit of the average velocity as t approaches 1, i.e.,

$$v(1) = \lim_{t \rightarrow 1} \frac{f(t) - f(1)}{t - 1} = \lim_{t \rightarrow 1} (t + 1) = 2.$$

2.2 Example: slope of tangent line

Consider the function $f(x) = \frac{1}{2}x^2$ and the tangent line of the graph of $y = f(x)$ at $x = 1$. We first consider the secant line of the graph at $x = 1$.



The secant line passes through two points $P_0 = (1, \frac{1}{2})$ and $P = (1 + h, \frac{1}{2}(1 + h)^2)$. The slope of the secant line is

$$S(h) = \frac{\frac{1}{2}(1 + h)^2 - \frac{1}{2}}{(1 + h) - 1} = \frac{h + \frac{1}{2}h^2}{h} = 1 + \frac{1}{2}h.$$

As the point P approaches P_0 , i.e., as h approaches 0, the secant line approaches the tangent line. In the language of limits, the tangent line is the limit of the secant line. Hence the limit of the slope of the secant line is the slope S of the tangent line, i.e.

$$S = \lim_{h \rightarrow 0} S(h) = \lim_{h \rightarrow 0} \frac{f(1 + h) - f(1)}{h} = \lim_{h \rightarrow 0} (1 + \frac{1}{2}h) = 1.$$

2.3 Remark: derivative - motivation

Given a function $y = f(x)$, take a point a in the domain of the function. The change in y from a to x is $f(x) - f(a)$. The average rate of change is $\frac{f(x) - f(a)}{x - a}$. The instantaneous rate of change at a is the limit of average rate of change when x approaches a , i.e.,

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

Since many important concepts such as velocity and slope of the tangent line of a function can be described in this way, we will call such kind of limit the derivative of the function at $x = a$.

2.4 Definition: derivative

Given a function f and a point a in the domain of f , the **derivative** of f at $x = a$ is

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

provided the limit exists. In this case, we say f **has a derivative** at $x = a$.

For single variable functions, the term “*being differentiable*” is used as synonymous with “*having a derivative*”.

2.5 Definition: notations of derivative

There are several widely used notations for derivatives.

- The derivative of $f(x)$ at $x = a$ is denoted by $f'(a)$.
- The limit $\lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a}$ can be written as

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

where we replace x by $a + h$. When x varies, h varies so that x approaches a if and only if h goes to zero.

- The limit $\lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a}$ can be written as

$$\lim_{\Delta x \rightarrow 0} \frac{f(a + \Delta x) - f(a)}{\Delta x},$$

where $\Delta x = x - a$ is the another symbol representing the change in x .

- We can also calculate the derivative of f at any point of the domain. Thus we define the derivative of the function at x :

$$f'(x) = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}.$$

- If we write the change in y as $\Delta y = f(x + \Delta x) - f(x)$, then the derivative can also be expressed as

$$f'(x) = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x}.$$

Hence the change in y is Δy , the change in x is Δx , the average rate of change is $\frac{\Delta y}{\Delta x}$, and the (instantaneous) rate of change of the function at x is the limit of the average rate of change as Δx approaches zero.

- Leibniz notation: $f'(x)$ can also be expressed as $\frac{df(x)}{dx}$. If we write $y = f(x)$, we can also write $\frac{dy}{dx}$ for $f'(x)$.
- We sometimes use $f'(x)|_{x=a}$, or $\left. \frac{df(x)}{dx} \right|_{x=a}$ to denote $f'(a)$.

2.6 Example: calculate derivatives by the definition

1. Find $f'(x)$ for $f(x) = x^2$.

Solution By the definition,

$$\begin{aligned} f'(x) &= \lim_{h \rightarrow 0} \frac{f(x + h) - f(x)}{h} = \lim_{h \rightarrow 0} \frac{(x + h)^2 - x^2}{h} \\ &= \lim_{h \rightarrow 0} \frac{2hx + h^2}{h} = \lim_{h \rightarrow 0} (2x + h) = 2x. \end{aligned}$$

In the above calculation, the changing variable is h , while x should be treated as a constant.

2. Find $f'(x)$ for $f(x) = \sqrt{x}$ where $x > 0$.

Solution By the definition,

$$\begin{aligned} f'(x) &= \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{\sqrt{x+h} - \sqrt{x}}{h} \\ &= \lim_{h \rightarrow 0} \frac{(\sqrt{x+h} - \sqrt{x})(\sqrt{x+h} + \sqrt{x})}{h(\sqrt{x+h} + \sqrt{x})} \\ &= \lim_{h \rightarrow 0} \frac{h}{h(\sqrt{x+h} + \sqrt{x})} \\ &= \lim_{h \rightarrow 0} \frac{1}{(\sqrt{x+h} + \sqrt{x})} = \frac{1}{2\sqrt{x}}. \end{aligned}$$

In the above calculation, again the changing variable is h , while x should be treated as a constant. When we calculate $\lim_{h \rightarrow 0} \sqrt{h+x} = \sqrt{x}$, we used the fact that the function $\sqrt{h+x}$ is a continuous function as a function of h .

Also from the calculation above, we see that $f'(0)$ does not exist.

2.7 MATLAB: derivative

In MATLAB, the derivative of a function can be calculated using command `diff`. Moreover, using the symbolic substitution command `subs`, we can obtain the value of the derivative at a certain point. The MATLAB code below shows the calculations of $f'(x)$, where $f(x) = x^2$, and the value of $f'(3)$.

```
>> syms x;
>> f = x^2;
>> d = diff(f)

d =
2*x

>> subs(d,x,3)

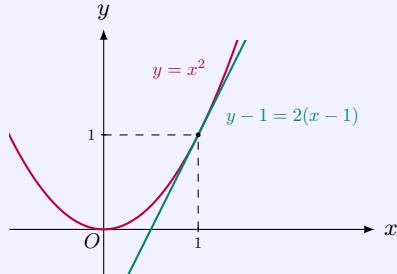
ans =
6
```

2.8 Example: tangent line

Find the equation of the tangent line of the graph of $y = x^2$ at $x = 1$.

Solution By using the definition, we can calculate the derivative $f'(1) = 2$, which is the slope of the tangent line. Since the tangent line passes through the point $(1, f(1)) = (1, 1)$, the equation of the tangent line is

$$y - 1 = 2 \cdot (x - 1).$$



2.9 Example: velocity

If we toss a ball straight up into the air with initial velocity 98 m/s, the distance (measured in meters) of the ball to the ground is given by the function $f(t) = -4.9t^2 + 98t$ where t is the time measured in seconds. Find the velocity of the ball at time t and find the time and the distance travelled when the ball reaches the highest height.

Solution Since the velocity $v(t)$ is the derivative of the distance function, we obtain

$$\begin{aligned} v(t) &= f'(t) = \lim_{h \rightarrow 0} \frac{[-4.9(h+t)^2 + 98(h+t)] - (-4.9t^2 + 98t)}{h} \\ &= \lim_{h \rightarrow 0} \frac{-4.9(2ht + h^2) + 98h}{h} \\ &= \lim_{h \rightarrow 0} [-4.9(2t+h) + 98] = -9.8t + 98. \end{aligned}$$

The ball reaches the highest height when the velocity is zero. So let $v(t) = 0$, we get $-9.8t + 98 = 0$ and $t = 10$. So in ten seconds, the ball reaches the highest height. The distance travelled during the ten seconds is $f(10) = -490 + 980 = 490$ meters.

Relationship Between Differentiability and Continuity

2.10 Theorem: differentiability and continuity

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.

Pitfall: continuous but not differentiable

Indeed, consider the function $f(x) = |x|$. It is continuous everywhere. However, since

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

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

we know that $\lim_{h \rightarrow 0} \frac{f(h) - f(0)}{h}$ does not exist. So, f is not differentiable at $x = 0$.

Justification:

If f is differentiable at a , then $\lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a} = f'(a)$ exists. Thus,

$$\begin{aligned} \lim_{x \rightarrow a} [f(x) - f(a)] &= \lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a} \cdot (x - a) \\ &= \lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a} \cdot \lim_{x \rightarrow a} (x - a) = f'(a) \cdot 0 = 0. \end{aligned}$$

Hence $\lim_{x \rightarrow a} f(x) = f(a)$, so that f is continuous at a .

2.11 Engineering Example: electric current

An electric current is the flow of electric charge. Electric currents can cause Joule heating which creates light in the incandescent light bulbs. Electric currents can also cause magnetic fields which are used in motor, inductors and generators.

In mathematics, the electric current can be defined as the rate at which the electric charge flows through a given surface, which is given as follows:

$$I = \frac{dQ}{dt},$$

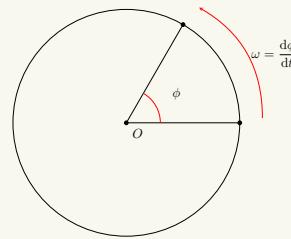
where I represents the electric current, Q represents the electric charge, and t represents the time.

2.12 Engineering Example: angular velocity

The angular velocity is a basic concept in physics. It is the rate at which a particle rotates around a center point. In other words, the angular velocity represents how fast an object goes around something. For example, the angular velocity of Earth represents how fast the Earth orbits the Sun.

Mathematically, the angular velocity is defined as the rate of a particle's angular displacement change relative to the origin and can be expressed as follows:

$$\omega = \frac{d\phi}{dt},$$



where ω represents the angular velocity, ϕ represents the angular displacement, and t represents the time. The angular velocity can be depicted in the figure.

Suppose a person is jogging on a circular track with angular displacement $\phi(t) = \frac{1}{20}t$ (radians). Calculate this person's angular velocity.

According to the definition of angular velocity, we have

$$\omega = \frac{d\phi(t)}{dt} = \frac{d(\frac{1}{20}t)}{dt} = \frac{1}{20} \text{ (radians/second)}.$$

2.13 Engineering Example: acceleration

Acceleration widely exists in our life. For example, when a car speeds up, the passengers might experience a force of pushing them back to seats. This feeling occurs because they are accelerating. In layman's words, acceleration means any process where the velocity (either the magnitude or the direction) changes. In particular, the magnitude of the velocity is called speed. Mathematically, the magnitude of the acceleration is defined as the rate of the speed change of an object with respect to time, and given as follows:

$$a = \frac{dv}{dt},$$

where a represents the acceleration, v represents the speed, and t represents the time.

Suppose a person is running on the track. At the starting stage, the speed of this person is $v(t) = t$ (meter/second). Calculate the acceleration of this person.

According to the definition of acceleration, we have

$$a = \frac{dv(t)}{dt} = \frac{dt}{dt} = 1 \text{ (meter/second}^2).$$

2.2 Rules of Differentiation

2.14 Theorem: rules of differentiation

Suppose $f'(x)$ and $g'(x)$ exist. Then the following rules of differentiation hold.

1. Sum Rule:

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

2. Difference Rule:

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

3. Product Rule:

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

4. Quotient Rule: If $g(a) \neq 0$, then

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

Justification:

By the definition, for the sum and difference rules, we have

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

For the product rule, we may re-write the derivative as follows:

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

If f is differentiable, f is continuous, so that

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

Hence,

$$\begin{aligned} & [f(x)g(x)]' \\ &= \lim_{h \rightarrow 0} f(x+h) \cdot \lim_{h \rightarrow 0} \frac{g(x+h) - g(x)}{h} + \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \cdot g(x) \\ &= f(x)g'(x) + g(x)f'(x). \end{aligned}$$

The quotient rule can be shown similarly.

2.15 Example: derivatives of polynomials

Show that

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

Solution We prove the equality in following steps.

1. If $f(x) \equiv c$, then $f'(x) = 0$.

In fact, by the definition,

$$f'(x) = \lim_{h \rightarrow 0} \frac{c - c}{h} = 0.$$

2. If $f(x) = x$, then $f'(x) = 1$.

In fact, by the definition,

$$f'(x) = \lim_{h \rightarrow 0} \frac{(x + h) - x}{h} = \lim_{h \rightarrow 0} \frac{h}{h} = 1.$$

3. If $f(x) = x^n$ for an positive integer n , then $f'(x) = nx^{n-1}$.

This can be proved by induction on n .

First, we know that the formula $(x^n)' = nx^{n-1}$ holds for $n = 1$.

Suppose the formula holds for $n = k$, i.e., $(x^k)' = kx^{k-1}$. Then, for $n = k + 1$, from the product rule,

$$\begin{aligned} f'(x) &= (x^{k+1})' = (x \cdot x^k)' \\ &= x' \cdot x^k + x \cdot (x^k)' \\ &= x^k + kx^k = (k+1)x^k. \end{aligned}$$

Hence, for any positive integer n , $(x^n)' = nx^{n-1}$.

4. If $f(x) = cg(x)$, then $f'(x) = cg'(x)$.

In fact, applying the product rule gives

$$(cg)' = c' \cdot g + c \cdot g' = 0 + c \cdot g' = cg'.$$

5. Let $f(x) = a_n x^n + \cdots + a_1 x + a_0$ be a polynomial. Then

$$f'(x) = na_n x^{n-1} + \cdots + a_1.$$

In fact,

$$\begin{aligned} f'(x) &= (a_n x^n)' + \cdots + (a_1 x)' + (a_0)' && \text{sum rule} \\ &= na_n x^{n-1} + \cdots + a_1 + 0 && \text{items 1, 3, 4} \\ &= na_n x^{n-1} + \cdots + a_1. \end{aligned}$$

2.16 Example: reciprocal rule

Suppose f is differentiable. By applying the quotient rule, we get the following *reciprocal rule*:

$$\left[\frac{1}{f(x)} \right]' = \frac{(1)' \cdot f(x) - 1 \cdot f'(x)}{f(x)^2} = -\frac{f'(x)}{f(x)^2}.$$

2.17 Example: horizontal tangent line

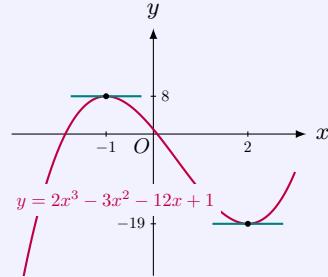
Find points where the tangent lines of $y = f(x) = 2x^3 - 3x^2 - 12x + 1$ are horizontal.

Solution A line is horizontal if and only if the slope of the line is zero. Since the slope of the tangent line is given by $f'(x)$, we just need to find points where $f'(x) = 0$.

Since

$$\begin{aligned} & (2x^3 - 3x^2 - 12x + 1)' \\ &= 2 \cdot 3x^2 - 3 \cdot 2x - 12 \\ &= 6(x^2 - x - 2) \\ &= 6(x - 2)(x + 1), \end{aligned}$$

we know that at $x = 2$ and $x = -1$, the tangent lines are horizontal.



2.18 MATLAB: rules of differentiation

With `diff`, we can easily calculate the derivatives of $f(x) + g(x)$, $f(x) - g(x)$, $f(x)g(x)$ and $\frac{f(x)}{g(x)}$. Following examples show the derivatives of $f(x)g(x)$ and $\frac{f(x)}{g(x)}$ in MATLAB.

```
>> syms f(x) g(x);
>> diff(f(x)*g(x),x)

ans =
f(x)*diff(g(x), x) + g(x)*diff(f(x), x)

>> diff(f(x)/g(x),x)

ans =
diff(f(x), x)/g(x) - (f(x)*diff(g(x), x))/g(x)^2
```

Differentiation of Trigonometric Functions

2.19 Theorem:

derivatives of trigonometric functions

- | | |
|---|--|
| 1. $(\sin x)' = \cos x;$
2. $(\cos x)' = -\sin x;$
3. $(\tan x)' = \sec^2 x;$ | 4. $(\csc x)' = -\csc x \cot x;$
5. $(\sec x)' = \sec x \tan x;$
6. $(\cot x)' = -\csc^2 x.$ |
|---|--|

Justification:

The derivatives of all six trigonometric functions can be found as follows.

1. By using some trigonometric identities (addition and double angle identities), continuity of $\sin x$, and the limit $\lim_{h \rightarrow 0} \frac{\sin h}{h} = 1$, we have

$$\begin{aligned}
 (\sin x)' &= \lim_{h \rightarrow 0} \frac{\sin(x+h) - \sin x}{h} \\
 &= \lim_{h \rightarrow 0} \frac{\sin x \cos h + \cos x \sin h - \sin x}{h} \\
 &= \lim_{h \rightarrow 0} \frac{\sin x(\cos h - 1) + \sin h \cos x}{h} \\
 &= \lim_{h \rightarrow 0} \left[\sin x \cdot \frac{-2\sin^2(h/2)}{h} + \cos x \cdot \frac{\sin h}{h} \right] \\
 &= \sin x \cdot \lim_{h \rightarrow 0} \left[-\sin(h/2) \cdot \frac{\sin(h/2)}{h/2} \right] + \cos x \cdot \lim_{h \rightarrow 0} \frac{\sin h}{h} \\
 &= 0 + \cos x = \cos x.
 \end{aligned}$$

2. By using some trigonometric identities (addition and double angle identities), continuity of $\sin x$, and the limit $\lim_{h \rightarrow 0} \frac{\sin h}{h} = 1$, we have

$$\begin{aligned}
 (\cos x)' &= \lim_{h \rightarrow 0} \frac{\cos(x+h) - \cos x}{h} \\
 &= \lim_{h \rightarrow 0} \frac{\cos x \cos h - \sin x \sin h - \cos x}{h} \\
 &= \lim_{h \rightarrow 0} \frac{\cos x(\cos h - 1) - \sin h \sin x}{h} \\
 &= \lim_{h \rightarrow 0} \left[\cos x \cdot \frac{-2\sin^2(h/2)}{h} - \sin x \cdot \frac{\sin h}{h} \right] \\
 &= \cos x \cdot \lim_{h \rightarrow 0} \left[-\sin(h/2) \cdot \frac{\sin(h/2)}{h/2} \right] - \sin x \cdot \lim_{h \rightarrow 0} \frac{\sin h}{h} \\
 &= 0 - \sin x = -\sin x.
 \end{aligned}$$

3. By applying the quotient rule, we have

$$\begin{aligned} (\tan x)' &= \left(\frac{\sin x}{\cos x} \right)' = \frac{(\sin x)' \cos x - \sin x(\cos x)'}{\cos^2 x} \\ &= \frac{\cos x \cos x - \sin x(-\sin x)}{\cos^2 x} = \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = \sec^2 x. \end{aligned}$$

4. By applying the quotient rule, we have

$$\begin{aligned} (\csc x)' &= \left(\frac{1}{\sin x} \right)' = \frac{(1)' \sin x - (\sin x)'}{\sin^2 x} \\ &= \frac{-\cos x}{\sin^2 x} = -\cot x \csc x. \end{aligned}$$

5. By applying the quotient rule, we have

$$\begin{aligned} (\sec x)' &= \left(\frac{1}{\cos x} \right)' = \frac{(1)' \cos x - (\cos x)'}{\cos^2 x} \\ &= \frac{\sin x}{\cos^2 x} = \tan x \sec x. \end{aligned}$$

6. By applying the quotient rule, we have

$$\begin{aligned} (\cot x)' &= \left(\frac{\cos x}{\sin x} \right)' = \frac{(\cos x)' \sin x - \cos x(\sin x)'}{\sin^2 x} \\ &= \frac{-\sin x \sin x - \cos x \cos x}{\sin^2 x} \\ &= \frac{-\cos^2 x - \sin^2 x}{\sin^2 x} = -\csc^2 x. \end{aligned}$$

2.20 MATLAB: derivatives of trigonometric functions

We can calculate the derivatives of trigonometric functions easily using command `diff`. Moreover, with the help of the variable precision arithmetic command `vpa` and the symbolic substitution command `subs`, we can calculate the value of the derivative of a trigonometric function at a certain point. The following example shows the calculations of $f'(x)$, where $f(x) = \sin(x)$, and $f'(1)$.

```
>> syms x;
>> f = sin(x);
>> d = diff(f)

d =
cos(x)
```

```

>> v1 = subs(d,x,1)
v1 =
cos(1)

>> vpa(v1)
ans =
0.54030230586813971740093660744298

```

2.21 Example: tangent line - trigonometric functions

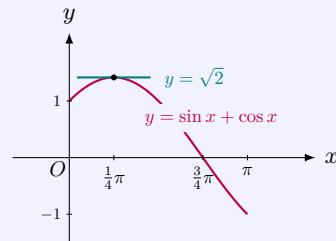
Find the equation of horizontal tangent line(s) of $y = f(x) = \sin x + \cos x$ in $(0, \pi)$.

Solution Horizontal tangent lines have zero slope. The slope of the tangent line is given by the derivative of f . Thus we calculate

$$f'(x) = (\sin x + \cos x)' = \cos x - \sin x = 0 \implies \tan x = 1.$$

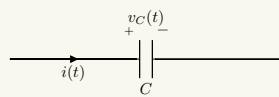
Solving the equation gives $x = \pi/4$. Since $f(\frac{\pi}{4}) = \sin \frac{\pi}{4} + \cos \frac{\pi}{4} = \sqrt{2}$, we get the equation of the horizontal tangent line

$$y = \sqrt{2}.$$



2.22 Engineering Example: electric current - capacitor

Consider a circuit with a capacitor below, where C is the capacitance of the capacitor, $i(t)$ is the current through the capacitor, and $v_C(t)$ is the voltage across the capacitor.



Given that $v_C(t) = 4 \cos t$, where $t \geq 0$, calculate the electric current $i(t)$ where $i(t) = C \frac{dv_C(t)}{dt}$.

First, we calculate $\frac{dv_C(t)}{dt}$, and we have

$$\frac{dv_C(t)}{dt} = -4 \sin t.$$

Therefore, the current $i(t)$ is given by

$$i(t) = -4C \sin t$$

where “-” specifies the direction of the current.

Chain Rule

2.23 Theorem: chain rule

If g is differentiable at a and f is differentiable at $g(a)$, then the composite function F , defined by $F(x) = f(g(x))$, is differentiable at a and the derivative $F'(a)$ is given by the product:

$$F'(a) = f'(g(a)) \cdot g'(a).$$

Or

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

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,

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

Thus,

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

Hence,

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

so that

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

Since g is differentiable at a , it must be continuous at a and thus $\lim_{\Delta x \rightarrow 0} g(a + \Delta x) = g(a)$ so that Δu goes to zero as Δx goes to zero. This gives that both $\varepsilon_1 \rightarrow 0$ and $\varepsilon_2 \rightarrow 0$ as $\Delta x \rightarrow 0$. Therefore,

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

This proves the chain rule.

2.24 Example: chain rule

1. Compute the derivative of $h(x) = \sin^2 x$.

Solution Write $y = f(u) = u^2$ and $u = g(x) = \sin x$. Then $h(x) = f(g(x))$. Thus, by the chain rule,

$$h'(x) = \frac{dy}{du} \Big|_{u=g(x)} \cdot g'(x) = 2u|_{u=g(x)} \cdot \cos x = 2 \sin x \cos x.$$

2. Compute the derivative of $h(x) = \sin(x^2)$.

Solution Write $y = f(u) = \sin u$ and $u = g(x) = x^2$. Then $h(x) = f(g(x))$. Thus, by the chain rule,

$$h'(x) = \frac{dy}{du} \Big|_{u=g(x)} \cdot g'(x) = \cos u|_{u=g(x)} \cdot 2x = 2x \cos(x^2).$$

2.25 MATLAB: chain rule

The following example shows the calculation of $F'(x)$, where $F(x) = f(g(x))$, in MATLAB. It verifies that the Chain Rule holds in general.

```
>> syms f(x) g(x);
>> F(x) = compose(f(x),g(x))

F(x) =

f(g(x))

>> diff(F(x))

ans =

D(f)(g(x))*diff(g(x), x)
```

2.26 Engineering Example: maximum wind speed

Consider the maximum wind speed at the site of the building. Denote the maximum wind speed as V (the wind direction is ignored for simplicity) and the probability that this maximum wind speed is no larger than v as $F_V(v)$. Empirical data indicate that, in many locations, $F_V(v)$ has the following form:

$$F_V(v) = e^{-\left(\frac{v}{u}\right)^{-k}}$$

where u and k are positive parameters.

Specifically, in Boston, plausible values of u and k are 49.4 miles per hour (mph) and 6.5, respectively. Hence, in Boston, $F_V(v)$ can be written as

$$F_V(v) = e^{-\left(\frac{v}{49.4}\right)^{-6.5}}$$

where the unit of the maximum wind speed V is mph.

Calculate $F'_V(v)$ by the chain rule.

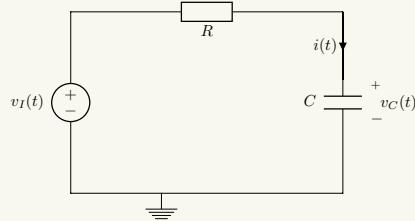
Let $f(v) = e^v$ and $g(v) = -\left(\frac{v}{49.4}\right)^{-6.5}$. Then, $F_V(v)$ can be rewritten as $F_V(v) = f(g(v))$. By the chain rule, we have

$$\begin{aligned} F'_V(v) &= f'(g(v))g'(v) \\ &= e^{g(v)}(-49.4^{6.5})(v^{-6.5})' \\ &= ae^{-\left(\frac{v}{49.4}\right)^{-6.5}}v^{-7.5}, \end{aligned}$$

where $a = 6.5 \times 49.4^{6.5} \approx 6.6395 \times 10^{11}$. Note that in probability theory and statistics, $F_V(v)$ is called the *cumulative distribution function* of V and $F'_V(v)$ is called the *probability density function* of V , which are both important functions to determine the statistical characteristics of V .

2.27 Engineering Example: electric current in RC circuit

Consider a resistor-capacitor (RC) circuit below, where R is the resistance of the resistor, C is the capacitance of the capacitor and $v_C(t)$ is the voltage across the capacitor.



Given that $v_C(t) = 1 - e^{-\frac{t}{RC}}$ ($t \geq 0$), calculate the electric current $i(t)$ where $i(t) = C \frac{dv_C(t)}{dt}$.

First, we calculate $\frac{dv_C(t)}{dt}$. Let $y = 1 - e^u$ and $u = -\frac{t}{RC}$. Then, $\frac{dv_C(t)}{dt}$ can be written as $\frac{dy}{dt}$. By the chain rule, we have

$$\frac{dy}{dt} = \frac{dy}{du} \cdot \frac{du}{dt} = -e^u \left(-\frac{1}{RC} \right) = \frac{1}{RC} e^{-\frac{t}{RC}}.$$

Substituting the above result into $i(t) = C \frac{dv_C(t)}{dt}$, we can obtain the electric current $i(t)$, as shown in the following:

$$i(t) = \frac{1}{R} e^{-\frac{t}{RC}}.$$

2.28 Engineering Example: reliability function

In reliability engineering, which is an interdisciplinary field of engineering and engineering management, the reliability function gives the probability that a component will survive beyond any given specified time. Suppose the reliability function of a component is given by

$$R(t) = e^{-\sqrt{t}} \quad (t \geq 0).$$

Calculate $\frac{d(1 - R(t))}{dt}$.

First, given the expression of $R(t)$, $\frac{d(1 - R(t))}{dt}$ can be written as

$$\frac{d(1 - R(t))}{dt} = \frac{d(1 - e^{-\sqrt{t}})}{dt}$$

Let $f(t) = 1 - e^t$ and $g(t) = -\sqrt{t}$. Then, $\frac{d(1 - R(t))}{dt}$ can be rewritten as

$$\frac{d(1 - R(t))}{dt} = \frac{df(g(t))}{dt}.$$

By the chain rule, we have

$$\frac{df(g(t))}{dt} = f'(g(t))g'(t) = -e^{g(t)} \left(-\frac{1}{2\sqrt{t}} \right) = \frac{1}{2\sqrt{t}}e^{-\sqrt{t}}.$$

Note that $R(t)$ and $\frac{d(1 - R(t))}{dt}$ are the *complementary cumulative distribution function* and the *probability density function* of the lifetime of the component, respectively. Besides the engineering field, the reliability function can also be used for the study in other fields, e.g., the duration of marriage, the death of a patient, etc. In these cases, the reliability function is commonly known as the survival function.

Differentiation of Inverse Functions

2.29 Theorem: differentiation of inverse function

If f is differentiable and has an inverse function f^{-1} over an interval I , then f^{-1} is also differentiable on I whose derivative equals

$$\frac{df^{-1}(x)}{dx} = \frac{1}{f'(f^{-1}(x))} = \frac{1}{\left. \frac{df(u)}{du} \right|_{u=f^{-1}(x)}}.$$

For short, the formula can be re-written as

$$\frac{dx}{dy} = \frac{1}{\frac{dy}{dx}}.$$

Justification:

We can apply the chain rule to prove the formula for differentiation of inverse functions. If applying the differentiation with respect to x to the identity

$$f(f^{-1}(x)) = x,$$

we obtain

$$\begin{aligned} \frac{df(f^{-1}(x))}{dx} &= 1 \implies \left. \frac{df(u)}{du} \right|_{u=f^{-1}(x)} \cdot \frac{df^{-1}(x)}{dx} = 1 \\ &\implies \frac{df^{-1}(x)}{dx} = \frac{1}{\left. \frac{df(u)}{du} \right|_{u=f^{-1}(x)}}. \end{aligned}$$

2.30 Example: derivative of $\sqrt[n]{x}$

Let $f(x) = x^n$ with n a positive integer, then $f^{-1}(x) = \sqrt[n]{x}$. We write $f(u) = u^n$. Thus $\frac{df(u)}{du} = nu^{n-1}$. Thus we get $\left. \frac{df(u)}{du} \right|_{u=f^{-1}(x)} = n(\sqrt[n]{x})^{n-1}$, and

$$\left(x^{\frac{1}{n}} \right)' = (\sqrt[n]{x})' = \frac{1}{n(\sqrt[n]{x})^{n-1}} = \frac{1}{n} x^{\frac{1}{n}-1}.$$

2.31 Example: derivative of x^a

Consider the function x^a where a is a rational number. If a is positive, we can write $a = n/m$, where n and m are positive integers, so that

$$x^a = (\sqrt[m]{x})^n.$$

Write $g(x) = \sqrt[m]{x}$ and $f(u) = u^n$. Then $x^a = f(g(x))$. Applying the chain rule, we get

$$\begin{aligned} (x^a)' &= \left. \frac{du^n}{du} \right|_{u=\sqrt[m]{x}} \cdot (\sqrt[m]{x})' = n(\sqrt[m]{x})^{n-1} \cdot \frac{1}{m} x^{\frac{1}{m}-1} \\ &= \frac{n}{m} x^{\frac{n-1}{m} + \frac{1}{m} - 1} = \frac{n}{m} x^{\frac{n}{m}-1} = ax^{a-1}. \end{aligned}$$

If a is negative, we can write $x^a = x^{-b}$, where $b = -a$ is a positive rational number. Then

$x^a = \frac{1}{x^b}$. Applying the quotient rule gives

$$(x^a)' = -\frac{(x^b)'}{(x^b)^2} = -\frac{bx^{b-1}}{x^{2b}} = -bx^{-b-1} = ax^{a-1}.$$

In summary, for any rational number a , we have

$$(x^a)' = ax^{a-1}.$$

2.32 Theorem:

derivatives of inverse trigonometric functions

$$1. (\sin^{-1} x)' = \frac{1}{\sqrt{1-x^2}};$$

$$4. (\csc^{-1} x)' = -\frac{1}{|x|\sqrt{x^2-1}};$$

$$2. (\cos^{-1} x)' = -\frac{1}{\sqrt{1-x^2}};$$

$$5. (\sec^{-1} x)' = \frac{1}{|x|\sqrt{x^2-1}};$$

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

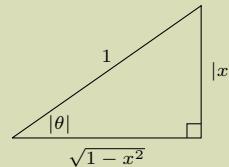
$$6. (\cot^{-1} x)' = -\frac{1}{1+x^2}.$$

Justification:

We can use the formula of derivative of inverse functions to calculate the derivatives of the inverse trigonometric functions. For example,

$$(\sin^{-1} x)' = \frac{1}{(\sin u)'|_{u=\sin^{-1}(x)}} = \frac{1}{\cos(\sin^{-1} x)}.$$

The last expression can be simplified as follows. Let $\theta = \sin^{-1} x$, then $\sin \theta = x$. Draw a right triangle with one angle $|\theta|$ whose opposite side has length $|x|$ and hypotenuse has length 1. Then the adjacent side has length $\sqrt{1-x^2}$. Thus, $\cos \theta = \sqrt{1-x^2}$.



Hence,

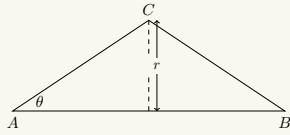
$$(\sin^{-1} x)' = \frac{1}{\cos \theta} = \frac{1}{\sqrt{1-x^2}}.$$

The other five derivatives can be calculated similarly.

2.33 Engineering Example:

roof truss design

There is a symmetrical roof truss shown below, where the span AB is 16 m.



Now, we need to design the rise r . Find the rate at which the angle θ between the rafter AC and the span AB changes with the rise r , i.e., find $\frac{d\theta}{dr}$.

First, let the length of the span AB be l_{AB} . We note that $\tan \theta = \frac{r}{\frac{1}{2}l_{AB}} = \frac{r}{8}$. So, we have

$$\theta = \tan^{-1}\left(\frac{r}{8}\right).$$

Let $u = \frac{r}{8}$, $f(x) = \tan x$, then $\theta = f^{-1}(u)$. According to the chain rule, we have

$$\begin{aligned}\frac{d\theta}{dr} &= \frac{df^{-1}(u)}{dr} \\ &= \frac{df^{-1}(u)}{du} \frac{du}{dr}.\end{aligned}$$

Moreover, according to the theorem for the differentiation of inverse function, we can calculate

$$\begin{aligned}\frac{df^{-1}(u)}{du} &= \frac{1}{\frac{df(x)}{dx} \Big|_{x=f^{-1}(u)}} \\ &= \frac{1}{\sec^2(\tan^{-1}(u))} \\ &= \frac{\cos^2(\tan^{-1}(u))}{\sin^2(\tan^{-1}(u)) + \cos^2(\tan^{-1}(u))} \\ &= \frac{1}{1 + \tan^2(\tan^{-1}(u))} \\ &= \frac{1}{1 + u^2}.\end{aligned}$$

Therefore, we can get

$$\begin{aligned}\frac{d\theta}{dr} &= \frac{1}{1+u^2} \frac{du}{dr} \\ &= \frac{1}{1+\left(\frac{r}{8}\right)^2} \frac{1}{8} \\ &= \frac{8}{r^2+64}.\end{aligned}$$

2.3 Differentiation of Exponential and Logarithmic Functions

2.34 Theorem: derivatives of exponential functions

$$(e^x)' = e^x.$$

In general, for any $a > 0$,

$$(a^x)' = a^x \ln a.$$

Justification:

Since $e = \lim_{x \rightarrow +\infty} \left(1 + \frac{1}{x}\right)^x$, by making a change of variable, we get

$$e = \lim_{y \rightarrow 0^+} (1 + y)^{1/y}.$$

From the continuity of the logarithmic function, we have

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

Denote $y = e^h - 1$. It is easy to see that $y \rightarrow 0^+$ is equivalent to $h \rightarrow 0^+$. Thus, we get

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

We further get

$$\lim_{h \rightarrow 0^-} \frac{h}{e^h - 1} = \lim_{h \rightarrow 0^+} \frac{-h}{e^{-h} - 1} = \lim_{h \rightarrow 0^+} \frac{he^h}{e^h - 1} = 1.$$

So, we know that $\lim_{h \rightarrow 0} \frac{h}{e^h - 1} = 1$. Hence,

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

In general,

$$\begin{aligned} (a^x)' &= \lim_{h \rightarrow 0} \frac{a^{x+h} - a^x}{h} = a^x \cdot \lim_{h \rightarrow 0} \frac{a^h - 1}{h} \\ &= a^x \ln a \cdot \lim_{h \rightarrow 0} \frac{e^{h \ln a} - 1}{h \ln a} = a^x \ln a. \end{aligned}$$

2.35 Theorem: derivatives of logarithmic functions

$$(\ln x)' = \frac{1}{x}.$$

In general, for any $a > 0$, $a \neq 1$,

$$(\log_a x)' = \frac{1}{x \ln a}.$$

Justification:

Let $f(u) = a^u$ and $u = \log_a x$. Then the equality $a^{\log_a x} = x$ can be re-written as $a^u = x$. Applying the chain rule, we get

$$a^u \ln a \cdot \frac{du}{dx} = 1.$$

Hence,

$$(\log_a x)' = \frac{du}{dx} = \frac{1}{a^u \ln a} = \frac{1}{x \ln a}.$$

2.36 Example: derivative of $\ln|x|$

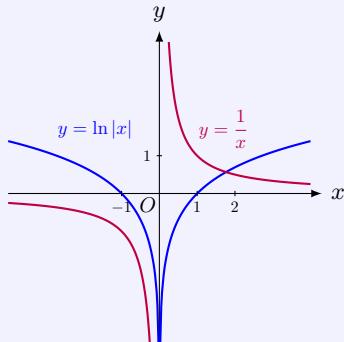
Show that $(\ln|x|)' = \frac{1}{x}$ for any $x \neq 0$.

Solution In fact, we know that for $x > 0$,

$$(\ln|x|)' = (\ln x)' = \frac{1}{x}.$$

If $x < 0$, then, by the chain rule,

$$(\ln|x|)' = [\ln(-x)]' = \frac{1}{-x} \cdot (-1) = \frac{1}{x}.$$



2.37 Example: derivative of composite function

Calculate the derivative of $\ln(a^x + 1)$.

Solution By applying the chain rule, we have

$$[\ln(a^x + 1)]' = \frac{1}{a^x + 1} \cdot (a^x + 1)' = \frac{a^x \ln a}{a^x + 1}.$$

Logarithmic Differentiation

2.38 Method: logarithmic differentiation

1. Take natural logarithms of both sides of an equation $y = f(x)$ and use the Laws of Logarithms to simplify.
2. Differentiate implicitly with respect to x .
3. Solve the resulting equation for y' .

2.39 Example: logarithmic differentiation

The “*logarithmic differentiation*” method can often be used to calculate the derivatives for exponential functions in the form $[u(x)]^{v(x)}$, or products $u_1(x)u_2(x)\cdots u_k(x)$ with relatively larger k .

1. Calculate the derivative of x^μ where $x > 0$ and μ is a real constant.

Solution Let $f(x) = x^\mu$. Taking logarithm of both sides gives

$$\ln f(x) = \ln x^\mu = \mu \ln x.$$

Then we can take derivative of both sides of the last equation and apply the chain rule to obtain

$$\frac{1}{f(x)} f'(x) = \mu \cdot \frac{1}{x} = \mu x^{-1}.$$

Thus,

$$f'(x) = f(x) \cdot \mu x^{-1} = \mu x^{\mu-1}.$$

2. Calculate the derivative of x^x where $x > 0$.

Solution Let $f(x) = x^x$. Taking logarithm of both sides gives

$$\ln f(x) = \ln x^x = x \ln x.$$

Then we can take derivative of both sides of the last equation and apply the chain rule to obtain

$$\frac{1}{f(x)} f'(x) = \ln x + x \cdot \frac{1}{x} = \ln x + 1.$$

Thus,

$$f'(x) = f(x) \cdot (\ln x + 1) = x^x (\ln x + 1).$$

3. Calculate the derivative of $f(x) = \frac{e^x \sqrt{1-x^2}}{1+x+x^2}$.

Solution We can calculate this derivative directly using the product rule and the quotient rule. Since there are many terms, it is quite involved. In the following, we make use of logarithm for a simpler calculation. In fact, we have

$$\begin{aligned}\ln f(x) &= \ln \frac{e^x \sqrt{1-x^2}}{1+x+x^2} \\ &= x + \frac{1}{2} \ln(1-x^2) - \ln(1+x+x^2).\end{aligned}$$

Taking derivative of both sides of the equation and applying the chain rule, we obtain

$$\frac{1}{f(x)} f'(x) = 1 - \frac{x}{1-x^2} - \frac{1+2x}{1+x+x^2},$$

so that

$$f'(x) = \frac{e^x \sqrt{1-x^2}}{1+x+x^2} \left(1 - \frac{x}{1-x^2} - \frac{1+2x}{1+x+x^2} \right).$$

2.40 Engineering Example: atmospheric pressure

The atmospheric pressure on earth changes with the altitude of the surface. The weather conditions, such as the temperature and the humidity, also affect the atmospheric pressure, so it is necessary to know these to calculate an accurate atmospheric pressure at a certain location. However, for general conditions, the National Aeronautics and Space Administration (NASA) has averaged the conditions for all parts of the earth year-round, and it is shown that the atmospheric pressure decreases as the altitude increases.

Specifically, at low altitudes above sea level (e.g., the first 1000 meters above sea level), the atmospheric pressure decreases by approximately 11.3 Pa per meter. For higher altitudes where the standard temperature lapse rate (i.e., the rate at which the temperature changes with the altitude) is zero, the atmospheric pressure is shown in the following equation:

$$P(h) = P_b e^{-\frac{gM}{RT_b}(h-h_b)}$$

where

Parameter	Explanation
P_b	static pressure of layer b (Pa)
g	gravitational acceleration : 9.80665 (m/s ²)
M	molar mass of Earth's air : 0.0289644 (kg/mol)
R	universal gas constant : 8.3144598 (J/mol/K)
T_b	standard temperature of layer b (K)
h_b	height at bottom of layer b (m)

with $b \in [0, 6]$ ($b \in \mathbb{Z}$) in accordance with each of seven successive layers of the atmosphere.

Calculate $\frac{dP(h)}{dh}$.

In $P(h)$, we treat h as the independent variable and all the other parameters as constants. Then, by the chain rule, we have

$$\begin{aligned}\frac{dP(h)}{dh} &= P_b e^{-\frac{gM}{RT_b}(h-h_b)} \left(-\frac{gM}{RT_b} (h - h_b) \right)' \\ &= -\frac{gM}{RT_b} P_b e^{-\frac{gM}{RT_b}(h-h_b)}.\end{aligned}$$

Note that the result above represents the rate at which the atmospheric pressure changes with the altitude.

For example, in the stratosphere (i.e., $b = 1$), the standard temperature lapse rate is 0. We thus can use the above expressions of $P(h)$ and $\frac{dP(h)}{dh}$ to calculate the atmospheric pressure and its changing rate, respectively. In particular, when $b = 1$, we have

Parameter	Value
P_1	22632.10 (Pa)
T_1	216.65 (K)
h_1	11000 (m)

Therefore, $P(h)$ and $\frac{dP(h)}{dh}$ in the stratosphere can be calculated as follows:

$$\begin{aligned}P(h) &= 22632.10 e^{-\frac{9.80665 \times 0.0289644}{8.3144598 \times 216.65} (h-11000)} \\ &\approx 2.2 \times 10^4 e^{-1.6 \times 10^{-4} (h-11000)} \text{ (Pa)}, \\ \frac{dP(h)}{dh} &= -\frac{9.80665 \times 0.0289644}{8.3144598 \times 216.65} \times 22632.10 \times e^{-\frac{9.80665 \times 0.0289644}{8.3144598 \times 216.65} (h-11000)} \\ &\approx -3.6 e^{-1.6 \times 10^{-4} (h-11000)} \text{ (Pa/m)}.\end{aligned}$$

Note that the sign “-” in $\frac{dP(h)}{dh}$ means that the atmospheric pressure *decreases* with the altitude.

2.41 Engineering Example: entropy - information theory

First introduced by Claude Shannon in 1948, in information theory, the entropy is used to *quantitatively* measure the information which is produced by a stochastic source of data. Specifically, when a low-probability event occurs at the source, then the event carries more information than when a high-probability event occurs. The amount of information conveyed by each event becomes a random variable whose expected value is the entropy. Hence, the low-probability event has more entropy. In general, the more uncertain a event is, the more information the event carries and the more the entropy of this event is. In other words, the entropy refers to uncertainty or disorder.

For a communication system, if it has two inputs 0 and 1 and they occur with probability

p and $1 - p$ ($0 < p < 1$), respectively, then the entropy of the source is defined as

$$H(p) = -p \log_2 p - (1-p) \log_2(1-p).$$

Calculate $\frac{dH(p)}{dp}$ in this communication system.

By applying the product rule and the chain rule, we have

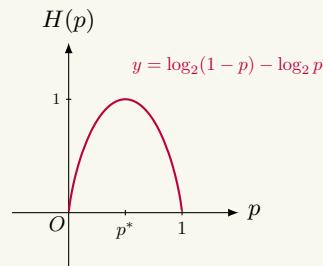
$$\begin{aligned} \frac{dH(p)}{dp} &= -\log_2 p - p(\log_2 p)' - (-1)\log_2(1-p) - (1-p)(\log_2(1-p))' \\ &= -\log_2 p - p \frac{1}{p \ln 2} + \log_2(1-p) - (1-p) \frac{-1}{(1-p) \ln 2} \\ &= \log_2(1-p) - \log_2 p. \end{aligned}$$

Note that $\frac{dH(p)}{dp}$ shows the changing rate of the entropy $H(p)$ with respect to the probability of the input p . Now, we investigate how $H(p)$ changes with p .

In particular, since $\log_2(1-p)$ is a decreasing function of p and $\log_2 p$ is an increasing function of p , $\frac{dH(p)}{dp}$ is thus a decreasing function of p . Moreover, we notice that when p is small, $\frac{dH(p)}{dp} > 0$; while when p gets bigger and approaches 1, $\frac{dH(p)}{dp} < 0$. We thus know that there exists a point $p^* \in (0, 1)$ such that $H(p^*) = 0$. Furthermore, since $\frac{dH(p)}{dp} > 0$ means $H(p)$ increases with p and $\frac{dH(p)}{dp} < 0$ means $H(p)$ decreases with p , the entropy $H(p)$ is thus maximized at p^* . This shows that when p is neither too small nor too big, i.e., the input is more uncertain, more entropy can be achieved.

Actually, we can plot the figure of $H(p)$ as follows:

Indeed, we see from the figure above that $H(p)$ increases with p up to a certain point, denoted as p^* , then it decreases with p . In other words, more entropy is obtained when p is neither too big nor too small, i.e., when the input is more uncertain.



2.4 Implicit Differentiation

2.4.2 Example: implicit differentiation - circle

For the function $y = f(x)$ defined by $y^2 + x^2 = 1$, we can take the derivative of y with respect to x of both sides of the equation. By treating y as a function of x and applying the chain rule, we get

$$\frac{d(x^2 + y^2)}{dx} = 0 \implies 2x + 2y \cdot \frac{dy}{dx} = 0,$$

so that

$$\frac{dy}{dx} = -\frac{x}{y}.$$

This function $y = f(x)$ is defined implicitly. We keep in mind that y is a function of x . We can find the derivative $\frac{dy}{dx}$ by “*implicit differentiation*”.

2.43 Method: implicit differentiation

1. Replace y by $y(x)$ in the given equation which defines the function $y = y(x)$ implicitly.
2. Differentiate both sides of the equation with respect to x , by applying the rules of differentiation.
3. Solve for $\frac{dy}{dx}$ from the resulting equation.

2.44 Example: implicit differentiation

The equation $2y^4 - x^2 - \sin y + 1 = 0$ ($x > 0$) gives a curve on the plane. Use implicit differentiation to find $y' = \frac{dy}{dx}$ and then determine the tangent line of the curve at the point $P = (1, 0)$.

Solution By treating y as a function of x , we differentiate both sides of the equation $2y^4 - 2x^2 - \sin y + 1 = 0$ with respect to x and get

$$8y^3y' - 2x - y'\cos y = 0.$$

Solving the equation for y' gives

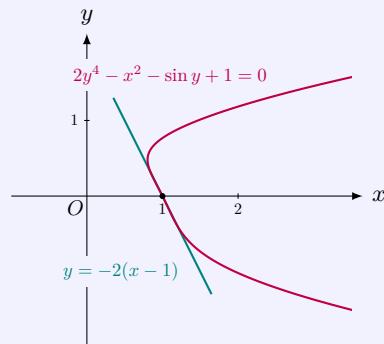
$$y' = \frac{2x}{8y^3 - \cos y}.$$

The point $P = (1, 0)$ satisfies the equation and thus lies on the curve. The tangent line of the curve at the point has slope

$$y'|_P = \frac{2(1)}{8(0^3) - \cos 0} = -2.$$

Hence the tangent line at P is given by the equation

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



2.45 MATLAB: implicit differentiation

MATLAB supports the implicit differentiation if the symbolic function $y(x)$ is appropriately created. The following example shows the calculation of y' in MATLAB, where x and y satisfy the equation $2y^4 - x^2 - \sin y + 1 = 0$.

```
>> syms x y(x) t
>> f = 2*y(x)^4 - x^2 - sin(y(x)) + 1;
>> d = diff(f,x)

d =
8*y(x)^3*diff(y(x), x) - 2*x - cos(y(x))*diff(y(x), x)

>> d0 = subs(d,diff(y(x),x),t)
d0 =
8*t*y(x)^3 - 2*x - t*cos(y(x))

>> s = solve(d0,t)

s =
-(2*x)/(cos(y(x)) - 8*y(x)^3)
```

where $d0 = \text{subs}(d, \text{diff}(y(x), x), t)$ is to replace $\text{diff}(y(x), x)$ with t , so that the command `solve` can work to obtain the expression of t (i.e., the expression of $\text{diff}(y(x), x)$ or y'). Note that if we solve $d = 0$ with respect to $\text{diff}(y(x), x)$ directly, i.e., using command `solve(d, diff(y(x), x))`, MATLAB is unable to find the explicit solution.

Moreover, based on the expression of $\text{diff}(y(x), x)$ or y' , which was calculated above, we can easily determine the slope of the tangent line at point $(1, 0)$.

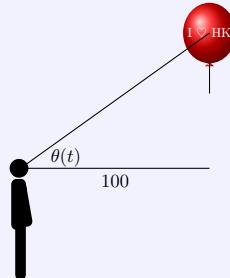
```
>> s_x = subs(s,x,1)           % substitute x = 1
s_x =
-2/(cos(y(1)) - 8*y(1)^3)
>> s_xy = subs(s_x,y,0)       % substitute y = 0
s_xy =
-2
```

MATLAB shows that the slope of the tangent line at point $(1, 0)$ is -2 .

2.5 Related Rates

2.46 Example: application - related rates

John observes a rising helium balloon. He stands 100 meters away from the launch site. The balloon rises vertically at the constant speed 2 meter/second. How fast is the angle of the elevation of the balloon increasing 50 seconds after the launch? (The angle of elevation is the angle between the ground and John's line of sight to the balloon.)



Solution Let $y(t)$ be the distance traveled by the balloon from the launch site at time t . Let $\theta(t)$ be the angle of elevation at time t .

The speed of the balloon is $y'(t)$, which equals to 2 m/s. “How fast is the angle of the elevation” means that we need to find the rate of change of $\theta(t)$, i.e., the derivative $\theta'(t)$.

The relation between $y(t)$ and $\theta(t)$ is given by

$$y(t) = 100 \cdot \tan \theta(t).$$

Taking the differentiation of both sides of the equation above, we obtain

$$y'(t) = \frac{100 \cdot \theta'(t)}{\cos^2 \theta(t)}.$$

Since we need to find $\theta'(t)$, we obtain

$$\theta'(t) = \frac{y'(t)}{100} \cos^2 \theta(t) = \frac{2}{100} \cos^2 \theta(t).$$

We need to find $\cos \theta(t)$ at $t = 50$. In fact,

$$y(50) = 2 \times 50 = 100.$$

From the relation between $y(t)$ and $\theta(t)$, we have $100 = 100 \cdot \tan \theta(50)$, so that $\theta(50) = \frac{\pi}{4}$. Thus we obtain

$$\cos \theta(50) = \cos\left(\frac{\pi}{4}\right) = \frac{\sqrt{2}}{2}.$$

Therefore, we have

$$\theta'(50) = \frac{2}{100} \cdot \left(\frac{\sqrt{2}}{2}\right)^2 = 0.01.$$

Since two rates of change, i.e., derivatives $\theta'(t)$ and $y'(t)$, are related by the equation

$$\theta'(t) = \frac{y'(t)}{100} \cos^2 \theta(t).$$

This is a typical problem of **related rates**. In general, so-called related rates problems involve finding a rate at which a quantity changes by relating that quantity to other quantities whose rates of change are known.

2.47 Example: related rates - area

An oil tank springs an oil leak on the ground, and the oil spreads in a circular patch around the tank. If the radius of the oil patch increases at a rate of 10 meters/hour, how fast is the area of the patch increasing when the patch has a radius of 20 meters?

Solution

Step 1: Label all variables:

Let $r(t)$ be the radius of the oil patch at time t and $A(t)$ be the area of the oil patch at time t .

The rate of the area change is the quantity $A'(t)$.

Step 2: Write the relation between concerned variables:

The area of the patch is

$$A(t) = \pi r(t)^2.$$

Step 3: Write the relations between rates:

We take the differentiation with respect to t on both sides of the equation above,

$$A'(t) = 2\pi r(t) \cdot r'(t).$$

Step 4: List all known quantities:

The radius of the oil patch increases at a rate of 10 meters/hour, that is,

$$r'(t) = 10.$$

At the time t_0 when the patch has a radius of 20 meters, we have

$$r(t_0) = 20, \quad r'(t_0) = 10.$$

Step 5: Find the desired rate:

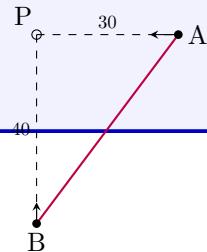
Combining all the results above gives

$$A'(t_0) = 2\pi r(t_0) \cdot r'(t_0) = 2\pi \cdot 20 \cdot 10 = 400\pi \approx 1256.64 \text{ (m}^2/\text{h}\text{).}$$

2.48 Example: related rates - distance

Two sailing boats sail towards a docking port. Boat A sails due west at the constant speed 8 miles/hour and Boat B sails due north at the constant speed 10 miles/hour.

How fast is the distance between the boats changing when the westbound boat is 30 miles from the port and the northbound boat is 40 miles from the port?



Solution**Step 1:** Label all variables:

Let $x(t)$ be the distance of the westbound boat to the port at time t . Let $y(t)$ be the distance of the northbound boat to the port at time t . Let $z(t)$ be the distance between two boats at time t .

The rate of the distance change is the quantity $z'(t)$.

Step 2: Write the relation between concerned variables:

From the Pythagorean theorem,

$$z(t)^2 = x(t)^2 + y(t)^2.$$

Step 3: Write the relations between the rates:

We take the differentiation with respect to t on both sides of the equation above,

$$2z(t) \cdot z'(t) = 2x(t) \cdot x'(t) + 2y(t) \cdot y'(t) \implies z'(t) = \frac{x(t) \cdot x'(t) + y(t) \cdot y'(t)}{z(t)}.$$

Step 4: List all given quantities:

At the time t_0 when westbound boat is 30 miles from the port, we have

$$x(t_0) = 30, \quad y(t_0) = 40, \quad x'(t_0) = -8, \quad y'(t_0) = -10.$$

Notice that the derivatives $x'(t)$ and $y'(t)$ are negative, since the distances from both boats to the port decrease in time.

Step 5: Determine other quantities from the given quantities:

We can determine the quantity $z(t_0)$ by using the Pythagorean identity:

$$z(t_0) = \sqrt{x(t_0)^2 + y(t_0)^2} = \sqrt{30^2 + 40^2} = 50.$$

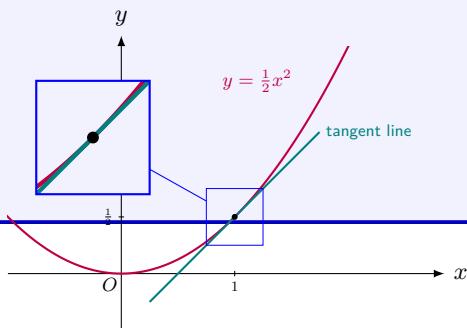
Step 6: Find the desired rate:

Combining all the results above gives

$$z'(t_0) = \frac{x(t_0) \cdot x'(t_0) + y(t_0) \cdot y'(t_0)}{z(t_0)} = \frac{-8 \cdot 30 - 10 \cdot 40}{50} = -12.8 \text{ (mi/h).}$$

2.6 Linear Approximation

2.49 Example: linear approximation



Geometrically, the tangent line of a graph $y = f(x)$ at $x = a$ is close to the graph when x is very close to a . Algebraically, the value $f(a) + f'(a)(x - a)$ is very close to $f(x)$ when x is close to a , that is, when $x \approx a$,

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

2.50 Theorem: linear approximation

Let f be differentiable at a . Then the linear function $y = f(a) + f'(a)(x - a)$ approximates $y = f(x)$ near $x = a$, that is, when $x \approx a$,

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

This linear function $y = f(a) + f'(a)(x - a)$ is called the **linear approximation** (or **linearization**) of f at $x = a$.

2.51 Example: linear approximation - numerical approximation

Consider the function $f(x) = \sqrt{1+x}$. Its derivative is $f'(x) = \frac{1}{2\sqrt{1+x}}$. The tangent line at $x = 0$ is $y = f(0) + f'(0)x = 1 + \frac{1}{2}x$. We know that, for small x , $y = 1 + \frac{1}{2}x$ approximates $\sqrt{1+x}$. Numerically, we can verify this. For different values of small x , we make the following table

x	$\sqrt{1+x}$	$1 + \frac{1}{2}x$
0.1	1.04880884817	1.05
0.01	1.00498756211	1.005
0.001	1.00049987506	1.0005

We see that the approximation is getting better when x is closer to 0.

2.52 Example: linear approximation - numerical approximation

Find approximate value of $\sqrt[3]{1.02}$.

Solution Consider the function $f(x) = \sqrt[3]{1+x}$. Since $f'(x) = \frac{1}{3\sqrt[3]{(1+x)^2}}$, the linear approximation of f is

$$y = 1 + \frac{1}{3}x.$$

Thus, we have an approximate value of $\sqrt[3]{1.02}$:

$$\sqrt[3]{1.02} \approx 1 + \frac{1}{3}(0.02) \approx 1.006667.$$

The real value of $\sqrt[3]{1.02}$ is $1.0066227\dots$

Differential

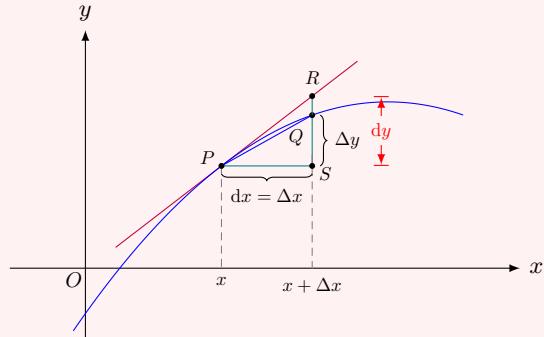
2.53 Definition:

The **differential** of $y = f(x)$ at x , denoted as dy or $df(x)$, is

$$dy = f'(x) \Delta x.$$

We often call the change in x , Δx , as the differential of the independent variable, denoted as dx , that is $dx = \Delta x$. Hence,

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



2.54 Example:

Compute the differential of the function $f(x) = \cos x$. Use it to find approximate value of $\cos 94^\circ$.

Solution For the function $y = \cos x$, we have $y' = -\sin x$, so that $dy = -\sin x dx$. Thus,

$$\cos(a + \Delta x) - \cos a = \Delta y \approx dy = -\sin a dx = -\sin a \Delta x.$$

Notice that $94^\circ = 90^\circ + 4^\circ = \frac{1}{2}\pi + \frac{4}{180}\pi = \frac{1}{2}\pi + \frac{1}{45}\pi$. By taking $a = \frac{1}{2}\pi$ and $\Delta x = \frac{1}{45}\pi$, we have

$$\begin{aligned} \cos 94^\circ &= \cos(a + \Delta x) \approx \cos a - \sin a \cdot \Delta x \\ &= \cos \frac{1}{2}\pi - \sin \frac{1}{2}\pi \cdot \left(\frac{1}{45}\pi\right) \\ &= 0 - 1 \cdot \frac{1}{45}\pi \approx -0.0698132. \end{aligned}$$

The real value of $\cos 94^\circ$ is $-0.0697565 \dots$

2.55 MATLAB:

In MATLAB, the linear approximation of a function can be calculated using the command **taylor**. The following examples show the calculations of the linear approximations of $f(x) = \sqrt[3]{1+x}$ at $x = 0$ and $f(x) = \cos x$ at $x = a$, respectively.

```
>> syms x;
>> f = (1+x)^(1/3);
>> taylor(f,x,0, 'Order',2)

ans =
```

```
x/3 + 1
>> syms x a;
>> f = cos(x);
>> taylor(f,x,a,'Order',2)

ans =

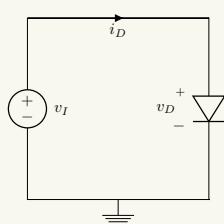
cos(a) + sin(a)*(a-x)
```

Because the linear approximation $f(x) \approx f(a) + f'(a)(x - a)$ has two terms, we specify '**Order**' and 2 in the command **taylor** in the MATLAB code above.

2.56 Engineering Example: small signal analysis

In electronic circuits, many of the components that make up of the circuits are nonlinear, e.g., diodes, transistors and vacuum tubes. For these nonlinear components, the current through them is not proportional to the voltage, which makes calculations and analysis complicated. Moreover, there are many applications where nonlinear devices only operate in a very restricted range of voltage or current, such as many sensor applications and audio amplifiers. For easy calculations and analysis of the voltage and current, a method called *small signal analysis* (or *incremental analysis*) is proposed to approximate circuits containing nonlinear elements with linear equations over a restricted region. Specifically, if the time-varying alternating current (AC) signal of an electronic current is *small* compared to the deterministic direct current (DC) bias, then using **linear approximation**, the nonlinear circuit elements are replaced by linear elements over a restricted range in the resultant AC-equivalent circuit. The small signal analysis method widely applies to electronic circuits where the AC signals are small compared to the DC ones, e.g., radio receivers, sensors, telecommunications and signal processing circuits.

A nonlinear circuit example is shown in the following, where the diode is a nonlinear component.



Parameter	Description
V_D	DC voltage
I_D	DC current
v_d	AC voltage
i_d	AC current
$v_D = V_D + v_d$	total voltage
$i_D = I_D + i_d$	total current
v_I	input voltage

For the total voltage v_D and the total current i_D of the above circuit, we have $i_D = f(v_D)$. Note that i_D is a *nonlinear* function of v_D due to the nonlinear diode. We will then use the small signal analysis method to approximate the behaviour of this circuit near the DC bias point.

In particular, at $v_D = V_D$, using linear approximation, the function $i_D = f(v_D)$ can be

approximated as

$$\begin{aligned} i_D &\approx f(V_D) + \frac{df(v_D)}{dv_D} \Big|_{v_D=V_D} \cdot (v_D - V_D) \\ &= f(V_D) + \frac{df(v_D)}{dv_D} \Big|_{v_D=V_D} \cdot v_d. \end{aligned}$$

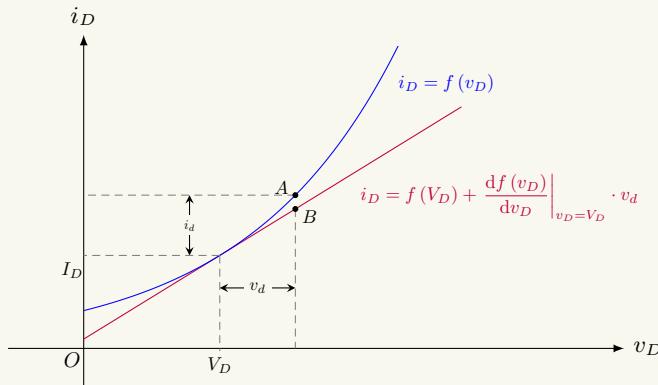
Note that $i_D = I_D + i_d$ and I_D can be written as $I_D = f(V_D)$. Hence, the term $\frac{df(v_D)}{dv_D} \Big|_{v_D=V_D} \cdot v_d$ can be treated as the AC current i_d , and we have

$$i_d = \frac{df(v_D)}{dv_D} \Big|_{v_D=V_D} \cdot v_d.$$

Further, we note that $\frac{df(v_D)}{dv_D} \Big|_{v_D=V_D}$ is independent of v_d , so i_d is in proportion with v_d after linear approximation, i.e.,

$$i_d \propto v_d.$$

A graphical interpretation of the small signal analysis is shown below.



The point (V_D, I_D) representing the steady DC voltage and current is called the operating point. Note that the small signal analysis is dependent on the operating point, i.e., the DC bias voltage and current. From the figure above, we see that over a restricted range around $v_D = V_D$, the time-varying AC current i_d becomes a linear function of the AC voltage v_d (or graphically, we approximate A with B in the restricted region), which facilitates the analysis of the nonlinear circuit over this range.

2.7 Higher Derivatives

2.57 Definition: higher derivatives

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

$$f''(x) = (f')'(x).$$

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

$$f'''(x) = (f'')'(x).$$

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

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

2.58 Example: higher derivatives

Calculate the second derivative of $f(x) = \sin x^2$.

Solution We apply the chain rule to calculate the first derivative:

$$f'(x) = (\cos x^2) \cdot 2x = 2x \cos x^2.$$

Then we need to use product rule and the chain rule to calculate the second derivative:

$$\begin{aligned} f''(x) &= (f'(x))' = (2x \cos x^2)' \\ &= (2x)' \cdot \cos x^2 + 2x \cdot (\cos x^2)' = 2 \cos x^2 - 4x^2 \sin x^2. \end{aligned}$$

2.59 Example: higher derivatives - polynomials

For any polynomial $P_n(x)$ of degree n , show that $\frac{d^k}{dx^k} P_n(x) = 0$ for $k > n$.

Solution Since $(x^m)' = mx^{m-1}$ for integer m , we know that $\frac{d^m}{dx^m}(x^m) = m!$ and hence $\frac{d^{m+1}}{dx^{m+1}}(x^m) = (m!) = 0$. Thus we obtain $\frac{d^k}{dx^k}(x^m) = 0$ if $k > m$. If $k > n$, then

$$\begin{aligned} \frac{d^k}{dx^k} P_n(x) &= \frac{d^k}{dx^k} (a_0 + a_1 x + \cdots + a_n x^n) \\ &= a_0 \frac{d^k}{dx^k}(1) + a_1 \frac{d^k}{dx^k}(x) + \cdots + a_n \frac{d^k}{dx^k}(x^n) \\ &= 0 + 0 + \cdots + 0 = 0. \end{aligned}$$

2.60 Example: higher derivatives - $\sin x$ and $\cos x$

Prove that

$$\frac{d^n}{dx^n} \sin x = \sin(x + \frac{1}{2}n\pi), \quad \frac{d^n}{dx^n} \cos x = \cos(x + \frac{1}{2}n\pi).$$

Solution Both equalities can be proved by induction. Here we only demonstrate the proof of the first equality.

Clearly, since $(\sin x)' = \cos x = \sin(x + \frac{1}{2}\pi)$, the equality holds for $n = 1$.

Suppose the equality holds for $n = k$, that is, $\frac{d^k}{dx^k} \sin x = \sin(x + \frac{1}{2}k\pi)$. Thus,

$$\begin{aligned} \frac{d^{k+1}}{dx^{k+1}} \sin x &= \left(\frac{d^k}{dx^k} \sin x \right)' \\ &= [\sin(x + \frac{1}{2}k\pi)]' \\ &= \sin(x + \frac{1}{2}k\pi + \frac{1}{2}\pi) \\ &= \sin \left[x + \frac{1}{2}(k+1)\pi \right]. \end{aligned}$$

So, the equality holds for $n = k + 1$.

Hence, the first equality holds for all $n \geq 1$.

2.61 Example: higher derivatives - implicit differentiation

For the function $y = f(x)$ defined implicitly by $y^2 + x^2 = 1$, by implicit differentiation, we get

$$2y \frac{dy}{dx} + 2x = 0 \implies \frac{dy}{dx} = -\frac{x}{y}.$$

For finding the second derivative, we can adopt one of the following two approaches.

(1) By differentiating $\frac{dy}{dx} = -\frac{x}{y}$ once more, we have

$$\begin{aligned} \frac{d^2y}{dx^2} &= \frac{d}{dx} \left(-\frac{x}{y} \right) = -\frac{1 \cdot y - x \cdot \frac{dy}{dx}}{y^2} \\ &= -\frac{1 \cdot y - x \cdot \left(-\frac{x}{y} \right)}{y^2} = -\frac{x^2 + y^2}{y^3} = -\frac{1}{y^3}. \end{aligned}$$

(2) By implicitly differentiating $2y \frac{dy}{dx} + 2x = 0$, we have

$$2 \left(\frac{dy}{dx} \right)^2 + 2y \frac{d^2y}{dx^2} + 2 = 0,$$

so that

$$\begin{aligned}\frac{d^2y}{dx^2} &= -\frac{1 + \left(\frac{dy}{dx}\right)^2}{y} = -\frac{1 + \left(-\frac{x}{y}\right)^2}{y} \\ &= -\frac{x^2 + y^2}{y^3} = -\frac{1}{y^3}.\end{aligned}$$

2.62 Example: acceleration

Let $f(t) = \frac{g}{2}t^2$ be the distance function of a free-fall object and t be the time. Its derivative is the velocity $v(t) = f'(t) = gt$. The rate of change of the velocity is the acceleration $a(t)$: $a(t) = v'(t) = f''(t)$. Thus the acceleration is the second derivative of $f(t)$.

In general, if $y = f(t)$ is a distance function, then the first derivative of f is the velocity $v(t) = f'(t)$ and the second derivative of f is the acceleration $a(t) = v'(t) = f''(t)$.

2.63 Engineering Example: Newton's second law of motion

Newton's second law of motion presents the relationship between the net force acting on an object and the rate of the object's momentum change. This is one of the most important laws in physics. It was first proposed by Isaac Newton in his book "*Philosophiae Naturalis Principia Mathematica (Mathematical Principles of Natural Philosophy)*", which was first published in 1687. In terms of magnitude, Newton's second law is given in the following:

$$F = \frac{dp}{dt} = m \frac{dv}{dt},$$

where F is the net force acting on an object, p is the momentum of the object, v is the speed, and t is the time.

Note that the speed v equals to the rate of an object's displacement change with respect to time, i.e., $v = \frac{dx}{dt}$, where x is the displacement of an object. Then, we can rewrite Newton's second law as

$$F = m \frac{d^2x}{dt^2}.$$

Suppose that a ball of mass 0.25 kg is rolling on the ground, and it starts to decelerate at $t = 0$. The displacement of this ball from its original position at $t = 0$ is $x(t) = 2t - 0.2t^2$ ($0 \leq t \leq 5$). What is the resistive force acting on the ball?

According to Newton's second law, we have

$$F = m \frac{d^2x(t)}{dt^2} = 0.25 \times \frac{d^2(2t - 0.2t^2)}{dt^2} = 0.25 \times \frac{d(2 - 0.4t)}{dt} = -0.1 \text{ (N).}$$

The result above shows that the magnitude of the resistive force acting on the ball is 0.1 N. The sign “−” means the direction of the resistive force is opposite to the rolling direction of the ball.

Leibniz Formula

2.64 Theorem:

Leibniz Formula

If $f^{(n)}$ and $g^{(n)}$ exists, so is $(f \cdot g)^{(n)}$ and the **Leibniz Formula** holds:

$$[f(x) \cdot g(x)]^{(n)} = \sum_{i=0}^n \binom{n}{i} f^{(i)}(x) \cdot g^{(n-i)}(x).$$

This is also called the Leibniz Rule.

Justification:

The formula can be proved by induction. In fact, it is clear that the formula holds for $n = 1$:

$$[f(x) \cdot g(x)]' = f'(x)g(x) + f(x)g'(x) = \sum_{i=0}^1 \binom{1}{i} f^{(i)}(x) \cdot g^{(1-i)}(x).$$

Suppose the formula holds for $n = k$, that is,

$$[f(x) \cdot g(x)]^{(k)} = \sum_{i=0}^k \binom{k}{i} f^{(i)}(x) \cdot g^{(k-i)}(x).$$

Thus,

$$\begin{aligned} & [f(x) \cdot g(x)]^{(k+1)} \\ &= \left[\sum_{i=0}^k \binom{k}{i} f^{(i)}(x) \cdot g^{(k-i)}(x) \right]' \\ &= \sum_{i=0}^k \binom{k}{i} \left[f^{(i+1)}(x) \cdot g^{(k-i)}(x) + f^{(i)}(x) \cdot g^{(k+1-i)}(x) \right] \\ &= f^{(k+1)}(x)g(x) + \sum_{i=1}^k \binom{k}{i-1} f^{(i)}(x) \cdot g^{(k+1-i)}(x) \\ &\quad + \sum_{i=1}^k \binom{k}{i} f^{(i)}(x) \cdot g^{(k+1-i)}(x) + f(x)g^{(k+1)}(x) \\ &= f^{(k+1)}(x)g(x) \\ &\quad + \sum_{i=1}^k \left[\binom{k}{i-1} + \binom{k}{i} \right] f^{(i)}(x) \cdot g^{(k+1-i)}(x) + f(x)g^{(k+1)}(x) \\ &= \sum_{i=0}^{k+1} \binom{k+1}{i} f^{(i)}(x) \cdot g^{(k+1-i)}(x) \end{aligned}$$

So, the Leibniz Formula holds for all $n \geq 1$.

2.65 Example: Leibniz Formula

Compute $\frac{d^n}{dx^n} (x^2 e^x)$ for $n \geq 1$.

Solution We know that

$$(x^2)' = 2x, \quad (x^2)'' = 2, \quad (x^2)^{(i)} = 0 \text{ for all } i > 2.$$

Thus, by the Leibniz Formula, we have

$$\begin{aligned}\frac{d^n}{dx^n} (x^2 e^x) &= x^2 (e^x)^{(n)} + n(x^2)' (e^x)^{(n-1)} + \frac{n(n-1)}{2} (x^2)'' (e^x)^{(n-2)} \\ &= [x^2 + 2nx + n(n-1)] e^x.\end{aligned}$$

Summary

Definition: basic concepts

derivative : $f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$

higher derivative : $f^{(n)}(x) = (f^{(n-1)})'(x), \quad n \geq 2$

differential : $dy = f'(x) dx$

linear approximation : $f(x) \approx f(a) + f'(a)(x - a)$

Theorem: rules of differentiation

Suppose $f'(x)$ and $g'(x)$ exist. Then the following rules of differentiation hold.

1. Sum Rule:

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

2. Difference Rule:

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

3. Product Rule:

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

4. Quotient Rule: If $g(a) \neq 0$, then

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

5. Chain Rule: Suppose f' and g' exist. Then

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

or in Leibniz notation,

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

6. Inverse Function: Suppose f is one-to-one and differentiable with non-zero derivative. Then

$$[f^{-1}(x)]' = \frac{1}{f'(f^{-1}(x))}.$$

7. Leibniz Formula: Suppose $f^{(n)}$ and $g^{(n)}$ exist. Then

$$[f(x) \cdot g(x)]^{(n)} = \sum_{i=0}^n \binom{n}{i} f^{(i)}(x) \cdot g^{(n-i)}(x).$$

Theorem: derivatives of some elementary functions

$$1. (c)' = 0$$

$$2. (x^\mu)' = \mu x^{\mu-1}$$

$$3. (e^x)' = e^x,$$

$$(a^x)' = a^x \ln a$$

$$4. (\ln x)' = \frac{1}{x},$$

$$(\log_a x)' = \frac{1}{x \ln a}$$

$$5. (\sin x)' = \cos x$$

$$6. (\cos x)' = -\sin x$$

$$7. (\tan x)' = \sec^2 x$$

$$8. (\csc x)' = -\csc x \cot x$$

$$9. (\sec x)' = \sec x \tan x$$

$$10. (\cot x)' = -\csc^2 x.$$

$$11. (\sin^{-1} x)' = \frac{1}{\sqrt{1-x^2}}$$

$$12. (\cos^{-1} x)' = -\frac{1}{\sqrt{1-x^2}}$$

$$13. (\tan^{-1} x)' = \frac{1}{1+x^2}$$

$$14. (\csc^{-1} x)' = -\frac{1}{|x|\sqrt{x^2-1}}$$

$$15. (\sec^{-1} x)' = \frac{1}{|x|\sqrt{x^2-1}}$$

$$16. (\cot^{-1} x)' = -\frac{1}{1+x^2}$$

Method: logarithmic differentiation

1. Take natural logarithms of both sides of an equation $y = f(x)$ and use the Laws of Logarithms to simplify.
2. Differentiate implicitly with respect to x .
3. Solve the resulting equation for y' .

Method: implicit differentiation

1. Replace y by $y(x)$ in the given equation which defines the function $y = y(x)$ implicitly.
2. Differentiate both sides of the equation with respect to x , by applying the rules of differentiation.
3. Solve for $\frac{dy}{dx}$ from the resulting equation.

Applications of Differentiation

3.1 Maximum and Minimum Values

3.1 Definition: extreme values

Suppose f is a function defined over a domain D and c is a number in D .

Definition: absolute extreme values

The number $f(c)$ is the

- **absolute maximum** value of f if $f(c) \geq f(x)$ for all x in D ;
- **absolute minimum** value of f if $f(c) \leq f(x)$ for all x in D .

Definition: local extreme values

The number $f(c)$ is a

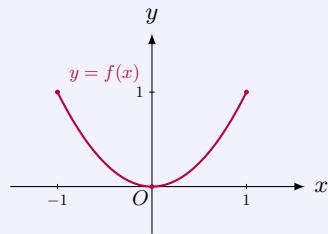
- **local maximum** value of f if $f(c) \geq f(x)$ when x is near c ;
- **local minimum** value of f if $f(c) \leq f(x)$ when x is near c .

When f attains its extremum at a point $x = c$, we call $x = c$ an **extreme point** of f .

3.2 Example: absolute maximum and absolute minimum

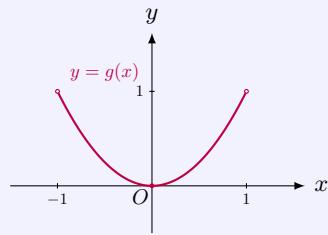
1. Consider $f(x) = x^2$ defined over the interval $[-1, 1]$. From the graph of the function, we see that

- f has the absolute maximum value $f(\pm 1) = 1$;
- f has the absolute minimum value $f(0) = 0$;
- f has two local maximum values $f(\pm 1) = 1$;
- f has one local minimum value $f(0) = 0$.



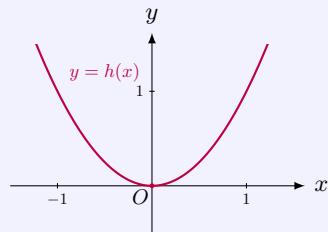
2. Consider $g(x) = x^2$ defined over the interval $(-1, 1)$. From the graph of the function, we see that

- g has no absolute maximum value;
- g has the absolute minimum value $g(0) = 0$;
- g has no local maximum value;
- g has one local minimum value $g(0) = 0$.



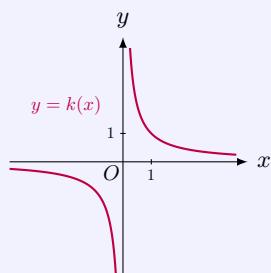
3. Consider $h(x) = x^2$ defined over $(-\infty, +\infty)$. From the graph of the function, we see that

- h has no absolute maximum value;
- h has the absolute minimum value $h(0) = 0$;
- h has no local maximum value;
- h has one local minimum value $h(0) = 0$.



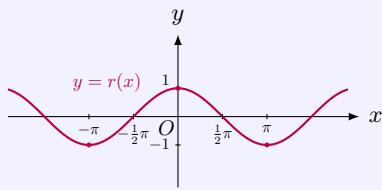
4. Consider $k(x) = \frac{1}{x}$ defined over $(0, +\infty)$. From the graph of the function, we see that

- k has no absolute maximum value;
- k has no absolute minimum value;
- k has no local maximum value;
- k has no local minimum value.



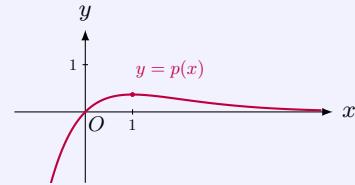
5. Consider $r(x) = \cos x$ defined over $(-\infty, +\infty)$. From the graph, we see that, for $n = 0, \pm 1, \pm 2, \dots$,

- r has the absolute maximum value $r(2n\pi) = 1$;
- r has the absolute minimum value $r((2n+1)\pi) = -1$;
- r has local maximum values $r(2n\pi) = 1$;
- r has local minimum values $r((2n+1)\pi) = -1$.



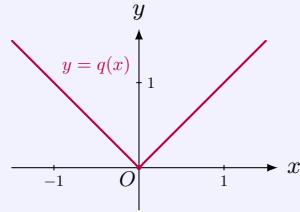
6. Consider the function $p(x) = xe^{-x}$ with the graph given below. We see that

- p has the absolute maximum value $p(1) = 1/e$;
- p has no absolute minimum value;
- p has a local maximum values $p(1) = 1/e$;
- p has no local minimum value.



7. For the function $q(x) = |x|$, we see that

- q has no absolute maximum value;
- q has the absolute minimum value $q(0) = 0$;
- q has no local maximum value;
- q has one local minimum value $q(0) = 0$.



3.3 Theorem: Fermat's Theorem

If f has a local maximum or minimum value at an *interior point* c in the domain D of f , and if $f'(c)$ exists, then $f'(c) = 0$.

A point c is an **interior point** of D if there is an open interval (a, b) such that $c \in (a, b) \subset D$.

Justification:

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

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

This implies that

$$f(c+h) - f(c) \leq 0,$$

for all h sufficiently close to 0, since c is an interior point. Thus, if $h > 0$ and h is sufficiently small, we have

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

Since f is differentiable at c , we have

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

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

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

Since f is differentiable at c , we have

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

Since both $f'(c) \leq 0$ and $f'(c) \geq 0$ hold, we get $f'(c) = 0$.

The case of a local minimum can be proved in a similar manner.

Remark: interior point \iff not a boundary point

That a point is an interior point means that it is not a boundary point. For example, the point 0 is an interior point of the interval $[-1, 1]$ since $0 \in (-1, 1) \subset [-1, 1]$. The point -1 is not an interior point of $[-1, 1]$, but a boundary point.

The point 0 is not an interior point of the interval $(0, 1)$ since it is even not a point *in* $(0, 1)$. In fact, 0 is a boundary point of $(0, 1)$. The number 10^{-100} is an interior point of $(0, 1)$ since $10^{-100} \in (0.5 \times 10^{-100}, 1.5 \times 10^{-100}) \subset (0, 1)$.

Remark: locations of local extrema

Fermat's Theorem concludes that a local extreme value $f(c)$, where c is in the domain of f , only occurs at one of the following three types of locations:

- an interior point where $f'(c) = 0$;
- an interior point where $f'(c)$ does not exist;
- a boundary point.

3.4 Definition: critical point

Let f be defined over a domain D . A number c in D is called a **critical point** of f if either $f'(c) = 0$ or $f'(c)$ does not exist.

3.5 Theorem: Fermat's Theorem - rephrase

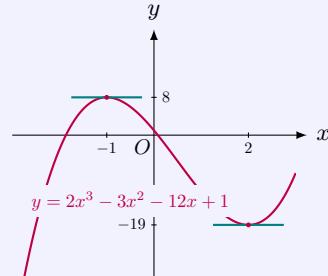
If f has a local maximum or minimum value at an *interior point* c in the domain D of f , then c is a critical point of f .

3.6 Example: local extreme occur at critical points

1. Consider the function $p(x) = 2x^3 - 3x^2 - 12x + 1$ with the graph given below. The derivative of p is

$$p'(x) = 6x^2 - 6x - 12 = 6(x^2 - x - 2) = 6(x - 2)(x + 1).$$

Solving $p'(x) = 0$ gives $x = 2$ or $x = -1$. Thus these two points are all the critical points of p . From the graph, we see that the function p attains local extreme values at these points.

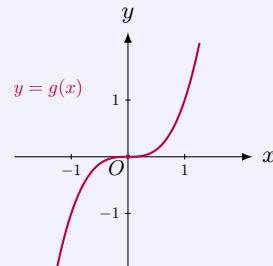


2. Consider the function $q(x) = |x|$. Since $q(x)$ does not have derivative at $x = 0$, so that $x = 0$ is a critical point. The function also attains a local minimum value as well as an absolute minimum value at $x = 0$.

3.7 Pitfall: critical points do not necessarily yield local extreme values

Fermat's Theorem states that a local extremum value only occurs at an interior critical point or at a boundary point. However, not all critical points will yield local extreme values.

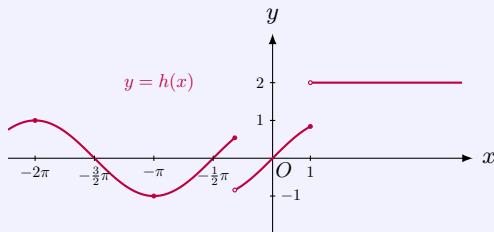
1. Consider $g(x) = x^3$. Since $g'(x) = 3x^2$, $x = 0$ is a critical point of g . From the graph of function, we see that $x = 0$ is not a local extreme point.



2. Consider the function

$$h(x) = \begin{cases} \cos x, & \text{if } x \leq -1; \\ \sin x, & \text{if } -1 < x \leq 1; \\ 2, & \text{if } x > 1. \end{cases}$$

It is not differentiable at points $x = -1$ and 1 , so that they are critical points. From the graph, we see that $x = -1$ is a local maximum point and 1 is not a local extreme point.

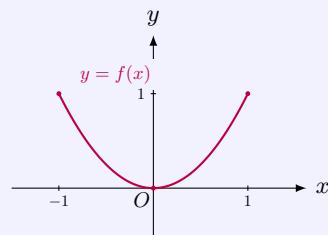


3.8 Theorem: Extreme Value Theorem

Let f be a *continuous* function defined over a *finite closed* interval $[a, b]$. Then $f(x)$ has at least one absolute maximum point and one absolute minimum point in the interval $[a, b]$.

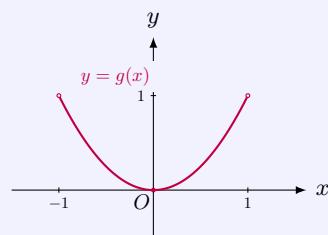
3.9 Example: continuous function on a finite closed interval

The function $f(x) = x^2$ is continuous on the finite closed interval $[-1, 1]$. So the Extreme Value Theorem applies. In fact, the function has the absolute maximum value $f(-1) = f(1) = 1$ and the absolute minimum value $f(0) = 0$.



3.10 Pitfall: continuous function may not have absolute maximum

The function $g(x) = x^2$ is continuous on the finite interval $(-1, 1)$. However, the Extreme Value Theorem does not apply to the function g , since the interval $(-1, 1)$ is not closed. In fact, it is easy to see that the function g does not have absolute maximum value.

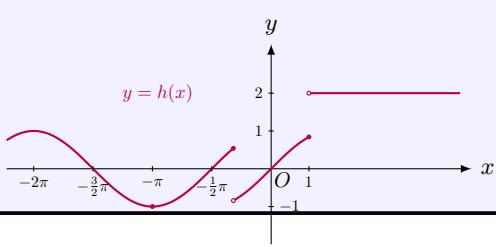


3.11 Pitfall: discontinuous function may have absolute extrema

Consider the function h :

$$h(x) = \begin{cases} \cos x, & \text{if } x \leq -1; \\ \sin x, & \text{if } -1 < x \leq 1; \\ 2, & \text{if } x > 1. \end{cases}$$

The function is not continuous on the domain $(-\infty, +\infty)$, and the domain is not closed nor finite. Nevertheless, it has an absolute maximum and absolute minimum.



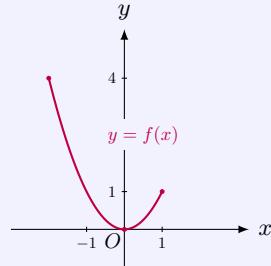
3.12 Method: finding absolute extrema

Given a continuous function f over a finite closed interval $[a, b]$.

1. Find the critical points and evaluate the values of f at critical points.
2. Evaluate the values of f at the boundary points $x = a$ and $x = b$.
3. The largest value and smallest value from steps 1 and 2 are the absolute maximum value and the absolute minimum value respectively.

3.13 Example: absolute maximum and absolute minimum

Consider $f(x) = x^2$ defined over the interval $[-2, 1]$. The function f is continuous, and the domain $[-2, 1]$ is a finite closed interval. Since the function f satisfies all the hypotheses in the Extreme Value Theorem, the function f must have the absolute maximum value and the absolute minimum value.



To find the absolute extrema, we proceed as follows:

- (1) First try to determine the critical points. Indeed, by solving the equation $f'(x) = 2x = 0$, we see that the critical point of f is $x = 0$. Thus, we get the value $f(0) = 0$.
- (2) Next we evaluate the values of f at two boundary points: $f(-2) = (-2)^2 = 4$ and $f(1) = 1$.
- (3) Among these three values, the largest one is $f(-2) = 4$ and the smallest one is $f(0) = 0$. So the absolute maximum value occurs at $x = -2$ and the absolute minimum value occurs at $x = 0$.

3.14 MATLAB: maximum and minimum values

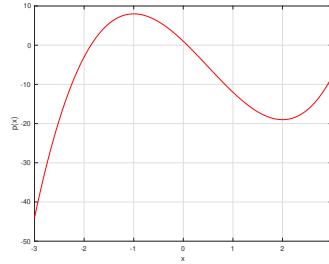
1. To obtain the local maximum or minimum value of a function, we can first plot the function in MATLAB to see the rough location of the local maximum or minimum value. The following code shows how to plot the figure of the function $p(x) = 2x^3 - 3x^2 - 12x + 1$.

```
>> syms x;
>> p(x) = 2*x^3 - 3*x^2 - 12*x + 1;
>> x_array = -3:0.05:3;
>> y_array = p(x_array);
>> figure
>> plot(x_array,y_array, 'r')
```

```
>> grid
>> xlabel('x')
>> ylabel('p(x)')
```

After running the MATLAB code above, we have the figure for the function p .

From this figure, we see that, for the function p , the local maximum value is around $x = -1$ and the local minimum value is around $x = 2$. According to Fermat's Theorem, we know that $p'(x_{\max}) = p'(x_{\min}) = 0$, where x_{\max} and x_{\min} are the points obtaining the local maximum value and the local minimum value, respectively.



To obtain the exact locations of the local maximum and minimum values of p , we need the following MATLAB code.

```
>> d = diff(p);
>> d_fun = matlabFunction(d)

d_fun =

function_handle with value:

@(x)x.*-6.0+x.^2.*6.0-1.2e1

>> max_p = fzero(d_fun,-1)

max_p =

-1

>> min_p = fzero(d_fun,2)

min_p =

2
```

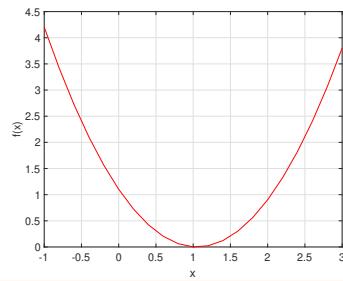
where `matlabFunction(d)` is to convert the symbolic function `d` to a *function handle* which can be analyzed by the command `fzero`. Note that the command `solve` is not used here, because MATLAB may not be able to obtain explicit solutions for some equations; while `fzero` can numerically find the zero of a function near the initial guess. MATLAB returns that the local maximum value can be achieved at $x = -1$, and the local minimum value can be achieved at $x = 2$.

2. The following example shows how to find the absolute minimum value of the function $f(x) = x^2 - 2.1x + 1.1025$. First, we plot the curve of f . The MATLAB code is shown below.

```
>> syms x;
>> f(x) = x^2 - 2.1*x + 1.1025;
>> x_array = -1:0.2:3;
>> y_array = f(x_array);
>> figure
>> plot(x_array,y_array, 'r')
>> xlabel('x')
>> ylabel('f(x)')
>> grid
```

Then, we have the figure showing the graph of f .

From this figure, we see that the absolute minimum value of $f(x)$ is around $x = 1$. To find the exact location of the absolute minimum value, we need the following MATLAB code.



```
>> d = diff(f);
>> d_fun = matlabFunction(d)

d_fun =
function_handle with value:

@(x)x.*2.0-2.1e1./1.0e1

>> min_p = fzero(d_fun,1)

min_p =
1.0500
```

MATLAB returns that the absolute minimum value of $f(x)$ can be achieved at $x = 1.05$.

3.15 Engineering Example: fugacity coefficient

In chemical thermodynamics, the fugacity is an *effective* partial pressure of a real gas. It equals the pressure of an ideal gas which has the same temperature and molar Gibbs free energy as this real gas. For example, the Nitrogen gas (N_2) at $0^\circ C$ and a pressure of $P = 100$ atmospheres (atm) has a fugacity of $f = 97.03$ atm. This means that the molar Gibbs free energy of this real Nitrogen gas is the same with the ideal gas at $0^\circ C$ and a pressure of 97.03 atm. The pressure and the fugacity of a gas is related through a parameter called **fugacity coefficient**, which is defined as $\varphi = \frac{f}{P}$, where f is the fugacity and P is the pressure. For example, the fugacity coefficient of the Nitrogen gas at $0^\circ C$ and a pressure of $P = 100$ atm is 0.9703. Note that for the ideal gas, the fugacity coefficient is always 1.

The fugacity coefficient can be rewritten in the form of the compressibility factor which is a correction factor describing the deviation of a real gas from ideal gas behaviour. In particular, under certain conditions, the fugacity coefficient in terms of the compressibility factor is provided as follows:

$$\varphi = f(x) = x - 1 - \ln(x - 1) - \ln\left(1 + \frac{1}{x}\right)$$

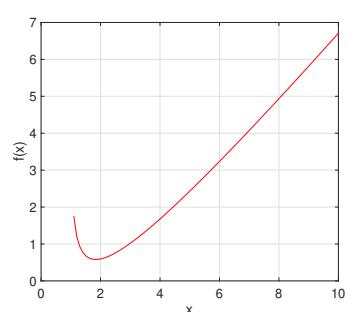
where x ($x > 1$) is the compressibility factor.

Use MATLAB to calculate the minimum value of the fugacity coefficient function and the corresponding compressibility factor.

First, we plot the function $f(x) = x - 1 - \ln(x - 1) - \ln\left(1 + \frac{1}{x}\right)$ in MATLAB to see the rough location of the minimum value. The code is shown in the following.

```
>> syms x
>> f(x) = x - 1 - log(x - 1) - log(1 + 1/x);
>> x_array = 1.1:0.1:10;
>> y_array = f(x_array);
>> figure
>> plot(x_array, y_array, 'r')
>> grid
>> xlabel('x')
>> ylabel('f(x)')
```

From the graph of f , we see that the minimum value is around $x = 2$. According to Fermat's Theorem, to obtain the minimum value of $f(x)$, we need to calculate the value of x which satisfies $f'(x) = 0$.



Next, we run the following MATLAB code.

```

>> df = diff(f)

df(x) =

1/(x^2*(1/x + 1)) - 1/(x - 1) + 1

>> df_fun = matlabFunction(df)

df_fun =

function_handle with value:

@(x)-1.0./(x-1.0)+1.0./x.^2./(1.0./x+1.0)+1.0

>> x_min = fzero(df_fun,2)

x_min =

1.8393

```

It shows that $f'(1.8393) = 0$, i.e., the minimum value of $f(x)$ is achieved when the compressibility factor $x = 1.8393$.

Substituting $x = 1.8393$ into the fugacity coefficient function $f(x)$, we have the minimum value of the fugacity coefficient shown as follows.

```

>> f_min = vpa(f(x_min))

f_min =

0.58031458931953300345263227407065

```

The result shows that the minimum value of the fugacity coefficient is around 0.5803.

3.2 Monotone Functions and Mean Value Theorem

3.16 Definition: monotone functions

A function f is said to be

- **increasing** on an interval I if

$$f(x_1) < f(x_2) \quad \text{whenever } x_1 < x_2 \text{ in } I;$$

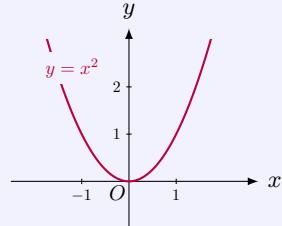
- **decreasing** on an interval I if

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

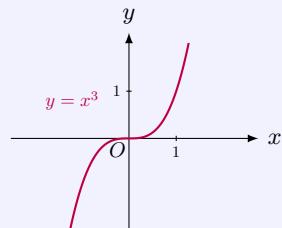
Such functions are also said to be **monotone** (or **monotonic**) functions.

3.17 Example: monotone functions

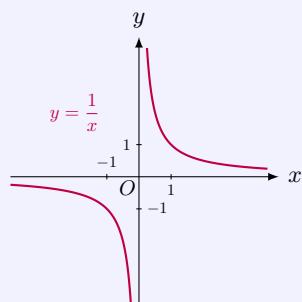
- The function $f(x) = x^2$ is decreasing on the interval $(-\infty, 0)$ and increasing on the interval $(0, +\infty)$.



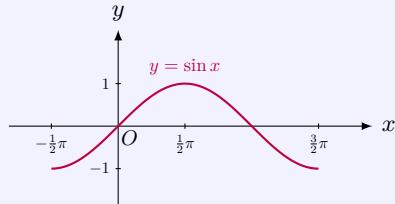
- The function $g(x) = x^3$ is increasing on the whole real line $(-\infty, +\infty)$.



- The function $h(x) = \frac{1}{x}$ is decreasing on the interval $(-\infty, 0)$ as well as on the interval $(0, +\infty)$.



- The function $k(x) = \sin x$ is increasing on the interval $(-\frac{1}{2}\pi, \frac{1}{2}\pi)$ and decreasing on the interval $(\frac{1}{2}\pi, \frac{3}{2}\pi)$.



3.18 Theorem: Mean Value Theorem

Suppose f is a function that satisfies the following:

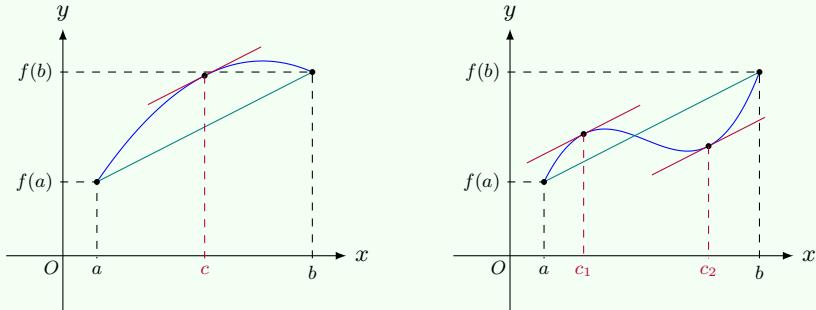
- f is continuous on the *closed* interval $[a, b]$.
- f is differentiable in the *open* interval (a, b) .

Then there exists at least one number c in (a, b) such that

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

or, equivalently,

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



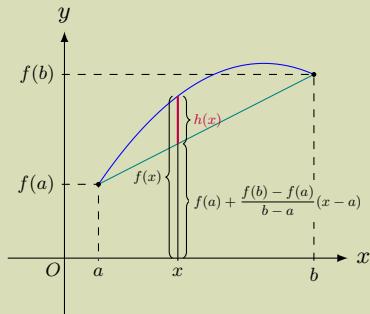
Justification:

An equation of the line segment connecting $(a, f(a))$ and $(b, f(b))$ is

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

As shown in the figure, at x with $a \leq x \leq b$, consider the auxiliary function

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



Since f is continuous on $[a, b]$, so is h . By the Extreme Value Theorem, there exist absolute maximum point M and absolute minimum point m .

If one of m and M is attained at an interior point of (a, b) , since h is differentiable in (a, b) , by Fermat's Theorem, there exists $c \in (a, b)$ such that $h'(c) = 0$. Since

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

we get

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

If both m and M are attained at the boundary points of $[a, b]$, since

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

we must have $m = M = 0$. This means that h must be a constant function. Thus, for any $c \in (a, b)$, we have $h'(c) = 0$. Similarly as above, we again have

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

3.19 Pitfall: when the hypotheses in the Mean Value Theorem fail

The hypotheses of the Mean Value Theorem are that the function f is continuous on the closed interval $[a, b]$ and differentiable in the open interval (a, b) . We will see that if any of these fails, then the conclusion of the Mean Value Theorem does not necessarily hold.

1. Consider the function f defined on $[0, 1]$:

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

Clearly, f is discontinuous at $x = 1$, so that f is not continuous on $[0, 1]$. Obviously, for any $x \in (0, 1)$, $f'(x) = 0$ and

$$\frac{f(1) - f(0)}{1 - 0} = \frac{1 - 0}{1 - 0} = 1.$$

So there exists no c in $(0, 1)$ such that

$$f'(c) = \frac{f(1) - f(0)}{1 - 0}.$$

2. Consider the function $g(x) = |x|$ defined over $[-1, 1]$. We know that $g'(0)$ does not exist, so g is not differentiable in $(-1, 1)$. It is easy to see that

$$\frac{g(1) - g(-1)}{1 - (-1)} = \frac{1 - 1}{2} = 0,$$

and

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

Clearly, there exists no c inside $(-1, 1)$ such that

$$g'(c) = \frac{g(1) - g(-1)}{1 - (-1)}.$$

3.20 Theorem: Monotone Test

Suppose f is differentiable on an interval I .

1. If $f'(x) > 0$ on I , then f is increasing on I .
2. If $f'(x) < 0$ on I , then f is decreasing on I .

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) .

The other case can be proved similarly.

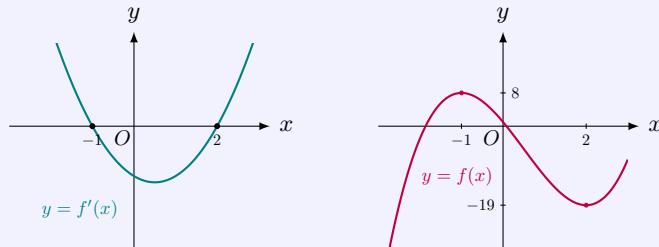
3.21 Example: determine intervals of monotonicity

Consider the function $f(x) = 2x^3 - 3x^2 - 12x + 1$. To find the intervals where f is monotone, by the Monotone Test, we need to determine the intervals in which f' keeps its sign unchanged.

In fact,

$$f'(x) = 6x^2 - 6x - 12 = 6(x^2 - x - 2) = 6(x - 2)(x + 1),$$

so that it has two real roots $x = -1$ and $x = 2$. These two numbers divide the interval $(-\infty, +\infty)$ into three open intervals $(-\infty, -1)$, $(-1, 2)$, and $(2, +\infty)$. Since the function f' is continuous and nonzero over each of these intervals, by the Intermediate Value Theorem, it must be either positive or negative on each interval. In $(-\infty, -1)$, since f' is positive for large negative x , $f'(x)$ is positive on $(-\infty, -1)$. The function f' changes its sign on two sides of $x = -1$ and $x = 2$, so that $f'(x)$ is negative on $(-1, 2)$ and positive on $(2, +\infty)$.



Thus the function $f(x)$ is increasing over the intervals $(-\infty, -1)$ and $(2, +\infty)$, and decreasing over the interval $(-1, 2)$.

3.22 Method: solving inequalities by using the number line

There is a method that works for solving inequalities of the form $f(x) > 0$, or inequalities involving \geq , $<$, or \leq , 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.

3.23 Theorem: zero derivative implies constant function

Suppose f is *differentiable* in (a, b) . If $f'(x) = 0$ for all x in (a, b) , then f is a constant function.

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) .

3.24 Example: identity proving

Show that when $|x| \leq \frac{1}{2}$, the following identity holds:

$$3 \arccos x - \arccos(3x - 4x^3) = \pi.$$

Solution When $|x| < \frac{1}{2}$,

$$\begin{aligned} & [3 \arccos x - \arccos(3x - 4x^3)]' \\ &= -\frac{3}{\sqrt{1-x^2}} + \frac{1}{\sqrt{1-(3x-4x^3)^2}}(3-12x^2) \\ &= -\frac{3}{\sqrt{1-x^2}} + \frac{1}{\sqrt{1-x^2 \cdot (1-4x^2)}}(3-12x^2) = 0, \end{aligned}$$

so we have

$$3 \arccos x - \arccos(3x - 4x^3) = C, \quad -\frac{1}{2} < x < \frac{1}{2},$$

where C is a constant. Setting $x = 0$ gives $C = \pi$. Hence,

$$3 \arccos x - \arccos(3x - 4x^3) = \pi, \quad -\frac{1}{2} < x < \frac{1}{2}.$$

It is easy to verify that the identity also holds when $x = \pm\frac{1}{2}$. Therefore,

$$3 \arccos x - \arccos(3x - 4x^3) = \pi, \quad |x| \leq \frac{1}{2}.$$

3.25 MATLAB: monotone function

The following example shows how to determine the monotone intervals of the function $f(x) = 2x^3 - 3x^2 - 12x + 1$ in MATLAB.

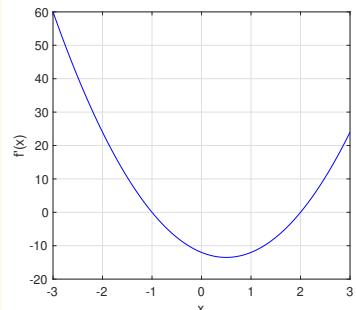
```
>> syms x;
>> f(x) = 2*x^3 - 3*x^2 - 12*x + 1;
>> d(x) = diff(f(x),x);
>> rts_d = solve(d(x) == 0,x)

rts_d =
-1
2

>> x_array = -3:0.05:3;
>> figure
>> plot(x_array,d(x_array), 'b')
>> xlabel('x')
>> ylabel('f''(x)')
>> grid
```

Note that for a string (e.g., `x` inside `'x'` and `f"(x)` inside `'f''(x)'`), we need to use *double single quotes* to obtain a single quote (simply using one single quote represents conjugate transpose in MATLAB), e.g., we use `'f"(x)'` to represent $f'(x)$.

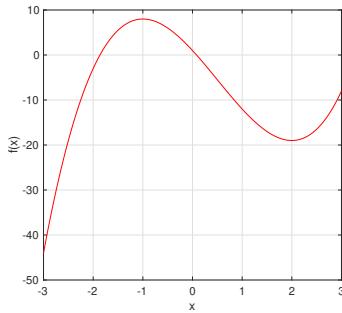
From the MATLAB code above, we have the graph of f' and see that $f'(x) = 0$ has two roots: $x = -1$ and $x = 2$. We also see that f' is negative on $(-1, 2)$, and positive on $(-\infty, -1)$ and $(2, +\infty)$. According to the Monotone Test, we know that f is decreasing on the interval $(-1, 2)$, and f is increasing on the intervals $(-\infty, -1)$ and $(2, +\infty)$.



Actually, we can plot f to check the above results.

```
>> f(x) = 2*x^3 - 3*x^2 - 12*x + 1;
>> figure
>> plot(x_array,f(x_array), 'r')
>> xlabel('x')
>> ylabel('f(x)')
>> grid
```

We see from the graph of f that indeed f is decreasing on $(-1, 2)$, and increasing on $(-\infty, -1)$ and $(2, +\infty)$.



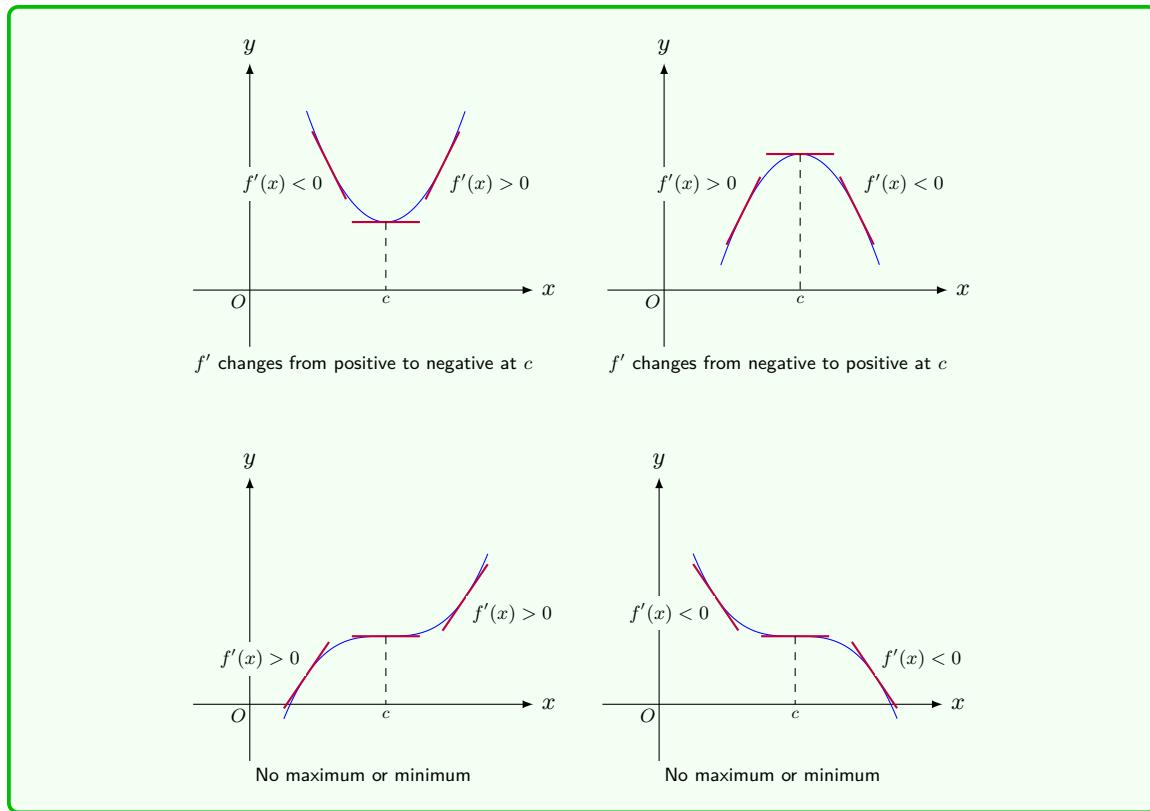
3.3 Tests for Local Extrema and Concavity

Local Extrema

3.26 Theorem: First Derivative Test

Suppose that c is a *critical point* of a continuous function f .

1. If f' changes *from positive to negative* at c , then f has a *local maximum* value at c .
2. If f' changes *from negative to positive* at c , then f has a *local minimum* value at c .
3. If f' *does not change sign* at c , then f has *no local extreme* value at c .

**Justification:**

Near c , suppose $f'(x)$ is positive for $a < x < c$ and negative for $c < x < b$. Then $f(x)$ is continuous on $[a, c]$ and differentiable on (a, c) . Thus $f(x)$ is increasing over $[a, c]$, i.e. for any x with $a < x < c$, $f(x) < f(c)$. By the similar argument, $f(x)$ is decreasing over $[c, b]$, i.e. for any x with $c < x < b$, $f(c) > f(x)$. Therefore c is a local maximum point.

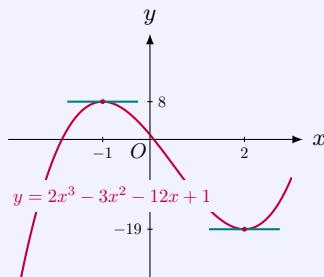
The same argument applies to the other two cases.

3.27 Example: local extrema using the First Derivative Test

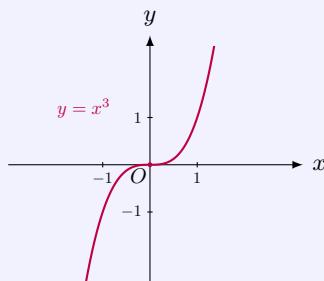
1. Consider the function $f(x) = 2x^3 - 3x^2 - 12x + 1$. Its derivative is

$$f'(x) = 6x^2 - 6x - 12 = 6(x - 2)(x + 1).$$

By solving the equation $f'(x) = 0$, we see that f has critical points $x = -1, 2$. It is easy to see that $f'(x)$ is positive for $x < -1$, negative for $-1 < x < 2$, and positive for $x > 2$. Thus, by the First Derivative Test, we know that $x = -1$ is a local maximum point and $x = 2$ is a local minimum point.



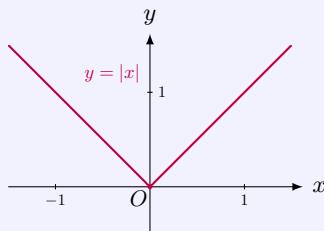
2. Consider the function $g(x) = x^3$. Its derivative is $g'(x) = 3x^2$. By solving the equation $g'(x) = 0$, we see that g has one critical point $x = 0$. It is easy to see that near $x = 0$, $g'(x)$ does not change signs. Thus, by the First Derivative Test, we know that $x = 0$ is not a local extreme point.



3. Consider the function

$$h(x) = |x|.$$

It is not differentiable at $x = 0$, so that h has a critical point at $x = 0$. When $x < 0$, $h'(x) = -1 < 0$; when $x > 0$, $h'(x) = 1 > 0$. Thus, by the First Derivative Test, $x = 0$ is a local minimum point.



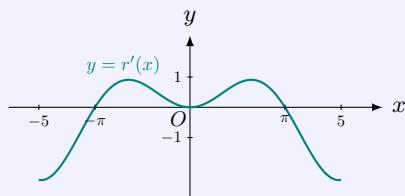
4. Consider $r(x) = \frac{1}{2} \sin x - \frac{1}{2}x \cos x$ over $(-5, 5)$. Its derivative is

$$r'(x) = \frac{1}{2} \cos x - \frac{1}{2}(\cos x - x \sin x) = \frac{1}{2}x \sin x.$$

Solving $r'(x) = \frac{1}{2}x \sin x = 0$ in the interval $(-5, 5)$ gives three critical points $-\pi, 0, \pi$.

In fact, the derivative $r'(x) = \frac{1}{2}x \sin x$ has simple root at $x = -\pi$ and π , and a double root at $x = 0$. The later is because

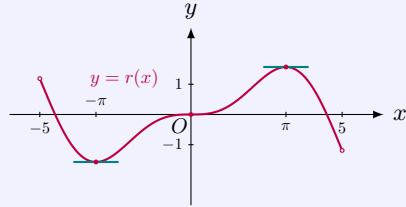
$$\lim_{x \rightarrow 0} \frac{r'(x)}{x^2} = \frac{1}{2} \lim_{x \rightarrow 0} \frac{\sin x}{x} = \frac{1}{2} \neq 0.$$



Since $\sin x$ is positive for $-5 < x < -\pi$, we know that $r'(x)$ is negative near $x = -5$. Thus, we can sketch the graph of r' , as shown.

From the graph of r' , we see that $r'(x)$ changes from negative to positive at the critical point $x = -\pi$, does not change sign at the critical point $x = 0$, and changes from positive to negative at the critical point $x = \pi$.

Hence, by the First Derivative Test, f has a local minimum value at $x = -\pi$, not a local extreme at $x = 0$, and a local maximum value at $x = \pi$.



3.28 Theorem: Second Derivative Test

Suppose $f''(c)$ exists.

1. If $f'(c) = 0$ and $f''(c) > 0$, then f has a *local minimum* value at c .
2. If $f'(c) = 0$ and $f''(c) < 0$, then f has a *local maximum* value at c .
3. If $f'(c) = 0$ and $f''(c) = 0$, it is *inconclusive* about local extreme of f at c .

Justification:

1. Assume $f'(c) = 0$ and $f''(c) > 0$. Since $f''(c)$ exists and is positive, we know that, for x near c , $\frac{f'(x) - f'(c)}{x - c} > 0$. Hence $f'(x) < f'(c) = 0$ for $x < c$, and $f'(x) > f'(c) = 0$ for $x > c$. By the First Derivative Test, c is a local minimum point.

2. The proof for this case is similar as case 1.

3. We use two functions to demonstrate this. They both satisfy the hypotheses, but the first function has no local extreme value at c while the second has a local maximum value at c .

(1) Consider $f(x) = x^3$. It is easy to see that $f'(0) = 0$, $f''(0) = 0$, and $f''(x) = 6x$ is continuous near $x = 0$. Since $f(x) < 0 = f(0)$ when $x < 0$ and $f(x) > 0 = f(0)$, $f(0)$ is not a local extreme value.

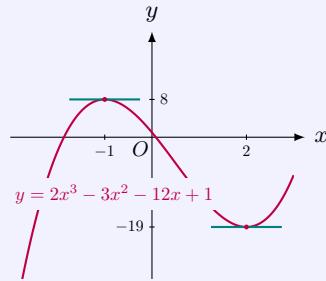
(2) Consider $g(x) = x^4$. It is easy to see that $g'(0) = 0$, $g''(0) = 0$, and $g''(x) = 12x^2$ is continuous near $x = 0$. Since $g(x) \geq 0 = g(0)$ for all real x , $g(0)$ is a local minimum value.

3.29 Example: local extrema using the Second Derivative Test

1. Consider $f(x) = 2x^3 - 3x^2 - 12x + 1$. Its derivatives are

$$\begin{aligned}f'(x) &= 6x^2 - 6x - 12 \\&= 6(x-2)(x+1), \\f''(x) &= 12x - 6.\end{aligned}$$

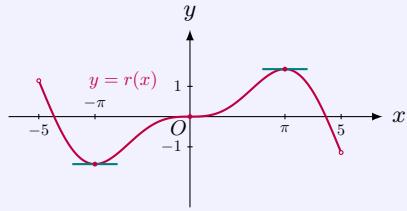
Solving $f'(x) = 0$ gives $x = -1, 2$. Since $f''(-1) = -18 < 0$, $f(-1) = 8$ is a local maximum value; $f''(2) = 18 > 0$, thus $f(2) = -19$ is a local minimum value.



2. Consider $r(x) = \frac{1}{2}\sin x - \frac{1}{2}x \cos x$ over $(-5, 5)$. Its derivatives are

$$\begin{aligned}r'(x) &= \frac{1}{2}\cos x - \frac{1}{2}(\cos x - x \sin x) = \frac{1}{2}x \sin x, \\r''(x) &= \frac{1}{2}(\sin x + x \cos x).\end{aligned}$$

Solving $r'(x) = \frac{1}{2}x \sin x = 0$ in the interval $(-5, 5)$ gives three critical points $-\pi, 0, \pi$. Since $r''(-\pi) = \frac{1}{2}\pi > 0$, r has a local minimax value at $x = -\pi$; $r''(\pi) = -\frac{1}{2}\pi < 0$, r has a local minimum value at $x = \pi$.



However, since $r''(0) = 0$, the Second Derivative Test is *inconclusive* in this case. In fact, we can apply the First Derivative Test to conclude that r has not a local extreme at $x = 0$, since r' does not change sign at the critical point $x = 0$.

3.30 Pitfall: the Second Derivative Test may not be applicable

Consider the function $f(x) = x|x|$. When $x \neq 0$,

$$f'(x) = \begin{cases} -2x, & \text{if } x < 0; \\ 2x, & \text{if } x > 0. \end{cases}$$

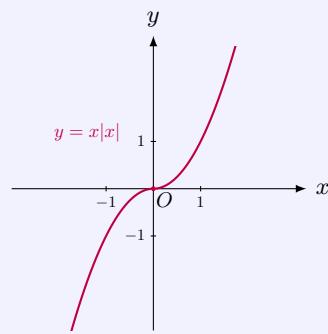
To calculate the derivative $f'(0)$, we use the definition of derivative.

$$\begin{aligned}\lim_{x \rightarrow 0^-} \frac{f(x) - f(0)}{x - 0} &= \lim_{h \rightarrow 0^-} \frac{-2x^2}{x} = \lim_{h \rightarrow 0^-} (-2x) = 0, \\\lim_{x \rightarrow 0^+} \frac{f(x) - f(0)}{x - 0} &= \lim_{h \rightarrow 0^+} \frac{2x^2}{x} = \lim_{h \rightarrow 0^+} (2x) = 0.\end{aligned}$$

Since the left limit and right limit are equal, $f'(0) = 0$. Thus $f'(x) = 2|x|$ for all $x \in \mathbb{R}$.

Clearly $f'(x)$ has no derivative at $x = 0$. Hence f does not have the second derivative at $x = 0$, so that the Second Derivative Test cannot be applied to the function f .

However, we can still apply the First Derivative Test to show that $f(0)$ is not a local extreme value, since $f'(0) = 0$ and f' does not change sign at $x = 0$.



3.31 MATLAB: tests for local extrema

The following example shows the tests for local extrema of the function $f(x) = 2x^3 - 3x^2 - 12x + 1$.

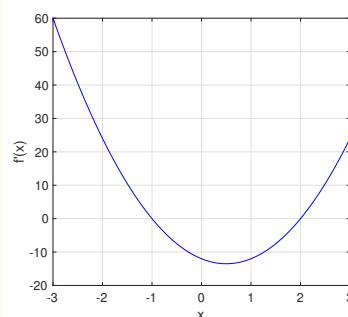
We first test the local extrema of f using First Derivative Test.

```
>> syms x;
>> f(x) = 2*x^3 - 3*x^2 - 12*x + 1;
>> d(x) = diff(f(x),x);
>> rts_d = solve(d(x) == 0,x)

rts_d =
-1
2

>> x_array = -3:0.05:3;
>> figure
>> plot(x_array,d(x_array),'b')
>> xlabel('x')
>> ylabel('f''(x)')
>> grid
```

From the MATLAB code above, we have the graph of f' and see that $f'(x) = 0$ has two roots: $x = -1$ and $x = 2$. We also see that f' is negative on $(-1, 2)$, and positive on $(-\infty, -1)$ and $(2, +\infty)$. According to First Derivative Test, we know that f has a local maximum value at $x = -1$ and a local minimum value at $x = 2$.



Next, we test the local extrema of f using Second Derivative Test.

```
>> d2(x) = diff(f(x),x,2)

d2(x) =

12*x - 6

>> d2(-1)

ans =

-18

>> d2(2)

ans =

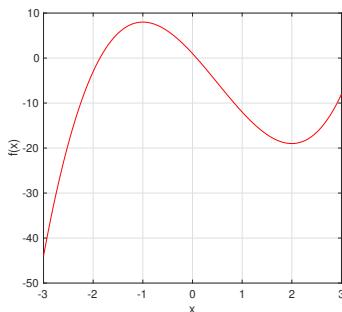
18
```

We see from the MATLAB code above that $f''(-1) = -18 < 0$ and $f''(2) = 18 > 0$. Hence, according to Second Derivative Test, f has a local maximum value at $x = -1$ and a local minimum value at $x = 2$.

Actually, we can plot f to check the above results.

```
>> figure
>> plot(x_array,f(x_array),'r')
>> xlabel('x')
>> ylabel('f(x)')
>> grid
```

We see from the figure that f indeed has a local maximum value at $x = -1$ and a local minimum value at $x = 2$.



Concavity

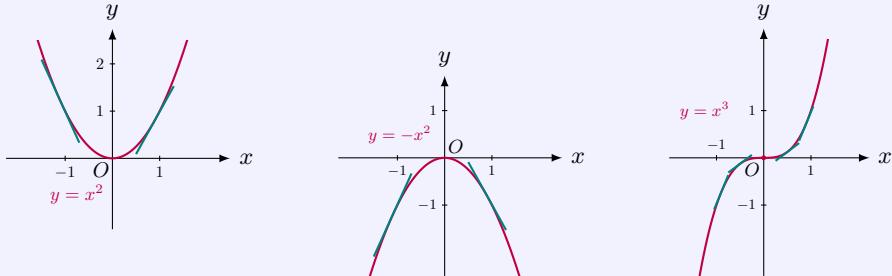
3.32 Definition: concave up and concave down

Suppose f is defined over an interval I . The function f is said to be

- **concave up** if its graph is above all its tangent lines over the interval;
- **concave down** if its graph is below all its tangent lines over the interval.

A point P is an **inflection point** of f if f is continuous at P and changes concavity at P .

3.33 Example: concave up and concave down



1. The function $y = x^2$ is concave up on \mathbb{R} .
2. The function $y = -x^2$ is concave down on \mathbb{R} .
3. The function $y = x^3$ is concave down over the interval $(-\infty, 0)$ and concave up over the interval $(0, \infty)$. The point 0 is an inflection point of f .

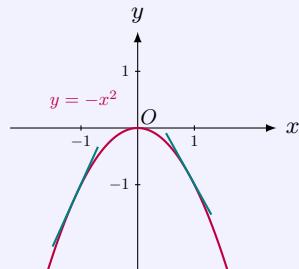
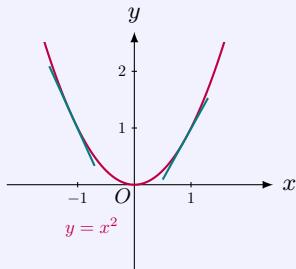
3.34 Theorem: Concavity Test

Suppose f'' exists on an interval I .

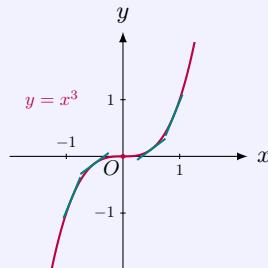
1. If $f''(x) > 0$ on I , then f is concave up on I .
2. If $f''(x) < 0$ on I , then f is concave down on I .

3.35 Example: Concavity Test

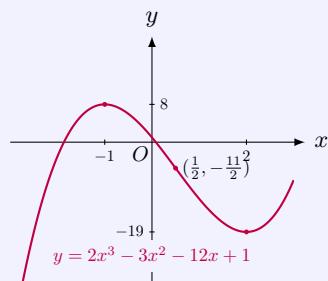
1. For the function $y_1 = x^2$, we have $y_1'' = 2 > 0$. Thus, by the Concavity Test, this function is concave up on \mathbb{R} . For the function $y_2 = -x^2$, we have $y_2'' = -2 < 0$ and the function is concave down on \mathbb{R} .



2. For the function $f(x) = x^3$, $f''(x) = 6x$. Since $f''(x) < 0$ when $x < 0$, by the Concavity Test, f is concave down on $(-\infty, 0)$. Similarly, since $f''(x) > 0$ when $x > 0$, f is concave up on $(0, \infty)$. Thus, 0 is an inflection point.



3. Consider $g(x) = 2x^3 - 3x^2 - 12x + 1$. The second derivative $g''(x) = 12x - 6 = 6(2x - 1)$. Since $g''(x) < 0$ when $x < \frac{1}{2}$, by the Concavity Test, g is concave down on $(-\infty, \frac{1}{2})$. Similarly, since $g''(x) > 0$ when $x > \frac{1}{2}$, g is concave up on $(\frac{1}{2}, +\infty)$. Thus, $x = \frac{1}{2}$ is an inflection point.



3.36 MATLAB: concavity

The example below shows the concavity test of the function $f(x) = 2x^3 - 3x^2 - 12x + 1$.

```
>> syms x;
>> f(x) = 2*x^3 - 3*x^2 - 12*x + 1;
>> d2(x) = diff(f(x),x,2)

d2(x) =
12*x - 6

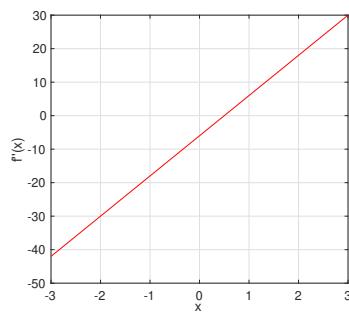
>> rts_d2 = solve(d2(x) == 0,x)

rts_d2 =
```

1/2

```
>> x_array = -3:0.05:3;
>> figure
>> plot(x_array,d2(x_array), 'r')
>> grid
```

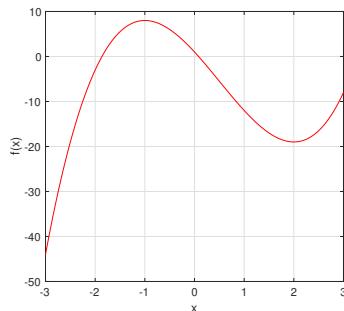
From the MATLAB code, we know that f'' is a linear function and $f''(x) = 0$ has one root: $x = \frac{1}{2}$. Furthermore, from the figure of f'' above, we see that $f''(x) < 0$ if $x < \frac{1}{2}$ and $f''(x) > 0$ if $x > \frac{1}{2}$. Hence, according to the Concavity Test, f is concave down on $(-\infty, \frac{1}{2})$ and concave up on $(\frac{1}{2}, +\infty)$, while $x = \frac{1}{2}$ is thus an inflection point.



Actually, we can plot f to check the above results.

```
>> figure
>> plot(x_array,f(x_array), 'r')
>> xlabel('x')
>> ylabel('f(x)')
>> grid
```

We can see from the graph of f that f is indeed concave down on $(-\infty, \frac{1}{2})$ and concave up on $(\frac{1}{2}, +\infty)$.



3.37 Engineering Example: genetics

According to genetics, if a parent provides a gene A with probability p ($0 \leq p \leq 1$) and a gene a with probability $1-p$, then the offspring is of the genotype AA with probability p^2 , the genotype Aa with probability $2p(1-p)$ and the genotype aa with probability $(1-p)^2$. Suppose that there are 100 people. Among them, 25 of them is of the genotype AA , 47 of them is of the

genotype Aa , and 28 of them is of the genotype aa . Let $f(p)$ be the probability of observing this specific sequence among these 100 people. Then, we can calculate the probability $f(p)$ as follows:

$$\begin{aligned} f(p) &= (p^2)^{25} [2p(1-p)]^{47} [(1-p)^2]^{28} \\ &= 2^{47} p^{97} (1-p)^{103}. \end{aligned}$$

Calculate the probability p such that $f(p)$ is maximized.

First, using the rules of differentiation, we calculate the first derivative of f in the following:

$$\begin{aligned} f'(p) &= 2^{47} [97p^{96}(1-p)^{103} - 103p^9(1-p)^{102}] \\ &= 2^{47} p^{96}(1-p)^{102}(97 - 200p). \end{aligned}$$

Next, we calculate the value of p in the open interval $(0, 1)$ at which f can achieve the local maximum value. Note that when $p \in (0, 1)$, we have $p^{96}(1-p)^{102} > 0$. Hence, from the expression above, we know that $f'(p) = 0$ when $p = \frac{97}{200} = 0.485$ and

$$\begin{cases} f'(p) > 0, & \text{if } p < 0.485; \\ f'(p) < 0, & \text{if } p > 0.485. \end{cases}$$

According to First Derivative Test, f can achieve the local maximum value at $p = 0.485$.

To verify whether f can achieve its absolute maximum value at $p = 0.485$, we need to calculate the values of $f(p)$ at the boundary points $p = 0$ and $p = 1$, respectively. We can easily calculate that $f(0) = f(1) = 0 < f(0.485)$. Therefore, we conclude that f can achieve the absolute maximum value at $p = 0.485$.

MATLAB solution:

Alternatively, this problem can also be solved using MATLAB.

First, we calculate f' and the root of $f'(p) = 0$ using the following MATLAB code.

```
>> syms p
>> f(p) = 2^47*p^97*(1-p)^103;
>> df(p) = simplify(diff(f(p)))

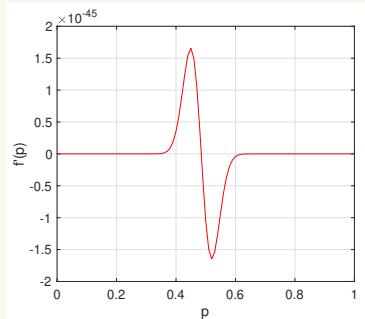
df(p) =
-140737488355328*p^96*(200*p - 97)*(p - 1)^102

>> df_root = solve(df(p) == 0, p>0, p<1) % specify the range of p
df_root =
97/200
```

In order to verify whether $p = \frac{97}{200}$ is a local maximum point or a local minimum point, we plot the figure of f' using the following MATLAB code.

```
>> p_array = 0:0.01:1;
>> df_array = df(p_array);
>> figure
>> plot(p_array,df_array,'r')
>> grid
>> xlabel('p')
>> ylabel('f''(p)')
```

The figure is plotted as shown. We see from the figure that $f'(p) > 0$ if $p < \frac{97}{200}$ and $f'(p) < 0$ if $p > \frac{97}{200}$. According to First Derivative Test, we thus know that at $p = \frac{97}{200}$, f achieves a local maximum value.



At last, to verify whether $p = \frac{97}{200}$ is an absolute maximum point, we need to calculate the value of $f(0)$ and $f(1)$, respectively. It can be easily shown that $f(0) = f(1) = 0$ and $f\left(\frac{97}{200}\right) > 0$. Therefore, $p = \frac{97}{200}$ is an absolute maximum point and f achieves the absolute maximum value at $p = \frac{97}{200}$.

3.38 Engineering Example: beam testing problem

Quality control is a process of reviewing the quality of all factors to confirm the fulfillment of the quality requirements. A quality control engineer is testing the failure rate of a certain type of beams under stress. To conduct the testing, beams are put under steady stress for some time. Suppose that the beam can pass the test with probability p , then it cannot pass the test with probability $1 - p$. If the quality control engineer conducts trials for 100 beams and 5 of them do not pass the test, then the probability of observing this sequence of successes and failures is

$$f(p) = p^{95}(1 - p)^5.$$

Calculate the probability p such that $f(p)$ is maximized.

First, using the rules of differentiation, we calculate the first derivative of $f(p)$ in the following:

$$\begin{aligned} f'(p) &= 95p^{94}(1 - p)^5 - 5p^{95}(1 - p)^4 \\ &= p^{94}(1 - p)^4(95 - 100p). \end{aligned}$$

Next, we calculate the value of the local maximum point p in the open interval $(0, 1)$. Note that when $p \in (0, 1)$, we have $p^{94}(1 - p)^4 > 0$. Hence, from the expression above, we know

that $f'(p) = 0$ when $p = \frac{95}{100} = 0.95$ and

$$\begin{cases} f'(p) > 0, & \text{if } p < 0.95; \\ f'(p) < 0, & \text{if } p > 0.95. \end{cases}$$

According to First Derivative Test, we know that $f(p)$ can achieve a local maximum value at $p = 0.95$.

To verify whether $p = 0.95$ is an absolute maximum point, we now calculate the values of $f(p)$ at $p = 0$ and $p = 1$, respectively. It can be easily calculated that $f(0) = f(1) = 0$ and $f(0.95) > 0$. Therefore, we conclude that $f(p)$ is maximized at $p = 0.95$.

MATLAB solution:

Alternatively, this problem can also be solved using MATLAB.

First, we calculate $f'(p)$ and the root of $f'(p) = 0$ using the following MATLAB code.

```
>> syms p
>> f(p) = p^95*(1-p)^5;
>> df(p) = simplify(diff(f(p)))

df(p) =
-5*p^94*(20*p - 19)*(p - 1)^4

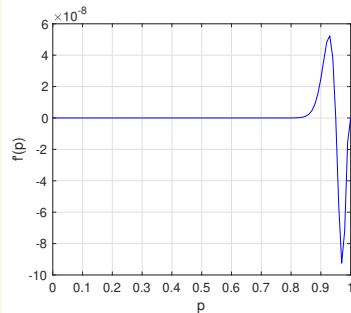
>> df_root = solve(df(p) == 0, p>0, p<1) % specify the range of p
df_root =
19/20
```

In order to check whether $p = \frac{19}{20}$ is a local maximum point or a local minimum point, we plot the figure of $f'(p)$ using the following MATLAB code:

```
>> p_array = 0:0.01:1;
>> df_array = df(p_array);
>> figure
>> plot(p_array, df_array, 'b')
>> grid
>> xlabel('p')
>> ylabel('f''(p)')
```

The figure is plotted as shown.

From the figure, we see that $f'(p) > 0$ if $p < \frac{19}{20}$ and $f'(p) < 0$ if $p > \frac{19}{20}$. Hence, according to First Derivative Test, we know that $p = \frac{19}{20}$ is a local maximum point.



At last, to check whether $p = \frac{19}{20}$ is an absolute maximum point, we need to calculate the values of $f(p)$ at $p = 0$ and $p = 1$, respectively. It can be easily shown that $f(0) = f(1) = 0$ and $f\left(\frac{19}{20}\right) > 0$. Therefore, $p = \frac{19}{20}$ is an absolute maximum point and f is maximized at $p = \frac{19}{20}$.

3.39 Engineering Example: profit maximization

In economics, profit maximization is a short run or long run process by which a firm can determine the price, input and output levels to make the greatest profit. Given price $p \in \mathbb{R}$, the demand of a product is $d(p)$. If the unit cost is c , then to maximize the profit, the seller's objective is

$$\max_p d(p)(p - c)$$

Calculate the value of p which satisfies the seller's objective in the following cases:

1. If $d(p)$ is a *linear demand*, i.e., $d(p) = a - bp$.
2. If $d(p)$ is an *exponential demand*, i.e., $d(p) = \lambda \exp(-\frac{p}{\theta})$.
3. In the constant elasticity of substitution case, i.e., $d(p) = ap^{-b}$ ($b > 1$).

Let the objective function be $f(p) = d(p)(p - c)$. First, by the Product Rule, we can easily calculate $f'(p)$ as follows:

$$f'(p) = d'(p)(p - c) + d(p).$$

Next, we calculate the value of p which maximizes $f(p)$ in different cases, respectively.

1. If $d(p) = a - bp$, we have

$$\begin{aligned} f'(p) &= (-b)(p - c) + a - bp \\ &= -2bp + bc + a. \end{aligned}$$

Then, the root of $f'(p) = 0$ is $p = \frac{1}{2}(c + \frac{a}{b})$.

It can be easily verified that

$$\begin{cases} f'(p) > 0, & \text{if } p < \frac{1}{2}(c + \frac{a}{b}); \\ f'(p) < 0, & \text{if } p > \frac{1}{2}(c + \frac{a}{b}). \end{cases}$$

Hence, according to First Derivative Test, we know that $p = \frac{1}{2}(c + \frac{a}{b})$ maximizes $f(p)$, i.e., it satisfies the seller's objective. Note that the optimal price $p = \frac{1}{2}(c + \frac{a}{b})$ increases with the unit cost c .

2. If $d(p) = \lambda \exp\left(-\frac{p}{\theta}\right)$, we have

$$\begin{aligned} f'(p) &= -\frac{\lambda}{\theta} \exp\left(-\frac{p}{\theta}\right)(p - c) + \lambda \exp\left(-\frac{p}{\theta}\right) \\ &= \exp\left(-\frac{p}{\theta}\right) \left[-\frac{\lambda}{\theta} p + \left(1 + \frac{c}{\theta}\right) \lambda \right]. \end{aligned}$$

Then, we can calculate that the root of $f'(p) = 0$ is $p = \theta + c$.

It can be easily verified that

$$\begin{cases} f'(p) > 0, & \text{if } p < \theta + c; \\ f'(p) < 0, & \text{if } p > \theta + c. \end{cases}$$

Hence, according to First Derivative Test, we know that $p = \theta + c$ maximizes $f(p)$, i.e., it satisfies the seller's objective. Note that $p = \theta + c > c$, which means the optimal price is greater than and increasing with the unit cost c .

3. If $d(p) = ap^{-b}$, we have

$$\begin{aligned} f'(p) &= (-ab)p^{-(b+1)}(p - c) + ap^{-b} \\ &= p^{-b} [(a - ab) + abcp^{-1}]. \end{aligned}$$

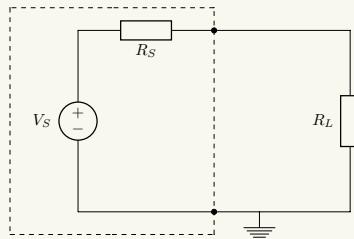
Then, the root of $f'(p) = 0$ is $p = \frac{bc}{b-1}$.

It can be easily verified that

$$\begin{cases} f'(p) > 0, & \text{if } p < \frac{bc}{b-1}; \\ f'(p) < 0, & \text{if } p > \frac{bc}{b-1}. \end{cases}$$

Hence, according to First Derivative Test, we know that $p = \frac{bc}{b-1}$ maximizes $f(p)$, i.e., it satisfies the seller's objective. Note that $p = \frac{bc}{b-1} > c$, which means the optimal price is greater than and increasing with the unit cost c .

3.40 Engineering Example: maximum power transfer



Consider the circuit in the following figure. A load of resistance R_L is connected over a power source of voltage V_S and internal resistance R_S .

According to Ohm's Law, we have

$$\text{Electric Current : } I = \frac{V_S}{R_S + R_L};$$

$$\text{Output Power : } P = \frac{V_S^2 R_L}{(R_S + R_L)^2}.$$

Calculate the the value of R_L such that the output power P is maximized.

Let $f(R_L) = P = \frac{V_S^2 R_L}{(R_S + R_L)^2}$, then by Quotient Rule, we have

$$\begin{aligned} f'(R_L) &= \frac{V_S^2 (R_S + R_L)^2 - 2V_S^2 R_L (R_S + R_L)}{(R_S + R_L)^4} \\ &= \frac{V_S^2 (R_S - R_L)}{(R_S + R_L)^3}. \end{aligned}$$

Then, the root of $f'(R_L) = 0$ is $R_L = R_S$. Moreover, we can easily verify that

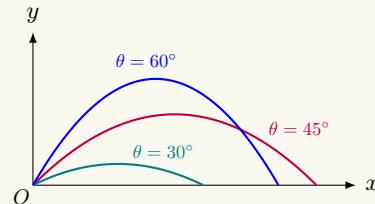
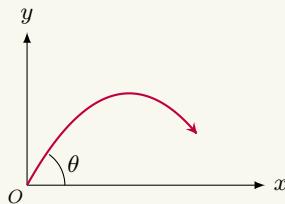
$$\begin{cases} f'(R_L) > 0, & \text{if } 0 < R_L < R_S; \\ f'(R_L) < 0, & \text{if } R_L > R_S. \end{cases}$$

Hence, according to First Derivative Test, we know that $R_L = R_S$ maximizes $f(R_L)$, i.e., the maximum power transfer occurs in the load when the load resistance (R_L) equals to the source resistance (R_S). This is called a “matched condition”, and it illustrates the theorem that the maximum power would be transferred to the output when the impedance of the external device matches that of the source. Note that improper impedance matching would lead to excessive power use and dissipation.

3.41 Engineering Example: projectile motion

Consider the case where a ball is thrown with initial velocity u and an angle θ ($0^\circ \leq \theta \leq 90^\circ$) against the ground.

- Find the relationship between the horizontal distance that the ball can reach s_x , and the variables u and θ .
- Prove that in order to throw a ball as far as possible, we need to start the motion with an angle of 45° .



1. First, we note that for the motion, the only force that affects the system is the gravitational force in the vertical direction. For the vertical movement, we have

$$u_y = u \sin \theta$$

where u_y is the velocity in the y -axis direction.

With the velocity in the y -axis direction u_y , we can calculate the displacement of the ball in the y -direction at time t , which is shown in the following:

$$\begin{aligned} s_y &= u_y t - \frac{1}{2} g t^2 \\ &= (u \sin \theta) t - \frac{1}{2} g t^2. \end{aligned}$$

When the ball hits the ground again, i.e., $s_y = 0$, we have

$$\begin{aligned} (u \sin \theta) t - \frac{1}{2} g t^2 &= 0. \\ \Rightarrow t \left(\frac{1}{2} g t - u \sin \theta \right) &= 0. \end{aligned}$$

Solving the above equation, we can obtain two roots: $t = 0$ and $t = \frac{2u \sin \theta}{g}$. Note that at $t = 0$, the ball is at its initial position. Hence, the ball hits the ground again at $t = \frac{2u \sin \theta}{g}$. In other words, the projectile motion ends at $t = \frac{2u \sin \theta}{g}$ and the ball hits the ground again.

Next, we consider the horizontal movement of the ball. The velocity in the x -axis direction can be calculated as follows:

$$u_x = u \cos \theta.$$

With the velocity in the x -axis direction u_x , we can calculate the displacement of the ball in the x -direction at time t , which is shown in the following:

$$s_x = u_x t = (u \cos \theta) t.$$

At $t = \frac{2u \sin \theta}{g}$, the displacement of the ball in the x -direction is

$$s_x = u \frac{2u \sin \theta}{g} \cos \theta = \frac{2u^2}{g} \sin \theta \cos \theta.$$

2. Let $f(\theta) = s_x = \frac{2u^2}{g} \sin \theta \cos \theta$. To calculate the maximum horizontal movement of the ball, i.e., to maximize s_x , we first need to calculate $f'(\theta)$. According to Product Rule, the calculation is shown in the following:

$$f'(\theta) = \frac{2u^2}{g} (\cos^2 \theta - \sin^2 \theta).$$

In the range of $(0, \frac{\pi}{2})$, $f'(\theta) = 0$ has a solution at $\theta = \frac{\pi}{4}$. Moreover, it can be easily verified that

$$\begin{cases} f'(\theta) > 0, & \text{if } 0 \leq \theta < \frac{\pi}{4}; \\ f'(\theta) < 0, & \text{if } \frac{\pi}{4} < \theta \leq \pi. \end{cases}$$

Hence, according to First Derivative Test, we know that $\theta = \frac{\pi}{4}$ is a local maximum point. In order to prove that $\theta = \frac{\pi}{4}$ is also an absolute maximum point, we need to calculate the values of $f(0)$ and $f(\pi)$. It can be easily calculated that $f(0) = f(\pi) = 0 < f\left(\frac{\pi}{4}\right)$. Therefore, $\theta = \frac{\pi}{4}$ is also an absolute maximum point, i.e., in order to throw a ball as far as possible, we need to start the motion with an angle of 45° .

MATLAB solution:

We now turn to the use of MATLAB. Let $f(\theta) = s_x = \frac{2u^2}{g} \sin \theta \cos \theta$. Specifically, we calculate $f'(\theta)$ and the roots of $f'(\theta) = 0$ using MATLAB.

```
>> syms a x % a:2u^2/g, x:theta
>> f(x) = a*sin(x)*cos(x);
>> df(x) = diff(f(x),x)

df(x) =
a*cos(x)^2 - a*sin(x)^2

>> root_df = solve(df(x) == 0,x)
root_df =
pi/4
```

From the MATLAB code above, we see that $f'\left(\frac{\pi}{4}\right) = 0$, i.e., $f'(45^\circ) = 0$. In order to check whether $f'\left(\frac{\pi}{4}\right) = 0$ is a local maximum point or a local minimum point, we need to know the sign of $f'(\theta)$ (i.e., positive or negative) when $0 \leq \theta < \frac{\pi}{4}$ and $\frac{\pi}{4} < \theta \leq \pi$, respectively. Since $a = \frac{2u^2}{g} > 0$, we plot the figures of $g(\theta) = \cos^2 \theta - \sin^2 \theta$ for $0 \leq \theta < \frac{\pi}{4}$ and $\frac{\pi}{4} < \theta \leq \pi$, respectively, using the following MATLAB code.

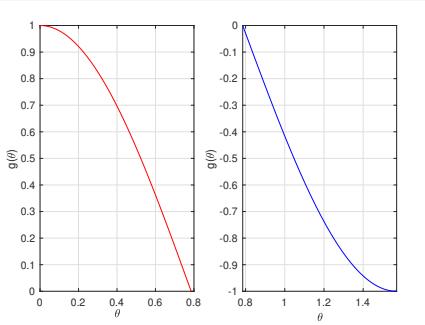
```
>> syms x % x:theta
>> g(x) = cos(x)^2 - sin(x)^2;
>> x_array1 = linspace(0,pi/4,100);
>> g_array1 = g(x_array1);
>> subplot(1,2,1)
>> plot(x_array1,g_array1,'r')
>> xlabel('\theta')
>> ylabel('g(\theta)')
```

```

>> grid
>> x_array2 = linspace(pi/4,pi/2,100);
>> g_array2 = g(x_array2);
>> subplot(1,2,2)
>> plot(x_array2,g_array2,'b')
>> xlabel('\theta')
>> ylabel('g(\theta)')
>> grid

```

The figure plotted is shown as follows:



Note that $\frac{\pi}{4} \approx 0.7854$. We thus see from the figure above that

$$\begin{cases} g(\theta) > 0, & \text{if } 0 \leq \theta < \frac{\pi}{4}; \\ g(\theta) < 0, & \text{if } \frac{\pi}{4} < \theta \leq \pi \end{cases} \implies \begin{cases} f'(\theta) > 0, & \text{if } 0 \leq \theta < \frac{\pi}{4}; \\ f'(\theta) < 0, & \text{if } \frac{\pi}{4} < \theta \leq \pi. \end{cases}$$

According to First Derivative Test, we know that $\theta = \frac{\pi}{4}$ is a local maximum point. In order to prove that $\theta = \frac{\pi}{4}$ is also an absolute maximum point, we need to calculate the values of $f(0)$ and $f(\pi)$. It can be easily obtained that $f(0) = f(\pi) = 0 < f\left(\frac{\pi}{4}\right)$. Therefore, $\theta = \frac{\pi}{4}$ is also an absolute maximum point, i.e., in order to throw a ball as far as possible, we need to start the motion with an angle of 45° .

3.42 Example: indeterminate forms

- In calculating limits, we often encounter indeterminate forms such as $\lim_{x \rightarrow 0} \frac{\sin x}{x}$ where it becomes the indeterminate form $\frac{0}{0}$ if we naively plug $x = 0$ into the expression. In order to calculate the limit, we used some geometry of the sine function to get $\lim_{x \rightarrow 0} \frac{\sin x}{x} = 1$.

2. For the limit $\lim_{x \rightarrow 1} \frac{x^2 - 1}{x - 1}$, if we directly plug $x = 1$ into the expression, we will get the indeterminate form $\frac{0}{0}$. To calculate the actual limit, we can perform the cancellation $\frac{x^2 - 1}{x - 1} = x + 1$ and $\lim_{x \rightarrow 1} \frac{x^2 - 1}{x - 1} = \lim_{x \rightarrow 1} (x + 1) = 2$.
3. For the limit $\lim_{x \rightarrow 1} \frac{\ln x}{x - 1}$, it is an indeterminate form $\frac{0}{0}$ if we plug $x = 1$ into the expression. We will develop a general approach to find such a limit.

3.43 Definition: indeterminate forms

The following are typical **indeterminate forms**:

$$\left. \begin{array}{l} \text{indeterminate form of type } \frac{0}{0} \\ \text{indeterminate form of type } \frac{\infty}{\infty} \\ \text{indeterminate form of type } 0 \cdot \infty \\ \text{indeterminate form of type } \infty - \infty \\ \text{indeterminate form of type } 0^0 \\ \text{indeterminate form of type } \infty^0 \\ \text{indeterminate form of type } 1^\infty \end{array} \right\} \begin{array}{l} \text{two standard indeterminate forms} \\ \\ \text{indeterminate product} \\ \text{indeterminate difference} \\ \\ \text{indeterminate powers} \end{array}$$

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

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

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

3.44 Example: indeterminate forms

The following are some examples of indeterminate forms:

- $\frac{0}{0}$: $\lim_{x \rightarrow 1} \frac{\ln x}{x - 1}$;
- $\frac{\infty}{\infty}$: $\lim_{x \rightarrow \infty} \frac{e^x}{x}$;
- $0 \cdot \infty$: $\lim_{x \rightarrow 0^+} (\sin x) \ln x$;
- $\infty - \infty$: $\lim_{x \rightarrow +\infty} (\sqrt{x+1} - \sqrt{x})$;
- 0^0 : $\lim_{x \rightarrow 0^+} x^x$;
- ∞^0 : $\lim_{x \rightarrow +\infty} x^{1/x}$;
- 1^∞ : $\lim_{x \rightarrow \infty} \left(1 - \frac{1}{x}\right)^x$.

3.45 Theorem: L'Hospital's Rule

Suppose

1. f and g are differentiable functions;
2. $g'(x)$ is not zero on an open interval I containing a except possibly at a ;
3. $\lim_{x \rightarrow a} f(x) = 0 = \lim_{x \rightarrow a} g(x)$ or $\lim_{x \rightarrow a} f(x) = \infty = \lim_{x \rightarrow a} g(x)$;
4. $\lim_{x \rightarrow a} \frac{f'(x)}{g'(x)}$ exists or is ∞ .

Then

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

The rule also holds when limits are one-sided limits or when $x \rightarrow a$ is replaced by $x \rightarrow \infty$.

3.46 Example: $\frac{0}{0}$ form

Find the limit $\lim_{x \rightarrow 1} \frac{\ln x}{x - 1}$.

Solution For the given limit, we can verify the hypotheses of l'Hospital's Rule:

1. Both functions $\ln x$ and $x - 1$ are differentiable;
2. $(x - 1)' = 1$ is nowhere zero near $x = 1$;
3. $\lim_{x \rightarrow 1} \ln x = 0$ and $\lim_{x \rightarrow 1} (x - 1) = 0$;
4. $\lim_{x \rightarrow 1} \frac{(\ln x)'}{(x - 1)'} = \lim_{x \rightarrow 1} \frac{1}{x} = 1$.

Thus, legitimately we can apply l'Hospital's Rule and get

$$\lim_{x \rightarrow 1} \frac{\ln x}{x - 1} \stackrel{H}{=} \lim_{x \rightarrow 1} \frac{(\ln x)'}{(x - 1)'} = \lim_{x \rightarrow 1} \frac{1}{x} = 1.$$

3.47 Example: $\frac{\infty}{\infty}$ form

Find the limit $\lim_{x \rightarrow +\infty} \frac{e^x}{x}$.

Solution For the given limit, we can verify the hypotheses of l'Hospital's Rule:

1. Both functions e^x and x are differentiable;
2. $x' = 1$ is nowhere zero;
3. $\lim_{x \rightarrow +\infty} e^x = +\infty$ and $\lim_{x \rightarrow +\infty} x = +\infty$;
4. $\lim_{x \rightarrow +\infty} \frac{(e^x)'}{(x)'} = \lim_{x \rightarrow +\infty} e^x = +\infty$.

Thus, legitimately we can apply l'Hospital's Rule and get

$$\lim_{x \rightarrow +\infty} \frac{e^x}{x} \stackrel{\text{H}}{=} \lim_{x \rightarrow +\infty} \frac{(e^x)'}{(x)'} = \lim_{x \rightarrow +\infty} e^x = +\infty.$$

3.48 Example: using l'Hospital's Rule repeatedly

Find the limit $\lim_{x \rightarrow +\infty} \frac{e^x}{x^2}$.

Analysis: If we apply l'Hospital's Rule, we get

$$\lim_{x \rightarrow +\infty} \frac{e^x}{x^2} \stackrel{\text{H}}{=} \lim_{x \rightarrow +\infty} \frac{(e^x)'}{(x^2)'} = \lim_{x \rightarrow +\infty} \frac{e^x}{2x}.$$

For the last limit, we again apply l'Hospital's Rule and get

$$\lim_{x \rightarrow +\infty} \frac{e^x}{2x} \stackrel{\text{H}}{=} \lim_{x \rightarrow +\infty} \frac{(e^x)'}{(2x)'} = \lim_{x \rightarrow +\infty} \frac{e^x}{2} = +\infty.$$

These two steps need to be justified by checking the hypotheses of l'Hospital's Rule.

Solution For the limit $\lim_{x \rightarrow +\infty} \frac{e^x}{2x}$, we can verify the hypotheses of l'Hospital's Rule:

1. Both functions e^x and $2x$ are differentiable;
2. $(2x)' = 2$ is nowhere zero when x is large;
3. $\lim_{x \rightarrow +\infty} e^x = +\infty$ and $\lim_{x \rightarrow +\infty} 2x = +\infty$;
4. $\lim_{x \rightarrow +\infty} \frac{(e^x)'}{(2x)'} = \lim_{x \rightarrow +\infty} \frac{e^x}{2} = +\infty$.

Thus, legitimately we can apply l'Hospital's Rule and get

$$\lim_{x \rightarrow +\infty} \frac{e^x}{2x} \stackrel{\text{H}}{=} \lim_{x \rightarrow +\infty} \frac{(e^x)'}{(2x)'} = \lim_{x \rightarrow +\infty} \frac{e^x}{2} = +\infty.$$

For the original limit $\lim_{x \rightarrow +\infty} \frac{e^x}{x^2}$, we can verify the hypotheses of l'Hospital's Rule:

1. Both functions e^x and x^2 are differentiable;
2. $(x^2)' = 2x$ is nowhere zero when x is large;
3. $\lim_{x \rightarrow +\infty} e^x = +\infty$ and $\lim_{x \rightarrow +\infty} x^2 = +\infty$;
4. As shown above, $\lim_{x \rightarrow +\infty} \frac{(e^x)'}{(x^2)'} = \lim_{x \rightarrow +\infty} \frac{e^x}{2x} = +\infty$.

Thus, legitimately we can apply l'Hospital's Rule and get

$$\lim_{x \rightarrow +\infty} \frac{e^x}{x^2} \stackrel{\text{H}}{=} \lim_{x \rightarrow +\infty} \frac{(e^x)'}{(x^2)'} = +\infty.$$

Remark:

Since the verification of the hypotheses of l'Hospital's Rule is straightforward, often we omit this step. Thus, for finding the limit $\lim_{x \rightarrow +\infty} \frac{e^x}{x^2}$, typically, we simply write:

$$\begin{aligned}\lim_{x \rightarrow +\infty} \frac{e^x}{x^2} &\stackrel{\text{H}}{=} \lim_{x \rightarrow +\infty} \frac{(e^x)'}{(x^2)'} = \lim_{x \rightarrow +\infty} \frac{e^x}{2x} \\ &\stackrel{\text{H}}{=} \lim_{x \rightarrow +\infty} \frac{(e^x)'}{(2x)'} = \lim_{x \rightarrow +\infty} \frac{e^x}{2} = +\infty.\end{aligned}$$

3.49 Pitfall: l'Hospital's Rule

1. Consider $\lim_{x \rightarrow 0} \frac{1 - \sin x}{1 + \sin x}$.

If one calculates the limit by using l'Hospital's Rule without verifying the hypotheses, one would get

$$\lim_{x \rightarrow 0} \frac{1 - \sin x}{1 + \sin x} \stackrel{\text{H}}{=} \lim_{x \rightarrow 0} \frac{(1 - \sin x)'}{(1 + \sin x)'} = \lim_{x \rightarrow 0} \frac{-\cos x}{\cos x} = -1.$$

However, this is a wrong value! The correct value can be obtained by the continuity:

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

The reason for the failure is that one of the hypotheses of l'Hospital's Rule is not satisfied. In fact, we see that $\lim_{x \rightarrow 0} (1 - \sin x) = 1$, which is neither zero nor ∞ .

2. Consider $\lim_{x \rightarrow +\infty} \frac{x + \sin x}{x}$.

If one applies l'Hospital's Rule, one would get

$$\lim_{x \rightarrow +\infty} \frac{x + \sin x}{x} \stackrel{\text{H}}{=} \lim_{x \rightarrow +\infty} \frac{(x + \sin x)'}{(x)'} = \lim_{x \rightarrow +\infty} (1 + \cos x).$$

Clearly, the last limit does not exist. However, we have

$$\lim_{x \rightarrow +\infty} \frac{x + \sin x}{x} = \lim_{x \rightarrow +\infty} \left(1 + \frac{\sin x}{x}\right) = 1.$$

This contradiction shows that not all indeterminate form can apply l'Hospital's Rule. The failure in this example is due to the fact that the limit $\lim_{x \rightarrow +\infty} (1 + \cos x)$ does not exist. This is one of the hypotheses of l'Hospital's Rule.

3.50 Example: $0 \cdot \infty$ form

Find the limit $\lim_{x \rightarrow 0^+} (\sin x) \ln x$.

Solution The given limit is indeterminate form of $0 \cdot \infty$. We may first convert it to a form of $\frac{\infty}{\infty}$ then apply l'Hospital's Rule, as shown in the following:

$$\begin{aligned}\lim_{x \rightarrow 0^+} (\sin x) \ln x &= \lim_{x \rightarrow 0^+} \frac{\ln x}{(\sin x)^{-1}} \\ &\stackrel{\text{H}}{=} \lim_{x \rightarrow 0^+} \frac{(\ln x)'}{[(\sin x)^{-1}]'} \\ &= \lim_{x \rightarrow 0^+} \frac{x^{-1}}{-(\sin x)^{-2} \cdot \cos x} \\ &= \lim_{x \rightarrow 0^+} \frac{\sin^2 x}{-x \cos x} = 0.\end{aligned}$$

Remark:

We may first convert it to a form of $\frac{0}{0}$ then apply l'Hospital's Rule, as shown in the following:

$$\begin{aligned}\lim_{x \rightarrow 0^+} (\sin x) \ln x &= \lim_{x \rightarrow 0^+} \frac{\sin x}{(\ln x)^{-1}} \\ &\stackrel{\text{H}}{=} \lim_{x \rightarrow 0^+} \frac{(\sin x)'}{[(\ln x)^{-1}]'} \\ &= \lim_{x \rightarrow 0^+} \frac{\cos x}{-(\ln x)^{-2} \cdot x^{-1}} \\ &= \lim_{x \rightarrow 0^+} [-x \ln^2 x \cdot \cos x].\end{aligned}$$

We see that the last limit seems more complicated than the original one. This shows that sometimes it is an art to convert indeterminate form to $\frac{0}{0}$ or $\frac{\infty}{\infty}$ in order to rightfully apply l'Hospital's Rule.

3.51 Example: $\infty - \infty$ form

Find the limit $\lim_{x \rightarrow +\infty} (\sqrt{x+1} - \sqrt{x})$.

Solution The given limit is indeterminate form of $\infty - \infty$. We may first convert it to a

form of $\frac{0}{0}$ then apply l'Hospital's Rule, as shown in the following:

$$\begin{aligned} \lim_{x \rightarrow +\infty} (\sqrt{x+1} - \sqrt{x}) &= \lim_{x \rightarrow +\infty} \sqrt{x} \left(\frac{\sqrt{x+1}}{\sqrt{x}} - 1 \right) \\ &= \lim_{x \rightarrow +\infty} \frac{\sqrt{1+x^{-1}} - 1}{x^{-1/2}} \\ &\stackrel{\text{H}}{=} \lim_{x \rightarrow +\infty} \frac{(\sqrt{1+x^{-1}} - 1)'}{(x^{-1/2})'} \\ &= \lim_{x \rightarrow +\infty} \frac{\frac{1}{2}(1+x^{-1})^{-1/2} \cdot (-x^{-2})}{-\frac{1}{2}x^{-3/2}} \\ &= \lim_{x \rightarrow +\infty} \frac{1}{\sqrt{x} \cdot \sqrt{1+x^{-1}}} = 0. \end{aligned}$$

3.52 Example: 0^0 form

Find the limit $\lim_{x \rightarrow 0^+} x^x$.

Solution Let $y = x^x$. If taking logarithm of both sides, we get $\ln y = \ln x^x = x \ln x$. Clearly, $x \ln x$ is $0 \cdot \infty$ form. We can change it into $\frac{\infty}{\infty}$ form $\frac{\ln x}{x^{-1}}$ and apply l'Hospital's Rule to get

$$\begin{aligned} \lim_{x \rightarrow 0^+} x \ln x &= \lim_{x \rightarrow 0^+} \frac{\ln x}{x^{-1}} \\ &\stackrel{\text{H}}{=} \lim_{x \rightarrow 0^+} \frac{(\ln x)'}{(x^{-1})'} \\ &= \lim_{x \rightarrow 0^+} \frac{x^{-1}}{-x^{-2}} \\ &= \lim_{x \rightarrow 0^+} (-x) = 0. \end{aligned}$$

Since $y = e^{\ln y}$, by the continuity of the exponential function, we get

$$\lim_{x \rightarrow 0^+} x^x = \lim_{x \rightarrow 0^+} e^{\ln y} = e^0 = 1.$$

3.53 Example: ∞^0 form

Find the limit $\lim_{x \rightarrow +\infty} x^{1/x}$.

Solution Let $y = x^{1/x}$. If taking logarithm of both sides, we get $\ln y = \ln x^{1/x} = \frac{\ln x}{x}$.

Clearly, $\frac{\ln x}{x}$ is $\frac{\infty}{\infty}$ form. Applying l'Hospital's Rule gives

$$\begin{aligned}\lim_{x \rightarrow +\infty} \frac{\ln x}{x} &\stackrel{\text{H}}{=} \lim_{x \rightarrow +\infty} \frac{(\ln x)'}{(x)'} \\ &= \lim_{x \rightarrow +\infty} \frac{x^{-1}}{1} = 0.\end{aligned}$$

Since $y = e^{\ln y}$, by the continuity of the exponential function, we get

$$\lim_{x \rightarrow +\infty} x^{1/x} = \lim_{x \rightarrow +\infty} e^{\ln y} = e^0 = 1.$$

3.54 Example: 1^∞ form

Find the limit $\lim_{x \rightarrow \infty} \left(1 - \frac{1}{x}\right)^x$.

Solution Let $y = \left(1 - \frac{1}{x}\right)^x$. If taking logarithm of both sides, we get $\ln y = x \ln \left(1 - \frac{1}{x}\right)$.

Clearly, it is $0 \cdot \infty$ form. We can change it into $\frac{0}{0}$ form

$$\frac{\ln \left(1 - \frac{1}{x}\right)}{x^{-1}}.$$

If we let $z = x^{-1}$, then $x \rightarrow \infty \iff z \rightarrow 0$. Thus, by applying l'Hospital's Rule, we get

$$\begin{aligned}\lim_{x \rightarrow \infty} \frac{\ln \left(1 - \frac{1}{x}\right)}{x^{-1}} &= \lim_{z \rightarrow 0} \frac{\ln(1-z)}{z} \\ &\stackrel{\text{H}}{=} \lim_{z \rightarrow 0} \frac{[\ln(1-z)]'}{(z)'} \\ &= \lim_{z \rightarrow 0} \frac{-1}{1-z} = -1.\end{aligned}$$

Since $y = e^{\ln y}$, by the continuity of the exponential function, we get

$$\lim_{x \rightarrow \infty} \left(1 - \frac{1}{x}\right)^x = \lim_{x \rightarrow \infty} e^{\ln y} = e^{-1}.$$

3.55 MATLAB: indeterminate forms

We use MATLAB to calculate the following limits:

- $\lim_{x \rightarrow 1} \frac{\ln x}{x - 1};$
- $\lim_{x \rightarrow \infty} \frac{e^x}{x};$
- $\lim_{x \rightarrow 0^+} (\sin x) \ln x;$
- $\lim_{x \rightarrow +\infty} (\sqrt{x+1} - \sqrt{x});$
- $\lim_{x \rightarrow 0^+} x^x;$
- $\lim_{x \rightarrow +\infty} x^{1/x};$
- $\lim_{x \rightarrow \infty} \left(1 - \frac{1}{x}\right)^x.$

```

>> syms x;
>> f = log(x)/(x-1);
>> limit(f,x,1)

ans =
1

>> f = exp(x)/x;
>> limit(f,x,inf)

ans =
Inf

>> f = sin(x)*log(x);
>> limit(f,x,0,'right')

ans =
0

>> f = sqrt(x + 1) - sqrt(x);
>> limit(f,x,inf)

ans =
0

>> f = x^x;
>> limit(f,x,0,'right')

ans =
1

```

```

>> f = x^(1/x);
>> limit(f,x,inf)

ans =
1

>> f = (1 - 1/x)^x;
>> limit(f,x,inf)

ans =
exp(-1)

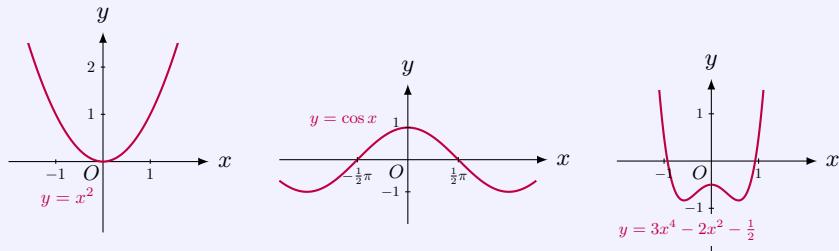
```

3.56 Definition: symmetry

- (A) Suppose the domain D of f is symmetric about the origin.
- f is **even** if $f(-x) = f(x)$ for all x in D .
 - f is **odd** if $f(-x) = -f(x)$ for all x in D .
- (B) A function f is said to be **periodic** 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**.

3.57 Example: even and odd functions

1. $f(x) = x^2$, $g(x) = \cos x$, and $h(x) = 3x^4 - 2x^2 - \frac{1}{2}$ are even functions.

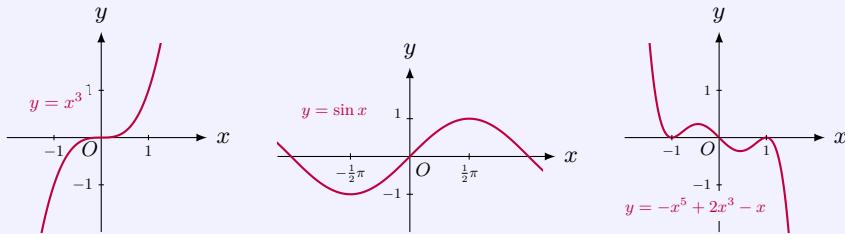


Suppose (a, b) is a point on the graph of an even function $y = F(x)$. Since

$$F(-a) = F(a) = b,$$

we know that $(-a, b)$ is also on the graph of F . This means that the graph of every even function is *symmetric with respect to the y-axis*.

2. $f(x) = x^3$, $g(x) = \cos x$, and $h(x) = -x^5 + 2x^3 - x$ are odd functions.



Suppose (a, b) is a point on the graph of an odd function $y = F(x)$. Since

$$F(-a) = -F(a) = -b,$$

we know that $(-a, -b)$ is also on the graph of F . This means that the graph of every odd function is *symmetric with respect to the origin*.

3.58 Method: guidelines for curve sketching

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 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 even function and the curve is symmetric about the y -axis.
- (ii) If $f(-x) = -f(x)$ for all $x \in D$, then f is an odd function 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 periodic function and the smallest such number p is the period.

D. Asymptotes:

- (i) The line $y = L$ is a horizontal asymptote of the curve $y = f(x)$ if either

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

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

$$\begin{aligned} \lim_{x \rightarrow a} f(x) &= +\infty, & \lim_{x \rightarrow a^-} f(x) &= +\infty, & \lim_{x \rightarrow a^+} f(x) &= +\infty, \\ \lim_{x \rightarrow a} f(x) &= -\infty, & \lim_{x \rightarrow a^-} f(x) &= -\infty, & \lim_{x \rightarrow a^+} f(x) &= -\infty. \end{aligned}$$

E. Intervals of Increase and Decrease:

Apply the Increasing/Decreasing Test:

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

F. Local Maximum and Minimum Values:

First find the critical points, 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) 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) 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) Inflection Points: 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.**3.59 Example:** curve sketching

Sketch the function $f(x) = 2x^3 - 3x^2 - 12x + 1$.

A. Domain: The function f is defined at every point, so that the domain is the whole real line $(-\infty, +\infty)$.

B. Intercepts: The y -intercept is $f(0) = 1$. The x -intercepts are the real solutions of the equation $f(x) = 0$, that is,

$$2x^3 - 3x^2 - 12x + 1 = 0.$$

The exact values of solutions can be approximated by using the Intermediate Value Theorem. In fact, we can obtain the following bounds for the three roots of f :

$$\begin{cases} f(-2) = -3, & f(-1.5) = 5.5 \implies -2 < x_1 < -1.5; \\ f(0) = 1, & f(0.5) = -5.5 \implies 0 < x_2 < 0.5; \\ f(3) = -8, & f(3.5) = 8 \implies 3 < x_3 < 3.5. \end{cases}$$

C. Symmetry: The function f is neither even nor odd, and it is not a periodic function either.

D. Asymptotes: The function f is defined everywhere and thus has no vertical asymptote. Since

$$\lim_{x \rightarrow \pm\infty} (2x^3 - 3x^2 - 12x + 1) = \lim_{x \rightarrow \pm\infty} x^3 \left(2 - \frac{3}{x} - \frac{12}{x^2} + \frac{1}{x^3} \right) = \pm\infty,$$

the function f does not have any horizontal asymptote either.

E. Intervals of Increase or Decrease: Since

$$f'(x) = 6x^2 - 6x - 12 = 6(x^2 - x - 2) = 6(x + 1)(x - 2),$$

we see that the derivative f' has two simple roots $x = -1$ and $x = 2$. These two points divide the interval $(-\infty, +\infty)$ into three open intervals $(-\infty, -1)$, $(-1, 2)$, and $(2, +\infty)$. We solve the inequalities $f'(x) \geq 0$ by using the number line and know that f' is positive on $(-\infty, -1)$, negative in $(-1, 2)$, and positive in $(2, +\infty)$. Thus, by the Monotone Test, the function f is increasing over the intervals $(-\infty, -1)$ and $(2, +\infty)$, and decreasing over the interval $(-1, 2)$.

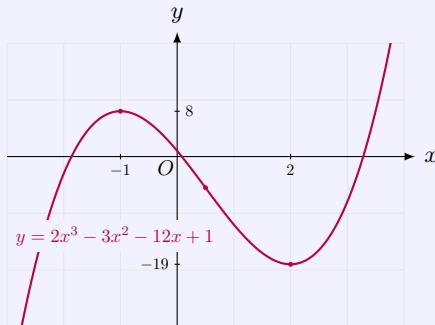
F. Local Maximum and Minimum Values: The roots of f' , $x = -1$ and $x = 2$, are critical points of f . The second derivative

$$f''(x) = 12x - 6$$

gives $f''(-1) < 0$ and $f''(2) > 0$. By the Second Derivative Test, f has a local maximum value $f(-1) = 8$ and a local minimum value $f(2) = -19$.

G. Concavity and Inflection Points: Since $f''(x) < 0$ when $x < \frac{1}{2}$, by the Concavity Test, f is concave down on $(-\infty, \frac{1}{2})$. Similarly, since $f''(x) > 0$ when $x > \frac{1}{2}$, f is concave up on $(\frac{1}{2}, +\infty)$. Thus, $x = \frac{1}{2}$ is an inflection point, with $f(\frac{1}{2}) = -\frac{11}{2}$.

H. Sketch the Curve: Use the results obtained in items A-G, we can sketch the graph of $f(x) = 2x^3 - 3x^2 - 12x + 1$.



3.60 Example: curve sketching

Sketch the function $f(x) = \frac{x}{x^2 + 1}$.

A. Domain: The function f is defined at every point, so that the domain is the whole real line $(-\infty, +\infty)$.

B. Intercepts: The y -intercept is $f(0) = 0$. The x -intercept(s) is the real solutions of the

equation $f(x) = 0$, that is,

$$\frac{x}{x^2 + 1} = 0,$$

which gives the unique solution $x = 0$. Thus, the curve passes through the x -axis at the origin.

C. Symmetry: Since

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

the function f is odd, so the graph is symmetric about the origin.

The function f is not a periodic function.

D. Asymptotes: The function f is defined everywhere and thus has no vertical asymptote. Since

$$\lim_{x \rightarrow \pm\infty} \frac{x}{x^2 + 1} = \lim_{x \rightarrow \pm\infty} \frac{1}{x + \frac{1}{x}} = 0,$$

the function f has the x -axis as its horizontal asymptote.

E. Intervals of Increase or Decrease, and

F. Local Maximum and Minimum Values: Since

$$f'(x) = -\frac{(x+1)(x-1)}{(x^2+1)^2},$$

we see that the derivative f' has two simple roots $x = -1$ and $x = 1$. These two points divide the interval $(-\infty, +\infty)$ into three open intervals $(-\infty, -1)$, $(-1, 1)$, and $(1, +\infty)$. We solve the inequalities $f'(x) \geq 0$ by using the number line, we know that f' is negative on $(-\infty, -1)$, positive in $(-1, 1)$, and negative in $(1, +\infty)$. Thus, by the Monotone Test and the First Derivative Test, we can have the following table:

interval	sign of f'	monotonicity
$(-\infty, -1)$	—	\searrow
-1	0	local min
$(-1, 1)$	+	\nearrow
1	0	local max
$(1, +\infty)$	—	\searrow

This table gives the intervals in which the function is monotone and the locations of the local extrema. In fact, the local minimum value is $f(-1) = -\frac{1}{2}$ and the local maximum value is $f(1) = \frac{1}{2}$.

G. Concavity and Inflection Points: The second derivative f'' is

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

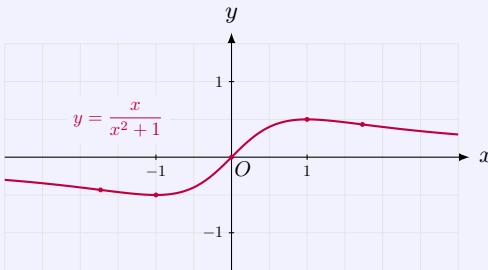
It has three simple roots $x = -\sqrt{3}$, $x = 0$, and $x = \sqrt{3}$. These three points divide the interval $(-\infty, +\infty)$ into four open intervals $(-\infty, -\sqrt{3})$, $(-\sqrt{3}, 0)$, $(0, \sqrt{3})$, and $(\sqrt{3}, +\infty)$. We solve the inequalities $f''(x) \geq 0$ by using the number line and know that f'' is negative

in $(-\infty, -\sqrt{3})$, positive in $(-\sqrt{3}, 0)$, negative in $(0, \sqrt{3})$, and positive in $(\sqrt{3}, +\infty)$. Thus, by the Concavity Test, we can have the following table:

interval	sign of f''	concavity
$(-\infty, -\sqrt{3})$	—	⌞
$-\sqrt{3}$	0	inflection
$(-\sqrt{3}, 0]$	+	⌞
0	0	inflection
$(0, \sqrt{3})$	—	⌞
$\sqrt{3}$	0	inflection
$(\sqrt{3}, +\infty)$	+	⌞

This table gives the intervals in which the function is concave up/down and the locations of the inflection points. In fact, the inflection points are at $x = 0$ and $x = \pm\sqrt{3}$.

H. Sketch the Curve: Use the results obtained in items **A-G**, we can sketch the graph of $f(x) = \frac{x}{x^2 + 1}$.



3.61 Example: curve sketching

Sketch the function $f(x) = \frac{x^3}{x^2 - 1}$.

A. Domain: The function $f(x) = \frac{x^3}{x^2 - 1}$ is defined everywhere except ± 1 , so that the domain of f is $(-\infty, -1) \cup (-1, 1) \cup (1, +\infty)$.

B. Intercepts: The y -intercept is $f(0) = 0$. The x -intercept(s) is the real solutions of the equation $f(x) = 0$, that is,

$$\frac{x^3}{x^2 - 1} = 0,$$

which gives the unique solution $x = 0$. Thus, the curve passes through the x -axis at the origin.

C. Symmetry: Since

$$f(-x) = \frac{(-x)^3}{(-x)^2 - 1} = -\frac{x^3}{x^2 - 1} = -f(x),$$

the function f is odd, so the graph is symmetric about the origin.

The function f is not a periodic function.

D. Asymptotes: Since

$$\lim_{x \rightarrow -1^-} \frac{x^3}{x^2 - 1} = \lim_{x \rightarrow 1^-} \frac{x^3}{x^2 - 1} = -\infty,$$

$$\lim_{x \rightarrow -1^+} \frac{x^3}{x^2 - 1} = \lim_{x \rightarrow 1^+} \frac{x^3}{x^2 - 1} = +\infty,$$

Moreover, by long division, we have

$$\frac{x^3}{x^2 - 1} = x + \frac{x}{x^2 - 1},$$

so that

$$\lim_{x \rightarrow \infty} \left(\frac{x^3}{x^2 - 1} - x \right) = \lim_{x \rightarrow \infty} \frac{x}{x^2 - 1} = 0$$

This shows that $y = x$ is a *slant asymptote* along both infinity directions, i.e., the graph of $f(x)$ gets closer to the line $y = x$ when $x \rightarrow \infty$. We also see that the graph of f is above the asymptote for $x > 0$ and below the asymptote for $x < 0$.

E. Intervals of Increase or Decrease, and

F. Local Maximum and Minimum Values: The derivative f' is

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

Its denominator has two double roots at $x = \pm 1$, while the numerator has two simple roots $x = -\sqrt{3}$ and $x = \sqrt{3}$, and a double root $x = 0$. These five points divide the interval $(-\infty, +\infty)$ into six open intervals $(-\infty, -\sqrt{3})$, $(-\sqrt{3}, -1)$, $(-1, 0)$, $(0, 1)$, $(1, \sqrt{3})$, and $(\sqrt{3}, +\infty)$. We solve the inequalities $f'(x) \geq 0$ by using the number line and know that f' changes its signs only at $x = \pm\sqrt{3}$. By the Monotone Test and the First Derivative Test, we can have the following table:

interval	sign of f'	monotonicity
$(-\infty, -\sqrt{3})$	+	\nearrow
$-\sqrt{3}$	0	local max
$(-\sqrt{3}, -1)$	-	\searrow
-1	not exist	
$(-1, 0)$	-	\searrow
0	0	no extreme
$(0, 1)$	-	\searrow
1	not exist	
$(1, \sqrt{3})$	-	\searrow
$\sqrt{3}$	0	local min
$(\sqrt{3}, +\infty)$	+	\nearrow

This table gives the intervals in which the function is monotone and the locations of the local extrema. In fact, the local maximum value is $f(-\sqrt{3}) = -\frac{3\sqrt{3}}{2}$ and the local minimum value is $f(\sqrt{3}) = \frac{3\sqrt{3}}{2}$.

G. Concavity and Inflection Points: The second derivative f'' is

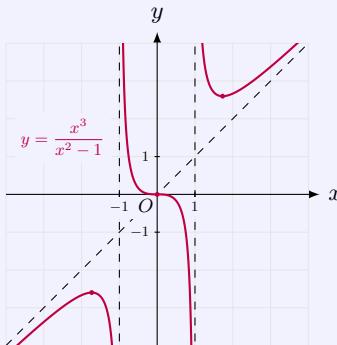
$$f''(x) = \frac{2x(x^2 + 3)}{(x^2 - 1)^3}.$$

Its denominator has two triple roots at $x = \pm 1$, while the numerator has one simple root $x = 0$. These three points divide the interval $(-\infty, +\infty)$ into four open intervals $(-\infty, -1)$, $(-1, 0)$, $(0, 1)$, and $(1, +\infty)$. We solve the inequalities $f''(x) \geq 0$ by using the number line and know that f'' changes its signs at $x = \pm 1$ and $x = 0$. By the Concavity Test, we can have the following table:

interval	sign of f''	concavity
$(-\infty, -1)$	-	\smile
-1	not exist	
$(-1, 0)$	+	\frown
0	0	inflection
$(0, 1)$	-	\smile
1	not exist	
$(1, +\infty)$	+	\frown

The function has one inflection point at $x = 0$, with $f(0) = 0$. The other two points $x = \pm 1$ are not inflection points, since they are not in the domain of f .

H. Sketch the Curve: Use the results obtained in items A-G, we can sketch the graph of $f(x) = \frac{x^3}{x^2 - 1}$.



3.62 MATLAB: curve sketching

There are several commands to plot figures in MATLAB.

1. The following example shows the plots of $f(x) = 2x^3 - 3x^2 - 12x + 1$ and $g(x) = \frac{x}{x^2 + 1}$ in the same figure.

```
| syms x;
```

```

f(x) = 2*x^3 - 3*x^2 - 12*x + 1;
g(x) = x/(x^2 + 1);

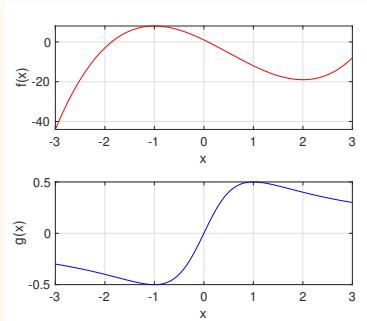
x_array = -3:0.05:3; % x-axis ranges from -3 to 3 with ...
    step size 0.05
subplot(2,1,1) % create 2 rows and 1 column of graphs and ...
    place the current figure in the 1st position
plot(x_array,f(x_array),'r')
xlabel('x')
ylabel('f(x)')
grid

subplot(2,1,2)
x_array2 = linspace(-3,3,121); % a vector with 121 evenly ...
    spaced points from -3 to 3
plot(x_array2,g(x_array2),'b')
xlabel('x')
ylabel('g(x)')
grid

```

Note that we use different commands to generate vectors for x-axis, but the vectors generated, i.e., `x_array` and `x_array2`, are the same.

The figure plotting f and g is shown below.



2. The following example shows the plot of $f(x) = \frac{x^3}{x^2 - 1}$ using the command `fplot`.

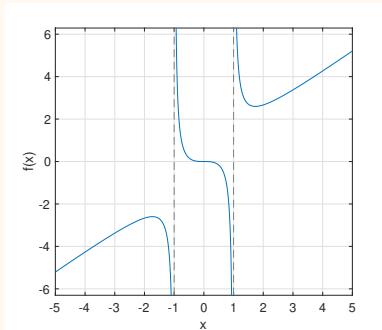
```

syms x;
f(x) = x^3/(x^2 - 1);

fplot(f(x))
xlabel('x')
ylabel('f(x)')
grid

```

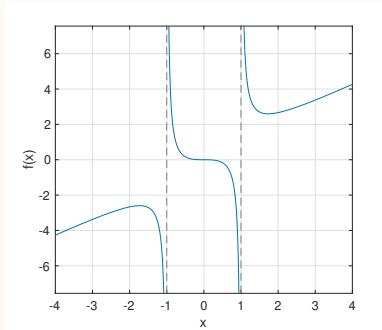
where `fplot(f,[0,5])` plots the figure of f on the default interval $[0,5]$ along the x -axis. The figure of f on $[-5,5]$ is shown below.



We can see from the figure above that $x = -1$ and $x = 1$ are vertical asymptotes of the curve $y = f(x)$.

Note that we can input the expression of f into `fplot` directly without the need of defining a symbolic function beforehand. In this way, f should be input in the form of an anonymous function. Moreover, we can also specify the interval along the x -axis for the figure to be plotted. The following example shows this alternative way of using the command `fplot`.

```
fplot(@(x)x.^3./(x.^2 - 1),[-4,4]) % 1st input: anonymous ...
    function; 2nd input: x-axis ranges from -4 to 4
xlabel('x')
ylabel('f(x)')
grid
```



The figure above shows f on the interval $[-4,4]$ (except $x \neq \pm 1$).

3.6 Parametric Curves

3.63 Definition:

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 so-called **parametric equations**:

$$x = f(t), \quad y = g(t), \quad a \leq t \leq b.$$

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.

3.64 Example:

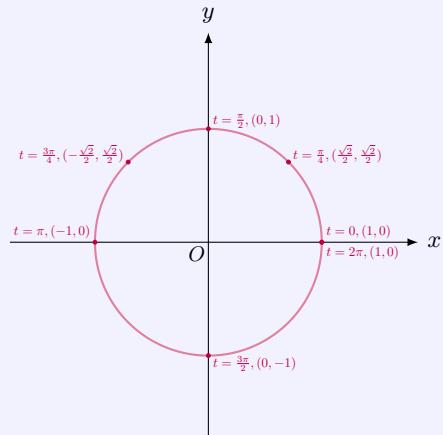
parametric equations and parametric curve

The parametric equations

$$x = \cos t, \quad y = \sin t, \quad 0 \leq t \leq 2\pi$$

describe the unit circle. In fact, for each value of t , the corresponding ordered pair (x, y) is a point on the circle centered at $(0, 0)$ with radius 1. To be more explicit, for some special values of t , we can generate the following table and the corresponding points (x, y) which are on the circle.

t	(x, y)
0	$(1, 0)$
$\frac{\pi}{4}$	$(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$
$\frac{\pi}{2}$	$(0, 1)$
$\frac{3\pi}{4}$	$(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$
π	$(-1, 0)$
$\frac{3\pi}{2}$	$(0, -1)$
2π	$(1, 0)$



From the figure, we see that as t increases from the initial value $t = 0$ to the final value of $t = 2\pi$, the curve is generated from the initial point $(1, 0)$ to the final point $(1, 0)$. The parametric equations trace out the unit circle in the counterclockwise direction. The initial point and the final point coincide at the same location.

3.65 Example: parametric equations of a line

The parametric equations

$$x = x_0 + at, \quad y = y_0 + bt, \quad -\infty < t < +\infty,$$

where x_0 , y_0 , a , and b are constants, describe a line with slope $\frac{b}{a}$, if $a \neq 0$. If $a = 0$ and $b \neq 0$, the line is vertical.

3.66 Example: parametric equations of cycloid

Suppose a point P starting at the origin on the circumference of a circle of radius r traces out a curve as the circle rolls along the x -axis. Show that the parametric equations of the curve are

$$x = r(t - \sin t), \quad y = r(1 - \cos t),$$

where the parameter t is the angle of the rotation of the circle with $t = 0$ when P is at the origin.

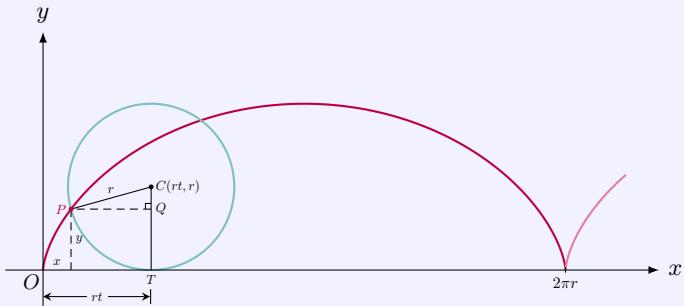
Solution We take t , the angle of the rotation of the circle, as the parameter. Suppose the circle has rotated through t radians. From the figure, we see that the distance it has rolled from the origin is

$$|OT| = \text{arc } PT = rt,$$

so that the coordinates of the center of the circle is $C(rt, r)$. Let the coordinates of P be (x, y) . Then

$$\begin{aligned} x &= |OT| - |PQ| = rt - r \sin t = r(t - \sin t), \\ y &= |TC| - |QC| = r - r \cos t = r(1 - \cos t). \end{aligned}$$

These give the parametric equations of the cycloid.

**3.67 Remark:** non-uniqueness of parametrization

A curve can have more than one parametrizations. For example, it is easy to verify that

the unit circle $x^2 + y^2 = 1$ admits the following parametrizations:

- (1) $x = \cos t, \quad y = \sin t, \quad 0 \leq t < 2\pi;$
- (2) $x = \sin t, \quad y = \cos t, \quad \frac{\pi}{2} < t \leq \frac{5\pi}{2}.$

3.68 Theorem: tangent to parametric curve

Suppose a parametric curve is given by the parametric equations

$$x = f(t), \quad y = g(t).$$

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 $(x, y) = (f(t), g(t))$:

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt} = \frac{g'(t)}{f'(t)}.$$

Theorem: horizontal tangent

The parametric curve has a horizontal tangent at $(x, y) = (f(t), g(t))$ if

$$\frac{dy}{dt} = g'(t) = 0, \quad \frac{dx}{dt} = f'(t) \neq 0.$$

Theorem: vertical tangent

The parametric curve has a vertical tangent at $(x, y) = (f(t), g(t))$ if

$$\frac{dx}{dt} = f'(t) = 0, \quad \frac{dy}{dt} = g'(t) \neq 0.$$

3.69 MATLAB: parametric curves

In MATLAB, parametric curves can be plotted using the command `plot`. The following examples show the plot of two parametric curves.

1. The parametric equations are

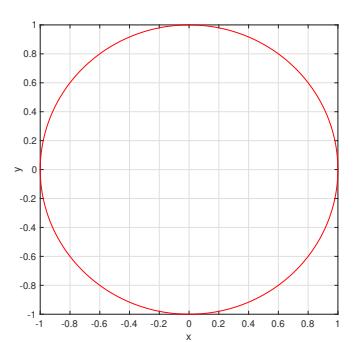
$$x = \cos t, \quad y = \sin t, \quad 0 \leq t \leq 2\pi$$

The parametric curve is plotted using the following code:

```
>> t = linspace(0, 2*pi, 200);
>> x = cos(t);
>> y = sin(t);
>> figure
```

```
>> plot(x,y, 'r')
>> grid
>> xlabel('x')
>> ylabel('y')
```

The parametric curve is shown below.



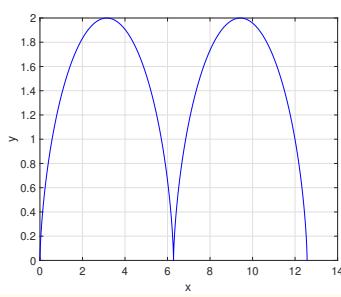
2. The parametric equations are

$$x = t - \sin t, \quad y = 1 - \cos t, \quad 0 \leq t \leq 4\pi$$

The parametric curve is plotted using the following code:

```
>> t = linspace(0,4*pi,200);
>> x = t - sin(t);
>> y = 1 - cos(t);
>> figure
>> plot(x,y, 'b')
>> grid
>> xlabel('x')
>> ylabel('y')
```

The parametric curve is shown below.



Polar Coordinates and Polar Curves

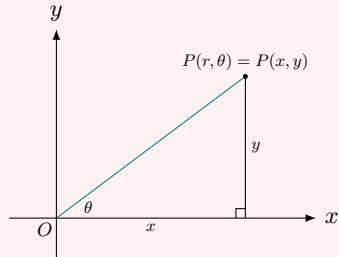
3.70 Definition: Cartesian coordinates and polar coordinates

A point (x, y) in Cartesian coordinates can be expressed in polar coordinates by

$$x = r \cos \theta, \quad y = r \sin \theta.$$

Conversely, we can express r and θ in terms of x and y :

$$r^2 = x^2 + y^2, \quad \tan \theta = \frac{y}{x}.$$



3.71 Definition: polar curve

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.

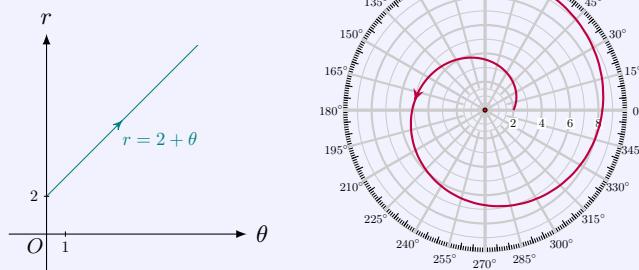
3.72 Example: polar curve - Archimedean spiral

The general Archimedean spiral is a polar curve described by the equation

$$r = a + b\theta, \quad \theta \geq 0,$$

where a and b are real numbers. Sketch the curve with $a = 2$ and $b = 1$.

Solution For $\theta \geq 0$, in Cartesian coordinates, the equation $r = 2 + \theta$ gives a line of slope 1 in the first quadrant, starting from the point $(0, 2)$, as in the following left figure. As θ increases from 0, it shows that r increases monotonically from 2. It gives a spiral curve, as shown in the following right figure.



3.73 Example: polar curve - three-leaved rose

Sketch the curve with the given polar equation

$$r = 2 \sin 3\theta.$$

Solution The function is periodic of period $\frac{2}{3}\pi$, and a full revolution is 2π , so the least common multiple is 2π . Thus, we sketch $r = 2 \sin 3\theta$, $0 \leq \theta \leq 2\pi$, in Cartesian coordinates in the following left figure.

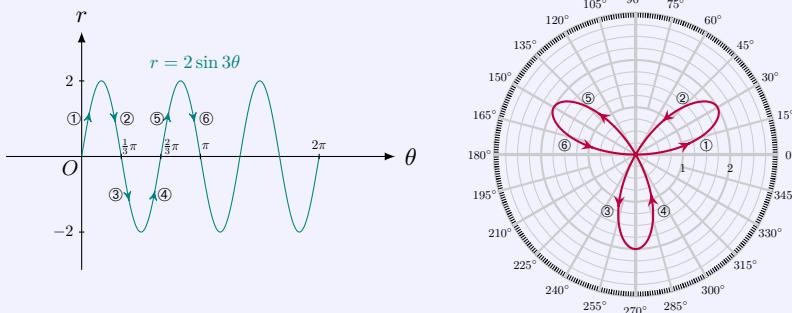
(1) As θ increases from 0 to $\frac{\pi}{6}$, r increases from 0 to 2; as θ increases from $\frac{\pi}{6}$ to $\frac{\pi}{3}$, r decreases from 2 to 0. Since r is positive, the curve is in the first quadrant.

(2) As θ increases from $\frac{\pi}{3}$ to $\frac{\pi}{2}$, r decreases further from 0 to -2; as θ increases from $\frac{\pi}{2}$ to $\frac{2\pi}{3}$, r increases from -2 to 0. Since r is negative, the curve is in the lower-half plane.

(3) As θ increases from $\frac{2\pi}{3}$ to $\frac{5\pi}{6}$, r increases from 0 to 2; as θ increases from $\frac{5\pi}{6}$ to π , r decreases from 2 to 0. Since r is positive, the curve is in the second quadrant.

Thus, as θ varies from 0 to π , the curve traces a three-leaved rose.

(4) As θ increases from π to 2π , since $r = -2 \sin(3\theta - \pi)$, we know that the curve traces the same three-leaved rose once more.



3.74 Example: polar curve sketching

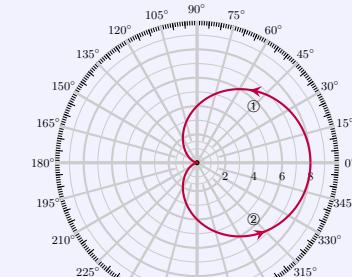
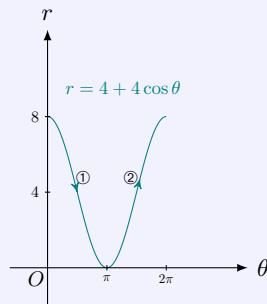
Sketch the curve with the given polar equation

$$r = a + b \cos \theta,$$

where (A) $a = b = 4$; (B) $a = 3$ and $b = 5$.

Solution For any given a and b , the function $r = a + b \cos \theta$ is periodic of period 2π . Since a full revolution is 2π , we only need to sketch the polar curve for $0 \leq \theta \leq 2\pi$.

(A) For $0 \leq \theta \leq 2\pi$, we sketch $r = 4 + 4 \cos \theta$ in Cartesian coordinates in the following left figure. As θ increases from 0 to π , it shows that r decreases from 4 to 0. Since r is positive, the curve is in the upper-half plane. As θ increases from π to 2π , r increases from 0 to 4. Again since r is positive, the curve is in the lower-half plane. Thus, as θ varies from 0 to 2π , the curve completes one loop, as shown in the following right figure. This is a *cardioid*.



(B) For $0 \leq \theta \leq 2\pi$, we sketch $r = 3 + 5 \cos \theta$ in Cartesian coordinates in the following left figure.

(1) As θ increases from 0 to $\cos^{-1}(-\frac{3}{5}) \approx 2.2143$ (rad) $\approx 126.87^\circ$, r decreases from 8 to 0. Since r is positive, the curve is in the upper-half plane.

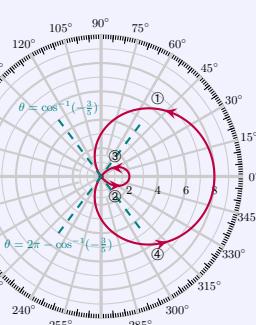
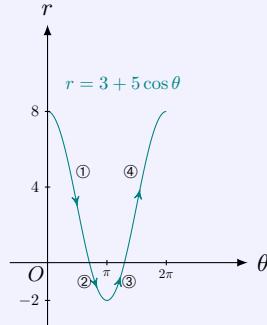
(2) As θ increases from $\cos^{-1}(-\frac{3}{5})$ to π , r decreases further from 0 to -3. Since r is negative, the curve is in the fourth quadrant.

(3) As θ increases from π to $2\pi - \cos^{-1}(-\frac{3}{5}) \approx 4.0689$ (rad) $\approx 233.13^\circ$, r increases from -2 to 0. Since r is negative, the curve is in the first quadrant.

In fact, when θ varies from $\cos^{-1}(-\frac{3}{5})$ to $2\pi - \cos^{-1}(-\frac{3}{5})$, the curve completes the small loop.

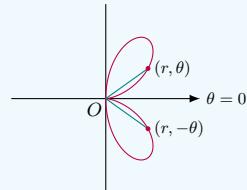
(4) As θ increases from $2\pi - \cos^{-1}(-\frac{3}{5})$ to 2π , r increases from 0 to 8, and the curve is in the lower-half plane.

The curve also completes a large loop, as shown in the following right figure. We get a *limaçon*.

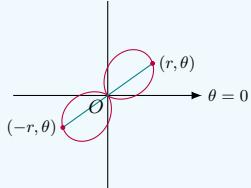


3.75 Method: determine symmetry of a polar curve

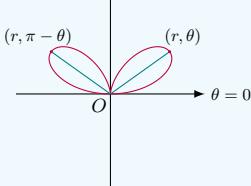
- (a) If a polar equation is unchanged when θ is replaced by $-\theta$, the curve is symmetric about the polar axis.



- (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 $\theta = \frac{\pi}{2}$.



3.76 Theorem: tangent to polar curve

Suppose a polar curve is given by the equation $r = f(\theta)$. Then we can write its parametric equations as

$$x = r \cos \theta = f(\theta) \cos \theta, \quad y = r \sin \theta = f(\theta) \sin \theta.$$

If f is differentiable, and if $\frac{dx}{d\theta} = f'(\theta) \cos \theta - f(\theta) \sin \theta \neq 0$, then the chain rule gives the slope of the tangent line at $(x, y) = (f(\theta) \cos \theta, f(\theta) \sin \theta)$:

$$\frac{dy}{dx} = \frac{dy/d\theta}{dx/d\theta} = \frac{f'(\theta) \sin \theta + f(\theta) \cos \theta}{f'(\theta) \cos \theta - f(\theta) \sin \theta}.$$

Theorem: horizontal tangent

The parametric curve has a horizontal tangent at $(x, y) = (f(\theta) \cos \theta, f(\theta) \sin \theta)$ if

$$\frac{dy}{d\theta} = f'(\theta) \sin \theta + f(\theta) \cos \theta = 0, \quad \frac{dx}{d\theta} = f'(\theta) \cos \theta - f(\theta) \sin \theta \neq 0.$$

Theorem: vertical tangent

The parametric curve has a vertical tangent at $(x, y) = (f(\theta) \cos \theta, f(\theta) \sin \theta)$ if

$$\frac{dx}{d\theta} = f'(\theta) \cos \theta - f(\theta) \sin \theta = 0, \quad \frac{dy}{d\theta} = f'(\theta) \sin \theta + f(\theta) \cos \theta \neq 0.$$

3.77 Example: highest point of limaçon

Find the highest point of limaçon $r = 3 + 5 \cos \theta$.

Solution The function $r = 3 + 5 \cos \theta$ is periodic of period 2π . Since a full revolution is 2π , we only need to determine the highest point for $0 \leq \theta \leq 2\pi$.

The parametric equations for the limaçon are

$$x = (3 + 5 \cos \theta) \cos \theta, \quad y = (3 + 5 \cos \theta) \sin \theta, \quad 0 \leq \theta \leq 2\pi.$$

At the highest point, we must have $\frac{dy}{d\theta} = \frac{dy}{dx} \cdot \frac{dx}{d\theta} = 0 \cdot \frac{dx}{d\theta} = 0$. It is easy to have

$$\begin{aligned}\frac{dy}{d\theta} &= (-5 \sin \theta) \sin \theta + (3 + 5 \cos \theta) \cos \theta \\ &= -5(1 - \cos^2 \theta) + (3 + 5 \cos \theta) \cos \theta \\ &= 10 \cos^2 \theta + 3 \cos \theta - 5.\end{aligned}$$

Solving $\frac{dy}{d\theta} = 0$ gives

$$\cos \theta = \frac{1}{20}(-3 + \sqrt{209}) \approx 0.5728 \quad \text{or} \quad \frac{1}{20}(-3 - \sqrt{209}) \approx -0.8728.$$

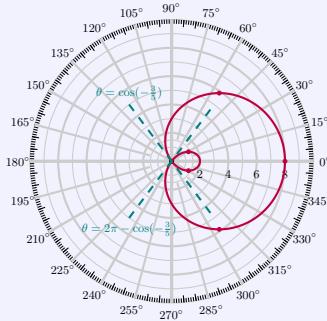
Thus, for $0 < \theta < 2\pi$, there are four critical points:

$$\begin{aligned}\theta_1 &= \cos^{-1}(\frac{1}{20}(-3 + \sqrt{209})) \approx 0.9608, \\ \theta_2 &= \cos^{-1}(\frac{1}{20}(-3 - \sqrt{209})) \approx 2.6318, \\ \theta_3 &= 2\pi - \cos^{-1}(\frac{1}{20}(-3 + \sqrt{209})) \approx 3.6514, \\ \theta_4 &= 2\pi - \cos^{-1}(\frac{1}{20}(-3 - \sqrt{209})) \approx 5.3224,\end{aligned}$$

at which the corresponding Cartesian coordinates (x, y) are

$$\begin{aligned}(x_1, y_1) &= (3.35926, 4.8067), \\ (x_2, y_2) &= (1.19074, -0.6657), \\ (x_3, y_3) &= (1.19074, 0.6657), \\ (x_4, y_4) &= (3.35926, -4.8067).\end{aligned}$$

The Cartesian coordinates of two endpoints when $\theta = 0$ and $\theta = 2\pi$ coincide at $(8, 0)$. By comparing the values of the y -coordinates at these five points, we see that the highest point of the curve attains at (x_1, y_1) , approximately at $(3.35926, 4.8067)$.



3.78 MATLAB: polar curves

In MATLAB, the command `polarplot` facilitates the plot of polar curves. Two examples are provided below to show how to plot polar curves in MATLAB.

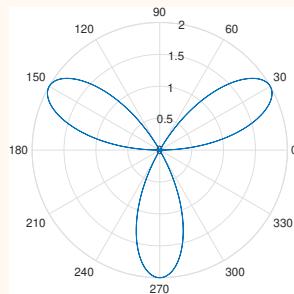
1. The polar curve is described by the following equation:

$$r = 2 \sin 3\theta.$$

This polar curve can be plotted using the following code:

```
>> theta = linspace(0,2*pi,200);
>> r = 2*sin(3*theta);
>> figure
>> polarplot(theta,r)
```

The polar curve plotted is shown below.



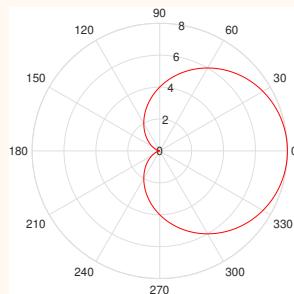
2. The polar curve is described by the following equation:

$$r = 4 + 4 \cos \theta,$$

This polar curve can be plotted using the following code:

```
>> theta = linspace(0,2*pi,200);
>> r = 4 + 4*cos(theta);
>> figure
>> polarplot(theta,r,'r')
```

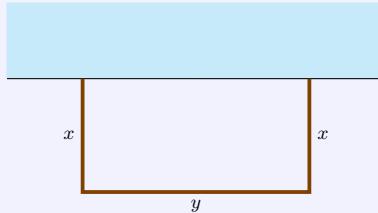
The polar curve plotted is shown below.



3.7 Optimization

3.79 Example: optimization - largest area

John has 120 meters of fencing to build a rectangular fence along a straight river. He needs to build three sides, and the side along the river does not need fencing. What are the lengths of sides of the field to have the largest area?



Solution This is a typical example of optimization problems. Its solution involves a few steps, shown below.

Step 1: Introduce variables.

Let x be the length of the side of the rectangle perpendicular to the river and y be the length of the other side.

Step 2: Write down the function to be optimized and the constraints on variables. Simplify the function under the constraints.

(1) We need to maximize the area of the rectangle, which is $A = xy$.

(2) There are some constraints for variables x and y . Clearly, we must have $2x + y = 120$, or $y = 120 - 2x$. We also have $0 \leq x \leq 60$ and $0 \leq y \leq 120$.

(3) Substitute $y = 120 - 2x$ into the function we need to optimize, we get

$$A(x) = x(120 - 2x) = 120x - 2x^2,$$

where $0 \leq x \leq 60$.

Step 3: Solve the optimization problem.

This is a problem of finding maximum for the function $A(x) = 120x - 2x^2$ defined in a finite closed interval $[0, 120]$.

In fact, the function A is continuous on $[0, 120]$ and differentiable in $(0, 120)$. The derivative of A is

$$A'(x) = 120 - 4x.$$

Solving $A'(x) = 0$ gives only one critical point $x = 30$ in $[0, 120]$. We can use one of the following two approaches to determine the absolute maximum value of A in the interval $[0, 120]$.

(a) We evaluate the function A at two boundary points $x = 0$ and $x = 120$ and the critical point at $x = 30$:

$$A(0) = 0 = A(120), \quad A(30) = 30 \times 60 = 1800.$$

By choosing the largest value, we get the absolute maximum value $A(30) = 1800$.

(b) It is easy to see that

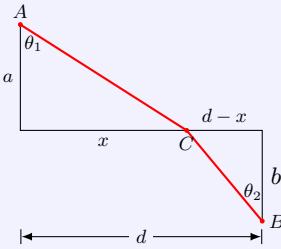
$$\begin{cases} A'(x) > 0, & \text{if } 0 < x < 30; \\ A'(x) < 0, & \text{if } 30 < x < 120. \end{cases}$$

Thus, by the First Derivative Test, the function A has a local maximum value at $x = 30$. Since A is increasing on $(0, 30)$, and decreasing on $(30, 120)$, the local maximum value is also the absolute maximum.

Hence, the function A attains the absolute maximum value $A(30) = 1800$.

3.80 Example: optimization - Snell's law of refraction

Fermat's principle says that light travels along the path of shortest traveling time. In the figure below, we consider a light ray traveling from point A located in one medium to point B in another medium. The media are separated by the x -axis. Assume that the speed of light is respectively c_1 and c_2 in the media. Denote by θ_1 the angle of incidence and θ_2 the angle of refraction.



The total time for light to travel along the path ACB is

$$\begin{aligned} T(x) &= (\text{time from } A \text{ to } C) + (\text{time from } C \text{ to } B) \\ &= \frac{\sqrt{a^2 + x^2}}{c_1} + \frac{\sqrt{b^2 + (d-x)^2}}{c_2}. \end{aligned}$$

We shall determine the value of x by applying Fermat's Principle to minimize the function T on $[0, d]$.

In fact, the derivative of T is

$$T'(x) = \frac{x}{c_1 \sqrt{a^2 + x^2}} - \frac{d-x}{c_2 \sqrt{b^2 + (d-x)^2}}.$$

We claim that the equation $T'(x) = 0$ has a unique solution on $[0, d]$. Indeed, the equation has a solution by the Intermediate Value Theorem, and

$$T'(0) = -\frac{d}{c_2 \sqrt{b^2 + d^2}} < 0, \quad T'(d) = \frac{d}{c_1 \sqrt{a^2 + d^2}} > 0.$$

The solution is unique since T' is increasing on $(0, d)$:

$$T''(x) = \frac{a^2}{c_1} (a^2 + x^2)^{-3/2} + \frac{b^2}{c_2} (b^2 + (d-x)^2)^{-3/2} > 0.$$

Let $x = x_0$ be the unique solution of the equation $T'(x) = 0$ on $(0, d)$. The above argument shows that

$$\begin{cases} T'(x) < 0, & \text{if } 0 < x < x_0; \\ T'(x) > 0, & \text{if } x_0 < x < d. \end{cases}$$

Thus, by the First Derivative Test, the function T has a local minimum at $x = x_0$. Since T is increasing on (x_0, d) , and decreasing on $(0, x_0)$, the local minimum is also the absolute minimum. Hence, along the path ACB when $x = x_0$, light travels from A to B in the shortest time.

When $x = x_0$, the equation $T'(x_0) = 0$ gives

$$\frac{x_0}{c_1 \sqrt{a^2 + x_0^2}} = \frac{d - x_0}{c_2 \sqrt{b^2 + (d - x_0)^2}}.$$

In terms of the angles of incidence and refraction, we have

$$\frac{\sin \theta_1}{c_1} = \frac{\sin \theta_2}{c_2}$$

which gives Snell's law

$$\frac{\sin \theta_1}{\sin \theta_2} = \frac{c_1}{c_2},$$

3.81 Example: optimization - shortest distance

Find points on the parabola $y = \frac{1}{2}x^2$ that is closest to the point $(4, 1)$.

Solution The distance from a point (x, y) on the parabola to the point $(4, 1)$ is

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

We need to minimize the function d . It is equivalent to minimize the function $f(x) = d^2(x)$, which is computationally easier.

In fact, the derivative f' is

$$f'(x) = [(x - 4)^2 + (\frac{1}{2}x^2 - 1)^2]' = 2(x - 4) + 2(\frac{1}{2}x^2 - 1)x = x^3 - 8.$$

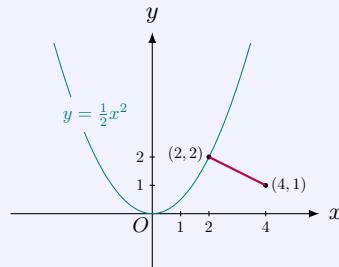
By solving $f'(x) = 0$, we see that f has only one critical point $x = 2$. It is easy to see that

$$\begin{cases} f'(x) < 0, & \text{if } x < 2; \\ f'(x) > 0, & \text{if } x > 2. \end{cases}$$

Thus, by the First Derivative Test, the function f has a local minimum value at $x = 2$. Since f is decreasing on $(-\infty, 2)$, and increasing on $(2, +\infty)$, the local minimum value is also the absolute minimum.

Hence, the shortest distance is attained when $x = 2$, with

$$d = \sqrt{(2 - 4)^2 + (2^2/2 - 1)^2} = \sqrt{5}.$$



3.82 MATLAB: optimization

John has 120 meters of fencing to build a rectangular fence along a straight river. He needs to build three sides, and the side along the river does not need fencing. What are the lengths of sides of the field to have the largest area?

The following MATLAB code solves the above optimization problem.

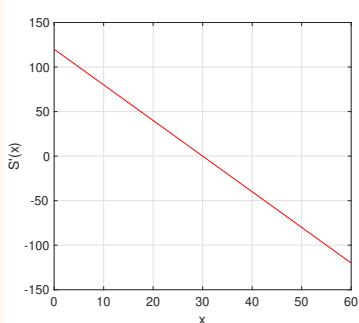
```
>> syms x y S(x);
>> y = 120 - 2*x; % the side parallel to the river
>> S(x) = x*y; % area
>> d(x) = diff(S(x),x)

d(x) =
120 - 4*x

>> rts_d = solve(d(x) == 0,x)

rts_d =
30

>> x_array = 0:0.5:60; % x < 60
>> figure
>> plot(x_array,d(x_array),'r')
>> xlabel('x')
>> ylabel('S''(x)')
>> grid
```



We see from the above MATLAB code that $S'(x) = 0$ has one root: $x = 30$. Moreover, we note from the above figure of $S'(x)$ that $S'(x) > 0$ on $(0, 30)$ and $S'(x) < 0$ on $(30, 60)$. According to the First Derivative Test, $S(x)$ has a local maximum value at $x = 30$. Moreover, since $S'(x) > 0$ on $(0, 30)$, i.e., $S(x)$ is increasing on $(0, 30)$, and $S'(x) < 0$ on $(30, 60)$, i.e., $S(x)$ is decreasing on $(30, 60)$, we know that $S(x)$ also has an absolute maximum value at

$x = 30$. Therefore, to have the largest area, the length perpendicular to the river should be 30 meters, and the side parallel to the river should be $120 - 2 \times 30 = 60$ meters.

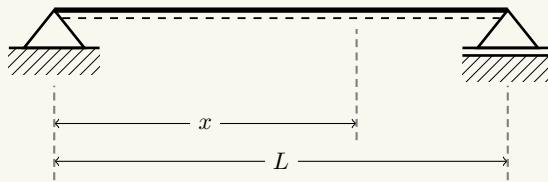
3.83 Engineering Example: bending moment

When an external force is applied to a structural element, e.g., beam, and it causes this element to bend, we say the bending moment is induced. The bending moment is an important concept in physics and engineering, especially civil and mechanical engineering.

Suppose there is a beam of length L with uniform distributed loading of q per unit length. At the distance x of the beam, the bending moment M is given by

$$M = -\frac{q}{2}x^2 + \frac{qL}{2}x.$$

This is illustrated in the figure below.



Find the value of x which gives the maximum bending moment and calculate this maximum bending moment.

First, let $f(x) = M = -\frac{q}{2}x^2 + \frac{qL}{2}x$ ($0 \leq x \leq L$). Then, we have

$$f'(x) = -qx + \frac{qL}{2}.$$

It is easy to see that

$$\begin{cases} f'(x) > 0, & \text{if } 0 \leq x < \frac{L}{2}; \\ f'(x) < 0, & \text{if } \frac{L}{2} < x \leq L. \end{cases}$$

Thus, by the First Derivative Test, we know that $f(x)$ has a local maximum value at $x = \frac{L}{2}$. Moreover, since $f'(x) > 0$ on $[0, \frac{L}{2})$, i.e., $f(x)$ is increasing on $[0, \frac{L}{2})$, and $f'(x) < 0$ on $(\frac{L}{2}, L]$, i.e., $S(x)$ is decreasing on $(\frac{L}{2}, L]$, we know that $f(x)$ also has an absolute maximum value at $x = \frac{L}{2}$. Hence, at $x = \frac{L}{2}$, the bending moment is the maximum.

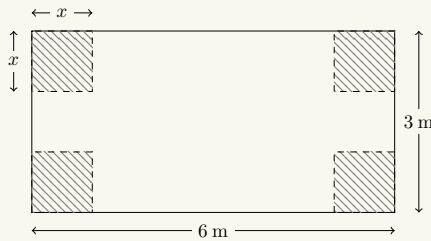
To calculate the corresponding maximum bending moment, we substitute $x = \frac{L}{2}$ into $f(x)$ and have

$$f\left(\frac{L}{2}\right) = -\frac{q}{2}\left(\frac{L}{2}\right)^2 + \frac{qL}{2}\frac{L}{2} = \frac{qL^2}{8}.$$

The result above shows that the maximum bending moment is $\frac{qL^2}{8}$.

3.84 Engineering Example: tank volume optimization

One piece of metal is provided to build a water tank. For this purpose, four corners of the metal should be cut, as illustrated in the figure below.



Find the value of x such that the volume of the water tank is maximized and calculate the maximum volume of the water tank.

First, the volume of the water tank can be calculated as follows:

$$\begin{aligned} V(x) &= (6 - 2x)(3 - 2x)x \\ &= 4x^3 - 18x^2 + 18x. \end{aligned}$$

The derivative of $V(x)$ is

$$V'(x) = 12x^2 - 36x + 18.$$

As we know, the roots of an equation $ax^2 + bx + c = 0$ are given by

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

We thus can calculate the roots of $V'(x) = 12x^2 - 36x + 18 = 0$ as

$$\begin{aligned} x &= \frac{36 \pm \sqrt{36^2 - 4 \times 12 \times 18}}{24} \\ &\approx 0.634 \text{ or } 2.366. \end{aligned}$$

Since $2x < 3$, the value of x should satisfy $0 < x < 1.5$, which shows that x cannot be 2.366. We will then check whether $x = 0.634$ is a maximum point.

The second derivative of $V(x)$ can be easily calculated and shown below:

$$V''(x) = 24x - 36.$$

Substituting $x = 0.634$ into $V''(x)$, we have

$$V''(0.634) = 15.216 - 36 < 0.$$

According to the Second Derivative Test, we know that $V(x)$ has a local maximum value at $x = 0.634$. To check whether $x = 0.634$ is also an absolute maximum point, we rewrite $V'(x)$ as

$$V'(x) = 2(x - 0.634)(x - 2.366).$$

Obviously, we know that

$$\begin{cases} V'(x) > 0, & \text{if } 0 < x < 0.634; \\ V'(x) < 0, & \text{if } 0.634 < x < 1.5, \end{cases}$$

which means $V(x)$ is increasing on $(0, 0.634)$ and decreasing on $(0.634, 1.5)$. Therefore, $x = 0.634$ is also an absolute maximum point, i.e., the volume of the water tank is maximized at $x = 0.634$ (m). Finally, substituting $x = 0.634$ into $V(x)$, we have the maximum volume of the water tank as

$$V(0.634) = 4 \times 0.634^3 - 18 \times 0.634^2 + 18 \times 0.634 \approx 5.196 \text{ (m}^3\text{)}.$$

3.85 Engineering Example: entropy optimization

In information theory, the entropy refers to the uncertainty or the disorder of an event. If an event has more uncertainty, then it carries more information and has more entropy.

For a communication system, if it has two inputs 0 and 1 and they occur with probability p and $1 - p$ ($0 < p < 1$), respectively, then the entropy of the source is given by

$$H(p) = -p \log_2 p - (1 - p) \log_2(1 - p).$$

Find the value of p such that the entropy $H(p)$ is maximized and calculate the maximum entropy.

By applying the product rule and the chain rule, we have

$$\begin{aligned} \frac{dH(p)}{dp} &= -\log_2 p - p(\log_2 p)' - (-1)\log_2(1-p) - (1-p)(\log_2(1-p))' \\ &= -\log_2 p - p \frac{1}{p \ln 2} + \log_2(1-p) - (1-p) \frac{-1}{(1-p) \ln 2} \\ &= \log_2(1-p) - \log_2 p. \end{aligned}$$

According to Fermat's Theorem, we first calculate the value of p such that $\frac{dH(p)}{dp} = 0$. It can be calculated that $\left.\frac{dH(p)}{dp}\right|_{p=0.5} = 0$. Moreover, since $\log_2 x$ is an increasing function, we notice that when $p < 0.5$, $\frac{dH(p)}{dp} > 0$; while when $p > 0.5$, $\frac{dH(p)}{dp} < 0$. Therefore, according to the first derivative test, the entropy $H(p)$ is maximized at $p = 0.5$.

Substituting $p = 0.5$ into the expression of $H(p)$, we have

$$\begin{aligned} H(p) &= -0.5 \times \log_2 0.5 - (1 - 0.5) \times \log_2(1 - 0.5) \\ &= -\log_2 0.5 = 1 \text{ (bit)}. \end{aligned}$$

The maximum entropy is thus 1 bit.

3.8 Newton's Method

3.86 Method: Newton's iteration

To find approximations to a solution r of the equation $f(x) = 0$, we start with an initial *guessing* value x_0 , which is not equal to r in general. To improve the approximation, consider the tangent line to the curve $y = f(x)$ at the point $(x_0, f(x_0))$. An equation of the line is

$$y - f(x_0) = f'(x_0)(x - x_0),$$

whose x -intercept gives the first approximation x_1 , obtained by putting $y = 0$ in the last equation:

$$\begin{aligned} 0 - f(x_0) &= f'(x_0)(x_1 - x_0) \implies x_1 - x_0 = -\frac{f(x_0)}{f'(x_0)} \\ &\implies x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}. \end{aligned}$$

The formula for x_1 is meaningful if $f'(x_0) \neq 0$.

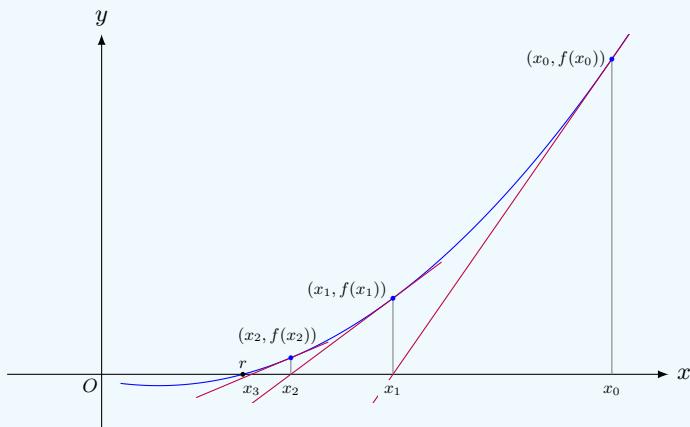
We repeat this procedure with x_0 replaced by the first approximation x_1 , using the tangent line at $(x_1, f(x_1))$. This gives a second approximation:

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

If we keep repeating this process, we generate a sequence of approximations $x_1, x_2, x_3, x_4, \dots$. In general, if x_n is the n th 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 iteration** and the sequence $\{x_n\}$ generated is called **Newton's approximations**.



Theorem: convergence of Newton's approximations

If x_0 is chosen to be *sufficiently close* to a solution r of $f(x) = 0$, and if $f'(x_n) \neq 0$ for $n = 1, 2, 3, \dots$, then the sequence $\{x_n\}$ generated by Newton's iteration converges to r :

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

3.87 Example: Newton's method

Given $f(x) = x^3 - 5x + 1$. Use six steps of Newton's method to approximate all three real roots of f .

1. Choose an initial guessing value for each root.

The function f is a polynomial of degree 3, so it has at most three real roots. Since

$$\begin{cases} f(-3) < 0, & f(-2) > 0; \\ f(0) > 0, & f(1) < 0; \\ f(2) < 0, & f(3) > 0, \end{cases}$$

by the Intermediate Value Theorem, f must have one root in each of the three intervals $(-3, -2)$, $(0, 1)$, and $(2, 3)$.

Hence, it is plausible to choose the initial guessing values -2.5 , 0.5 , 2.5 for Newton's iterations for approximating the three roots of f .

2. Write the recursive formula of Newton's method.

For $f(x) = x^3 - 5x + 1$, we have

$$f'(x) = (x^3 - 5x + 1)' = 3x^2 - 5.$$

Thus, the recursive formula of Newton's method is

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{x_n^3 - 5x_n + 1}{3x_n^2 - 5} = \frac{2x_n^3 - 1}{3x_n^2 - 5}.$$

3. Perform the iteration.

For the initial guessing value $x_0 = -2.5$,

$$x_1 = \frac{2x_0^3 - 1}{3x_0^2 - 5} = \frac{2(-2.5)^3 - 1}{3(-2.5)^2 - 5} = -2.345455,$$

$$x_2 = \frac{2x_1^3 - 1}{3x_1^2 - 5} = \frac{2(-2.345455)^3 - 1}{3(-2.345455)^2 - 5} = -2.330203,$$

\vdots

$$x_5 = \frac{2x_4^3 - 1}{3x_4^2 - 5} = -2.330059.$$

We list Newton's approximations of all three roots in the following table.

k	x_k	x_k	x_k
0	-2.500000	0.500000	2.500000
1	-2.345455	0.176471	2.200000
2	-2.330203	0.201568	2.131933
3	-2.330059	0.201640	2.128428
4	-2.330059	0.201640	2.128419
5	-2.330059	0.201640	2.128419

Remark: when to stop for Newton's iteration

There is a need for a good estimation of the error of the approximation. However, since the actual value of the root is not known, it is not easy to get the estimation of the error. The rule of thumb is that if we want n decimal accuracy, we should compute until two successive approximations agree up to n decimal places.

Remark: how many steps needed

For a given accuracy, the number of steps of Newton's iteration can vary. In the example, if the initial guessing value is $x_0 = 1.3$, then Newton's approximations are

k	x_k
0	1.300000
1	48.485718
2	32.346600
\vdots	\vdots
12	2.128441
13	2.128419

It takes 13 steps of Newton's iterations to achieve 6 decimal accuracy.

3.88 Pitfall: convergence of Newton's method

In Newton's method, $f'(x_n)$ appears in the denominator of the formula. When $f'(x_n)$ is zero, the method breaks down. When $f'(x_n)$ is close to zero, the method is slow to converge to the actual root or may fail to converge.

Consider $f(x) = \frac{8x^2}{3x^2 + 1}$. Then the recursive formula of Newton's method is

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

Clearly the function f has 0 as the only real root. For the following three initial guessing

values $x_0 = 1$, 0.15 , and 1.1 , we list the corresponding Newton's approximations:

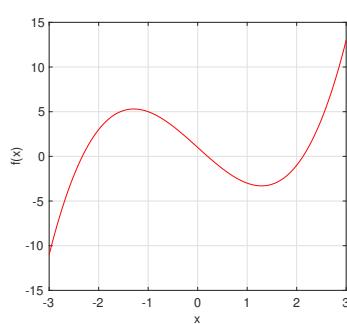
k	x_k	x_k	x_k
0	1.000000	0.150000	1.200000
1	-1.000000	0.069938	-1.992000
2	1.000000	0.034456	10.860575
3	-1.000000	0.017166	-1.92×10^3
4	1.000000	0.008576	1.06×10^{10}
5	-1.000000	0.004287	-1.76×10^{30}
6	1.000000	0.002143	8.21×10^{90}

From the computations, we see that the approximations with the initial guessing values $x_0 = 1$ and $x_0 = 1.2$ fail to converge.

3.89 MATLAB: Newton's method

The following example shows how to calculate the roots of $f(x) = x^3 - 5x + 1$ in MATLAB. We note that f is a polynomial of degree 3, so it has at most three real roots. To roughly estimate these roots, we first plot the figure of f as follows.

```
>> syms x;
>> f(x) = x^3 - 5*x + 1;
>> x_array = -3:0.05:3;
>> figure
>> plot(x_array,f(x_array), 'r')
>> xlabel('x')
>> ylabel('f(x)')
>> grid
```



From the figure above, we see that there are 3 roots of f , and these roots are on intervals $(-3, -2)$, $(0, 1)$ and $(2, 3)$, respectively. Hence, we choose the initial guessing values -2.5 , 0.5 and 2.5 for Newton's iterations for approximating the three roots of $f(x)$.

Next, we use Newton's method to approximate these three roots of f . The MATLAB code is provided below.

```

syms x;
f(x) = x^3 - 5*x + 1; % function
d(x) = diff(f(x),x);

n = 5;
N_max = 10; % maximum no. of iterations
x = zeros(1,N_max);
x0 = - 2.5; % initial guess
x(1) = x0;

for num = 1:N_max
    x(num + 1) = x(num) - f(x(num))/d(x(num));
    if abs(x(num + 1) - x(num))<10^(-(n + 1)) % to ensure n ...
        decimal accuracy
        break; % accuracy achieved, so exit!
    end
end

x_iter = x(1:(num + 1)) % show iteration results

```

By inputting different initial guessing values, we can obtain approximations of all three roots. Moreover, we can have 15-digit approximations by using the command `format long`. The Newton's approximations of all three roots with initial guessing values -2.5 , 0.5 and 2.5 , respectively, are listed in the following table.

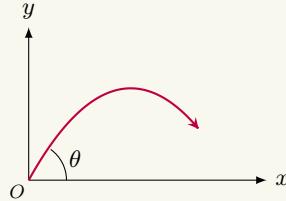
k	x_k	x_k	x_k
0	-2.500000000000000	0.500000000000000	2.500000000000000
1	-2.3454545454545	0.176470588235294	2.200000000000000
2	-2.330203407719357	0.201568074338339	2.131932773109244
3	-2.330058752527104	0.201639675087802	2.128428202971012
4	-2.330058739567982	0.201639675723405	2.128419063906659
5			2.128419063844577

Note that to ensure the 5 decimal accuracy, it requires the absolute difference of two successive approximations to be less than 10^{-6} in the MATLAB code. In this case, the initial guessing values of -2.5 and 0.5 result in 4 iterations, and the initial guessing value of 2.5 results in 5 iterations.

3.90 Engineering Example: projectile - Newton's method

Consider the case where a ball is thrown with initial velocity u and an angle θ ($0^\circ \leq \theta \leq 90^\circ$) against the ground, as shown in the following figure. Then, at time t , the displacement of the ball in the y -direction is $s_y = u \sin \theta t - \frac{1}{2}gt^2$, where $g \approx 9.8 \text{ m/s}^2$ is the standard gravity. Suppose that the initial velocity is 40 m/s and the angle against the ground is 45° . Use Newton's method to calculate the time t at which the ball is 10 meters

above the ground during the falling down stage.



When $u = 40$ m/s and $\theta = 45^\circ$, the displacement of the ball in the y -direction is

$$f(t) = 40 \sin(45^\circ)t - \frac{9.8}{2}t^2 = 20\sqrt{2}t - 4.9t^2.$$

Note that $\sqrt{2} \approx 1.414$, we thus have

$$f(t) \approx 28.28t - 4.9t^2.$$

First, we need to find an initial guessing value x_0 . We notice that

$$\begin{cases} f(5) \approx 18.9 > 10; \\ f(6) \approx -6.7 < 10, \end{cases}$$

so, by the Intermediate Value Theorem, one of the roots of $f(t) = 10$ is in the interval $(5, 6)$. Moreover, since $f(5) > f(6)$, the root of $f(t) = 10$ in the interval $(5, 6)$ is the time during the falling down stage of the ball. Hence, we set the initial guessing value as $x_0 = 5.5$.

Before performing the iterations, we first define a function $g(t) = f(t) - 10 = 20\sqrt{2}t - 4.9t^2 - 10$. To write the recursive formula of Newton's method, we need to calculate $g'(t)$, which is shown in the following

$$g'(t) = 20\sqrt{2} - 9.8t.$$

Then, the recursive formula of Newton's method can be written as

$$t_{n+1} = t_n - \frac{g(t_n)}{g'(t_n)} = t_n - \frac{20\sqrt{2}t_n - 4.9t_n^2 - 10}{20\sqrt{2} - 9.8t_n}.$$

From the initial guessing value $t_0 = 5.5$, we have

$$\begin{aligned} t_1 &= t_0 - \frac{g(t_0)}{g'(t_0)} = 5.5 - \frac{g(5.5)}{g'(5.5)} \\ &= 5.5 - \frac{-2.661508}{-25.615729} = 5.396099, \\ t_2 &= t_1 - \frac{g(t_1)}{g'(t_1)} = 5.396099 - \frac{g(5.396099)}{g'(5.396099)} \\ &= 5.396099 - \frac{-0.052906}{-24.597499} = 5.393948, \\ t_3 &= t_2 - \frac{g(t_2)}{g'(t_2)} = 5.393948 - \frac{g(5.393948)}{g'(5.393948)} \\ &= 5.393948 - \frac{-0.000019}{-24.576419} = 5.393947. \end{aligned}$$

We see that three steps are required to achieve 5 decimal accuracy. Moreover, since $t_3 = 5.393947$, we can conclude that at $t = 5.393947$, the ball is in the falling down stage and is 10 meters above the ground.

MATLAB solution:

We now turn to the use of MATLAB. Let $g(t) = f(t) - 10 = 20\sqrt{2}t - 4.9t^2 - 10$. We will run the following MATLAB code of the Newton's method to calculate the result.

```

syms t;
g(t) = 20*sqrt(2)*t - 4.9*t^2 - 10;
d(t) = diff(g(t),t);

n = 5; % n decimal accuracy
N_max = 10; % maximum no. of iterations
x = zeros(1,N_max);
x0 = 5.5; % initial guess
x(1) = x0;

for num = 1:N_max
    x(num + 1) = x(num) - g(x(num))/d(x(num));
    if abs(x(num + 1) - x(num))<10^(-(n + 1)) % to ensure n ...
        decimal accuracy
        break; % accuracy achieved, so exit!
    end
end

x_iter = x(1:(num + 1)) % show iteration results

```

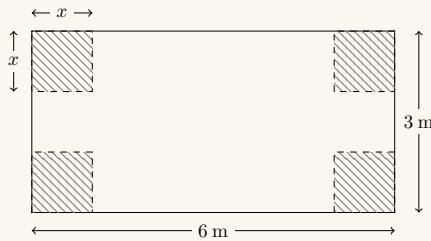
Starting from the initial guessing value $t_0 = 5.5$, we can obtain approximations of the root near 5.5. Note that by using the command `format long`, we can have 15-digit approximations. The Newton's approximations are listed in the following table.

k	t_k
0	5.500000000000000
1	5.396098675752262
2	5.393948136613312
3	5.393947214525773

We see that 3 iterations are needed to ensure 5 decimal accuracy. At $t = 5.393947214525773$, the ball is in the falling down stage and is 10 meters above the ground.

3.91 Engineering Example: water tank design

One piece of metal is provided to build a water tank. For this purpose, four corners of the metal should be cut, as illustrated in the figure below.



By Newton's method, find the value of x such that the volume of the water tank is 3.5 m^3 and the water tank covers the least area on the ground.

First, the volume of the water tank can be calculated as follows:

$$\begin{aligned} V(x) &= (6 - 2x)(3 - 2x)x \\ &= 4x^3 - 18x^2 + 18x. \end{aligned}$$

To find the value of x such that the volume of the water tank $V(x)$ is 3.5 m^3 , we actually need to find x which satisfies $4x^3 - 18x^2 + 18x = 3.5$. Let $g(x) = 4x^3 - 18x^2 + 18x - 3.5$. The derivative of $g(x)$ is

$$g'(x) = 12x^2 - 36x + 18.$$

Then, the recursive formula of Newton's method can be written as

$$\begin{aligned} x_{n+1} &= x_n - \frac{g(x_n)}{g'(x_n)} \\ &= x_n - \frac{4x_n^3 - 18x_n^2 + 18x_n - 3.5}{12x_n^2 - 36x_n + 18}. \end{aligned}$$

The function $g(x)$ is a polynomial of degree 3, so it has at most three real roots. Since

$$\begin{cases} g(0) = -3.5 < 0, & g(1) = 0.5 > 0; \\ g(1) = 0.5 > 0, & g(1.5) = -3.5 < 0; \\ g(3) = -3.5 < 0, & g(4) = 36.5 > 0, \end{cases}$$

by the Intermediate Value Theorem, $g(x)$ must have one root in each of the three intervals $(0, 1)$, $(1, 1.5)$, and $(3, 4)$. Note that $2x < 3$, so we have $0 < x < 1.5$. Moreover, the area that the water tank covers on the ground can be written as $(6 - 2x)(3 - 2x)$ which is a decreasing function of x . Hence, if we are looking for the value of x such that the water tank covers the least area on the ground, the value of x should be as large as possible, i.e., it is within the interval $(1, 1.5)$. Therefore, we choose the initial value x_0 as 1.25.

Now, we use the recursive formula of Newton's method to calculate the root of $g(x) = 0$

within $(1, 1.5)$. From the initial guessing value $x_0 = 1.25$, we have

$$\begin{aligned}x_1 &= x_0 - \frac{g(x_0)}{g'(x_0)} = 1.25 - \frac{g(1.25)}{g'(1.25)} \\&= 1.25 - \frac{-1.3125}{-8.25} = 1.090909, \\x_2 &= x_1 - \frac{g(x_1)}{g'(x_1)} = 1.090909 - \frac{g(1.090909)}{g'(1.090909)} \\&= 1.090909 - \frac{-0.092035}{-6.991735} = 1.077746, \\x_3 &= x_2 - \frac{g(x_2)}{g'(x_2)} = 1.077746 - \frac{g(1.077746)}{g'(1.077746)} \\&= 1.077746 - \frac{-0.000863}{-6.860419} = 1.077620.\end{aligned}$$

We see that three steps are required to achieve 3 decimal accuracy. Moreover, since $x_3 = 1.077620$, we can conclude that the volume of the water tank is 3.5 m^3 and the water tank covers the least area on the ground if $x = 1.078 \text{ m}$.

MATLAB solution:

We now turn to the use of MATLAB. Let $g(x) = 4x^3 - 18x^2 + 18x - 3.5$. We will run the following MATLAB code of the Newton's method to calculate the result.

```
syms x;
g(x) = 4*x^3 - 18*x^2 + 18*x - 3.5;
d(x) = diff(g(x),x);

n = 3; % n decimal accuracy
N_max = 10; % maximum no. of iterations
x_array = zeros(1,N_max);
x0 = 1.25; % initial guess
x_array(1) = x0;

for num = 1:N_max
    x_array(num + 1) = x_array(num) - ...
        g(x_array(num))/d(x_array(num));
    if abs(x_array(num + 1) - x_array(num)) < 10^(-(n + 1)) % ...
        to ensure n decimal accuracy
        break; % accuracy achieved, so exit!
    end
end

x_iter = x_array(1:(num + 1)) % show iteration results
```

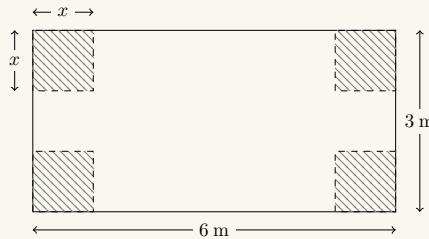
Starting from the initial guessing value $x_0 = 1.25$, we can obtain approximations of the root near 1.25. The Newton's approximations are listed in the following table.

k	t_k
0	1.250000000000000
1	1.090909090909091
2	1.077745540511498
3	1.077620217678657
4	1.077620206075159

We see that in order to ensure 3 decimal accuracy, 4 iterations are needed if running the MATLAB code above. At $x = 1.077620206075159$, i.e., $x = 1.078$ m, the volume of the water tank is 3.5 m^3 and the water tank covers the least area on the ground.

3.92 Engineering Example: tank optimization - Newton's method

One piece of metal is provided to build a water tank. For this purpose, four corners of the metal should be cut, as illustrated in the figure below.



By Newton's method, find the value of x such that the volume of the water tank is maximized.

First, the volume of the water tank can be calculated as follows:

$$\begin{aligned} V(x) &= (6 - 2x)(3 - 2x)x \\ &= 4x^3 - 18x^2 + 18x. \end{aligned}$$

The derivative of $V(x)$ is

$$V'(x) = 12x^2 - 36x + 18.$$

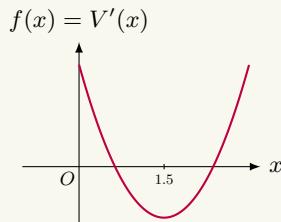
To find the value of x such that the volume of the water tank $V(x)$ is maximized, we need to find x which satisfies $V'(x) = 12x^2 - 36x + 18 = 0$, i.e., find the root of $V'(x) = 12x^2 - 36x + 18 = 0$. As such, we let $f(x) = V'(x) = 12x^2 - 36x + 18$, and the derivative of $f(x)$ is

$$f'(x) = 24x - 36.$$

Then, the recursive formula of Newton's method can be written as

$$\begin{aligned} x_{n+1} &= x_n - \frac{f(x_n)}{f'(x_n)} \\ &= x_n - \frac{12x_n^2 - 36x_n + 18}{24x_n - 36}. \end{aligned}$$

Before using Newton's method to find the value of x , we need to find an initial guessing value x_0 . Hence, we plot the figure of $f(x) = V'(x)$, as shown below.



Note that $2x < 3$, so we have $0 < x < 1.5$. To find the value of x which maximizes $V(x)$, we need to find the root of $V'(x) = 0$ within the interval $(0, 1.5)$. From the figure of $V'(x)$ above, we notice that there is one root within the interval $(0, 1.5)$ and it is roughly in the middle of this interval. Thus, we choose the initial guessing value x_0 around this root and set it as 0.75.

Now, we use the recursive formula of Newton's method to calculate the root of $f(x) = V'(x) = 0$ within $(0, 1.5)$. From the initial guessing value $x_0 = 0.75$, we have

$$\begin{aligned} x_1 &= x_0 - \frac{f(x_0)}{f'(x_0)} = 0.75 - \frac{f(0.75)}{f'(0.75)} \\ &= 0.75 - \frac{-2.25}{-18} = 0.625, \\ x_2 &= x_1 - \frac{f(x_1)}{f'(x_1)} = 0.625 - \frac{f(0.625)}{f'(0.625)} \\ &= 0.625 - \frac{0.1875}{-21.0} = 0.633928, \\ x_3 &= x_2 - \frac{f(x_2)}{f'(x_2)} = 0.633928 - \frac{f(0.633928)}{f'(0.633928)} \\ &= 0.633928 - \frac{0.000968}{-20.785728} = 0.633974. \end{aligned}$$

We see that three steps are required to achieve 3 decimal accuracy. Moreover, from the figure of $V'(x)$ above, we see that if x is smaller than the root of $V'(x)$ within $(0, 1.5)$, then $V'(x) > 0$, i.e., $V'(x)$ is an increasing function; if x is larger than this root, then $V'(x) < 0$, i.e., $V'(x)$ is a decreasing function. So this root is an absolute maximum point. Therefore, we can conclude that the volume of the water tank is maximized when $x = 0.634$ (m).

MATLAB solution:

We now turn to the use of MATLAB. Let $f(x) = V'(x) = 12x^2 - 36x + 18$. We will run the following MATLAB code of the Newton's method to calculate the result.

```
syms x;
f(x) = 12*x^2 - 36*x + 18;
d(x) = diff(f(x),x);
```

```

n = 3; % n decimal accuracy
N_max = 10; % maximum no. of iterations
x_array = zeros(1,N_max);
x0 = 0.75; % initial guess
x_array(1) = x0;

for num = 1:N_max
    x_array(num + 1) = x_array(num) - ...
        f(x_array(num))/d(x_array(num));
    if abs(x_array(num + 1) - x_array(num)) < 10^(-(n + 1)) % ...
        to ensure n decimal accuracy
        break; % accuracy achieved, so exit!
    end
end

x_iter = x_array(1:(num + 1)) % show iteration results

```

Starting from the initial guessing value $x_0 = 0.75$, we can obtain approximations of the root near 0.75. The Newton's approximations are listed in the following table.

k	t_k
0	0.7500000000000000
1	0.6250000000000000
2	0.633928571428571
3	0.633974594992636

We see that 3 iterations are needed to ensure 3 decimal accuracy. Moreover, as explained in the first part, the root of $V'(x)$ within $(0, 1.5)$ is an absolute maximum point. Therefore, we can conclude that when $x = 0.634$ (m), the volume of the water tank is maximized.

Remark:

The Newton's method is utilized to *find the root* of $f(x) = 0$ by iteratively approximating the root from an initial guessing value. Hence, any function whose root is required to be found can be set to be $f(x)$, e.g., $V'(x)$ instead of $V(x)$ is set to be $f(x)$ in this example.

Summary

Definition: basic concepts

Extrema :

absolute maximum : if $f(c) \geq f(x)$ for all x in D

absolute minimum : if $f(c) \leq f(x)$ for all x in D

local maximum : if $f(c) \geq f(x)$ when x is near c

local minimum : if $f(c) \leq f(x)$ when x is near c

SpecialPoints :

critical point c : either $f'(c) = 0$ or $f'(c)$ does not exist

interior point c of D : if there exists (a, b) such that $c \in (a, b) \subset D$

inflection point P : at P , f is continuous and changes concavity

Monotonicity :

increasing : if $f(x_1) < f(x_2)$ whenever $x_1 < x_2$ in (a, b)

decreasing : if $f(x_1) > f(x_2)$ whenever $x_1 < x_2$ in (a, b)

Concavity :

concave up : if the graph of f is above all its tangent lines over (a, b)

concave down : if the graph of f is below all its tangent lines over (a, b)

Theorem: Fermat's Theorem

If f has a local maximum or minimum value at an *interior point* c in the domain D of f , then c is a critical point of f .

Theorem: Extreme Value Theorem

Let f be a *continuous* function defined over a *finite closed* interval $[a, b]$. Then $f(x)$ has at least one absolute maximum point and one absolute minimum point in the interval $[a, b]$.

Method: find absolute extrema

Given a continuous function f over a finite closed interval $[a, b]$.

1. Find the critical points and evaluate the values of f at critical points.
2. Evaluate the values of f at the boundary points $x = a$ and $x = b$.
3. The largest value and smallest value from steps 1 and 2 are the absolute maximum

value and the absolute minimum value respectively.

Theorem: Mean Value Theorem

Suppose f is a function that satisfies the following:

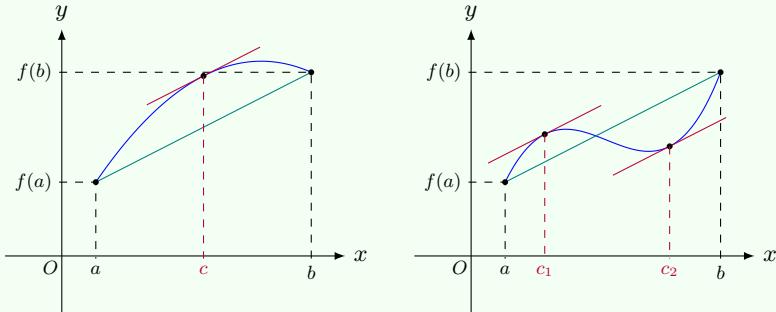
1. f is continuous on the *closed* interval $[a, b]$.
2. f is differentiable in the *open* interval (a, b) .

Then there exists at least one number c in (a, b) such that

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

or, equivalently,

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



Theorem: zero derivative implies constant function

Suppose f is *differentiable* in (a, b) . If $f'(x) = 0$ for all x in (a, b) , then f is a constant function.

Theorem: Monotone Test

Suppose f is *differentiable* on an interval I .

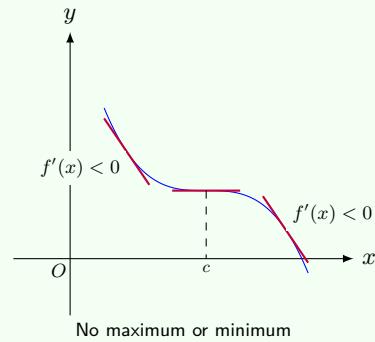
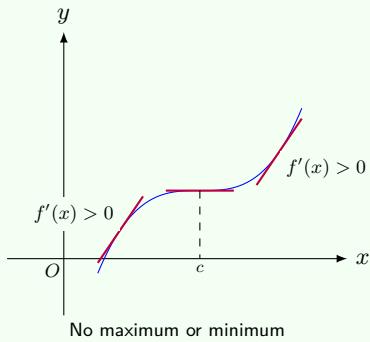
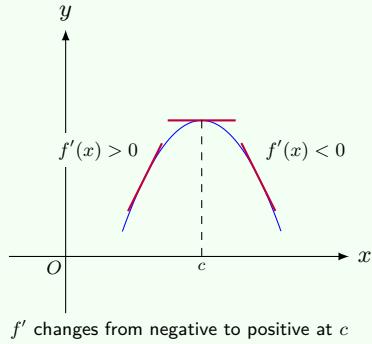
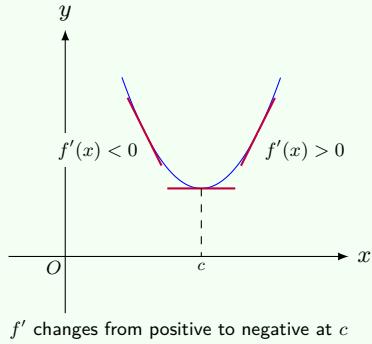
1. If $f'(x) > 0$ on I , then f is increasing on I .
2. If $f'(x) < 0$ on I , then f is decreasing on I .

Theorem: First Derivative Test

Suppose that c is a *critical point* of a continuous function f .

1. If f' changes *from positive to negative* at c , then f has a *local maximum* value at c .

2. If f' changes from negative to positive at c , then f has a local minimum value at c .
3. If f' does not change sign at c , then f has no local extreme value at c .



Theorem: Second Derivative Test

Suppose f'' is continuous near c .

1. If $f'(c) = 0$ and $f''(c) > 0$, then f has a local minimum value at c .
2. If $f'(c) = 0$ and $f''(c) < 0$, then f has a local maximum value at c .
3. If $f'(c) = 0$ and $f''(c) = 0$, it is inconclusive about local extreme of f at c .

Theorem: Concavity Test

Suppose f'' exists on an interval I .

1. If $f''(x) > 0$ on I , then f is concave up on I .
2. If $f''(x) < 0$ on I , then f is concave down on I .

Theorem: l'Hospital's Rule

Suppose

1. f and g are differentiable functions;
2. $g'(x)$ is not zero on an open interval I containing a except possibly at a ;
3. $\lim_{x \rightarrow a} f(x) = 0 = \lim_{x \rightarrow a} g(x)$ or $\lim_{x \rightarrow a} f(x) = \infty = \lim_{x \rightarrow a} g(x)$;
4. $\lim_{x \rightarrow a} \frac{f'(x)}{g'(x)}$ exists or is ∞ .

Then

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

The rule also holds when limits are one-sided limits or when $x \rightarrow a$ is replaced by $x \rightarrow \infty$.

Method: guidelines for curve sketching

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 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 even function and the curve is symmetric about the y -axis.
- (ii) If $f(-x) = -f(x)$ for all $x \in D$, then f is an odd function 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 periodic function and the smallest such number p is the period.

D. Asymptotes:

- (i) The line $y = L$ is a horizontal asymptote of the curve $y = f(x)$ if either

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

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

$$\begin{aligned} \lim_{x \rightarrow a} f(x) &= \infty, & \lim_{x \rightarrow a^-} f(x) &= \infty, & \lim_{x \rightarrow a^+} f(x) &= \infty, \\ \lim_{x \rightarrow a} f(x) &= -\infty, & \lim_{x \rightarrow a^-} f(x) &= -\infty, & \lim_{x \rightarrow a^+} f(x) &= -\infty. \end{aligned}$$

E. Intervals of Increase or Decrease:

Apply the Increasing/Decreasing Test:

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

F. Local Maximum and Minimum Values:

First find the critical points, 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) 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) 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) Inflection Points: 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**.

Theorem: convergence of Newton's approximations

If x_0 is chosen to be *sufficiently close* to a solution r of $f(x) = 0$, and if $f'(x_n) \neq 0$ for $n = 1, 2, 3, \dots$, then the sequence $\{x_n\}$ generated by Newton's iteration

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad n = 0, 1, 2, \dots$$

converges to r :

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

Integration

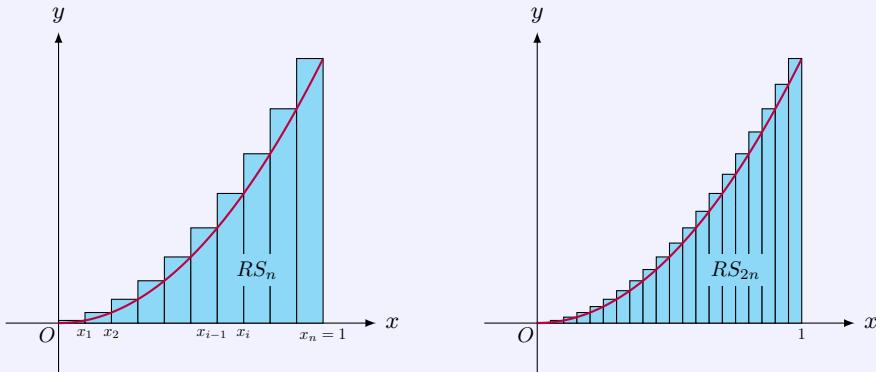
4.1 Areas under Curves

4.1 Example: area under a curve

To find the area A of the region bounded by the curve $y = f(x) = x^2$ and the x -axis over $[0, 1]$, we divide the interval $[0, 1]$ into n subintervals of equal length $\frac{1}{n}$. Let

$$x_0 = 0, \quad x_1 = \frac{1}{n}, \quad x_2 = \frac{2}{n}, \quad \dots, \quad x_{n-1} = \frac{n-1}{n}, \quad x_n = \frac{n}{n} = 1$$

be the endpoints of these subintervals. For each i ($1 \leq i \leq n$), over the i -th subinterval $[x_{i-1}, x_i]$, we use the area of the rectangle with height $f(x_i)$ to approximate the area of the region between the curve and the x -axis over the interval $[x_{i-1}, x_i]$.



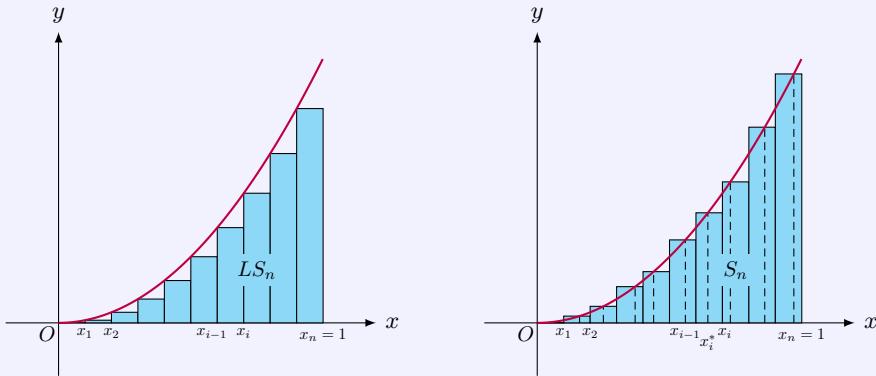
The sum of n rectangles is

$$\begin{aligned} RS_n &= f\left(\frac{1}{n}\right) \cdot \frac{1}{n} + f\left(\frac{2}{n}\right) \cdot \frac{1}{n} + \cdots + f\left(\frac{n}{n}\right) \cdot \frac{1}{n} \\ &= \left(\frac{1}{n^2} + \frac{2^2}{n^2} + \cdots + \frac{n^2}{n^2}\right) \cdot \frac{1}{n} \\ &= (1 + 2^2 + \cdots + n^2) \cdot \frac{1}{n^3} = \frac{n(n+1)(2n+1)}{6n^3}. \end{aligned}$$

Clearly RS_n approximates the area A , and when n get larger, the approximation gets better.

Since $\lim_{n \rightarrow \infty} RS_n = \lim_{n \rightarrow \infty} \frac{n(n+1)(2n+1)}{6n^3} = \frac{1}{3}$, we get $A = \frac{1}{3}$.

We can similarly choose the left endpoint of each subinterval instead of the right endpoint to make a new approximation to A . Denote it as LS_n . By the same manner, we also get $A = \lim_{n \rightarrow \infty} LS_n = \frac{1}{3}$.



For any fixed n , if we choose x_i^* in the interval $[x_{i-1}, x_i]$ *arbitrarily* and use the rectangle of height $f(x_i^*)$ over the base $[x_{i-1}, x_i]$ to make approximation, we will get the sum

$$S_n = [f(x_1^*) + \cdots + f(x_n^*)] \cdot \frac{1}{n}.$$

Clearly $LS_n \leq S_n \leq RS_n$. As $n \rightarrow \infty$, since both LS_n and RS_n converge to the same limit $\frac{1}{3}$, by using a similar argument as the Squeeze Theorem, we get

$$\lim_{n \rightarrow \infty} S_n = \frac{1}{3}.$$

4.2 Remark: sigma notation

We often encounter sums with multiple terms, for example,

$$1 + 2^2 + \cdots + n^2.$$

For convenience, we use a sigma notation to write

$$1 + 2^2 + \cdots + n^2 = \sum_{i=1}^n i^2.$$

Here i^2 is the general term of the summation. The symbol Σ is the Greek capital letter Sigma corresponding to our letter ‘S’, and refers to the initial letter of the word ‘Sum’. So this expression means the sum of all the terms i^2 , where i takes the values from 1 to n .

In general, for fixed numbers k and n with $k \leq n$, let a_i be a number depending on i . Then the sum $a_k + a_{k+1} + \cdots + a_n$ can be represented in terms of the following sigma sum notation:

$$\sum_{i=k}^n a_i = a_k + a_{k+1} + \cdots + a_n.$$

4.3 Definition: definite integral

Let f be a function defined over $[a, b]$. Divide the interval into n subintervals of equal length $\frac{b-a}{n}$:

$$[x_0, x_1], [x_1, x_2], \dots, [x_{n-1}, x_n].$$

Let $\Delta x = x_i - x_{i-1} = \frac{b-a}{n}$. Choose x_i^* in $[x_{i-1}, x_i]$ for each subinterval. The Riemann sum is

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

The function f is said to be **integrable** if the limit of the Riemann sum exists and has a unique limit, i.e.,

$$\lim_{\Delta x \rightarrow 0} \sum_{i=1}^n f(x_i^*) \Delta x = L$$

for all choices of x_i^* . The limit is called the **definite integral** of f (or Riemann integral) from a to b , denoted by

$$\int_a^b f(x) dx,$$

where the function f is called the **integrand**, a is the lower limit, and b is the upper limit.

Theorem: necessary condition

If f is integrable on $[a, b]$, then f must be bounded on $[a, b]$.

4.4 Theorem: integrable functions

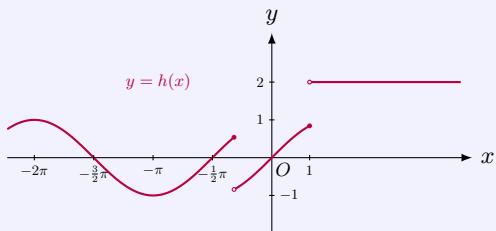
If f is continuous over $[a, b]$ or bounded on $[a, b]$ with a finite number of discontinuous points, then f is integrable on $[a, b]$.

4.5 Example: integrable functions

The function

$$h(x) = \begin{cases} \cos x, & \text{if } x \leq -1; \\ \sin x, & \text{if } -1 < x \leq 1; \\ 2, & \text{if } x > 1 \end{cases}$$

is integrable on $[-2, 2]$ since it is discontinuous only at the points $x = -1$ and $x = 1$. The function is bounded and continuous away from these two discontinuities.



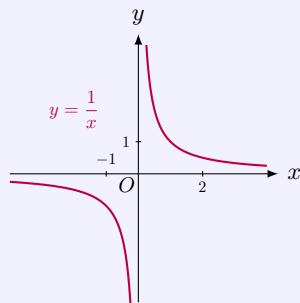
4.6 Pitfall: nonexistence of definite integrals

The integrand $\frac{1}{x}$ of the definite integral,

$$\int_{-1}^2 \frac{1}{x} dx,$$

is continuous everywhere except at $x = 0$. However, the integral does not exist since $f(x)$ is not bounded.

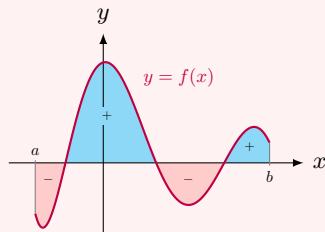
The function $\frac{1}{x}$ is not integrable on $[-1, 2]$.



4.7 Definition: area and net area

- Let \mathcal{R} be a region. The **net area** of \mathcal{R} is the area of the part of \mathcal{R} above the x -axis minus the area of the part of \mathcal{R} below the x -axis. If \mathcal{R} is bounded by the graph $y = f(x)$ and the x -axis over the interval $[a, b]$, then the net area of the region \mathcal{R} is

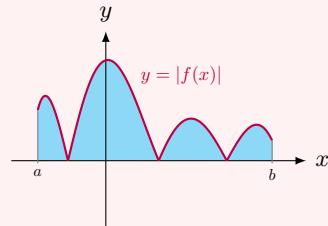
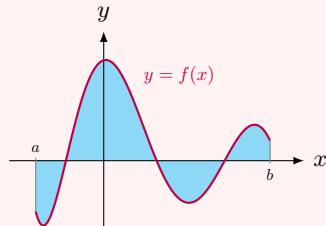
$$\int_a^b f(x) dx.$$



- Let \mathcal{R} be a region. The **area** of \mathcal{R} is the area of the part of \mathcal{R} above the x -axis plus the area of the part of \mathcal{R} below the x -axis. If \mathcal{R} is bounded by the graph $y = f(x)$ and the

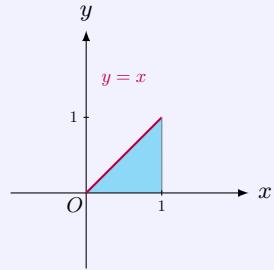
x -axis over the interval $[a, b]$, then the area of the region \mathcal{R} is

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

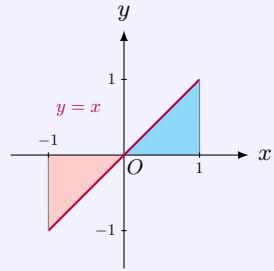


4.8 Example: area and net area

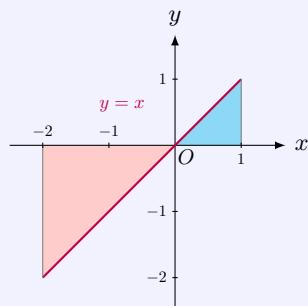
1. The integral $\int_0^1 x \, dx$ calculates the area of the region between the curve $y = x$ and the x -axis over the interval $[0, 1]$. This gives a triangular region above the x -axis, with the area $\frac{1}{2}$. Thus, the area equals the net area in this case.



2. The integral $\int_{-1}^1 x \, dx$ calculates the net area of the region between the curve $y = x$ and the x -axis over the interval $[-1, 1]$. The net area is the area of the region in the first quadrant, which is $\frac{1}{2}$, minus the area of the region in the third quadrant, which is also $\frac{1}{2}$. Hence the net area is 0. The area of the region is $\frac{1}{2} + \frac{1}{2} = 1$.



3. The integral $\int_{-2}^1 x \, dx$ calculates the net area of the region between the curve $y = x$ and the x -axis over the interval $[-2, 1]$. The net area is the area of the region in the first quadrant, which is $\frac{1}{2}$, minus the area of the region in the third quadrant, which is 2. Hence the net area is $-\frac{3}{2}$. However, the area of the region is $\frac{1}{2} + 2 = \frac{5}{2}$.



Properties of Definite Integrals

4.9 Theorem: properties of definite integrals

Let f and g be integrable functions over an interval I and let $a < b < c$ be points in I .

- If f and g are integrable over $[a, b]$, then $f + g$ is integrable over $[a, b]$, and we have

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

- If f is integrable over $[a, b]$, then $kf(x)$ is integrable over $[a, b]$, and we have

$$\int_a^b kf(x) \, dx = k \int_a^b f(x) \, dx.$$

- If f is integrable over $[a, b]$ and $[b, c]$, then f is integrable over $[a, c]$, and we have

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

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

- If f and g are both integrable over $[a, b]$, and $f(x) \leq g(x)$, then

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

6.

$$\left| \int_a^b f(x) \, dx \right| \leq \int_a^b |f(x)| \, dx.$$

- (Mean Value Theorem for Integrals) If f is continuous on $[a, b]$, then there is $c \in [a, b]$ such that

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

Justification: Part 1

The proof of these properties follows directly from the definition of the definite integral.

Let's prove the last two statements.

Consider the Riemann sum of the function f for any partition:

$$a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b,$$

which divides the interval $[a, b]$ into n subintervals. For each index i , choose x_i^* inside the

subinterval $[x_{i-1}, x_i]$ and make the Riemann sum

$$R_n(f) = \sum_{i=1}^n f(x_i^*) \Delta x_i,$$

where $\Delta x_i = x_i - x_{i-1} > 0$ is the length of the subinterval $[x_{i-1}, x_i]$. Using the same partition and same choice of x_i^* , we can also have the Riemann sum of the function $|f(x)|$,

$$R_n(|f|) = \sum_{i=1}^n |f(x_i^*)| \Delta x_i.$$

Recall the triangle inequality

$$|\alpha + \beta| \leq |\alpha| + |\beta|.$$

Apply the inequality to the Riemann sum, we get

$$\left| \sum_{i=1}^n f(x_i^*) \Delta x_i \right| \leq \sum_{i=1}^n |f(x_i^*)| \Delta x_i, \quad \text{i.e., } |R_n(f)| \leq R_n(|f|).$$

Take limits of both sides as $n \rightarrow \infty$, we get

$$\begin{aligned} \lim_{n \rightarrow \infty} |R_n(f)| &\leq \lim_{n \rightarrow \infty} R_n(|f|), && \text{order rule of limits} \\ \left| \lim_{n \rightarrow \infty} R_n(f) \right| &\leq \lim_{n \rightarrow \infty} R_n(|f|), && \text{continuity of the absolute value function} \\ \left| \int_a^b f(x) dx \right| &\leq \int_a^b |f(x)| dx. \end{aligned}$$

Justification: Part 2

If f is continuous on $[a, b]$, by the Extreme Value Theorem, the function f can attain its minimum m and maximum M :

$$f(c_1) = m, \quad f(c_2) = M,$$

where $c_1, c_2 \in [a, b]$. Since $m \leq f(x) \leq M$, we have

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

Thus,

$$f(c_1) \leq \frac{1}{b-a} \int_a^b f(x) dx \leq f(c_2).$$

By the Intermediate Value Theorem, there is a number c between c_1 and c_2 , such that

$$f(c) = \frac{1}{b-a} \int_a^b f(x) dx.$$

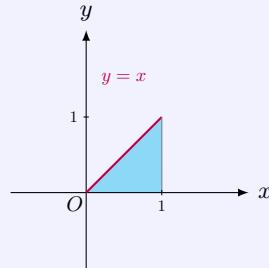
Obviously, $c \in [a, b]$ and the equality holds:

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

4.10 Example: simple definite integrals

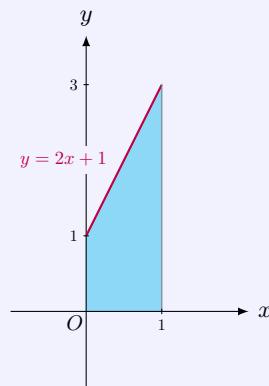
1. The integral $\int_0^1 x dx$ represents the area of the triangular region bounded by the graph $y = x$ and the x -axis over $[0, 1]$, which equals $\frac{1}{2}$. So,

$$\int_0^1 x dx = \frac{1}{2}.$$



2. From the properties of definite integrals, we have the net area of the region bounded by the graph $y = 2x + 1$ and the x -axis over $[0, 1]$:

$$\begin{aligned} \int_0^1 (2x + 1) dx &= \int_0^1 (2x) dx + \int_0^1 1 dx \\ &= 2 \int_0^1 x dx + \int_0^1 1 dx \\ &= 2 \cdot \frac{1}{2} + 1 = 2. \end{aligned}$$

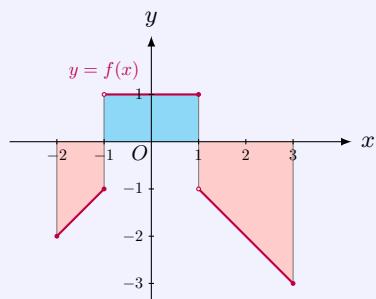


3. For the piecewise function

$$f(x) = \begin{cases} x, & \text{if } -2 \leq x \leq -1; \\ 1, & \text{if } -1 < x \leq 1; \\ -x, & \text{if } 1 < x \leq 3, \end{cases}$$

by using the different expressions over different intervals, we have

$$\begin{aligned} &\int_{-2}^3 f(x) dx \\ &= \int_{-2}^{-1} f(x) dx + \int_{-1}^1 f(x) dx + \int_1^3 f(x) dx \\ &= \int_{-2}^{-1} x dx + \int_{-1}^1 1 dx + \int_1^3 (-x) dx \\ &= -\frac{3}{2} + 2 - 4 = -\frac{7}{2}. \end{aligned}$$



We note that the contributions from the integrals with negative integrands are negative.

Average Value of a Function

4.11 Example: average temperature

Since temperatures fluctuate throughout the day, as well as varying substantially depending on the locations, average temperatures gives you a more accurate picture of the temperature in a specific location than a single measurement ever could. For example, knowing the average temperature of a given week in Hong Kong will certainly help a traveler to prepare his trip to the city.

The average temperature can be calculated from several temperature measurements using the formula:

$$\text{average temperature} = \frac{\text{sum of measured temperatures}}{\text{number of measurements}}.$$

For instance, for calculating the average temperature on January 1, 2019, one may first measure temperature every hour in Central, Hong Kong to get a dataset of 24 temperatures. Then using the formula above, one may obtain an approximation of the average temperature. Apparently, increasing the number of measurements should get a better result in reflecting Hong Kong's temperature situation on the new year day.

4.12 Definition: average value 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^*, \dots, x_n^* in successive subintervals and calculate the average of the numbers $f(x_1^*), \dots, f(x_n^*)$:

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

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

$$\begin{aligned} \frac{f(x_1^*) + \dots + f(x_n^*)}{(b - a)/\Delta x} &= \frac{1}{b - a} [f(x_1^*)\Delta x + \dots + f(x_n^*)\Delta x] \\ &= \frac{1}{b - a} \sum_{i=1}^n f(x_i^*)\Delta x, \end{aligned}$$

so that

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

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

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

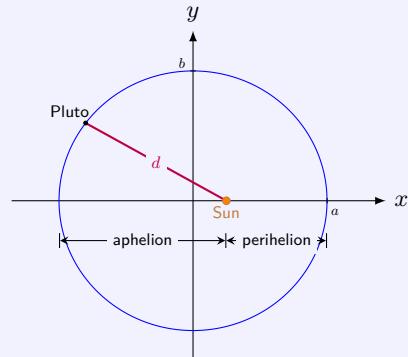
4.13 Example: average distance from the Pluto to the Sun

The perihelion is the shortest distance from a planet to the Sun, and aphelion is the longest distance. The sun is at one of the two foci. For the Pluto, the perihelion is 4.44 billion km and the aphelion is 7.38 billion km.

As shown in the figure, assume that the orbit of the Pluto is

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

where $0 < b < a$. It is known that the foci are located at $(\pm\sqrt{a^2 - b^2}, 0)$. Let us assume the coordinates of the Sun is $(\sqrt{a^2 - b^2}, 0)$.



Then the distance from the Pluto, located at (x, y) , to the Sun is

$$\begin{aligned} d &= \sqrt{(x - \sqrt{a^2 - b^2})^2 + y^2} \\ &= \sqrt{(x - \sqrt{a^2 - b^2})^2 + b^2 \left(1 - \frac{x^2}{a^2}\right)} \\ &= a^{-1} \sqrt{\left(a^2 - \sqrt{a^2 - b^2}x\right)^2} \\ &= a^{-1} \left(a^2 - \sqrt{a^2 - b^2}x\right). \quad \text{a}^2 > \sqrt{a^2 - b^2}x \text{ for } -a \leq x \leq a \end{aligned}$$

Thus, the average distance from the Pluto to the Sun is

$$\begin{aligned} D &= \frac{1}{2a} \cdot \int_{-a}^a a^{-1} \left(a^2 - \sqrt{a^2 - b^2}x\right) dx \\ &= \frac{1}{2a} \cdot \int_{-a}^a a dx - \frac{1}{2a} a^{-1} \sqrt{a^2 - b^2} \int_{-a}^a x dx \\ &= \frac{1}{2a} \cdot a \cdot 2a - 0 = a. \end{aligned}$$

In other words, the average distance equals to the average of the perihelion and aphelion:

$$D = \frac{\text{perihelion} + \text{aphelion}}{2}.$$

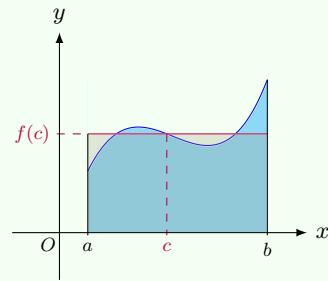
4.14 Theorem: geometric interpretation on average value

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

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

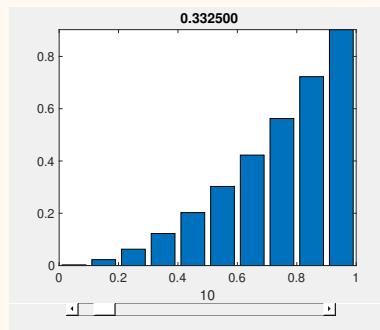


4.15 MATLAB: Riemann sums

In MATLAB, the command `rsums` is used to interactively approximate the integral of a function by Riemann sum. The following example shows the approximation of $\int_0^1 x^2 dx$ by using the command.

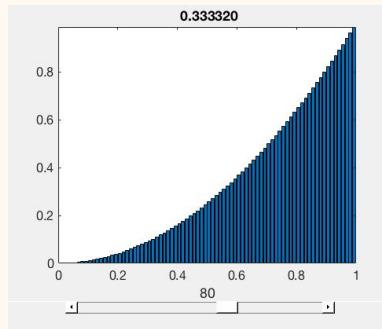
```
>> syms x;
>> f = x^2;
>> rsums(f, [0, 1])
```

After running the MATLAB code above, the figure shown in the following is generated.



The figure above displays a graph of $f(x) = x^2$ using 10 rectangles. In other words, the interval $[0, 1]$ is divided into 10 subintervals. Note that the height of each rectangle is determined by the value of $f(x) = x^2$ in the middle of each subinterval. In this case, $\int_0^1 x^2 dx \approx 0.332500$, which we can see from the top of the figure. Furthermore, we can

adjust the number of the rectangles (from 2 to 128) by using the slider below the figure. For example, if we choose the number of the rectangles to be 80, the graph of $f(x) = x^2$ is shown below.



From the figure above, we see that $\int_0^1 x^2 dx \approx 0.333320$ which is closer to the real value of $\int_0^1 x^2 dx$ (i.e., $\frac{1}{3}$).

4.16 MATLAB: definite integrals

To calculate definite integrals, we can use the commands `int` and `integral`. Specifically, `int` performs symbolic integration and `integral` performs numerical integration. The following example shows the calculation of $\int_{-1}^1 (1-x^2) dx$ using `int` and `integral`, respectively.

```
>> syms x;
>> f = 1 - x^2;
>> int(f, -1, 1)

ans =
4/3

>> integral(@(x)1 - x.^2, -1, 1)

ans =
1.3333
```

We see from the MATLAB code above that the command `int` returns a symbolic number $\frac{4}{3}$, while the command `integral` returns a floating-point number 1.3333. Note that the first input of the command `int` is a symbolic expression, but the first input of the command `integral` should be a function handle.

Besides `int` and `integral`, the command `quadgk` can also be used for numerical integration.

4.17 MATLAB: definite integrals - polynomials

If the integrand is a polynomial, besides `int` and `integral`, we can also use the command `polyint` to calculate the integral. The following example shows the calculation of $\int_{-1}^1 (1 - x^2) dx$ using `polyint`.

First, we calculate the coefficients of the new polynomial obtained after integration, and we have the following:

```
>> p = [-1 0 1]; % -x^2 + 1
>> q = polyint(p) % coefficients of the polynomial after ...
    integration

q =
    -0.3333      0      1.0000      0
```

We see from the MATLAB code above that after integration, the new polynomial is $-0.3333x^3 + x$. Next, we calculate the values of $-0.3333x^3 + x$ at the lower limit -1 and the upper limit 1 , respectively, and obtain the result of the integral using the following code.

```
>> a = -1; % lower limit
>> b = 1; % upper limit
>> q_val = polyval(q,[a b]) % evaluate the values of q at a ...
    and b, respectively

q_val =
    -0.6667      0.6667

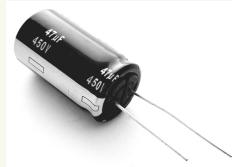
>> polyint_ans = diff(q_val) % q_val(2) - q_val(1)

polyint_ans =
    1.3333
```

The result shows that $\int_{-1}^1 (1 - x^2) dx = 1.3333$. Note that for `diff(X)`, if the input `X` is a vector, the command `diff(X)` returns the result of $[X(2) - X(1), X(3) - X(2), \dots, X(n) - X(n - 1)]$.

4.18 Engineering Example: energy stored in a capacitor

The capacitor is an electrical component that can store electrical energy in an electric field. Capacitors have been widely used as important parts of electrical circuits in many common electrical devices. The following figures show a real capacitor and a symbol for the capacitor in circuits, respectively.



The energy stored in a capacitor is equal to the work done in establishing the electric field from an uncharged state by moving charge from the negative to the positive plate of the capacitor. Mathematically, the energy W stored in a capacitor can be written as a definite integral, as shown in the following:

$$W = \int_0^Q V(q) dq,$$

where Q is the total charge stored in the capacitor, q is a variable representing the charge, and V is the voltage across the capacitor (which is a function of q).

4.19 Engineering Example: PID controller

The Proportional-Integral-Derivative (PID) controller is a closed-loop feedback component widely used in industrial control systems and different types of other applications, e.g., the cruise control on a car. The PID controller first calculates an error value as the difference between a desired setpoint and a measured output. Then, based on the error value, a correction term based on the proportional (P), integral (I) and derivative (D) terms of the PID controller is applied at the input. This process is executed continuously to achieve the desired setpoint in an optimal way as well as to keep the system in a steady state.

Mathematically, the overall control function of the PID controller can be written as

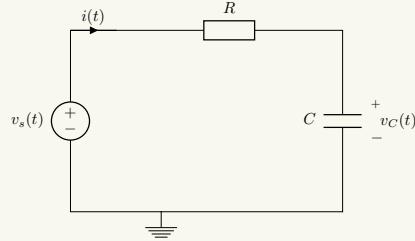
$$u(t) = \underbrace{K_p e(t)}_{\text{Proportional (P) term}} + \underbrace{K_i \int_0^t e(t') dt'}_{\text{Integral (I) term}} + \underbrace{K_d \frac{de(t)}{dt}}_{\text{Derivative (D) term}}$$

where $K_p \geq 0$, $K_i \geq 0$ and $K_d \geq 0$ are the coefficients for the proportional (P), integral (I), and derivative (D) terms, respectively, and $e(t)$ is the error value.

4.20 Engineering Example: electrical energy

Any object that carries charges will experience electrical force created by electrical potential difference (voltage) in electrical field. The force will drive charges to move orderly in one direction. Such orderly movement of charges will form current. All electrical devices work

based on the transformation of electrical energy to other types of energy, for example, heat, light, mechanical energy, etc. The power generated/consumed by a component (battery, resistor, etc.) is calculated by the product between the voltage across the component and the current going through the component. Consider a circuit as shown in the following:



The power generated by the power source at time t can be obtained by $v_s(t)i(t)$, where $v_s(t)$ is the voltage of the source and $i(t)$ is the electric current. As a result, the electrical energy E provided by the source during T can be calculated as

$$E = \int_0^T v_s(t)i(t) dt.$$

4.21 Engineering Example: electricity voltage in China

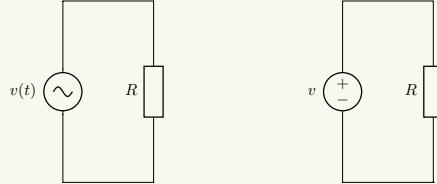
The signal we get from the wall socket is from the power plant. This signal fluctuates in terms of both the magnitude and the direction, we thus call it “Alternating Current (AC) signal”. Specifically, the voltage provided by the AC signal can be formulated as

$$v(t) = v_0 \sin(2\pi ft)$$

where f is the electricity frequency. Note that $f = 50$ Hz in China.

It is well known that China generally uses a 220 volts voltage. We may wonder whether $v_0 = 220$ or not. Actually, v_0 represents the peak value of the voltage and $v_0 = 311$. Then, why do we say “China uses a 220 volts voltage”?

Let's consider the following two circuits.



As we said, in practice, we use the AC signal, as shown in the left figure. Assume that the AC signal provides power P_{AC} for the resistance R . Can we find an equivalent Direct Current (DC) signal, as shown in the right figure, which provides power P_{DC} for R such that $P_{DC} = P_{AC}$?

In order to answer this question, we first need to calculate the power P_{AC} . Let the period

of the AC signal be denoted by T . Based on Ohm's law, the calculation is shown below:

$$\begin{aligned} P_{AC} &= \frac{1}{T} \int_0^T \frac{v^2(t)}{R} dt = \frac{1}{TR} \int_0^T v_0^2 \sin^2(\omega t) dt \\ &= \frac{1}{TR} \int_0^T v_0^2 \frac{1 - \cos(2\omega t)}{2} dt = \frac{1}{TR} \left(\frac{v_0^2}{2} \int_0^T dt - \frac{v_0^2}{2} \int_0^T \cos(2\omega t) dt \right) \\ &= \frac{1}{R} \left(\frac{v_0^2}{2} - \frac{1}{T} \frac{v_0^2}{2} \frac{1}{2\omega} \sin(2\omega t) \Big|_0^T \right) = \frac{v_0^2}{2R}. \end{aligned}$$

For the DC circuit in the right figure, based on Ohm's law, the power provided for R can be calculated as

$$P_{DC} = \frac{v^2}{R}.$$

If $P_{DC} = P_{AC}$, we have

$$\frac{v^2}{R} = \frac{v_0^2}{2R} \Rightarrow v = \frac{v_0}{\sqrt{2}} \approx 220 \text{ (V)}.$$

Therefore, the “220 volts” is a DC equivalent voltage such that the DC signal provides the same power as the AC signal $v(t) = 311 \sin(100\pi t)$.

4.2 Fundamental Theorem of Calculus

4.22 Theorem: Fundamental Theorem of Calculus

Let f be a continuous function on $[a, b]$.

1. Then the function

$$A(x) = \int_a^x f(t) dt$$

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

2. If F is a differentiable function $[a, b]$ such that $F'(x) = f(x)$, then

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

Justification: Part 1

Since f is continuous over $[a, b]$, it has an absolute maximum and an absolute minimum over $[a, b]$, and thus it is bounded, i.e., there exists B such that $|f(x)| \leq B$ for all x in $[a, b]$. Let c be a point in $[a, b]$. Let x be an arbitrary point and assume $x > c$. By the properties

of integration, we have

$$\begin{aligned}|A(x) - A(c)| &= \left| \int_a^x f(t) dt - \int_a^c f(t) dt \right| = \left| \int_c^x f(t) dt \right| \\&\leq \int_c^x |f(t)| dt \leq \int_c^x B dt = B(x - c).\end{aligned}$$

Thus when $x \rightarrow c^+$, by the Squeeze Theorem, $\lim_{x \rightarrow c^+} |A(x) - A(c)| = 0$, i.e., $\lim_{x \rightarrow c^+} A(x) = A(c)$. By the same argument, we get $\lim_{x \rightarrow c^-} A(x) = A(c)$. Therefore we conclude that $\lim_{x \rightarrow c} A(x) = A(c)$, i.e., A is continuous at c .

To show that f is differentiable in (a, b) , we assume c is any point in (a, b) . Without loss of generality, assume $x > c$. We write:

$$\begin{aligned}A(x) - A(c) - f(c)(x - c) &= \int_a^x f(t) dt - \int_a^c f(t) dt - \int_c^x f(c) dt \\&= \int_c^x f(t) dt - \int_c^x f(c) dt \\&= \int_c^x [f(t) - f(c)] dt.\end{aligned}$$

Let $M(x)$ be the maximum value of the function $|f(t) - f(c)|$ over the interval between c and x , i.e.,

$$|f(t) - f(c)| \leq M(x) \quad \text{for all } t \text{ between } c \text{ and } x.$$

It depends on x . Since $f(x)$ is continuous, $M(x)$ goes to zero as x goes to 0. Thus we get

$$\begin{aligned}\left| \frac{A(x) - A(c)}{x - c} - f(c) \right| &= \left| \int_c^x \frac{f(t) - f(c)}{x - c} dt \right| \\&\leq \int_c^x \left| \frac{f(t) - f(c)}{x - c} \right| dt \quad \text{properties of integrals} \\&\leq \int_c^x \frac{M(x)}{x - c} dt \quad \text{the order rule} \\&= \frac{M(x)}{x - c}(x - c) = M(x),\end{aligned}$$

which goes to zero as x goes to 0. Thus

$$\lim_{x \rightarrow c} \left[\frac{A(x) - A(c)}{x - c} - f(c) \right] = 0, \quad \text{i.e. } \lim_{x \rightarrow c} \frac{A(x) - A(c)}{x - c} = f(c).$$

Thus we obtain $A'(c) = f(c)$.

Justification: Part 2

Partition the interval $[a, b]$ with n subintervals of equal length $\frac{b-a}{n}$. By the Mean Value

Theorem, there exists x_i^* in $[x_{i-1}, x_i]$ such that

$$F(x_i) - F(x_{i-1}) = F'(x_i^*)(x_i - x_{i-1}) = f(x_i^*) \cdot \frac{b-a}{n}.$$

Then, we take the Riemann sum of the function f on $[a, b]$:

$$\sum_{i=1}^n f(x_i^*) \cdot \frac{b-a}{n} = \sum_{i=1}^n [F(x_i) - F(x_{i-1})] = F(x_n) - F(x_0) = F(b) - F(a).$$

Thus by the definition of definite integral, we have

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

4.23 Example: evaluate definite integrals

Evaluate

$$\int_0^1 x^2 dx, \quad \int_0^\pi \sin x dx, \quad \text{and} \quad \int_0^1 e^x dx.$$

Solution

1. Since $(\frac{1}{3}x^3)' = x^2$, by the Fundamental Theorem of Calculus, we have

$$\int_0^1 x^2 dx = \frac{1}{3}x^3 \Big|_0^1 = \frac{1}{3}.$$

2. Since $(-\cos x)' = \sin x$, by the Fundamental Theorem of Calculus, we have

$$\int_0^\pi \sin x dx = -\cos x \Big|_0^\pi = -\cos \pi - (-\cos 0) = 1 + 1 = 2.$$

3. Since $(e^x)' = e^x$, by the Fundamental Theorem of Calculus, we have

$$\int_0^1 e^x dx = e^x \Big|_0^1 = e^1 - e^0 = e - 1.$$

4.24 Example: area and net area

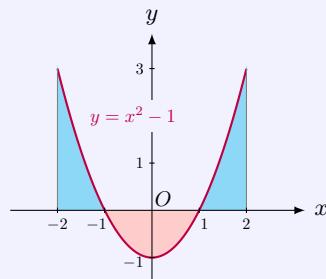
1. Find the *net* area of the region between the graph of $f(x) = x^2 - 1$ and the x -axis over $[-2, 2]$.

Solution The net area of the region is the integral of f over $[-2, 2]$. It is easy to see that

$$\left(\frac{1}{3}x^3 - x\right)' = x^2 - 1.$$

Thus, by the Fundamental Theorem of Calculus, the net area is

$$\begin{aligned}\int_{-2}^2 (x^2 - 1) dx &= \left(\frac{1}{3}x^3 - x\right)\Big|_{-2}^2 \\ &= \left[\frac{1}{3}(2^3) - 2\right] - \left[\frac{1}{3}(-2)^3 - (-2)\right] = \frac{4}{3}.\end{aligned}$$



2. Find the area of the region between the graph of $f(x) = x^2 - 1$ and the x -axis over $[-2, 2]$.

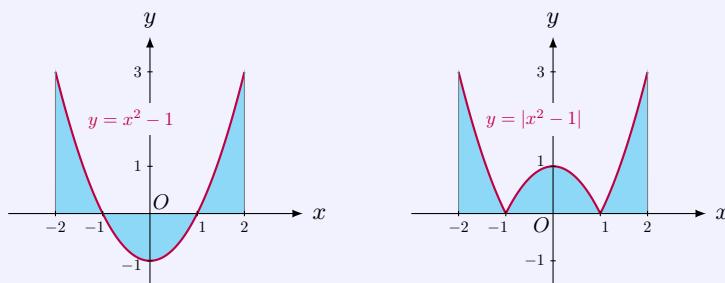
Solution The area of the region equals the sum of the area of the region above the x -axis and the area of the region below the x -axis.

We need to find the intervals over which f does not change its sign. Solve

$$x^2 - 1 = 0 \implies x = 1 \quad \text{or} \quad x = -1.$$

Thus $f(x)$ is positive over $[-2, -1]$, negative over $[-1, 1]$, and positive over $[1, 2]$. Thus, by the Fundamental Theorem of Calculus, the total area equals

$$\begin{aligned}&\int_{-2}^2 |x^2 - 1| dx \\ &= \int_{-2}^{-1} (x^2 - 1) dx - \int_{-1}^1 (x^2 - 1) dx + \int_1^2 (x^2 - 1) dx \\ &= \left(\frac{1}{3}x^3 - x\right)\Big|_{-3}^{-1} - \left(\frac{1}{3}x^3 - x\right)\Big|_{-1}^1 + \left(\frac{1}{3}x^3 - x\right)\Big|_1^3 \\ &= \left[\frac{1}{3}(-1)^3 - (-1)\right] - \left[\frac{1}{3}(-2)^3 - (-2)\right] \\ &\quad - \left[\frac{1}{3}(1^3) - 1\right] + \left[\frac{1}{3}(-1)^3 - (-1)\right] \\ &\quad + \left[\frac{1}{3}(2^3) - 2\right] - \left[\frac{1}{3}(1^3) - 1\right] \\ &= \frac{44}{3}.\end{aligned}$$



4.25 Example: Derivatives of Integrals

Find the derivatives

$$\frac{d}{dx} \int_0^x \sin t^2 dt \quad \text{and} \quad \frac{d}{dx} \int_1^{x^2} \sin t^2 dt.$$

Solution

1. We can use directly the Fundamental Theorem of Calculus to compute the first integral:

$$\frac{d}{dx} \int_0^x \sin t^2 dt = \sin x^2.$$

2. Here we need to apply the chain rule of differentiation. If we let $f(u) = \int_1^u \sin t^2 dt$ and $u = g(x) = x^2$, we get

$$\int_1^{x^2} \sin t^2 dt = f(g(x)).$$

So we apply the chain rule to get

$$\begin{aligned} \frac{d}{dx} \int_1^{x^2} \sin t^2 dt &= \left(\frac{d}{du} \int_1^u \sin t^2 dt \right) \Big|_{u=x^2} \cdot \frac{d(x^2)}{dx} \\ &= \sin u^2 \Big|_{u=x^2} \cdot 2x \\ &= \sin x^4 \cdot 2x = 2x \sin x^4. \end{aligned}$$

4.26 Pitfall: nonexistence of definite integrals

Consider the integral $\int_{-1}^2 \frac{1}{x^2} dx$.

Since $\left(-\frac{1}{x}\right)' = \frac{1}{x^2}$, if one had used the Fundamental Theorem of Calculus for this integral, one would get the integral equal to

$$\begin{aligned} \int_{-1}^2 \frac{1}{x^2} dx &= \left(-\frac{1}{x}\right) \Big|_{-1}^2 && \text{incorrect equality} \\ &= -\frac{1}{2} + (-1) = -\frac{3}{2}. \end{aligned}$$

The computation is incorrect because the integrand $\frac{1}{x^2}$ is not defined at $x = 0$ which is a point in $[-1, 2]$. In fact, if one uses Riemann sum to approximate this integral, the limit of the Riemann sum diverges.

4.3 Indefinite Integrals

4.27 Remark:

From the Fundamental Theorem of Calculus, in order to calculate a definite integral

$$\int_a^b f(x) dx,$$

it is important to know a function $F(x)$ whose derivative is $f(x)$, i.e., $F'(x) = f(x)$. This leads to the concept of indefinite integral, equivalently, antiderivative.

4.28 Definition: indefinite integral

The **indefinite integral** of f , denoted by

$$\int f(x) dx$$

is the collection of all functions F whose derivative equals f , i.e.,

$$F'(x) = f(x).$$

Any function F satisfying $F'(x) = f(x)$ is called an **antiderivative** of f .

4.29 Example: finding indefinite integral

Find $\int 2x dx$.

Solution We know that $(x^2)' = 2x$ and $(x^2 + 1)' = 2x$. In fact, for any constant C , $(x^2 + C)' = 2x$.

We know from the Mean Value Theorem that zero derivative implies constant function. Thus, any differentiable function F with $F'(x) = 2x$ must differ from x^2 at most by a constant. Hence, by the definition of indefinite integral, we have

$$\int 2x dx = x^2 + C.$$

4.30 Theorem: linearity of indefinite integral

$$1. \int [f(x) + g(x)] dx = \int f(x) dx + \int g(x) dx.$$

$$2. \int kf(x) dx = k \int f(x) dx, \text{ where } k \text{ is a constant.}$$

Justification:

The equalities follow from the linearity of differentiation.

4.31 Theorem: indefinite integrals of some elementary functions

$$1. \int k \, dx = kx + C$$

$$2. \int x^\mu \, dx = \frac{x^{\mu+1}}{\mu+1} \quad (\mu \neq -1)$$

$$3. \int \frac{dx}{x} = \ln|x| + C$$

$$4. \int \frac{dx}{x^2+1} = \arctan x + C$$

$$5. \int \frac{dx}{\sqrt{1-x^2}} = \arcsin x + C$$

$$10. \int \frac{dx}{\cos^2 x} = \int \sec^2 x \, dx = \tan x + C$$

$$11. \int \frac{dx}{\sin^2 x} = \int \csc^2 x \, dx = -\cot x + C$$

$$12. \int \sec x \tan x \, dx = \sec x + C$$

$$13. \int \csc x \cot x \, dx = -\csc x + C$$

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

$$7. \int \sin x \, dx = -\cos x + C$$

$$8. \int e^x \, dx = e^x + C$$

$$9. \int a^x \, dx = \frac{a^x}{\ln a} + C$$

Justification:

For formula 1, it is easy to see that $(kx + C)' = k$ for any constant C . We know from the Mean Value Theorem that zero derivative implies constant function. Thus, any differentiable function F with $F'(x) = k$ must differ from kx at most by a constant. Hence, by the definition of indefinite integral, we have

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

Similar argument applies to other formulae, except formula 3.

For formula 3, if $x > 0$, we have $\ln|x| = \ln x$ and

$$(\ln x)' = \frac{1}{x}.$$

Since, from the Mean Value Theorem, zero derivative implies constant function, we know that any differentiable function F with $F'(x) = \frac{1}{x}$ must differ from $\ln x$ at most by a constant.

Hence, by the definition of indefinite integral, we know that for $x > 0$, $\int x^{-1} \, dx = \ln x + C$.

If $x < 0$, we have $\ln|x| = \ln(-x)$ and, applying the chain rule,

$$(\ln(-x))' = \frac{1}{-x} \cdot (-x)' = \frac{1}{x}.$$

By a similar argument as above, we know that for $x < 0$, $\int x^{-1} dx = \ln(-x) + C$.

Combining two cases into one format, we obtain formula 3:

$$\int x^{-1} dx = \ln|x| + C.$$

4.32 Remark: constants in indefinite integrals

It is easy to see that

$$\begin{aligned}\int 1 dx &= x + C_1, \\ \int 2x dx &= x^2 + C_2, \\ \int (1 + 2x) dx &= x + x^2 + C,\end{aligned}$$

where C_1 , C_2 , and C are all arbitrary constants.

On the other hand, if we use the linearity of indefinite integrals, we have

$$\begin{aligned}\int (1 + 2x) dx &= \int 1 dx + \int 2x dx \\ &= x + C_1 + x^2 + C_2 = x + x^2 + C_1 + C_2.\end{aligned}$$

It seems that we get different results from different ways of computations.

However, the constant C in indefinite integrals means that it is an arbitrary constant. For our example, since C_1 and C_2 are arbitrary constants, so is $C_1 + C_2$. Therefore, it makes sense to write $x + x^2 + C_1 + C_2$ as $x + x^2 + C$.

4.33 Example: definite integral - Fundamental Theorem of Calculus

Evaluate $\int_0^1 \left(x^5 + 2 \cos x - \frac{1}{x^2 + 1} \right) dx$.

Solution We can first evaluate the indefinite integral:

$$\begin{aligned}&\int \left(x^5 + 2 \cos x - \frac{1}{x^2 + 1} \right) dx \\ &= \int x^5 dx + 2 \int \cos x dx - \int \frac{1}{x^2 + 1} dx && \text{applying linearity} \\ &= \frac{1}{6}x^6 + 2 \sin x - \tan^{-1} x + C && \text{using the known integrals}\end{aligned}$$

Then, by applying the Fundamental Theorem of Calculus, we get

$$\begin{aligned}\int_0^1 \left(x^5 + 2 \cos x - \frac{1}{x^2 + 1} \right) dx &= \left(\frac{1}{6}x^6 + 2 \sin x - \tan^{-1} x \right) \Big|_0^1 \\ &= \frac{1}{6} + 2 \sin 1 - \tan^{-1} 1 \\ &= \frac{1}{6} + 2 \sin 1 - \frac{\pi}{4}.\end{aligned}$$

4.34 Remark: displacement function and velocity function

If $s(t)$ is the displacement function, then $v(t) = s'(t)$ is the velocity function from the definition of derivatives.

Conversely, if we are given the velocity function $v(t)$, then the indefinite integral

$$\int v(t) dt$$

is a collection of displacement functions having $v(t)$ as its velocity.

4.35 Example: velocity - application

A particle moves on a line with the velocity at time t , $v(t) = t^2 - t - 6$ (measured in meters per second).

- Find the displacement function of the particle as a function of the time t . Assume the particle is at the position of the origin when time $t = 0$.

Solution Let f be the displacement function of the particle satisfying $f(0) = 0$. Since $f'(t) = v(t)$, in order to find f , we just need to calculate the indefinite integral

$$f(t) = \int v(t) dt = \int (t^2 - t - 6) dt = \frac{1}{3}t^3 - \frac{1}{2}t^2 - 6t + C.$$

The initial condition $f(0) = 0$ will determine the constant C . In fact, $C = 0$.

- Find the displacement traveled during the time period $1 \leq t \leq 4$.

Solution It requires us to calculate the following definite integral.

$$\int_1^4 v(t) dt = \left(\frac{1}{3}t^3 - \frac{1}{2}t^2 - 6t \right) \Big|_1^4 = -\frac{9}{2}.$$

- Find the distance traveled during the time period $1 \leq t \leq 4$.

Solution This amounts to calculate the definite integral $\int_1^4 |v(t)| dt$.

To do so, we need to find intervals where v has the same sign. In fact, let $v(t) = 0$, we get

$$t^2 - t - 6 = 0 \implies (t - 3)(t + 2) = 0 \implies t = 3 \text{ or } t = -2.$$

It is easy to see that $v(t)$ is positive on $(-\infty, -2)$, negative on $(-2, 3)$, and positive on $(3, \infty)$. Therefore the distance traveled during the time period $1 \leq t \leq 4$ is

$$\begin{aligned} \int_1^4 |v(t)| dt &= - \int_1^3 v(t) dt + \int_3^4 v(t) dt \\ &= -\left(\frac{1}{3}t^3 - \frac{1}{2}t^2 - 6t\right)\Big|_1^3 + \left(\frac{1}{3}t^3 - \frac{1}{2}t^2 - 6t\right)\Big|_3^4 = \frac{61}{6}. \end{aligned}$$

4.36 MATLAB: indefinite integrals

In MATLAB, the command `int` can be used to calculate indefinite integrals. The following examples show the calculation of $\int x^n dx$ and $\int (t^2 - t - 6) dt$.

```
>> syms x;
>> f = x^n;
>> int(f,x)

ans =

piecewise(n == -1, log(x), n != -1, x^(n + 1)/(n + 1))

>> syms t;
>> g = t^2 - t - 6;
>> m = int(g)

m =

-(t*(- 2*t^2 + 3*t + 36))/6

>> m_cl = collect(m) % rewrite m in terms of the powers of t
m_cl =

t^3/3 - t^2/2 - 6*t

>> pretty(m_cl) % rewrite m_cl in a format that resembles ...
   type-set mathematics
      2
      t   t
      -- - - - 6 t
      3   2
```

We see from the MATLAB code above that $\int x^n dx$ is a piecewise function. With the help of `collect` and `pretty`, we can rewrite the results calculated by `int`. Note that for all the results calculated by `int`, the constant term C is not shown.

4.4 Substitution Method

4.37 Example: substitution method

For the function e^{x^2} , we apply the chain rule and get

$$\frac{d}{dx} e^{x^2} = 2xe^{x^2}.$$

We can decompose the computation into several steps. First, let $u = x^2$. Then

$$\frac{d}{dx} e^{x^2} = \left. \frac{d}{du} e^u \right|_{u=x^2} \cdot \frac{du}{dx}.$$

By the definition of indefinite integral, we get

$$\int 2xe^{x^2} dx = e^{x^2} + C.$$

Or, if we let $u = x^2$, then we get $du = (x^2)' dx = 2x dx$, and the integral can be rewritten as

$$\int 2xe^{x^2} dx = \int e^{x^2} \underbrace{2x dx}_{du} = \int e^u du = e^u + C = e^{x^2} + C.$$

This process is a typical case of the substitution method.

4.38 Theorem: substitution method for indefinite integral

Let $u = g(x)$ be a differentiable function and f be a continuous function. Suppose the range of g is contained in the domain of f . Then

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

Justification:

It follows directly from the chain rule for differentiation. Let $F(x)$ be a function whose derivative is $f(x)$, i.e., $F'(x) = f(x)$. Equivalently, we have

$$\int f(u) du = F(u) + C.$$

Then by the chain rule of differentiation, we have

$$\frac{d}{dx} F(g(x)) = \frac{dF(u)}{du} \cdot g'(x) = f(u) \cdot g'(x) = f(g(x)) \cdot g'(x),$$

or equivalently

$$\begin{aligned} \int f(g(x)) \cdot g'(x) dx &= \int \frac{d}{dx} F(g(x)) dx \\ &= F(g(x)) + C = F(u) + C && u = g(x) \\ &= \int f(u) du. \end{aligned}$$

4.39 Example: substitution method for indefinite integrals

Using the substitution method to evaluate the integral

$$\int \frac{x}{\sqrt{x+1}} dx.$$

1. Let $u = x + 1$. Then $du = dx$. Thus, we have

$$\begin{aligned} &\int \frac{x}{\sqrt{x+1}} dx \\ &= \int \frac{u-1}{\sqrt{u}} du && u = x+1, du = dx \\ &= \int \left(u^{\frac{1}{2}} - u^{-\frac{1}{2}} \right) du && \text{simplifying} \\ &= \frac{2}{3}u^{\frac{3}{2}} - 2u^{\frac{1}{2}} + C && \text{evaluating} \\ &= \frac{2}{3}(x+1)^{\frac{3}{2}} - 2(x+1)^{\frac{1}{2}} + C && \text{replacing } u \text{ by } x+1 \\ &= \frac{2}{3}\sqrt{(x+1)^3} - 2\sqrt{x+1} + C. \end{aligned}$$

2. We can also make the following substitution $u = \sqrt{x+1}$. Then $x = u^2 - 1$ and $dx = 2u du$. Thus, we have

$$\begin{aligned} &\int \frac{x}{\sqrt{x+1}} dx \\ &= \int \frac{u^2-1}{u} 2u du && u = \sqrt{x+1}, dx = 2u du \\ &= \int (2u^2 - 2) du && \text{simplifying} \\ &= \frac{2}{3}u^3 - 2u + C && \text{evaluating} \\ &= \frac{2}{3}\sqrt{(x+1)^3} - 2\sqrt{x+1} + C. && \text{replacing } u \text{ by } \sqrt{x+1} \end{aligned}$$

From this example, we see that there might be more than one way to make substitution.

4.40 Theorem: substitution method for definite integral

Let $u = g(x)$, where $g'(x)$ is continuous on $[a, b]$, and let f be continuous on an interval containing the range of g . Then

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

4.41 Example: substitution method for definite integrals

Evaluate

$$\int_0^{\frac{\pi}{2}} \sin^3 x \cos x dx.$$

Solution If we let $u = \sin x$, then

$$du = (\sin x)' dx = \cos x dx.$$

When $x = 0$, $u = \sin 0 = 0$; when $x = \frac{\pi}{2}$, $u = \sin \frac{\pi}{2} = 1$. Thus the integral can be evaluated as follows:

$$\begin{aligned} & \int_0^{\frac{\pi}{2}} \sin^3 x \underbrace{\cos x dx}_{du} \\ &= \int_0^1 u^3 (du) && \text{substituting } u = \sin x, du = \cos x dx \\ &= \frac{1}{4}u^4 \Big|_0^1 && \text{and changing the upper and lower limits} \\ &= \left(\frac{1}{4} - 0\right) = \frac{1}{4}. \end{aligned}$$

4.42 Engineering Example: electrical power in a circuit

The electrical power can be supplied to business and homes by the electrical power industry through electrical power grids, and it can be converted into other forms of energy, such as heat, light and motion. The amount of power produced in a circuit is related to the electric current, the voltage, etc.

In particular, in the case of an alternating current (AC) circuit with Ohmic resistive loads, the Ohm's law can be combined, and thus the electrical power during one time period $(0, T)$ is given by

$$P = \frac{R}{T} \int_0^T i^2(t) dt,$$

where $i(t)$ is the alternating current.

Suppose that $i(t) = A \sin(\omega t)$, where A is the amplitude and ω is the angular frequency. Calculate P .

First, if $i(t) = A \sin(\omega t)$, the electrical power P can be calculated by solving the following

integral:

$$P = \frac{A^2 R}{T} \int_0^T \sin^2(\omega t) dt.$$

Let $u = \omega t$, then

$$du = \omega dt.$$

When $t = 0$, $u = \omega \times 0 = 0$; when $t = T$, $u = \omega T$. Thus the integral of P can be rewritten as follows:

$$P = \frac{A^2 R}{T \omega} \int_0^{\omega T} \sin^2 u du.$$

Moreover, we note that $\sin^2 u = \frac{1}{2} - \frac{1}{2} \cos(2u)$, the integral above thus can be rewritten as

$$\begin{aligned} P &= \frac{A^2 R}{T \omega} \int_0^{\omega T} \left(\frac{1}{2} - \frac{1}{2} \cos(2u) \right) du \\ &= \frac{A^2 R}{2T \omega} \int_0^{\omega T} du - \frac{A^2 R}{2T \omega} \int_0^{\omega T} \cos(2u) du \\ &= \frac{1}{2} A^2 R - \frac{A^2 R}{2T \omega} \int_0^{\omega T} \cos(2u) du. \end{aligned}$$

Furthermore, we let $v = 2u$, then

$$dv = 2du.$$

When $u = 0$, $v = 2 \times 0 = 0$; when $u = \omega T$, $v = 2 \times \omega T = 2\omega T$. We thus have

$$\begin{aligned} \int_0^{\omega T} \cos(2u) du &= \frac{1}{2} \int_0^{2\omega T} \cos v dv \\ &= \frac{1}{2} [\sin v]_0^{2\omega T} = \frac{1}{2} \sin(2\omega T). \end{aligned}$$

Therefore, we can obtain

$$P = \frac{1}{2} A^2 R - \frac{1}{4T \omega} A^2 R \sin(2\omega T).$$

Note that $T = \frac{2\pi}{\omega}$, so P can be further simplified as

$$P = \frac{1}{2} A^2 R - \frac{1}{8\pi} A^2 R \sin(4\pi) = \frac{1}{2} A^2 R.$$

4.43 Engineering Example: hazard function

The hazard function provides the instantaneous failure rate of a component at time t . Mathematically, the hazard function $h(t)$ can be defined as the ratio of the probability density function $f(t)$ and the complementary cumulative distribution function $R(t)$ of the lifetime of a component, i.e.,

$$h(t) = \frac{f(t)}{R(t)}.$$

Note that $R(t)$ can be calculated in terms of $f(t)$ as follows:

$$R(t) = 1 - \int_0^t f(u)du,$$

so hazard function can be rewritten as

$$h(t) = \frac{f(t)}{1 - \int_0^t f(u)du}.$$

Suppose that the probability density function of a component is $f(t) = \frac{1}{2\sqrt{t}}e^{-\sqrt{t}}$. Find $h(t)$.

In order to calculate $h(t)$, we first calculate the integral in the denominator:

$$\int_0^t f(u)du = \frac{1}{2} \int_0^t \frac{1}{\sqrt{u}} e^{-\sqrt{u}} du.$$

Let $v = \sqrt{u}$, then

$$dv = \frac{1}{2\sqrt{u}} du.$$

When $u = 0$, $v = \sqrt{0} = 0$; when $u = t$, $v = \sqrt{t}$. We thus have

$$\begin{aligned} \int_0^t f(u)du &= \int_0^{\sqrt{t}} e^{-v} dv \\ &= -[e^{-v}]_0^{\sqrt{t}} = 1 - e^{-\sqrt{t}}. \end{aligned}$$

Substituting the above result into the expression of $h(t)$ and after some algebraic manipulations, we have

$$h(t) = \frac{1}{2\sqrt{t}}.$$

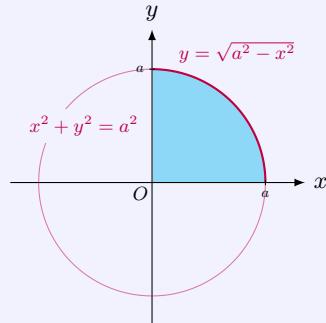
Trigonometric Substitutions

4.44 Example: integrals involving $\sqrt{a^2 - x^2}$

Find the area of a circle of radius a .

Solution The area of the circle equals four times the area of the part of the circle in the first quadrant, which can be calculated via the area below the curve $y = \sqrt{a^2 - x^2}$ over the interval $[0, a]$, i.e., we need to calculate the following integral

$$4 \int_0^a \sqrt{a^2 - x^2} dx.$$



In order to remove the radical in the integrals, we use the following trigonometric substitution for integrals involving $a^2 - x^2$. Let $x = a \sin \theta$. Then $dx = a \cos \theta d\theta$, and when $x = 0$, $\theta = 0$; when $x = a$, $\theta = \frac{\pi}{2}$. Hence,

$$\begin{aligned} & 4 \int_0^a \sqrt{a^2 - x^2} dx \\ &= 4 \int_0^{\frac{\pi}{2}} \sqrt{a^2 - a^2 \sin^2 \theta} \cdot a \cos \theta d\theta && \text{making substitution } x = a \sin \theta \\ &= 4 \int_0^{\frac{\pi}{2}} a \cos \theta \cdot a \cos \theta d\theta && \text{using } 1 - \sin^2 \theta = \cos^2 \theta \\ &= 4a^2 \int_0^{\frac{\pi}{2}} \frac{1}{2}(1 + \cos 2\theta) d\theta && \text{using the double-angle formula} \\ &= 2a^2 (\theta + \frac{1}{2} \sin 2\theta) \Big|_0^{\frac{\pi}{2}} && \text{evaluating} \\ &= 2a^2 (\frac{\pi}{2} + 0) = \pi a^2. && \text{simplifying} \end{aligned}$$

4.45 Method: trigonometric substitutions

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

4.46 Example: integrals involving $\sqrt{x^2 - a^2}$

Evaluate the integral

$$\int \frac{dx}{\sqrt{(x^2 - 1)^3}}.$$

Solution In order to remove the radical in the integrals, we use the trigonometric substitution $x = \sec \theta$. Then we have

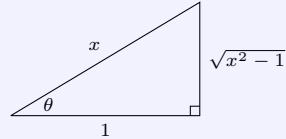
$$x^2 - 1 = \sec^2 \theta - 1 = \tan^2 \theta, \quad dx = \sec \theta \tan \theta d\theta.$$

Thus

$$\begin{aligned} \int \frac{dx}{\sqrt{(x^2 - 1)^3}} &= \int \frac{\sec \theta \tan \theta}{\tan^3 \theta} d\theta && \text{substituting } x = \sec \theta \\ &= \int \frac{\cos \theta}{\sin^2 \theta} d\theta && \text{simplifying} \\ &= \int \frac{1}{u^2} du && \text{substituting } u = \sin \theta \\ &= -\frac{1}{u} + C && \text{evaluating} \\ &= -\frac{1}{\sin \theta} + C && \text{changing the variable from } u \text{ to } \theta \\ &= -\frac{x}{\sqrt{x^2 - 1}} + C. && \text{changing the variable from } \theta \text{ to } x \end{aligned}$$

Here we use the following diagram to convert the variable from θ to x :

$$\sin \theta = \frac{\sqrt{x^2 - 1}}{x}.$$


4.47 Example: integrals involving $\sqrt{a^2 + x^2}$

Evaluate the integral $\int \frac{dx}{(1 + x^2)^2}$.

Solution Although the integrand contains no radicals, a trigonometric substitution can still apply to simplify the term $(1 + x^2)^k$. Indeed, using the trigonometric substitution $x = \tan \theta$, we have

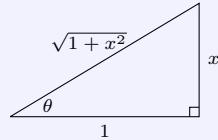
$$1 + x^2 = 1 + \tan^2 \theta = \sec^2 \theta, \quad d(\tan \theta) = \sec^2 \theta d\theta.$$

The integral is

$$\begin{aligned} \int \frac{dx}{(1 + x^2)^2} &= \int \frac{\sec^2 \theta}{\sec^4 \theta} d\theta = \int \cos^2 \theta d\theta \\ &= \int \frac{1}{2}(1 + \cos 2\theta) d\theta = \frac{1}{2}\theta + \frac{1}{4}\sin 2\theta + C. \end{aligned}$$

To convert θ back to x , we have $\theta = \tan^{-1} x$. From the triangle below, we have

$$\begin{aligned}\sin \theta &= \frac{x}{\sqrt{1+x^2}}, & \cos \theta &= \frac{1}{\sqrt{1+x^2}}, \\ \sin 2\theta &= 2 \sin \theta \cos \theta = \frac{2x}{1+x^2}.\end{aligned}$$



Thus the integral is

$$\int \frac{dx}{(1+x^2)^2} = \frac{1}{2} \tan^{-1} x + \frac{x}{2(1+x^2)} + C.$$

4.5 Integration by Parts

4.48 Example: product rule and integration by parts

Let $u(x)$ and $v(x)$ be differentiable functions of x . The product rule of differentiation gives that

$$(uv)' = u'v + uv'.$$

Rewrite the expression as

$$uv' = (uv)' - u'v.$$

As an example, consider the integral $\int xe^x dx$. If we let

$$u(x) = x, \quad v'(x) = e^x,$$

then, for the latter, we may choose $v = e^x$. From the equality $uv' = (uv)' - u'v$, we get

$$\begin{aligned}& \int xe^x dx \\&= \int x(e^x)' dx = \int u(x)v'(x) dx && u = x \text{ and } v = e^x \\&= \int [(u(x)v(x))' - u'(x)v(x)] dx && \text{applying the product rule} \\&= u(x)v(x) - \int u'(x)v(x) dx && \text{definition of indefinite integral} \\&= xe^x - \int e^x dx \\&= xe^x - e^x + C.\end{aligned}$$

This process is a typical case of so-called integration by parts.

4.49 Theorem: integration by parts

Suppose u and v are differentiable functions of x . Then

$$\int u(x) \cdot v'(x) dx = u(x) \cdot v(x) - \int v(x) \cdot u'(x) dx,$$

or

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

Furthermore, for definite integral,

$$\int_a^b u(x) \cdot v'(x) dx = [u(x) \cdot v(x)]_a^b - \int_a^b v(x) \cdot u'(x) dx.$$

Justification:

By the product rule of differentiation, we get

$$(uv)' = u'v + uv' \implies uv' = (uv)' - u'v.$$

Integrating the both sides of the latter equality gives

$$\int uv' dx = \int (uv)' dx - \int u'v dx = uv - \int u'v dx,$$

or equivalently

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

By further applying the Fundamental Theorem of Calculus, we get the formula of integration by parts for definite integral.

4.50 Example: repeatedly integrating by parts

Evaluate the integral

$$\int e^{2x} \sin x dx.$$

Solution We make the following choice of u and v :

Functions in original integral	$u = e^{2x}$	$dv = \sin x dx$
Functions in new integral	$du = 2e^{2x} dx$	$v = -\cos x$

We obtain

$$\begin{aligned}
 & \int e^{2x} \sin x \, dx \\
 &= \int e^{2x}(-\cos x)' \, dx = \int u(x)v'(x) \, dx \quad u = e^{2x} \text{ and } v = -\cos x \\
 &= \int [(u(x)v(x))' - u'(x)v(x)] \, dx \quad \text{applying the product rule} \\
 &= u(x)v(x) - \int u'(x)v(x) \, dx \quad \text{definition of indefinite integral} \\
 &= e^{2x}(-\cos x) - \int (-\cos x)2e^{2x} \, dx \\
 &= -e^{2x} \cos x + 2 \int e^{2x} \cos x \, dx.
 \end{aligned}$$

By comparing the last integral $\int e^{2x} \cos x \, dx$ with the original integral $\int e^{2x} \sin x \, dx$, we see that integration by parts converts the sine function to the cosine function in the integrands. This motivates us to apply integration by part once more to the integral

$$\int e^{2x} \cos x \, dx.$$

To do so, we make the following choice of u and v :

Functions in original integral	$u = e^{2x}$	$dv = \cos x \, dx$
Functions in new integral	$du = 2e^{2x} \, dx$	$v = \sin x$

We obtain

$$\begin{aligned}
 & \int e^{2x} \cos x \, dx \\
 &= \int e^{2x}(\sin x)' \, dx = \int u(x)v'(x) \, dx \quad u = e^{2x} \text{ and } v = \sin x \\
 &= \int [(u(x)v(x))' - u'(x)v(x)] \, dx \quad \text{applying the product rule} \\
 &= u(x)v(x) - \int u'(x)v(x) \, dx \quad \text{definition of indefinite integral} \\
 &= e^{2x} \sin x - \int \sin x \cdot 2e^{2x} \, dx \\
 &= e^{2x} \sin x - 2 \int e^{2x} \sin x \, dx.
 \end{aligned}$$

Combining two relations obtained above gives

$$\int e^{2x} \sin x \, dx = -e^{2x} \cos x + 2e^{2x} \sin x - 4 \int e^{2x} \sin x \, dx.$$

Therefore, by treating the desired integral as an unknown, we get

$$\int e^{2x} \sin x \, dx = -\frac{1}{5}e^{2x} \cos x + \frac{2}{5}e^{2x} \sin x + C.$$

4.51 Pitfall: choices of u and v

If we choose $u = e^x$, $dv = x \, dx$ and hence $v = \frac{1}{2}x^2$ to evaluate the integral

$$\int xe^x \, dx.$$

Integrating by parts, we get

$$\begin{aligned} \int xe^x \, dx &= \int e^x \left(\frac{1}{2}x^2\right)' \, dx \\ &= \frac{1}{2}x^2 \cdot e^x - \int \frac{1}{2}x^2 \cdot (e^x)' \, dx \\ &= \frac{1}{2}x^2 \cdot e^x - \int \frac{1}{2}x^2 \cdot e^x \, dx. \end{aligned}$$

Now we see the resulting integration on the right hand side is more complicated than the original one. Thus we made a wrong choice of u and v .

Instead, if we let $u = x$ and $v = e^x$, we get

$$\begin{aligned} \int xe^x \, dx &= \int x(e^x)' \, dx \\ &= xe^x - \int e^x \, dx = xe^x - e^x + C. \end{aligned}$$

4.52 Example: recursive formula

In many cases, integration by parts can be used to yield recursive formula for integrals, so that some complicated integrals can be evaluated through simpler integrals.

Consider the integral $I_p = \int_0^{\frac{\pi}{2}} \cos^p x \, dx$.

For $p > 1$, if taking $u(x) = \cos^{p-1} x$ and $v(x) = \sin x$, we integrate by parts to have

$$\begin{aligned} I_p &= \int_0^{\frac{\pi}{2}} \cos^{p-1} x \cdot (\sin x)' \, dx \\ &= [\cos^{p-1} x \cdot \sin x]_0^{\frac{\pi}{2}} - \int_0^{\frac{\pi}{2}} \sin x \cdot (p-1) \cos^{p-2} x \cdot (-\sin x) \, dx \\ &= (p-1) \int_0^{\frac{\pi}{2}} (1 - \cos^2 x) \cos^{p-2} x \, dx \\ &= (p-1)I_{p-2} - (p-1)I_p. \end{aligned}$$

Hence, we have the following recursive formula:

$$I_p = \frac{p-1}{p} I_{p-2}, \quad p > 1.$$

It is easy to have that

$$\begin{aligned} I_0 &= \int_0^{\frac{\pi}{2}} 1 \, dx = \frac{\pi}{2}; \\ I_1 &= \int_0^{\frac{\pi}{2}} \cos x \, dx = 1. \end{aligned}$$

Therefore, we have

$$\begin{aligned} \int_0^{\frac{\pi}{2}} \cos^{2n} x \, dx &= I_{2n} = \frac{2n-1}{2n} I_{2n-2} = \frac{2n-1}{2n} \cdot \frac{2n-3}{2n-2} I_{2n-4} \\ &= \cdots = \frac{2n-1}{2n} \cdot \frac{2n-3}{2n-2} \cdots \frac{1}{2} I_0 \\ &= \frac{(2n-1)(2n-3) \cdots 1}{2n(2n-2) \cdots 2} \cdot \frac{\pi}{2} \\ &= \frac{(2n)! \cdot \pi}{2^{2n+1}(n!)^2}; \\ \int_0^{\frac{\pi}{2}} \cos^{2n+1} x \, dx &= I_{2n+1} = \frac{2n}{2n+1} I_{2n-1} = \frac{2n}{2n+1} \cdot \frac{2n-2}{2n-1} I_{2n-3} \\ &= \cdots = \frac{2n}{2n+1} \cdot \frac{2n-2}{2n-1} \cdots \frac{2}{3} I_1 \\ &= \frac{(2n)(2n-2) \cdots 2}{(2n+1)(2n-1) \cdots 3} \cdot 1 \\ &= \frac{2^{2n}(n!)^2}{(2n+1)!}. \end{aligned}$$

These hold for all integers $n \geq 0$.

4.53 Example: use substitution or integration by parts?

Expressions of similar integrals sometimes need different integration techniques.

Consider the integral of the form

$$\int x^p \ln x \, dx.$$

- Assume $p = -1$. In this case, we use the substitution method. We take $u = \ln x$. Then $du = x^{-1} \, dx$. Thus,

$$\int x^{-1} \ln x \, dx = \int u \, du = \frac{1}{2}u^2 + C = \frac{1}{2}\ln^2 x + C.$$

2. Assume $p \neq -1$. In this case, we use integration by parts. In fact,

$$\begin{aligned} \int x^p \ln x \, dx &= \int \ln x \cdot \left(\frac{x^{p+1}}{p+1} \right)' \, dx \\ &= \frac{x^{p+1}}{p+1} \ln x - \int \frac{x^{p+1}}{p+1} (\ln x)' \, dx \\ &= \frac{x^{p+1}}{p+1} \ln x - \int \frac{x^{p+1}}{p+1} x^{-1} \, dx \\ &= \frac{x^{p+1}}{p+1} \ln x - \int \frac{x^p}{p+1} \, dx \\ &= \frac{x^{p+1}}{p+1} \ln x - \frac{x^{p+1}}{(p+1)^2} + C. \end{aligned}$$

4.54 Engineering Example: electric current - inductor

The inductor is a two-terminal electrical component that can store energy in a magnetic field and is characterized by its inductance. The inductance is defined as the ratio of the voltage to the rate of the change of the electric current. Mathematically, the inductance L is defined as

$$L = \frac{v(t)}{\frac{di(t)}{dt}}$$

where $v(t)$ is the voltage across the inductor and $i(t)$ is the electric current through the inductor.

Alternatively, according to the fundamental theorem of calculus, the electric current through the inductor $i(t)$ can be calculated as follows:

$$i(t) = \frac{1}{L} \int_0^t v(u) \, du.$$

Suppose that $L = 5 \times 10^{-3}$ (H) and $v(t) = 3te^{-t}$ (V), find $i(t)$.

First, we know that

$$i(t) = \frac{3}{5} \times 10^3 \int_0^t ue^{-u} \, du.$$

Integrating by parts, the integral above can be solved as

$$\begin{aligned} \int_0^t ue^{-u} \, dt &= - \int_0^t u (e^{-u})' \, du \\ &= - [ue^{-u}]_0^t + \int_0^t e^{-u} \, du \\ &= -te^{-t} - [e^{-u}]_0^t \\ &= 1 - e^{-t} - te^{-t}. \end{aligned}$$

Therefore, the electric current through the inductor is

$$i(t) = 600 (1 - e^{-t} - te^{-t}) \quad (\text{A}).$$

4.6 Some Common Integrals

Trigonometric Integrals

4.55 Example: integrating products of $\sin mx$ and $\cos nx$

1. Calculate

$$\int \sin 2x \cos 3x \, dx.$$

Solution By using the identity

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

we have

$$\begin{aligned} \int \sin 2x \cos 3x \, dx &= \int \frac{1}{2} (-\sin x + \sin 5x) \, dx \\ &= \frac{1}{2} \cos x - \frac{1}{10} \cos 5x + C. \end{aligned}$$

2. Calculate

$$\int \cos 4x \cos x \, dx.$$

Solution By using the identity

$$\cos \alpha \cos \beta = \frac{1}{2} [\cos(\alpha - \beta) + \cos(\alpha + \beta)],$$

we have

$$\begin{aligned} \int \cos 4x \cos x \, dx &= \int \frac{1}{2} (\cos 3x + \cos 5x) \, dx \\ &= \frac{1}{6} \sin 3x + \frac{1}{10} \sin 5x + C. \end{aligned}$$

4.56 Method: integrating products of $\sin mx$ and $\cos nx$

For evaluating

$$\int \sin mx \cos nx \, dx, \quad \int \sin mx \sin nx \, dx, \quad \int \cos mx \cos nx \, dx,$$

use the identities:

$$\sin \alpha \cos \beta = \frac{1}{2} [\sin(\alpha - \beta) + \sin(\alpha + \beta)]$$

$$\sin \alpha \sin \beta = \frac{1}{2} [\cos(\alpha - \beta) - \cos(\alpha + \beta)]$$

$$\cos \alpha \cos \beta = \frac{1}{2} [\cos(\alpha - \beta) + \cos(\alpha + \beta)]$$

4.57 Engineering Example: energy in a capacitor - calculation

The capacitor is a two-terminal electronic component that can store electrical energy in an electric field. The electric power is the rate (per unit time) at which the electrical energy is transferred by an electric circuit. Hence, to calculate the energy that is stored in a capacitor, we can calculate the electric power of the capacitor first. In particular, the instantaneous electric power of the capacitor $P_c(t)$ can be calculated by

$$P_c(t) = i_c(t)v_c(t)$$

where $i_c(t)$ is the electric current through the capacitor at time t and $v_c(t)$ is the voltage across the capacitor at time t . Then, the energy stored in a capacitor during the time interval $(0, T_c)$ is given by

$$W_c = \int_0^{T_c} P_c(t)dt = \int_0^{T_c} i_c(t)v_c(t)dt.$$

Suppose that the voltage across a capacitor with capacitance C and zero energy storage at time 0 is $v_c(t) = A \sin(\omega t)$, where A is the amplitude and ω is the angular frequency. Calculate W_c which is the energy stored during the time interval $(0, T_c)$.

First, the electric current through the capacitor can be calculated as

$$i_c(t) = C \frac{dv_c(t)}{dt} = C \frac{A \cdot d[\sin(\omega t)]}{dt} = CA\omega \cos(\omega t).$$

Hence, the energy stored in the capacitor during the time interval $(0, T_c)$ can be calculated by solving the following integral:

$$W_c = CA\omega \int_0^{T_c} \cos(\omega t) \sin(\omega t) dt.$$

Let $u = \cos(\omega t)$, then we have $du = -\omega \sin(\omega t) dt$. Moreover, when $t = 0$, $u = 1$; when $t = T_c$, $u = \cos(\omega T_c)$. So we can get

$$\begin{aligned} W_c &= -CA \int_1^{\cos(\omega T_c)} u du = CA \int_{\cos(\omega T_c)}^1 u du \\ &= CA \left[\frac{1}{2}u^2 \right]_{\cos(\omega T_c)}^1 = \frac{1}{2}CA (1 - \cos^2(\omega T_c)) \\ &= \frac{1}{2}CA \sin^2(\omega T_c). \end{aligned}$$

4.58 Example: integrating products of powers of $\sin x$ and $\cos x$

1. Calculate

$$\int \sin^2 x \cos x dx.$$

Solution We apply the substitution method and let $u = \sin x$. Then we have $du =$

$\cos x \, dx$ and get

$$\begin{aligned}\int \sin^2 x \cos x \, dx &= \int u^2 \, du \\ &= \frac{1}{3}u^3 + C = \frac{1}{3}\sin^3 x + C.\end{aligned}$$

2. Calculate

$$\int \sin^3 x \cos^2 x \, dx.$$

Solution If we let $u = \cos x$, then $du = -\sin x \, dx$ and

$$\sin^2 x \cos^2 x = (1 - \cos^2 x) \cos^2 x = (1 - u^2)u^2.$$

Applying the substitution method gives

$$\begin{aligned}\int \sin^3 x \cos^2 x \, dx &= \int (1 - u^2)u^2(-du) \\ &= \frac{1}{5}u^5 - \frac{1}{3}u^3 + C = \frac{1}{5}\cos^5 x - \frac{1}{3}\cos^3 x + C.\end{aligned}$$

3. Calculate

$$\int \sin^2 x \cos^2 x \, dx.$$

Solution By the double-angle formula, we write the integrand as

$$\begin{aligned}\sin^2 x \cos^2 x &= \frac{1}{2}(1 - \cos 2x) \cdot \frac{1}{2}(1 + \cos 2x) \\ &= \frac{1}{4}(1 - \cos^2 2x) \\ &= \frac{1}{4}[1 - \frac{1}{2}(1 + \cos 4x)] \\ &= \frac{1}{8}(1 - \cos 4x).\end{aligned}$$

Thus,

$$\begin{aligned}\int \sin^2 x \cos^2 x \, dx &= \int \frac{1}{8}(1 - \cos 4x) \, dx \\ &= \frac{1}{8}x - \frac{1}{32}\sin 4x + C.\end{aligned}$$

4.59 Method: integrating products of powers of $\sin x$ and $\cos x$

For integrals of the form

$$\int \sin x^m x \cos^n x \, dx,$$

one can integrate by the following approach:

1. When m odd, n real: split off $\sin x$, rewrite even power of $\sin x$ in terms of $\cos x$, then substitute $u = \cos x$.
2. When n odd, m real: split off $\cos x$, rewrite even power of $\cos x$ in terms of $\sin x$, then

substitute $u = \sin x$.

3. When m and n both even nonnegative integers: use half-angle identities to transform the integrand into a polynomial in terms of $\cos 2x$, and apply previous strategies.

4.60 Theorem: the universal trigonometric substitution

Consider any integral of the form

$$\int R(\sin x, \cos x) dx,$$

where R is rational in $\sin x$ and $\cos x$. Then the universal trigonometric substitution

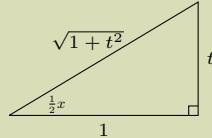
$$t = \tan\left(\frac{1}{2}x\right) \quad \text{or} \quad x = 2\tan^{-1} t$$

converts the integral into some integral of a rational function of t .

Justification:

For the given substitution $t = \tan\left(\frac{1}{2}x\right)$, from the triangle below, we have

$$\begin{aligned}\sin\left(\frac{1}{2}x\right) &= \frac{t}{\sqrt{1+t^2}}, \\ \cos\left(\frac{1}{2}x\right) &= \frac{1}{\sqrt{1+t^2}},\end{aligned}$$



so that

$$\begin{aligned}\sin x &= 2 \sin\left(\frac{1}{2}x\right) \cos\left(\frac{1}{2}x\right) = 2 \frac{t}{\sqrt{1+t^2}} \cdot \frac{1}{\sqrt{1+t^2}} = \frac{2t}{1+t^2}, \\ \cos x &= 2 \cos^2\left(\frac{1}{2}x\right) - 1 = 2 \cdot \frac{1}{1+t^2} - 1 = \frac{1-t^2}{1+t^2}.\end{aligned}$$

Thus, under the substitution, the function $R(\sin x, \cos x)$ becomes

$$R\left(\frac{2t}{1+t^2}, \frac{1-t^2}{1+t^2}\right),$$

which is a rational function in t . Furthermore, since

$$dt = \frac{1}{2} \sec^2\left(\frac{1}{2}x\right) dx \quad \text{or} \quad dx = \frac{2}{1+t^2} dt,$$

we get

$$\int R(\sin x, \cos x) dx = \int R\left(\frac{2t}{1+t^2}, \frac{1-t^2}{1+t^2}\right) \cdot \frac{2}{1+t^2} dt.$$

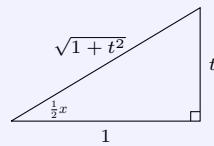
This shows that the universal trigonometric substitution converts the given integral into some integral of a rational function of t .

4.61 Example: integrating the secant function

Calculate $\int \sec x \, dx$.

Solution We notice that if we make the substitution $t = \tan(\frac{1}{2}x)$, then, from the triangle below, we have

$$\cos(\frac{1}{2}x) = \frac{1}{\sqrt{1+t^2}},$$



so that

$$\begin{aligned}\sec x &= \frac{1}{\cos x} = \frac{1}{2\cos^2(\frac{1}{2}x) - 1} \\ &= \frac{1}{2 \cdot \frac{1}{1+t^2} - 1} = \frac{1+t^2}{1-t^2}.\end{aligned}$$

It is easy to have

$$dt = \frac{1}{2} \sec^2(\frac{1}{2}x) dx \quad \text{or} \quad dx = \frac{2}{1+t^2} dt.$$

Thus, by applying the substitution rule, we get

$$\begin{aligned}\int \sec x \, dx &= \int \frac{1+t^2}{1-t^2} \cdot \frac{2}{1+t^2} dt \\ &= \int \frac{2}{1-t^2} dt. \quad \text{rational integrand}\end{aligned}$$

For the last integral, we write the integrand as its form of the partial fraction decomposition:

$$\frac{2}{1-t^2} = \frac{1}{1-t} + \frac{1}{1+t}.$$

Hence,

$$\begin{aligned}\int \sec x \, dx &= \int \left(\frac{1}{1-t} + \frac{1}{1+t} \right) dt \\ &= -\ln|1-t| + \ln|1+t| + C \\ &= \ln \left| \frac{1+t}{1-t} \right| + C = \ln \left| \frac{1+\tan(\frac{1}{2}x)}{1-\tan(\frac{1}{2}x)} \right| + C.\end{aligned}$$

Rational Functions

4.62 Example: partial fractions for rational functions

1. Calculate the integral $\int \left(\frac{1}{x-2} + \frac{1}{x+1} \right) dx$.

We can use substitution method.

$$\begin{aligned}\int \left(\frac{1}{x-2} + \frac{1}{x+1} \right) dx &= \int \frac{dx}{x-2} + \int \frac{dx}{x+1} && \text{splitting the integral} \\ &= \int \frac{du}{u} + \int \frac{dv}{v} && \text{letting } u = x - 2 \text{ and } v = x + 1 \\ &= \ln|u| + \ln|v| + C && \text{evaluating integrals} \\ &= \ln|x-2| + \ln|x+1| + C. && \text{converting variables back to } x\end{aligned}$$

2. Calculate the integral $\int \frac{2x-1}{x^2-x-2} dx$.

We notice that

$$\frac{2x-1}{x^2-x-2} = \frac{1}{x-2} + \frac{1}{x+1},$$

so that the integral equals the previous one.

It seems a mystery to have the identity. However, there is an algorithm to derive this, as demonstrated below.

Step 1 Factorize the denominator:

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

Step 2 Write the partial fraction decomposition with coefficients to be determined:

$$\frac{2x-1}{x^2-x-2} = \frac{A}{x-2} + \frac{B}{x+1}.$$

Step 3 Clear the denominators:

$$\begin{aligned}\frac{2x-1}{x^2-x-2} &= \frac{A}{x-2} + \frac{B}{x+1} \\ &= \frac{A(x+1) + B(x-2)}{x^2-x-2} = \frac{(A+B)x + A - 2B}{x^2-x-2}.\end{aligned}$$

Step 4 Solve for A and B by letting numerators of both sides equal:

$$\begin{aligned}2x-1 &= (A+B)x + A - 2B \implies A+B = 2, \quad A-2B = -1 \\ &\implies A = 1, \quad B = 1.\end{aligned}$$

The above process is a special case of the method of partial fractions.

4.63 Remark: integrals of rational functions

There are two facts worth knowing:

1. Any real rational function can be expressed as partial fractions;
2. Any partial fraction can be integrated.

Both can be effectively implemented on computers symbolically.

4.64 Method: integrating rational functions by partial fractions

For an improper rational function $P(x)/Q(x)$, where P and Q are polynomials with $\deg P \geq \deg Q$, 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 integrate proper rational functions.

For any proper rational function, we can integrate it by using partial fractions in one of the four cases:

1. The denominator Q has only distinct real roots.

In this case, we write

$$Q(x) = A(x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_k).$$

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

$$\frac{R(x)}{Q(x)} = \frac{A_1}{x - \lambda_1} + \frac{A_2}{x - \lambda_2} + \cdots + \frac{A_k}{x - \lambda_k}.$$

For each fraction, we can make a substitution $t = x - \lambda$ to integrate:

$$\int \frac{1}{x - \lambda} dx = \int \frac{1}{t} dt = \ln |t| + C.$$

2. The denominator Q is a product of linear factors and some of them are repeated.

Suppose the function $(x - \lambda)$ is repeated r times. Then instead of the single term $\frac{A_1}{x - \lambda}$, one needs to use the form

$$\frac{A_1}{x - \lambda} + \frac{A_2}{(x - \lambda)^2} + \cdots + \frac{A_r}{(x - \lambda)^r}.$$

For each fraction of the form $\frac{1}{(x - \lambda)^m}$, 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 $t = x - \lambda$, we get

$$\int \frac{1}{(x - \lambda)^m} dx = \int \frac{1}{t^m} dt = \frac{t^{-m+1}}{-m+1} + C.$$

- 3.** The denominator Q contains irreducible quadratic factors and none of them is repeated.

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

$$\frac{Ax + B}{x^2 + ax + b}$$

whose integration can be re-written as

$$\begin{aligned} \int \frac{Ax + B}{x^2 + ax + b} dx &= \int \frac{\frac{1}{2}A(2x + a) + (B - \frac{1}{2}aA)}{x^2 + ax + b} dx \\ &= \int \frac{\frac{1}{2}A(x^2 + ax + b)' + (B - \frac{1}{2}aA)}{x^2 + ax + b} dx \\ &= \int \frac{\frac{1}{2}A(x^2 + ax + b)'}{x^2 + ax + b} dx + \int \frac{(B - \frac{1}{2}aA)}{x^2 + ax + b} dx. \end{aligned}$$

For the last two integrals on the right-hand side, they can be computed by using different substitution. In fact, for the first integral, by using the substitution $t = x^2 + ax + b$, it can be reduced to

$$\int \frac{dt}{t} = \ln|t| + C;$$

for the second integral, by using the substitution $t = x + \frac{1}{2}a$, we can reduce the last integral to the form

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

- 4.** The denominator Q contains a repeated irreducible quadratic factor.

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

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

As in Case 3, we re-write the integral of a typical term as

$$\begin{aligned} &\int \frac{Ax + B}{(x^2 + ax + b)^m} dx \\ &= \int \frac{\frac{1}{2}A(x^2 + ax + b)'}{(x^2 + ax + b)^m} dx + \int \frac{(B - \frac{1}{2}aA)}{(x^2 + ax + b)^m} dx. \end{aligned}$$

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, by using the substitution $t = x^2 + ax + b$, it can be reduced to

$$\int \frac{dt}{t^m} = -\frac{1}{(m-1)t^{m-1}} + C;$$

for the second integral, by using the substitution $t = x + \frac{1}{2}a$, we can reduce it to the form

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

By integrating by parts, we have

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

So, we have the recursive formula

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

Using this formula recursively reduces the integral to the integral

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

4.65 Example: partial fractions for various rational functions

- Calculate the integral $\int \frac{x^6}{x^2 + 1} dx$.

Solution The integrand $\frac{x^6}{x^2 + 1}$ is an improper rational function. By long division, we have

$$\frac{x^6}{x^2 + 1} = (x^4 - x^2 + 1) + \frac{-1}{x^2 + 1}.$$

The denominator $x^2 + 1$ is an irreducible quadratic polynomial, so the term $\frac{-1}{x^2 + 1}$ is in the form of partial fraction already. Hence,

$$\begin{aligned} \int \frac{x^6}{x^2 + 1} dx &= \int \left[(x^4 - x^2 + 1) + \frac{-1}{x^2 + 1} \right] dx \\ &= \int (x^4 - x^2 + 1) dx - \int \frac{1}{x^2 + 1} dx \\ &= \frac{1}{5}x^5 - \frac{1}{3}x^3 + x - \tan^{-1} x = C. \end{aligned}$$

2. Calculate the integral $\int \frac{2x^2 + 1}{(x+1)(x-1)^3} dx$.

Solution We express the proper rational integrand as its partial fraction decomposition:

$$\frac{2x^2 + 1}{(x+1)(x-1)^3} = \frac{A_1}{x+1} + \frac{B_1}{x-1} + \frac{B_2}{(x-1)^2} + \frac{B_3}{(x-1)^3}.$$

By clearing the denominators, we get

$$2x^2 + 1 = A_1(x-1)^3 + B_1(x+1)(x-1)^2 + B_2(x+1)(x-1) + B_3(x+1).$$

So,

$$x = -1 : \quad 3 = A_1 \cdot (-2)^3 \implies A_1 = -\frac{3}{8};$$

$$x = 1 : \quad 3 = B_3 \cdot 2 \implies B_3 = \frac{3}{2};$$

$$\text{differentiating then } x = 1 : \quad 4 = B_2 \cdot 2 + B_3 \implies B_2 = \frac{5}{4};$$

$$\text{coefficients of } x^3 : \quad 0 = A_1 + B_1 \implies B_1 = \frac{3}{8}.$$

Thus, we have the partial fraction decomposition:

$$\frac{2x^2 + 1}{(x+1)(x-1)^3} = -\frac{3}{8} \cdot \frac{1}{x+1} + \frac{3}{8} \cdot \frac{1}{x-1} + \frac{5}{4} \cdot \frac{1}{(x-1)^2} + \frac{3}{2} \cdot \frac{1}{(x-1)^3}.$$

Hence,

$$\begin{aligned} & \int \frac{2x^2 + 1}{(x+1)(x-1)^3} dx \\ &= \int \left[-\frac{3}{8} \cdot \frac{1}{x+1} + \frac{3}{8} \cdot \frac{1}{x-1} + \frac{5}{4} \cdot \frac{1}{(x-1)^2} + \frac{3}{2} \cdot \frac{1}{(x-1)^3} \right] dx \\ &= -\frac{3}{8} \ln|x+1| + \frac{3}{8} \ln|x-1| - \frac{5}{4} \cdot \frac{1}{x-1} - \frac{3}{4} \cdot \frac{1}{(x-1)^2} + C. \end{aligned}$$

3. Calculate the integral $\int \frac{x}{x^3 + 1} dx$.

Solution We express the proper rational integrand as its partial fraction decomposition:

$$\frac{x}{x^3 + 1} = \frac{x}{(x+1)(x^2 - x + 1)} = \frac{A_1}{x+1} + \frac{B_1x + C_1}{x^2 - x + 1}.$$

By clearing the denominators, we get

$$x = A_1(x^2 - x + 1) + (B_1x + C_1)(x + 1).$$

So,

$$x = -1 : \quad -1 = A_1 \cdot 3 \implies A_1 = -\frac{1}{3};$$

$$x = 0 : \quad 0 = A_1 + C_1 \implies C_1 = \frac{1}{3};$$

$$\text{coefficients of } x^2 : \quad 0 = A_1 + B_1 \implies B_1 = \frac{1}{3}.$$

Thus, we have the partial fraction decomposition:

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

Note that $(x^2 - x + 1)' = 2x - 1$, so we further write the integrand as

$$\frac{x}{x^3 + 1} = -\frac{1}{3} \cdot \frac{1}{x+1} + \frac{\frac{1}{6}(2x-1) + \frac{1}{2}}{x^2 - x + 1}.$$

Hence,

$$\begin{aligned} & \int \frac{x}{x^3 + 1} dx \\ &= \int \left[-\frac{1}{3} \cdot \frac{1}{x+1} + \frac{\frac{1}{6}(2x-1) + \frac{1}{2}}{x^2 - x + 1} \right] dx \\ &= -\frac{1}{3} \ln|x+1| + \frac{1}{6} \ln(x^2 - x + 1) + \frac{1}{2} \int \frac{dx}{x^2 - x + 1}. \end{aligned}$$

The last integral can be integrated as

$$\begin{aligned} \int \frac{dx}{x^2 - x + 1} &= \int \frac{dx}{(x - \frac{1}{2})^2 + \frac{3}{4}} && \text{completing square} \\ &= \int \frac{\sqrt{\frac{3}{4}} \sec^2 t dt}{\frac{3}{4} \sec^2 t} && x - 1 = \sqrt{\frac{3}{4}} \tan t \\ &= \int \sqrt{\frac{4}{3}} dt \\ &= \sqrt{\frac{4}{3}} t + C = \sqrt{\frac{4}{3}} \tan^{-1} \sqrt{\frac{4}{3}}(x-1) + C. \end{aligned}$$

Therefore,

$$\int \frac{x}{x^3 + 1} dx = -\frac{1}{3} \ln|x+1| + \frac{1}{6} \ln(x^2 - x + 1) + \frac{1}{\sqrt{3}} \tan^{-1} \frac{2x-1}{\sqrt{3}} + C.$$

4. Calculate the integral $\int \frac{x^2 + 1}{x^4 + 1} dx$.

Solution The denominator $x^4 + 1$ can be factored as a product of irreducible quadratic factors:

$$\begin{aligned} x^4 + 1 &= (x^4 + 2x^2) + 1 - 2x^2 \\ &= (x^2 + 1)^2 - (\sqrt{2}x)^2 \\ &= (x^2 - \sqrt{2}x + 1)(x^2 + \sqrt{2}x + 1). \end{aligned}$$

Thus, we can express the proper rational integrand as its partial fraction decomposition:

$$\frac{x^2 + 1}{x^4 + 1} = \frac{A_1 x + B_1}{x^2 - \sqrt{2}x + 1} + \frac{A_2 x + B_2}{x^2 + \sqrt{2}x + 1}.$$

By clearing the denominators, we get

$$x^2 + 1 = (A_1x + B_1)(x^2 + \sqrt{2}x + 1) + (A_2x + B_2)(x^2 - \sqrt{2}x + 1).$$

So,

$$x = 0 : \quad 1 = B_1 + B_2;$$

$$\text{coefficients of } x^3 : \quad 0 = A_1 + A_2;$$

$$\text{coefficients of } x : \quad 0 = A_1 + \sqrt{2}B_1 + A_2 - \sqrt{2}B_2;$$

$$\text{coefficients of } x^2 : \quad 1 = B_1 + \sqrt{2}A_1 + B_2 - \sqrt{2}A_2.$$

Solving the system gives

$$A_1 = 0, \quad A_2 = 0, \quad B_1 = \frac{1}{2}, \quad B_2 = \frac{1}{2}.$$

Thus, we have the partial fraction decomposition:

$$\frac{x^2 + 1}{x^4 + 1} = \frac{\frac{1}{2}}{x^2 - \sqrt{2}x + 1} + \frac{\frac{1}{2}}{x^2 + \sqrt{2}x + 1}.$$

Since

$$\begin{aligned} \int \frac{dx}{x^2 - \sqrt{2}x + 1} &= \int \frac{dx}{(x - \frac{1}{\sqrt{2}})^2 + \frac{1}{2}} && \text{completing square} \\ &= \int \frac{\frac{1}{\sqrt{2}} \sec^2 t dt}{\frac{1}{2} \sec^2 t} && x - \frac{1}{\sqrt{2}} = \frac{1}{\sqrt{2}} \tan t \\ &= \int \sqrt{2} dt \\ &= \sqrt{2}t + C = \sqrt{2} \tan^{-1}(\sqrt{2}x - 1) + C, \end{aligned}$$

similarly,

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

Therefore,

$$\int \frac{x^2 + 1}{x^4 + 1} dx = \frac{1}{\sqrt{2}} \tan^{-1}(\sqrt{2}x - 1) + \frac{1}{\sqrt{2}} \tan^{-1}(\sqrt{2}x + 1) + C.$$

4.66 MATLAB: partial fractions for rational functions

For integrating rational functions, one important step is to perform the partial fraction decomposition for the rational functions. In MATLAB, the partial fraction decomposition can be executed by using the command `partfrac`. The following MATLAB code shows how to perform the partial fraction decomposition for the rational functions $\frac{2x^2 + 1}{(x + 1)(x - 1)^3}$ and

$\frac{x}{x^3 + 1}$, respectively.

```
>> syms x
>> f(x) = (2*x^2 + 1)/((x + 1)*(x - 1)^3);
>> partfrac(f(x),x)

ans =

3/(8*(x - 1)) - 3/(8*(x + 1)) + 5/(4*(x - 1)^2) + 3/(2*(x - ...
1)^3)

>> g(x) = x/(x^3 + 1);
>> partfrac(g(x),x)

ans =

(x/3 + 1/3)/(x^2 - x + 1) - 1/(3*(x + 1))
```

Based on the above results of the partial fraction decomposition, the integrals of $\int \frac{2x^2+1}{(x+1)(x-1)^3} dx$ and $\int \frac{x}{x^3+1} dx$ can be solved by utilizing some integration methods such as the substitution method.

4.7 Numerical Approximation of Integrals

4.67 Example: numerical integration

Consider the integral $\int_0^1 e^{x^2} dx$.

We know that there is a function $F(x)$ whose derivative is e^{x^2} and the integral above equals $F(1) - F(0)$. However, there is no explicit expression of $F(x)$ in terms of known functions such as polynomials, trigonometric functions and their inverses, and transcendental functions and their inverses. We can use the definition of definite integrals to approximate the integral.

Recall the definition of definite integrals. Let $f(x) = e^{x^2}$. Divide the interval $[0, 1]$ into n equal subintervals:

$$x_0 = 0, x_1 = \frac{1}{n}, \dots, x_{n-1} = \frac{n-1}{n}, x_n = \frac{n}{n} = 1.$$

To approximate the integral numerically, one can use the Riemann sum of left end points:

$$\int_0^1 e^{x^2} dx \approx L_n = [f(x_0) + f(x_1) + \dots + f(x_{n-1})] \cdot \frac{1}{n},$$

or the Riemann sum of right end points:

$$\int_0^1 e^{x^2} dx \approx R_n = [f(x_1) + f(x_2) + \cdots + f(x_n)] \cdot \frac{1}{n}.$$

Or, one can use the Riemann sum of middle points. In fact, if we let

$$x_1^* = \frac{x_0 + x_1}{2} = \frac{1}{2n}, \quad x_2^* = \frac{x_1 + x_2}{2} = \frac{3}{2n}, \quad \dots, \quad x_n^* = \frac{x_{n-1} + x_n}{2} = \frac{2n-1}{2n}$$

be the middle points of each subinterval, then the Riemann sum of middle points is

$$M_n = [f(x_1^*) + f(x_2^*) + \cdots + f(x_n^*)] \cdot \frac{1}{n}.$$

For instance, if one takes $n = 10$, the approximation gives:

$$\begin{aligned} \int_0^1 e^{x^2} dx &\approx M_{10} = [f(\frac{1}{20}) + f(\frac{3}{20}) + \cdots + f(\frac{19}{20})] \cdot \frac{1}{10} \\ &= (e^{0.0025} + e^{0.0225} + \cdots + e^{0.9025}) \cdot 0.1 \\ &\approx 1.460393. \end{aligned}$$

Or one can approximate the region under the graph of the function f on $[x_{i-1}, x_i]$ by a trapezoidal region, equivalently, the average of Riemann sums of left end points and right end points:

$$T_n = \frac{L_n + R_n}{2} = \frac{1}{2n} [f(x_0) + 2f(x_1) + \cdots + 2f(x_{n-1}) + f(x_n)].$$

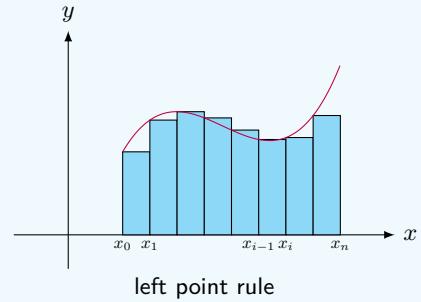
4.68 Method: numerical integration

Let f be integrable function over $[a, b]$. Divide the interval $[a, b]$ into n equal subintervals

$$x_0 = a, \quad x_1 = a + \frac{b-a}{n}, \quad \dots, \quad x_{n-1} = a + \frac{(n-1)(b-a)}{n}, \quad x_n = b.$$

- The **left point rule** is the Riemann sum of left end points of f on $[a, b]$. The region under the graph of the function f on $[x_{i-1}, x_i]$ is approximated by a rectangle.

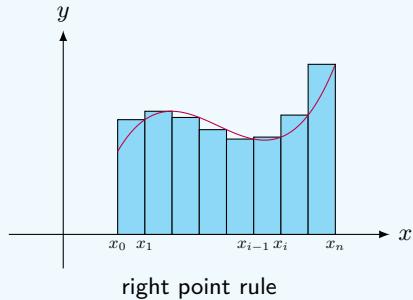
$$\begin{aligned} \int_a^b f(x) dx &\approx L_n \\ &= [f(x_0) + f(x_1) + \cdots + f(x_{n-1})] \cdot \frac{b-a}{n}. \end{aligned}$$



- The **right point rule** is the Riemann sum of right end points of f on $[a, b]$. The region under the graph of the function f on $[x_{i-1}, x_i]$ is approximated by a rectangle.

$$\int_a^b f(x) dx \approx R_n$$

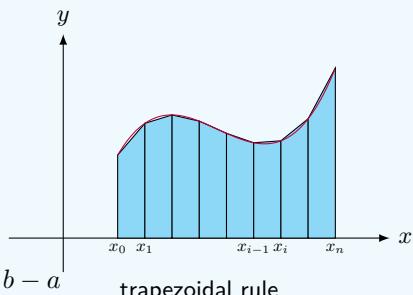
$$= [f(x_1) + f(x_2) + \cdots + f(x_n)] \cdot \frac{b-a}{n}.$$



- The **trapezoidal rule** is the average of the Riemann sums of left end points and right end points. The region under the graph of the function f on $[x_{i-1}, x_i]$ is approximated by a trapezoid.

$$\int_a^b f(x) dx \approx T_n$$

$$= [f(x_0) + 2f(x_1) + \cdots + 2f(x_{n-1}) + f(x_n)] \cdot \frac{b-a}{2n}.$$

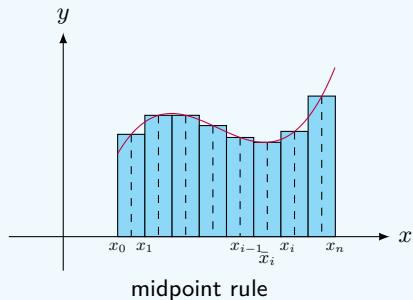


- The **midpoint rule** is the Riemann sum of middle points of f on $[a, b]$. The region under the graph of the function f on $[x_{i-1}, x_i]$ is approximated by a rectangle.

$$\int_a^b f(x) dx \approx M_n$$

$$= [f(x_1^*) + f(x_2^*) + \cdots + f(x_n^*)] \cdot \frac{b-a}{n},$$

where $x_i^* = \frac{x_{i-1} + x_i}{2}$.



4.69 Example: various numerical integrations and their errors

We use various rules to estimate the integral

$$I = \int_0^1 \frac{dx}{1+x^2}.$$

Take $n = 4$. The partition gives four subintervals

$$[0, 0.25], [0.25, 0.5], [0.5, 0.75], [0.75, 1].$$

Each subinterval has length 0.25. The values at the partition points are

$$\begin{aligned}y_0 &= 1.000000, & y_1 &= 0.941176, & y_2 &= 0.800000, \\y_3 &= 0.640000, & y_4 &= 0.500000.\end{aligned}$$

Then we get the approximate values of the integral I according to various rules

$$\begin{aligned}R_4 &= 0.25 \times (0.941176 + 0.800000 + 0.640000 + 0.500000) \approx 0.720294, \\L_4 &= 0.25 \times (1.000000 + 0.941176 + 0.800000 + 0.640000) \approx 0.845294, \\T_4 &= \frac{L_4 + R_4}{2} = \frac{0.25}{2} \times (1.000000 + 2 \times 0.941176 + 2 \times 0.800000 \\&\quad + 2 \times 0.640000 + 0.500000) \\&\approx 0.782794.\end{aligned}$$

The middle points are

$$0.125 < 0.375 < 0.625 < 0.875,$$

and the values at the middle points are

$$\bar{y}_1 = 0.984615, \quad \bar{y}_2 = 0.876712, \quad \bar{y}_3 = 0.719101, \quad \bar{y}_4 = 0.566372.$$

Then the midpoint rule gives

$$M_4 = 0.25 \times (0.984615 + 0.876712 + 0.719101 + 0.566372) \approx 0.786700.$$

If we double the number of partition points by taking $n = 8$, then the length of each subinterval is 0.125, and we have

i	x_i	y_i	\bar{x}_i	\bar{y}_i
0	0	1.000000		
1	0.125	0.984615	0.0625	0.996109
2	0.250	0.941176	0.1875	0.966038
3	0.375	0.876712	0.3125	0.911032
4	0.500	0.800000	0.4375	0.839344
5	0.625	0.719101	0.5625	0.759644
6	0.750	0.640000	0.6875	0.679045
7	0.875	0.566372	0.8125	0.602353
8	1.000	0.500000	0.9375	0.532225

Then we get the approximations

$$L_8 \approx 0.815997, \quad R_8 \approx 0.753497, \quad T_8 \approx 0.784747, \quad M_8 \approx 0.785721.$$

We actually know the exact value of the integral

$$I = \int_0^1 \frac{dx}{1+x^2} = \frac{\pi}{4} \approx 0.785398.$$

We may compare the approximate values with the actual value in the following table.

error	$n = 4$	$n = 8$
$I - L_n$	-0.059896	-0.030599
$I - R_n$	0.065104	0.031901
$I - T_n$	0.002604	0.000651
$I - M_n$	-0.001302	-0.000323

4.70 Theorem: error bounds of midpoint and trapezoidal rules

Suppose f'' is continuous and $|f''(x)| \leq K$ on $[a, b]$. Then

$$\left| \int_a^b f(x) dx - T_n \right| \leq \frac{K(b-a)^3}{12n^2}, \quad \left| \int_a^b f(x) dx - M_n \right| \leq \frac{K(b-a)^3}{24n^2}.$$

4.71 Remark: error bounds

We can make the following observations from the theorem.

- Both midpoint rule and trapezoidal rule produce better approximation as n increases.
- Generally, the error of the midpoint rule is about half of the error of the trapezoidal rule.

4.72 Example: error bound of trapezoidal rule

Consider the integral

$$\int_0^1 2 \cos x dx.$$

How large n should we take so that trapezoidal rule is accurate within 0.0001?

Since $|(2 \cos x)''| = |-2 \cos x| \leq 2$, from the error bound, we know that the number n should satisfy

$$\frac{2}{12n^2} < 0.0001,$$

so that $n > \frac{1}{\sqrt{0.0006}} \approx 40.8$.

Hence, if we take an integer $n \geq 41$, then the trapezoidal rule will yield a numerical value that approximates the given integral with error smaller than 0.0001.

4.73 Method: Simpson's Rule

Let f be integrable function over $[a, b]$. For an even integer n , divide the interval $[a, b]$ into n equal subintervals

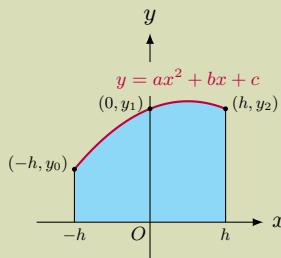
$$x_0 = a, \quad x_1 = a + \frac{b-a}{n}, \quad \dots, \quad x_{n-1} = a + \frac{(n-1)(b-a)}{n}, \quad x_n = b.$$

Simpson's rule is the approximation to the integral $\int_a^b f(x) dx$:

$$\int_a^b f(x) dx \approx S_n = \frac{b-a}{3n} \cdot [f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + \dots + 2f(x_{n-2}) + 4f(x_{n-1}) + f(x_n)].$$

Justification:

(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

$$\begin{aligned} A &= \int_{-h}^h (ax^2 + bx + c) dx \\ &= \left[\frac{1}{3}ax^3 + \frac{1}{2}bx^2 + cx \right]_{x=-h}^h \\ &= \frac{2}{3}ah^3 + 2ch = \frac{1}{3}h(2ah^2 + 6c). \end{aligned}$$

(2) Since the parabola passes through $(-h, y_0)$, $(0, y_1)$, and (h, y_2) , we have

$$\begin{aligned} y_0 &= a(-h)^2 + b(-h) + c, \\ y_1 &= c, \\ y_2 &= ah^2 + bh + c. \end{aligned}$$

It is easy to see that

$$y_0 + 4y_1 + y_2 = [a(-h)^2 + b(-h) + c] + 4c + (ah^2 + bh + c) = 2ah^2 + 6c,$$

which is the same the area under the parabola, so

$$A = \frac{1}{3}h(y_0 + 4y_1 + y_2).$$

This formula shows that the area under the parabola depends only on the values y_0, y_1, y_2 , and h . It is clear that shifting this parabola horizontally does not change the area under it.

(3) For an integral $\int_a^b f(x) dx$, we divide $[a, b]$ into an *even* number n of subintervals of equal length $\Delta x = \frac{b-a}{n}$ by using $n+1$ points

$$x_0 = a, \quad x_1 = a + \Delta x, \quad x_2 = a + 2\Delta x, \quad \dots, \quad x_n = a + n\Delta x = b.$$

For each consecutive pair of subintervals, we apply the area under the parabola. Adding all

the areas gives Simpson's rule:

$$\begin{aligned}\int_a^b f(x) dx &\approx \frac{1}{3}h(y_0 + 4y_1 + y_2) + \frac{1}{3}h(y_2 + 4y_3 + y_4) \\ &\quad + \cdots + \frac{1}{3}h(y_{n-2} + 4y_{n-1} + y_n) \\ &= \frac{1}{3}h(y_0 + 4y_1 + 2y_2 + 4y_3 + 2y_4 + \cdots + 2y_{n-2} + 4y_{n-1} + y_n) \\ &= S_n.\end{aligned}$$

4.74 Theorem: error bound of Simpson's rule

Suppose $f^{(4)}$ is continuous and $|f^{(4)}(x)| \leq K$ on $[a, b]$. Then

$$\left| \int_a^b f(x) dx - S_n \right| \leq \frac{K(b-a)^5}{180n^4}.$$

4.75 Remark: Simpson's rule

1. In Simpson's rule, the integrand f is approximated by a piecewise quadratic function. In comparison, in the left point rule, the right point rule, the trapezoidal rule, or the midpoint rule, the function f is approximated by a piecewise linear function.
2. Simpson's rule is related to other rules by

$$S_{2n} = \frac{1}{3}T_n + \frac{2}{3}M_n.$$

3. The error of Simpson's rule is smaller than the error in the trapezoidal rule. In other words, Simpson's rule generally yields a better result in evaluating integrals.

4.76 Example: error bound of Simpson's rule

Consider the integral

$$\int_0^1 2 \cos x dx.$$

How large n should we take so that Simpson's rule is accurate to within 0.0001?

Thus $|(2 \cos x)^{(4)}| = |2 \cos x| \leq 2$. The number n should satisfy

$$\frac{2}{180n^4} < 0.0001,$$

so that $n > \frac{1}{\sqrt[4]{0.009}} \approx 3.25$.

Hence, if we take an integer $n \geq 4$, then the Simpson's rule will yield a numerical value that approximates the given integral with error smaller than 0.0001.

4.77 MATLAB: left point rule

The left-point rule of approximating an integral can be executed in MATLAB via programming. In particular, the following MATLAB code shows the calculation of $\int_0^1 \frac{dx}{1+x^2}$ using the left-point rule.

```

Ln = 0;
a = 0; % interval start
b = 1; % interval end
n = 4; % number of subintervals
itv = (b - a)/n; % subinterval length

x = linspace(a,b,n + 1); % n subintervals, n + 1 points
y = 1./(1 + x.^2);

for k = 1:n % y value index
    Ln = Ln + itv*y(k);
end

```

In the MATLAB code above, the interval $[0, 1]$ is partitioned into 4 subintervals. After executing the above MATLAB code, the result can be seen by typing L_n in the command window, as shown in the following.

```

>> Ln
Ln =
0.8453

```

4.78 MATLAB: right point rule

The right-point rule of approximating an integral can be executed in MATLAB via programming. In particular, the following MATLAB code shows the calculation of $\int_0^1 \frac{dx}{1+x^2}$ using the right-point rule.

```

Rn = 0;
a = 0; % interval start
b = 1; % interval end
n = 4; % number of subintervals
itv = (b - a)/n; % subinterval length

x = linspace(a,b,n + 1); % n subintervals, n + 1 points

```

```

y = 1./(1 + x.^2);

for k = 1:n % y value index
    Rn = Rn + itv*y(k+1);
end

```

In the MATLAB code above, the interval $[0, 1]$ is partitioned into 4 subintervals. After executing the above MATLAB code, the result can be seen by typing R_n in the command window, as shown in the following.

```

>> Rn

Rn =
0.7203

```

4.79 MATLAB: midpoint rule

The midpoint rule of approximating an integral can be executed in MATLAB via programming. In particular, the following MATLAB code shows the calculation of $\int_0^1 \frac{dx}{1+x^2}$ using the midpoint rule.

```

Mn = 0;
a = 0; % interval start
b = 1; % interval end
n = 4; % number of subintervals
itv = (b - a)/n; % subinterval length

x = linspace(a,b,n + 1); % n subintervals, n + 1 points

for k = 1:n % y value index
    x_m = (x(k) + x(k + 1))/2; % midpoint
    y_m = 1/(1 + x_m.^2); % y value at midpoint
    Mn = Mn + itv*y_m;
end

```

In the MATLAB code above, the interval $[0, 1]$ is partitioned into 4 subintervals. After executing the above MATLAB code, the result can be seen by typing M_n in the command window, as shown in the following.

```

>> Mn

Mn =

```

0.7867

4.80 MATLAB: trapezoidal rule

In MATLAB, the command `trapz` can be used to compute the approximation of an integral using the trapezoidal rule. The following MATLAB code shows an example of how to use `trapz` to approximate $\int_0^1 \frac{dx}{1+x^2}$.

```
>> x = 0:0.25:1; % 4 subintervals with length 0.25 each
>> y = 1./(1 + x.^2);
>> T4 = trapz(x,y)

T4 =
0.7828
```

In the MATLAB code above, the interval $[0, 1]$ is partitioned into 4 subintervals of equal length 0.25. Note that if there is only one input in `trapz`, i.e., the command is written as `trapz(y)`, then the length of the subinterval (or the gap between two neighbouring points) will be set to the default value 1.

4.81 MATLAB: Simpson's rule

The following MATLAB code shows the calculation of $\int_0^1 \frac{dx}{1+x^2}$ using Simpson's rule.

```
Sn = 0;
a = 0; % interval start
b = 1; % interval end
n = 4; % even number of subintervals
itv = (b - a)/n; % subinterval length

x = linspace(a,b,n + 1); % n subintervals, n + 1 points
y = 1./(1 + x.^2);

for k = 1:2:(n-1) % y value index
    sum_term = y(k) + 4*y(k + 1) + y(k + 2);
    Sn = Sn + itv*sum_term/3;
end
```

In the MATLAB code above, the interval $[0, 1]$ is partitioned into 4 subintervals. After executing the above MATLAB code, the result can be seen by typing S_n in the command window, as shown in the following.

```
>> Sn
Sn =
0.7854
```

4.82 Engineering Example: error function

The error function is an important function in statistics, applied mathematics, physics, etc. For example, it is useful in determining the bit error rate (BER) of a digital communication system. The error function is defined as

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt.$$

Use the left point rule, the right point rule, the midpoint rule, the trapezoidal rule and the Simpson's rule, respectively, to approximate $\operatorname{erf}(3)$.

First, $\operatorname{erf}(3)$ can be written as

$$\operatorname{erf}(3) = \frac{2}{\sqrt{\pi}} \int_0^3 e^{-t^2} dt.$$

Take $n = 10$. The partition gives ten subintervals as follows:

$$\begin{aligned} & [0, 0.3], [0.3, 0.6], [0.6, 0.9], [0.9, 1.2], [1.2, 1.5], \\ & [1.5, 1.8], [1.8, 2.1], [2.1, 2.4], [2.4, 2.7], [2.7, 3]. \end{aligned}$$

Each subinterval has length 0.3. Let $y = f(t) = \frac{2}{\sqrt{\pi}} e^{-t^2}$. The values of $y = f(t)$ at the partition points are

$$\begin{aligned} y_0 &= 1.128379167095513, & y_1 &= 1.031260909618963, \\ y_2 &= 0.787243431714287, & y_3 &= 0.501968574240363, \\ y_4 &= 0.267344347003539, & y_5 &= 0.118930289223629, \\ y_6 &= 0.044191723332011, & y_7 &= 0.013715649999807, \\ y_8 &= 0.003555648680878, & y_9 &= 0.000769924759855, \\ y_{10} &= 0.000139253051947, \end{aligned}$$

where y_i denotes the value of $y = f(t)$ at the i th partition point. Then we get the approximate values of $\operatorname{erf}(3)$ according to various rules. In particular, by the left point rule, the right point

rule, the trapezoidal rule and the Simpson's rule, respectively, we have

$$\begin{aligned}L_{10} &= 0.3 \sum_{i=0}^9 y_i = 1.169207899700653, \\R_{10} &= 0.3 \sum_{i=1}^{10} y_i = 0.830735925487584, \\T_{10} &= \frac{0.3}{2} \sum_{i=0}^9 (y_i + y_{i+1}) = 0.999971912594118, \\S_{10} &= \frac{0.3}{3} (y_0 + 4y_1 + 2y_2 + 4y_3 + \cdots + 2y_8 + 4y_9 + y_{10}) \\&= 0.999977011297936.\end{aligned}$$

Before using the midpoint rule to approximate $\text{erf}(3)$, we first need to get the middle points of each interval:

$$\begin{aligned}x_1^* &= 0.15, \quad x_2^* = 0.45, \quad x_3^* = 0.75, \quad x_4^* = 1.05, \quad x_5^* = 1.35, \\x_6^* &= 1.65, \quad x_7^* = 1.95, \quad x_8^* = 2.25, \quad x_9^* = 2.55, \quad x_{10}^* = 2.85.\end{aligned}$$

The corresponding values of $y = f(t)$ at each middle point are

$$\begin{aligned}f(x_1^*) &= 1.103274126650824, \quad f(x_2^*) = 0.921532013012220, \\f(x_3^*) &= 0.642931069195207, \quad f(x_4^*) = 0.374666956970448, \\f(x_5^*) &= 0.182369986539023, \quad f(x_6^*) = 0.074146103374068, \\f(x_7^*) &= 0.025179684949841, \quad f(x_8^*) = 0.007142319022018, \\f(x_9^*) &= 0.001692213637679, \quad f(x_{10}^*) = 0.000334886877468.\end{aligned}$$

The approximation of $\text{erf}(3)$ by the midpoint rule is thus

$$M_{10} = 0.3 \sum_{i=1}^{10} f(x_i^*) = 0.999980808068639.$$

Actually, the exact value of $\text{erf}(3)$ can be calculated by MATLAB using command `erf`, as shown below:

```
>> y = erf(3)
y =
0.999977909503001
```

We may compare the approximate values with the actual value in the following table.

error	$n = 10$
$y - L_n$	-0.169229990197652
$y - R_n$	0.169241984015417
$y - T_n$	$5.996908883365570 \times 10^{-6}$
$y - S_n$	$8.982050653250084 \times 10^{-7}$
$y - M_n$	$-2.898565637621253 \times 10^{-6}$

4.8 Improper Integrals

4.83 Example: improper integrals over unbounded intervals

Consider the area of the graph of the function $f(x) = \frac{1}{x^2}$ above the x -axis over the interval $[1, +\infty)$. Since the interval $[1, +\infty)$ is unbounded, we can consider the area over a finite interval $[1, T]$, and let T get larger and larger to approximate the required area. We need to calculate

$$\int_1^T \frac{1}{x^2} dx = \left(-\frac{1}{x} \right) \Big|_1^T = 1 - \frac{1}{T}.$$

Thus the required area is

$$\lim_{T \rightarrow +\infty} \int_1^T \frac{1}{x^2} dx = \lim_{T \rightarrow +\infty} \left(1 - \frac{1}{T} \right) = 1.$$

We can use the following notation to denote this area, called an improper integral

$$\int_1^{+\infty} \frac{1}{x^2} dx = \lim_{T \rightarrow +\infty} \int_1^T \frac{1}{x^2} dx.$$

If we want to know the area of the graph of the function $f(x) = \frac{1}{x}$ above the x -axis over the interval $[1, +\infty)$, we need to calculate the improper integral

$$\int_1^{+\infty} \frac{1}{x} dx = \lim_{T \rightarrow +\infty} \int_1^T \frac{1}{x} dx = \lim_{T \rightarrow +\infty} (\ln T - 1) = +\infty.$$

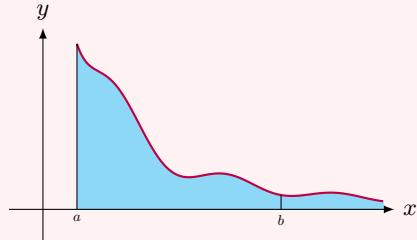
Thus the required area is infinite.

Comparing these two improper integrals, the former is a finite number and the latter is infinite. Since the improper integrals are in fact the limits of some integrals, we call the former convergent and the latter divergent.

4.84 Definition: improper integrals over infinite intervals

Suppose f is a function defined over an interval $[a, +\infty)$ such that f is integrable over $[a, b]$ for any $b > a$. The improper integral is defined as a limit of a definite integral as follows:

$$\int_a^{+\infty} f(x) dx = \lim_{b \rightarrow +\infty} \int_a^b f(x) dx.$$



We say the improper integral is **convergent** if the limit exists, otherwise divergent.

Similarly, we define

$$\int_{-\infty}^a f(x) dx = \lim_{b \rightarrow -\infty} \int_b^a f(x) dx.$$

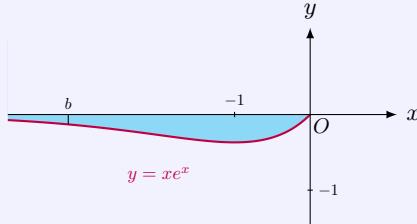
If $\int_a^{\infty} f(x) dx$ and $\int_{-\infty}^a f(x) dx$ are convergent, then we define

$$\int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^a f(x) dx + \int_a^{\infty} f(x) dx,$$

where a is any fixed real number.

4.85 Example: improper integral over unbounded interval

Consider the improper integral $\int_{-\infty}^0 xe^x dx$.



By definition, we first calculate the definite integral $\int_b^0 xe^x dx$ on a finite interval. From integration by parts, we have

$$\int_b^0 xe^x dx = \int_b^0 x d(e^x) = xe^x|_b^0 - \int_b^0 e^x dx = -be^b - e^x|_b^0 = -be^b - 1 + e^b.$$

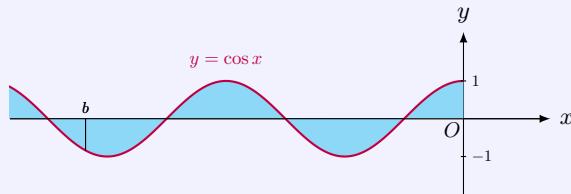
Thus,

$$\int_{-\infty}^0 xe^x dx = \lim_{b \rightarrow -\infty} (-be^b - 1 + e^b) = -1.$$

Hence this improper integral is convergent to -1 .

4.86 Example: divergent improper integral over unbounded interval

Consider the improper integral $\int_{-\infty}^0 \cos x \, dx$.



By definition, we first calculate the definite integral $\int_b^0 \cos x \, dx$ on a finite interval. From integration by parts, we have

$$\int_b^0 \cos x \, dx = \sin x \Big|_b^0 = -\sin b.$$

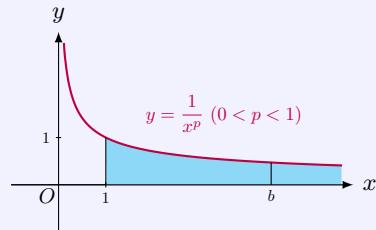
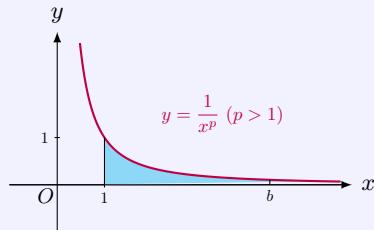
Thus,

$$\int_{-\infty}^0 \cos x \, dx = - \lim_{b \rightarrow -\infty} \sin b,$$

that is, the given improper integral is divergent.

4.87 Example: convergence/divergence of improper integral

Consider the improper integral $\int_1^{+\infty} \frac{1}{x^p} \, dx$ ($p > 0, p \neq 1$).



By definition, we first calculate the definite integral $\int_1^b \frac{1}{x^p} \, dx$ on a finite interval:

$$\int_1^b \frac{1}{x^p} \, dx = \frac{x^{1-p}}{1-p} \Big|_1^b = \frac{b^{1-p}}{1-p} - \frac{1}{1-p}.$$

Thus,

$$\int_1^{+\infty} \frac{1}{x^p} dx = \lim_{b \rightarrow +\infty} \left(\frac{b^{1-p}}{1-p} - \frac{1}{1-p} \right) = \frac{1}{p-1}, \quad \text{when } p > 1;$$

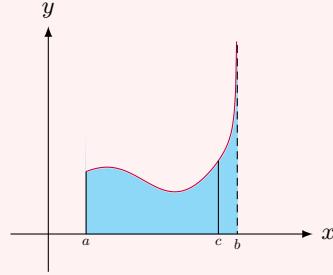
$$\int_1^{+\infty} \frac{1}{x^p} dx = \lim_{b \rightarrow +\infty} \left(\frac{b^{1-p}}{1-p} - \frac{1}{1-p} \right) = \infty, \quad \text{when } p < 1.$$

So, when $p > 1$, the given improper integral is convergent; when $p < 1$, it is divergent.

4.88 Definition: improper integrals with unbounded integrands

Suppose f is a function defined over an interval $[a, b)$ such that f is integrable over $[a, c]$ for any c with $a < c < b$, and suppose f is unbounded in $[c, b)$. We define the improper integral as a limit of a definite integral as follows:

$$\int_a^b f(x) dx = \lim_{c \rightarrow b} \int_a^c f(x) dx.$$



We say the improper integral is **convergent** if the limit exists, otherwise divergent.

Similarly, if f is unbounded near a , we define

$$\int_a^b f(x) dx = \lim_{c \rightarrow a} \int_c^b f(x) dx.$$

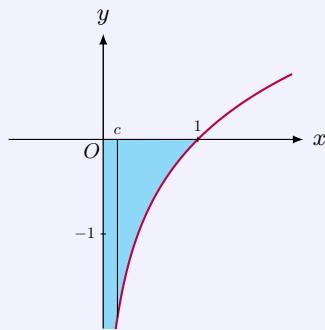
If f is unbounded near a and b , and if $\int_a^c f(x) dx$ and $\int_c^b f(x) dx$ are convergent, then we define

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

where c is any fixed real number with $a < c < b$.

4.89 Example: improper integral with unbounded integrand

Consider the improper integral $\int_0^1 \ln x dx$. The integrand $\ln x$ becomes unbounded near $x = 0$.



By definition, we first calculate the definite integral $\int_c^1 \ln x \, dx$ on $[c, 1]$. From integration by parts, we have

$$\int_c^1 \ln x \, dx = x \ln x|_c^1 - \int_c^1 1 \, dx = -c \ln c - x|_c^1 = -c \ln c - 1 + c.$$

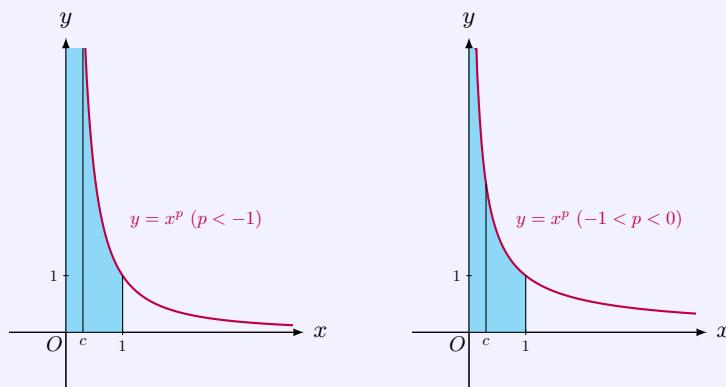
Thus,

$$\int_0^1 \ln x \, dx = \lim_{c \rightarrow 0} (-c \ln c - 1 + c) = -1.$$

Here we used l'Hospital's Rule and had $\lim_{c \rightarrow 0} (-c \ln c) = 0$. Hence, the given improper integral is convergent to -1 .

4.90 Example: convergence/divergence of improper integrals

Consider the improper integral $\int_0^1 x^p \, dx$ ($p < 0, p \neq -1$). The integrand x^p becomes unbounded near $x = 0$.



By definition, we first calculate the definite integral $\int_c^1 x^p dx$ on $[c, 1]$:

$$\int_c^1 x^p dx = \frac{x^{p+1}}{p+1} \Big|_c^1 = \frac{1}{p+1} - \frac{c^{p+1}}{p+1}.$$

Thus,

$$\begin{aligned}\int_0^1 x^p dx &= \lim_{c \rightarrow 0} \left(\frac{1}{p+1} - \frac{c^{p+1}}{p+1} \right) = \frac{1}{p+1}, \quad \text{when } -1 < p < 0; \\ \int_0^1 x^p dx &= \lim_{c \rightarrow 0} \left(\frac{1}{p+1} - \frac{c^{p+1}}{p+1} \right) = +\infty, \quad \text{when } p < -1.\end{aligned}$$

So, when $-1 < p < 0$, the given improper integral is convergent; when $p < -1$, it is divergent.

4.91 Pitfall: improper integrals

Consider the integral

$$\int_{-1}^1 \frac{1}{x} dx.$$

If one mistakenly thought this was a definite integral, he would have calculated as follows

$$\int_{-1}^1 \frac{1}{x} dx = \ln|x| \Big|_{-1}^1 = \ln 1 - \ln 1 = 0.$$

This is incorrect since the integral is an improper integral, for $\frac{1}{x}$ being unbounded near $0 \in [-1, 1]$.

We could write a sum of two improper integrals

$$\int_{-1}^1 \frac{1}{x} dx = \int_{-1}^0 \frac{1}{x} dx + \int_0^1 \frac{1}{x} dx.$$

One may naively calculate the improper integral as limits of the following definite integrals, for $b > 0$,

$$\int_{-1}^{-b} \frac{1}{x} dx + \int_b^1 \frac{1}{x} dx = \ln|x| \Big|_{-1}^{-b} + \ln|x| \Big|_b^1 = \ln b - \ln 1 + \ln 1 - \ln b = 0.$$

So, one might conclude that the improper integral is zero. In fact, this is a wrong answer again.

The correct way is to apply the definition. For each of the two improper integrals,

$$\int_{-1}^0 \frac{1}{x} dx \quad \text{and} \quad \int_0^1 \frac{1}{x} dx,$$

one should first evaluate the corresponding definite integral. Indeed, for $-1 < a < 0$ and $0 < b < 1$, one gets

$$\int_{-1}^a \frac{1}{x} dx + \int_b^1 \frac{1}{x} dx = \ln|x| \Big|_{-1}^a + \ln|x| \Big|_b^1 = \ln|a| - \ln 1 + \ln 1 - \ln b.$$

Clearly when $a \rightarrow 0$ and $b \rightarrow 0$, the expression above diverges. Therefore, the original improper integral $\int_{-1}^1 \frac{1}{x} dx$ is divergent.

Similar results also hold for other types of improper integrals.

4.92 Theorem: Comparison Test for improper integrals

Let f and g be continuous over $[a, +\infty)$ and $0 \leq f(x) \leq g(x)$. Then

1. If $\int_a^{+\infty} g(x) dx$ converges, then $\int_a^{+\infty} f(x) dx$ converges.
2. If $\int_a^{+\infty} f(x) dx$ diverges, then $\int_a^{+\infty} g(x) dx$ diverges.

Similar results hold for other types of improper integrals.

4.93 Example: Comparison Test for convergence

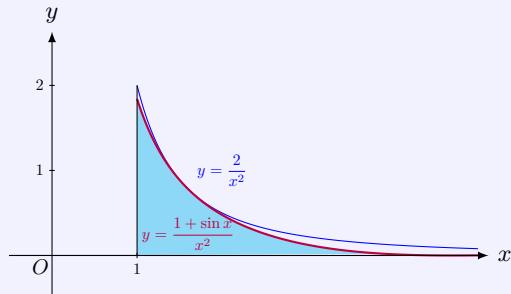
Consider the improper integral $\int_1^{+\infty} \frac{1 + \sin x}{x^2} dx$.

Since we have the inequality

$$\frac{1 + \sin x}{x^2} \leq \frac{2}{x^2},$$

and we know the improper integral $\int_1^{+\infty} \frac{2}{x^2} dx$ converges, the improper integral

$\int_1^{+\infty} \frac{1 + \sin x}{x^2} dx$ converges.



4.94 MATLAB: improper integrals over infinite intervals

In MATLAB, the commands `integral`, `quadgk` and `int` support calculating improper integrals over infinite intervals, i.e., the input which specifies the interval of the integral can be `-Inf` or `Inf`. The following example shows the calculation of $\int_{-\infty}^0 xe^x dx$ using `integral`, `quadgk` and `int`, respectively.

```
>> integral_ans = integral(@(x)x.*exp(x),-Inf,0)
integral_ans =
-1

>> quadgk_ans = quadgk(@(x)x.*exp(x),-Inf,0)
quadgk_ans =
-1.0000

>> syms t
>> f(t) = t*exp(t);
>> int_ans = int(f(t),-Inf,0)

int_ans =
-1
```

Note that the first input of the commands `integral` and `quadgk` should be a function handle, and the first input of the command `int` is a symbolic expression.

4.95 MATLAB: improper integral with unbounded integrand

For *convergent* improper integrals with unbounded integrands, we can use the commands `integral`, `quadgk` and `int` to calculate the results. The following example shows the calculation of $\int_0^1 x^{-0.5} dx$ in MATLAB.

```
>> integral_ans = integral(@(x)x.^(-0.5),0,1)
integral_ans =
2.0000

>> quadgk_ans = quadgk(@(x)x.^(-0.5),0,1)
quadgk_ans =
2.0000

>> syms t
>> f(t) = t.^(-0.5);
>> int_ans = int(f(t),0,1)
```

```
int_ans =
2
```

However, for *divergent* improper integrals with unbounded integrands, we need to use the command `int`. The following example shows the calculations of $\int_{-1}^1 \frac{1}{x} dx$ and $\int_0^1 x^{-2} dx$ using `int`, respectively.

```
>> syms t
>> f(t) = 1/t;
>> int_ans = int(f(t), -1, 1)

int_ans =

NaN

>> f(t) = t^(-2);
>> int_ans = int(f(t), 0, 1)

int_ans =

Inf
```

It shows that the result of $\int_{-1}^1 \frac{1}{x} dx$ is `NaN`, i.e., Not-a-Number, which means $\int_{-1}^1 \frac{1}{x} dx$ is divergent. The result of $\int_0^1 x^{-2} dx$ is `Inf`, i.e., ∞ , which also means $\int_0^1 x^{-2} dx$ is divergent.

4.96 Engineering Example: Fourier transform

Mathematically, the Fourier transform $\hat{f}(\xi)$ ($\xi \in \mathbb{R}$) of an integrable function $f(x)$ is defined as

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i x \xi} dx,$$

where $f(x)$ is in the time domain, and $\hat{f}(\xi)$ is in the frequency domain.

The Fourier transform decomposes a function in the time domain into the frequency-domain components which make it up. It allows us to analyze a signal in the time domain from a different perspective, i.e., in the frequency domain. The Fourier transform has wide applications in communication engineering, differentiation equation analysis area, quantum mechanics, etc. For example, the Fourier transform is used to help with the deduction of noise in a communication system. Moreover, the Fourier transform can facilitate the analysis of partial differential equations.

4.97 Engineering Example: Laplace transform

The Laplace transform takes a function of a real variable t (e.g., time) to a function of a complex variable s (e.g., frequency). Mathematically, the Laplace transform $f(s)$ of an integrable function $f(t)$ ($\xi \in \mathbb{R}^+$) is defined as follows

$$\hat{f}(s) = \int_0^\infty f(t)e^{-st}dt,$$

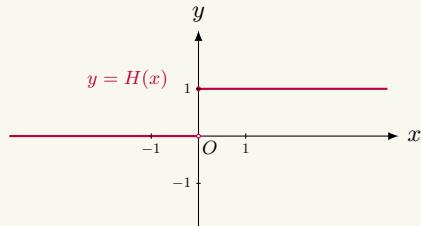
where $f(t)$ is in the time domain, and $f(s)$ is in the frequency domain.

Similar to the Fourier transform, the Laplace transform provides an alternative way of analyzing a signal in the time domain, i.e., in the frequency domain. The Laplace transform has been widely used in engineering and physics. For example, the Laplace transform has been used to analyze the linear time-invariant system in control theory by transforming the convolution to a multiplication which is easier to solve due to the algebraic form. Another important application of the Laplace transform is to solve partial differential equations, which has been extensively used in mechanical engineering and electrical engineering. In particular, the Laplace transform can turn a linear differential equation to an algebraic equation which can be easily solved by the formal rules of algebra, and the original linear differential equation can then be solved using the inverse Laplace transform.

4.98 Engineering Example: unit-step function - integral form

In mathematics, the unit-step function can be defined in the following:

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



Alternatively, $H(x)$ can also be defined as an integral of the Dirac delta function $\delta(s)$, which is provided as follows:

$$H(x) = \int_{-\infty}^x \delta(s) ds$$

where $\delta(s)$ is the Dirac delta function defined as

$$\delta(s) = \begin{cases} x, & \text{if } x \leq 0; \\ \sin x, & \text{if } x > 0, \end{cases}$$

and $\int_{-\infty}^{\infty} \delta(s) ds = 1$.

In communication theory, sum of the unit-step functions can be used for representations of waveforms. Moreover, the unit-step response, which is the time behaviour of the output of a general system when its input is a unit-step function, can be used to describe the characteristics of a system and is thus widely used in electronic engineering and control theory.

4.99 Engineering Example: continuous-time convolution

In mathematics, the continuous-time convolution is defined as an improper function as follows:

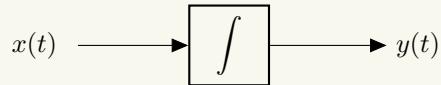
$$(f \star g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t - \tau) d\tau.$$

The function produced by convolution, i.e., $(f \star g)(t)$, shows how the shape of one function (e.g., $f(t)$) is modified by the other one (e.g., $g(t)$). The continuous-time convolution has wide applications in probability, statistics, signal processing and many other engineering areas. For example, the convolution is an important computation tool to characterize the linear time invariant (LTI) systems.

4.100 Engineering Example: integrator - signal processing

The integrator is a device whose output signal is the time integral of its input signal. It is an important part in many engineering and scientific applications, such as signal processing circuits and mechanical engineering devices.

In signal processing area, block diagrams are commonly seen to illustrate the signal flow paths. One example is shown below, where the quadrate block represents an integrator.



Given that the input signal is $x(t)$ and the output signal is $y(t)$, then the integrator represents the following integral:

$$y(t) = \int_{-\infty}^t x(\tau) d\tau.$$

Summary

Definition: basic concepts

definite integral : $\int_a^b f(x) dx = \lim_{\Delta x \rightarrow 0} \sum_{i=1}^n f(x_i^*) \Delta x$

indefinite integral : $\int f(x) dx = F(x) + C$, where $F'(x) = f(x)$

Theorem: properties of definite integrals

Let f and g be integrable functions over an interval I and let $a < b < c$ be points in I .

- If f and g are integrable over $[a, b]$, then $f + g$ is integrable over $[a, b]$, and we have

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

- If f is integrable over $[a, b]$, then $kf(x)$ is integrable over $[a, b]$, and we have

$$\int_a^b kf(x) dx = k \int_a^b f(x) dx.$$

- If f is integrable over $[a, b]$ and $[b, c]$, then f is integrable over $[a, c]$, and we have

$$\int_a^b f(x) dx + \int_b^c f(x) dx = \int_a^c f(x) dx.$$

- $\int_a^a f(x) dx = 0.$

- If f and g are both integrable over $[a, b]$, and $f(x) \leq g(x)$, then

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

-

$$\left| \int_a^b f(x) dx \right| \leq \int_a^b |f(x)| dx.$$

- (Mean Value Theorem for Integrals)** If f is continuous on $[a, b]$, then there is $c \in [a, b]$ such that

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

Theorem: linearity of indefinite integral

1. $\int [f(x) + g(x)] \, dx = \int f(x) \, dx + \int g(x) \, dx.$
2. $\int kf(x) \, dx = k \int f(x) \, dx,$ where k is a constant.

Theorem: Fundamental Theorem of Calculus

Let f be a continuous function on $[a, b].$

1. Then the function

$$A(x) = \int_a^x f(t) \, dt$$

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

2. If F is a differentiable function $[a, b]$ such that $F'(x) = f(x),$ then

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

Theorem: Indefinite Integrals of Some Elementary Functions

- | | |
|--|---|
| <ol style="list-style-type: none"> 1. $\int k \, dx = kx + C$ 2. $\int x^\mu \, dx = \frac{x^{\mu+1}}{\mu+1} \quad (\mu \neq -1)$ 3. $\int \frac{dx}{x} = \ln x + C$ 4. $\int \frac{dx}{x^2 + 1} = \arctan x + C$ 5. $\int \frac{dx}{\sqrt{1-x^2}} = \arcsin x + C$ 10. $\int \frac{dx}{\cos^2 x} = \int \sec^2 x \, dx = \tan x + C$ 11. $\int \frac{dx}{\sin^2 x} = \int \csc^2 x \, dx = -\cot x + C$ 12. $\int \sec x \tan x \, dx = \sec x + C$ 13. $\int \csc x \cot x \, dx = -\csc x + C$ | <ol style="list-style-type: none"> 6. $\int \cos x \, dx = \sin x + C$ 7. $\int \sin x \, dx = -\cos x + C$ 8. $\int e^x \, dx = e^x + C$ 9. $\int a^x \, dx = \frac{a^x}{\ln a} + C$ |
|--|---|

Theorem: substitution method**Theorem:** for indefinite integral

Let $u = g(x)$ be a differentiable function and f be a continuous function. Suppose the range of g is contained in the domain of f . Then

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

Theorem: for definite integral

Let $u = g(x)$, where $g'(x)$ is continuous on $[a, b]$, and let f be continuous on an interval containing the range of g . Then

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

Theorem: integration by parts

Suppose u and v are differentiable functions of x . Then

$$\int u(x) \cdot \underbrace{v'(x)}_{dv} dx = u(x) \cdot v(x) - \int v(x) \cdot \underbrace{u'(x)}_{du} dx,$$

or

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

Furthermore, for definite integral,

$$\int_a^b u(x) \cdot v'(x) dx = [u(x) \cdot v(x)]_a^b - \int_a^b v(x) \cdot u'(x) dx.$$

Method: numerical integration

Let f be integrable function over $[a, b]$. Divide the interval $[a, b]$ into n equal subintervals

$$x_0 = a, \quad x_1 = a + \frac{b-a}{n}, \quad \dots, \quad x_{n-1} = a + \frac{(n-1)(b-a)}{n}, \quad x_n = b.$$

- The **left point rule** is the Riemann sum of left end points of f on $[a, b]$:

$$\int_a^b f(x) dx \approx L_n = [f(x_0) + f(x_1) + \dots + f(x_{n-1})] \cdot \frac{b-a}{n}.$$

- The **right point rule** is the Riemann sum of right end points of f on $[a, b]$:

$$\int_a^b f(x) dx \approx R_n = [f(x_1) + f(x_2) + \cdots + f(x_n)] \cdot \frac{b-a}{n}.$$

- The **trapezoidal rule** is the average of Riemann sum of left end points and right end points:

$$\int_a^b f(x) dx \approx T_n = [f(x_0) + 2f(x_1) + \cdots + 2f(x_{n-1}) + f(x_n)] \cdot \frac{b-a}{2n}.$$

The region under the graph of the function f on $[x_{i-1}, x_i]$ is approximated by a trapezoid.

- The **midpoint rule** is the Riemann sum of middle points of f on $[a, b]$:

$$\int_a^b f(x) dx \approx M_n = [f(x_1^*) + f(x_2^*) + \cdots + f(x_n^*)] \cdot \frac{b-a}{n},$$

where $x_i^* = \frac{x_{i-1} + x_i}{2}$.

- Simpson's rule** is the following approximation

$$\begin{aligned} \int_a^b f(x) dx \approx S_n &= \frac{b-a}{3n} \cdot [f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) \\ &\quad + \cdots + 2f(x_{n-2}) + 4f(x_{n-1}) + f(x_n)], \end{aligned}$$

where n is an *even* integer.

Theorem: error bounds of numerical integration

Theorem: error bounds of the midpoint and trapezoidal rules

Suppose f'' is continuous and $|f''(x)| \leq K$ on $[a, b]$. Then

$$\begin{aligned} \left| \int_a^b f(x) dx - T_n \right| &\leq \frac{K(b-a)^3}{12n^2}, \\ \left| \int_a^b f(x) dx - M_n \right| &\leq \frac{K(b-a)^3}{24n^2}. \end{aligned}$$

Theorem: error bounds of Simpson's rule

Suppose $f^{(4)}$ is continuous and $|f^{(4)}(x)| \leq K$ on $[a, b]$. Then

$$\left| \int_a^b f(x) dx - S_n \right| \leq \frac{K(b-a)^5}{180n^4}.$$

Definition: improper integrals

- $\int_a^{+\infty} f(x) dx = \lim_{b \rightarrow +\infty} \int_a^b f(x) dx$.
- $\int_{-\infty}^a f(x) dx = \lim_{b \rightarrow -\infty} \int_b^a f(x) dx$.
- $\int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^a f(x) dx + \int_a^{\infty} f(x) dx$.
- $\int_a^b f(x) dx = \lim_{c \rightarrow b} \int_a^c f(x) dx$, if f is unbounded near b .
- $\int_a^b f(x) dx = \lim_{c \rightarrow a} \int_c^b f(x) dx$, if f is unbounded near a .
- $\int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx$, if f is unbounded near a and b (c is any fixed real number between a and b).

Theorem: Comparison Test for improper integrals

Let f and g be continuous over $[a, +\infty)$ and $0 \leq f(x) \leq g(x)$. Then

1. If $\int_a^{+\infty} g(x) dx$ converges, then $\int_a^{+\infty} f(x) dx$ converges.
2. If $\int_a^{+\infty} f(x) dx$ diverges, then $\int_a^{+\infty} g(x) dx$ diverges.

Similar results hold for other types of improper integrals.

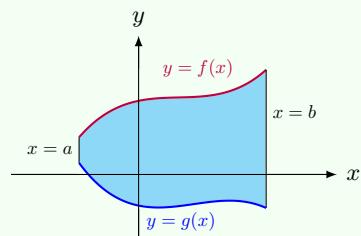
Applications of Integration

5.1 Areas of Regions Between Curves

5.1 Theorem: area of a region bounded by two curves

Let f and g be two continuous functions with $f(x) \geq g(x)$ over the interval $[a, b]$. The area of the region between the graphs of f and g over $[a, b]$ is

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



5.2 Example: area bounded by two curves

Find the area of the closed region bounded by the graphs of the functions $f(x) = \frac{1}{x^2 + 1}$ and $g(x) = \frac{1}{2}$.

Solution We sketch the graphs of the given functions.

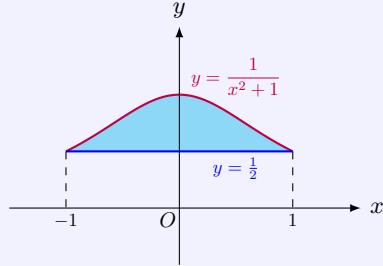
We need first to find the intersections of the two curves so that we know the interval over which the two curves bound a closed region. In fact, by solving the equation

$$\frac{1}{x^2 + 1} = \frac{1}{2},$$

we have $x = \pm 1$.

Over the interval $[-1, 1]$, $f(x) \geq g(x)$. Thus, the area of the region bounded by two curves is

$$\begin{aligned} \int_{-1}^1 \left(\frac{1}{x^2 + 1} - \frac{1}{2} \right) dx &= [\tan^{-1} x]_{x=-1}^1 - \left[\frac{1}{2}x \right]_{x=-1}^1 \\ &= \tan^{-1} 1 - \tan^{-1}(-1) - \left(\frac{1}{2} - \frac{-1}{2} \right) \\ &= \frac{\pi}{4} - \frac{-\pi}{4} - 1 = \frac{\pi}{2} - 1. \end{aligned}$$



5.3 Example: multiple regions

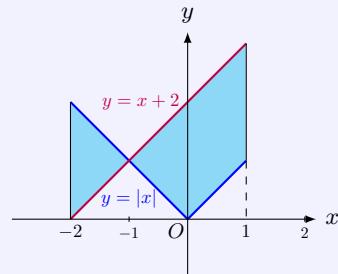
Find the area of the closed region bounded by the graphs of the functions $f(x) = x + 2$ and $g(x) = |x|$ over the interval $[-2, 1]$.

Solution Even though the interval $[-2, 1]$ is given, the integral

$$\int_{-2}^1 [f(x) - g(x)] dx$$

is *not* the area of the region bounded by the curves over $[-2, 1]$. This is because that neither $f(x) \geq g(x)$ hold over the whole interval $[-2, 1]$, nor $g(x) \geq f(x)$ over the whole interval. Hence, we need to determine the interval(s) over which $f(x) \geq g(x)$ and $f(x) \leq g(x)$, respectively.

In fact, by solving the equation $f(x) = g(x)$, or $x + 2 = |x|$, we get a unique solution $x = -1$. We see that $g(x) \geq f(x)$ when $x \leq -1$; $f(x) \geq g(x)$ when $x \geq -1$, as shown in the figure.



Thus, the area of the region bounded by the two curves is

$$\begin{aligned}
 & \int_{-2}^{-1} [g(x) - f(x)] dx + \int_{-1}^1 [f(x) - g(x)] dx \\
 &= \int_{-2}^{-1} [|x| - (x+2)] dx + \int_{-1}^1 [x+2 - |x|] dx \\
 &= \int_{-2}^{-1} [-x - (x+2)] dx + \int_{-1}^0 [x+2 - (-x)] dx + \int_0^1 [x+2 - x] dx \\
 &= (-x^2 - 2x)|_{-2}^{-1} + (x^2 + 2x)|_{-1}^0 + 2x|_0^1 \\
 &= (-1 + 2) - (-4 + 4) + 0 - (1 - 2) + 2 - 0 = 4.
 \end{aligned}$$

Geometrically, we see that the area is the same as

$$\int_{-2}^1 |f(x) - g(x)| dx.$$

5.4 Theorem: area between two curves

In general, the area between the curve $y = f(x)$ and $y = g(x)$ over $[a, b]$ is

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

5.5 Example: areas between two curves

Find the area between $y = \cos x$ and $y = \sin x$ over $[0, \frac{\pi}{2}]$.

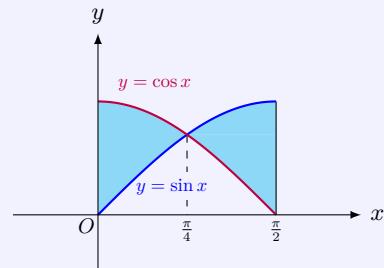
Solution We need to calculate the integral

$$\int_0^{\frac{\pi}{2}} |\sin x - \cos x| dx.$$

Firstly we find intervals where $\sin x \geq \cos x$. To do so, in the interval $[0, \frac{\pi}{2}]$, we solve the equation

$$\sin x = \cos x$$

and get $x = \frac{\pi}{4}$. We know that $\sin x \leq \cos x$ when $0 \leq x \leq \frac{\pi}{4}$; $\sin x \geq \cos x$ when $\frac{\pi}{4} \leq x \leq \frac{\pi}{2}$, as shown in the figure.



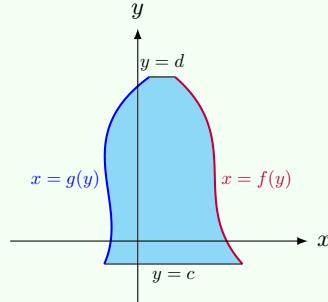
Hence, the area between $y = \cos x$ and $y = \sin x$ over $[0, \frac{\pi}{2}]$ is

$$\begin{aligned} & \int_0^{\frac{\pi}{2}} |\sin x - \cos x| dx \\ &= \int_0^{\frac{\pi}{4}} (\cos x - \sin x) dx + \int_{\frac{\pi}{4}}^{\frac{\pi}{2}} (\sin x - \cos x) dx \\ &= (\sin x + \cos x)|_0^{\frac{\pi}{4}} + (-\cos x - \sin x)|_{\frac{\pi}{4}}^{\frac{\pi}{2}} \\ &= (\sin \frac{\pi}{4} + \cos \frac{\pi}{4}) - (\sin 0 + \cos 0) + (-\cos \frac{\pi}{2} - \sin \frac{\pi}{2}) - (-\cos \frac{\pi}{4} - \sin \frac{\pi}{4}) \\ &= \sqrt{2} - 1 - 1 + \sqrt{2} = 2\sqrt{2} - 2. \end{aligned}$$

5.6 Theorem: area bounded by two curves along the y -axis

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

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



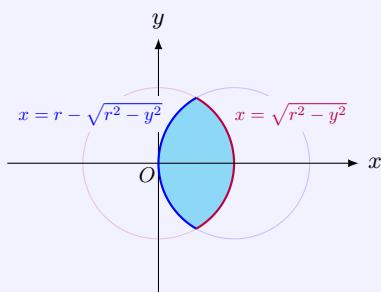
5.7 Example: areas between two curves

Find the common area of two disks $x^2 + y^2 \leq r^2$ and $(x - r)^2 + y^2 \leq r^2$.

Solution Solving two simultaneous equations

$$x^2 + y^2 = r^2, \quad (x - r)^2 + y^2 = r^2$$

gives $y = \pm \frac{\sqrt{3}}{2}r$. It is easy to get the two arcs that enclosed the region are $x = \sqrt{r^2 - y^2}$ and $x = r - \sqrt{r^2 - y^2}$, as shown in the figure.



Thus, the common region is

$$\int_{-\frac{\sqrt{3}}{2}r}^{\frac{\sqrt{3}}{2}r} \left[\sqrt{r^2 - y^2} - (r - \sqrt{r^2 - y^2}) \right] dy = \int_{-\frac{\sqrt{3}}{2}r}^{\frac{\sqrt{3}}{2}r} (r - 2\sqrt{r^2 - y^2}) dy.$$

To evaluate the last integral, we make a trigonometric substitution $y = r \sin t$. Then the area

is

$$\begin{aligned}
 \int_{-\frac{\sqrt{3}}{2}r}^{\frac{\sqrt{3}}{2}r} \left(2\sqrt{r^2 - y^2} - r\right) dy &= \int_{-\frac{\pi}{3}}^{\frac{\pi}{3}} (2r \cos t - r) \cdot r \cos t dt \\
 &= 2r^2 \int_0^{\frac{\pi}{3}} (2 \cos^2 t - \cos t) dt \\
 &= 2r^2 \int_0^{\frac{\pi}{3}} [(1 + \cos 2t) - \cos t] dt \\
 &= 2r^2 \cdot [t + \frac{1}{2} \sin 2t - \sin t] \Big|_0^{\frac{\pi}{3}} \\
 &= 2r^2 \cdot \left(\frac{\pi}{3} + \frac{1}{2} \cdot \frac{\sqrt{3}}{2} - \frac{\sqrt{3}}{2}\right) = (\frac{2\pi}{3} - \frac{\sqrt{3}}{2})r^2.
 \end{aligned}$$

5.8 MATLAB: area bounded by two curves along the x -axis

Find the area between $y = \cos x$ and $y = \sin x$ over $[0, \frac{\pi}{2}]$.

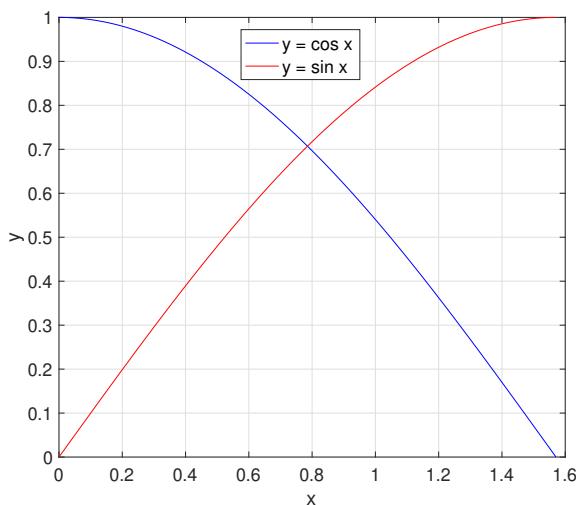
Solution First, we plot the curves of given functions.

```

>> syms x
>> y1(x) = cos(x);
>> y2(x) = sin(x);
>> x_array = linspace(0,pi/2,200);
>> y1_array = y1(x_array);
>> y2_array = y2(x_array);
>> figure
>> plot(x_array,y1_array,'b',x_array,y2_array,'r') % plot two ...
    curves on the same figure
>> legend('y = cos x','y = sin x')
>> xlabel('x')
>> ylabel('y')
>> grid

```

Note that to plot two curves on the same figure, we can write specifications of the two curves using one `plot` command instead of two `plot` commands. The figure plotted is shown below.



We see that $y = \cos x$ and $y = \sin x$ intersect at a point around $x = 0.8$. So, we try to find a point around $x = 0.8$ such that $\cos x = \sin x$. This is shown in the following MATLAB code.

```
>> Delta_y = y1 - y2
Delta_y(x) =
cos(x) - sin(x)

>> Delta_y_hd = matlabFunction(Delta_y) % convert symbolic ...
expression to function handle

Delta_y_hd =
function_handle with value:
@(x)cos(x)-sin(x)

>> x0 = fzero(Delta_y_hd, 0.8)
x0 =
0.7854
```

We see from the result above that $y = \cos x$ and $y = \sin x$ intersect at $x_0 = 0.7854$. Hence, we can determine the two intervals over which the two curves bound a closed region, namely, $[0, x_0]$ and $[x_0, \frac{\pi}{2}]$. According to these two intervals, we calculate the area bounded by the two curves as follows:

```
>> area1 = int(Delta_y,x,0,x0);
>> area2 = int(-Delta_y,x,x0,pi/2);
>> area = area1 + area2

area =

2*2^(1/2) - 2
```

The result shows that the area bounded by $y = \cos x$ and $y = \sin x$ over $[0, \frac{\pi}{2}]$ is $2\sqrt{2} - 2$.

5.9 MATLAB: area bounded by two curves along the y -axis

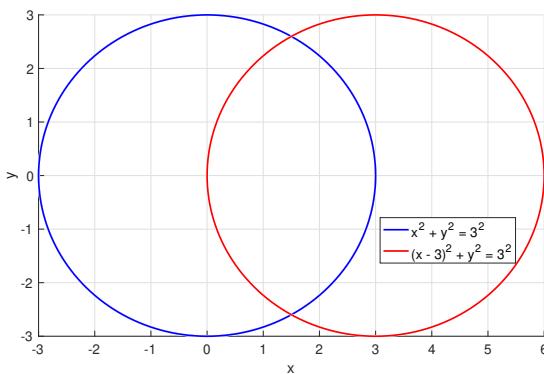
Find the common area of two disks $x^2 + y^2 \leq 3^2$ and $(x - 3)^2 + y^2 \leq 3^2$.

Solution First, we notice that $x^2 + y^2 = 3^2$ and $(x - 3)^2 + y^2 = 3^2$ are two circles. Specifically, $x^2 + y^2 = 3^2$ is a circle with the center at $(0, 0)$ and radius 3, and $(x - 3)^2 + y^2 = 3^2$ is a circle with the center at $(3, 0)$ and radius 3. We thus can plot a figure with these two circles using the following MATLAB code:

```
>> cla % clear the axes
>> c1 = viscircles([0 0],3,'Color','b'); % center: [0 0], ...
   radius: 3
>> c2 = viscircles([3 0],3,'Color','r'); % center: [3 0], ...
   radius: 3
>> grid
>> legend([c1,c2],{'x^2 + y^2 = 3^2','(x - 3)^2 + y^2 = 3^2'});
>> xlabel('x')
>> ylabel('y')
```

Note that the command `viscircles` can plot a circle with given center and radius. Moreover, to specify the colour of the circle, besides the parameter specifying the colour, e.g., '`b`' and '`r`', we also need to type in '`Color`' in the command `viscircles`. Furthermore, since the command `viscircles` does not clear the target axes before plotting circles, we need to first remove any graphs that have been previously plotted in the axes using the `cla` command.

The figure plotted is shown below.



From the figure above, we see that the intersections of the two circles are the ones of the two functions $x = \sqrt{3^2 - y^2}$ and $x = 3 - \sqrt{3^2 - y^2}$. In particular, these two functions intersect at points around $y = -2.5$ and $y = 2.5$, respectively. Therefore, we try to find the points around $y = -2.5$ and $y = 2.5$, respectively, such that $\sqrt{3^2 - y^2} = 3 - \sqrt{3^2 - y^2}$ by using the following MATLAB code.

```
>> syms y
>> x1 = sqrt(3^2 - y^2);
>> x2 = 3 - sqrt(3^2 - y^2);
>> Delta_x = x1 - x2

Delta_x =
2*(9 - y^2)^(1/2) - 3

>> Delta_x_hd = matlabFunction(Delta_x) % convert symbolic ...
    expression to function handle

Delta_x_hd =
function_handle with value:
@(y)sqrt(-y.^2+9.0).*2.0-3.0

>> y1 = fzero(Delta_x_hd,-2.5)

y1 =
-2.5981

>> y2 = fzero(Delta_x_hd,2.5)
```

```
y2 =
2.5981
```

The results above show that $x = \sqrt{3^2 - y^2}$ and $x = 3 - \sqrt{3^2 - y^2}$ intersect at $y_1 = -2.5981$ and $y_2 = 2.5981$. Hence, we can determine the interval over which these two functions bound a closed region, i.e., $[y_1, y_2]$. Then, we calculate the area bounded by the two curves as follows:

```
>> area = int(Delta_x,y1,y2)
area =
6*pi - (9*3^(1/2))/2
```

We see from the result above that the common area of two disks $x^2 + y^2 \leq 3^2$ and $(x-3)^2 + y^2 \leq 3^2$ is $6\pi - \frac{9\sqrt{3}}{2}$.

5.10 Theorem: area under a parametric curve

If a curve is traced out once by the parametric equations $x = f(t)$ and $y = g(t)$, $\alpha \leq t \leq \beta$, then the area under the curve is

$$\int_{\alpha}^{\beta} g(t) f'(t) dt.$$

Justification:

The area under a curve $y = F(x)$ over $[a, b]$ is $A = \int_a^b F(x) dx$. By applying the substitution rule, we get

$$A = \int_{\alpha}^{\beta} F(f(t)) f'(t) dt,$$

where $F(a) = \alpha$ and $F(b) = \beta$. Since the curve $y = F(x)$, $a \leq x \leq b$, is traced out by the given parametric equations, we have

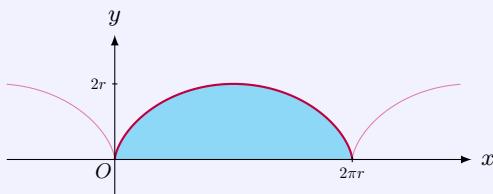
$$y = F(x) = F(f(t)),$$

and $y = g(t)$, so that $F(f(t)) = g(t)$. Hence, we obtain the area formula in parametric equations.

5.11 Example: area under one arch of the cycloid

Find the area under one arch of the cycloid

$$x = r(t - \sin t), \quad y = r(1 - \cos t), \quad 0 \leq t \leq 2\pi.$$



Solution It is known that, for a given curve above the x -axis, the area under the curve over an interval $[a, b]$ is $A = \int_a^b y(x) dx$. Thus, if the curve is given by parametric equations $x = x(t)$, $y = y(t)$, then these equations can be served as a substitution.

Hence, the area under one arch of the cycloid is

$$\begin{aligned} A &= \int_0^{2\pi r} y \, dx \\ &= \int_0^{2\pi} r(1 - \cos t) \cdot [r(t - \sin t)]' \, dt && \left\{ \begin{array}{l} x = r(t - \sin t) \\ y = r(1 - \cos t) \end{array} \right. \\ &= r^2 \int_0^{2\pi} (1 - 2\cos t + \cos^2 t) \, dt \\ &= r^2 \int_0^{2\pi} \left[1 - 2\cos t + \frac{1}{2}(1 + \cos 2t) \right] \, dt && \text{double-angle formula} \\ &= r^2 \left[\frac{3}{2}t - 2\sin t + \frac{1}{4}\sin 2t \right]_{t=0}^{2\pi} = 3\pi r^2. \end{aligned}$$

5.12 MATLAB: area under one arch of the cycloid

Find the area under one arch of the cycloid

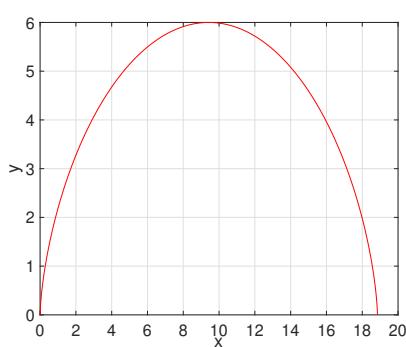
$$x = 3(t - \sin t), \quad y = 3(1 - \cos t), \quad 0 \leq t \leq 2\pi.$$

Solution First, we plot the graph of the cycloid when $0 \leq t \leq 2\pi$.

```
>> t_array = linspace(0, 2*pi, 200);
>> x_array = 3*(t_array - sin(t_array));
>> y_array = 3*(1 - cos(t_array));
>> figure
>> plot(x_array, y_array, 'r')
>> grid
>> xlabel('x')
```

```
>> ylabel('y')
```

The figure plotted is shown below.



Then, the area under one arch of the cycloid can be calculated as follows:

```
>> syms t
>> x(t) = 3*(t - sin(t));
>> y(t) = 3*(1 - cos(t));
>> dx(t) = diff(x(t),t);
>> area = int(y(t)*dx(t),0,2*pi)

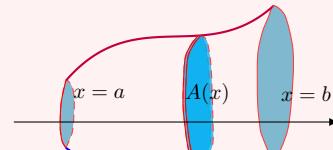
area =
27*pi
```

The result above shows that the area is 27π .

5.13 Definition: 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

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



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

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

We normally say the above method as finding **volume by using slices**.

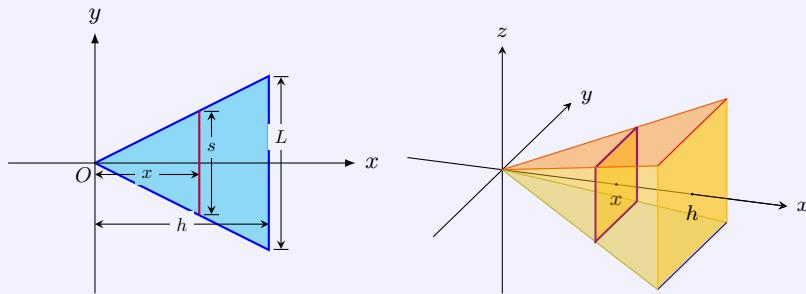
5.14 Example: volume of a pyramid - using slices

Find the volume of a pyramid whose base is a square with side L and whose height is h .

Solution We place the origin O of the coordinates at the vertex of the pyramid and the x -axis along its central axis. Any plane P_x that passes through x and is perpendicular to the x -axis intersects the pyramid in a square with side of length s , say. We can express s in terms of x by observing from the similar triangles:

$$\frac{x}{s} = \frac{h}{L}$$

and so $s = \frac{xL}{h}$. Thus, the cross-sectional area is $A(x) = s^2 = \frac{L^2}{h^2}x^2$.



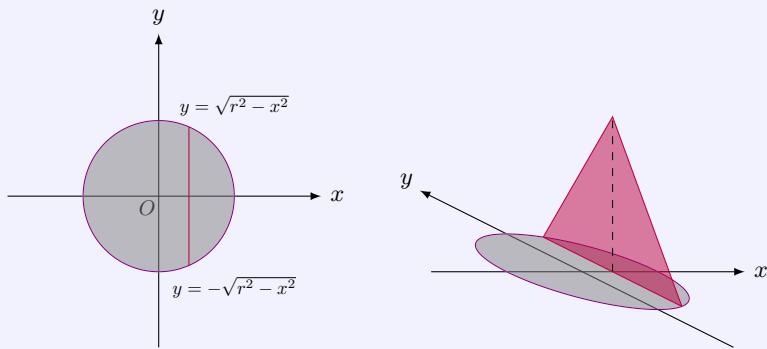
The pyramid lies between $x = 0$ and $x = h$, so its volume is

$$V = \int_0^h A(x) dx = \int_0^h \frac{L^2}{h^2} x^2 dx = \frac{L^2}{h^2} \left(\frac{1}{3} x^3 \right)_0^h = \frac{1}{3} L^2 h.$$

5.15 Example: volume of a solid - using slices

Consider a solid whose base is a circular disk with radius r . Find the volume of the solid if parallel cross-sections perpendicular to the base are isosceles triangles with height h and unequal side in the base.

Solution We sketch the diagram of the situation as in the figure.



Solving the equation $x^2 + y^2 = r^2$ for x , we have the expressions of the top-half and the bottom-half of the circle,

$$x = \pm\sqrt{r^2 - y^2}.$$

It is clear that the circle intersects the x -axis at $x = \pm r$.

By symmetry, without loss of generality, we can assume that cross-sections are perpendicular to the x -axis. For $-r \leq x \leq r$, the area of the cross-section is

$$A(x) = \frac{1}{2} \cdot 2\sqrt{r^2 - x^2} \cdot h = h\sqrt{r^2 - x^2}.$$

Hence, the volume of the solid is

$$V = \int_{-r}^r A(x) dx = \int_{-r}^r h\sqrt{r^2 - x^2} dx = h \int_{-r}^r \sqrt{r^2 - x^2} dx.$$

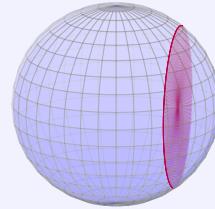
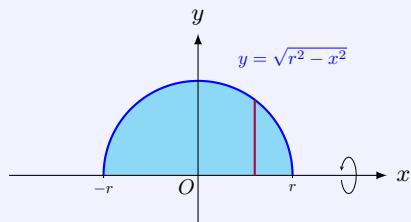
We observe that the last integral is the area of semi-circular disk of radius r , so

$$V = h \cdot \frac{1}{2}\pi r^2 = \frac{1}{2}\pi r^2 h.$$

5.16 Example: volume of a ball - using slices

It is known that the area of circular disk of radius r is πr^2 . Find the volume of the ball of radius r .

Solution A ball of radius r can be generated by rotating the upper semicircle $y = \sqrt{r^2 - x^2}$ about the x -axis. For $-r \leq x \leq r$, the cross-section is a circular disk, with area $\pi(r^2 - x^2)$, as shown in the figure.



Thus, the volume of the ball is

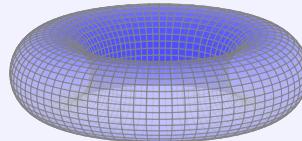
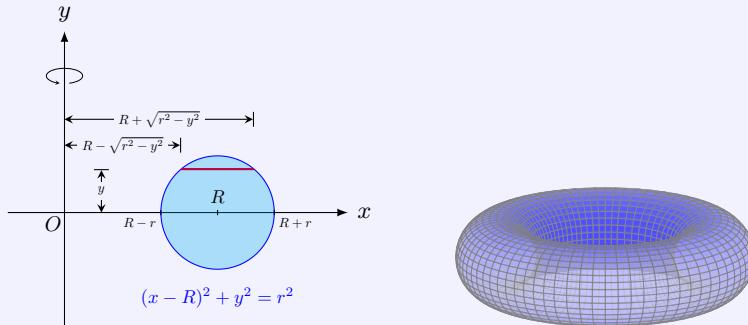
$$\begin{aligned} V &= \int_{-r}^r \pi(r^2 - x^2) dx \\ &= 2 \int_0^r \pi(r^2 - x^2) dx = 2\pi (r^2 x - \frac{1}{3}x^3) \Big|_0^r = \frac{4}{3}\pi r^3. \end{aligned}$$

5.17 Example: volume of a solid torus - using slices

Consider the region bounded by the curves

$$(x - R)^2 + y^2 = r^2, \quad 0 < r < R.$$

The solid obtained by rotating the region about the y -axis is the solid torus, as shown in the figure.



We shall use the method of slices to compute the volume of the solid.

The region and a typical slice are shown in the figure.

We see that the slice is a washer, with outer radius $R + \sqrt{r^2 - y^2}$, and inner radius $R - \sqrt{r^2 - y^2}$. So the cross-sectional area is

$$\pi \left(R + \sqrt{r^2 - y^2} \right)^2 - \pi \left(R - \sqrt{r^2 - y^2} \right)^2.$$

Thus, the volume of the solid torus is

$$\begin{aligned}
 V &= \int_{-r}^r \left[\pi \left(R + \sqrt{r^2 - y^2} \right)^2 - \pi \left(R - \sqrt{r^2 - y^2} \right)^2 \right] dx \\
 &= \pi \int_{-r}^r 2R \cdot 2\sqrt{r^2 - y^2} dy \quad \text{factoring} \\
 &= 4\pi R \int_{-r}^r \sqrt{r^2 - y^2} dy.
 \end{aligned}$$

Since the last integral represents the area of the semicircular region of radius r , we have

$$V = 4\pi R \cdot \frac{1}{2}\pi r^2 = 2\pi^2 Rr^2.$$

5.18 MATLAB: volume of a ball - using slices

Find the volume of a ball with radius 3.

Solution First, to well visualize the ball and the cross-section, we plot the ball and a cross-section for a fixed x value in MATLAB by the following code:

```
% plot the sphere with radius r
N = 40;
r = 3; % sphere radius

[XX,YY,ZZ] = sphere(N); % unit sphere generated by three ...
    % (N+1)-by-(N+1) matrices XX, YY and ZZ
X = r*XX; % expand the sphere radius
Y = r*YY;
Z = r*ZZ;

figure
surf(X,Y,Z, 'FaceAlpha', 0.2, 'FaceColor', 'b') % 'FaceAlpha': ...
    % face transparency-ranges from 0 (fully transparent) to 1 ...
    % (fully opaque)
xlabel('x')
ylabel('y')
zlabel('z')

% plot the cross-section
x0 = 2; % cross-section is at x = 2
r0 = sqrt(r^2 - x0^2); % radius of the cross-sectional circle

y1 = linspace(0,r0,200);
z1 = zeros(1,length(y1));

for ind = 1:length(y1) % first quadrant
```

```

y = y1(ind);
z1(ind) = sqrt(r0^2 - y^2);
end

y2 = -y1(1:length(y1)); % second quadrant
z2 = z1(1:length(z1));

y3 = -y1(1:(length(y1))); % third quadrant
z3 = -z1(1:(length(z1)));

y4 = y1(1:(length(y1))); % fourth quadrant
z4 = -z1(1:(length(z1)));

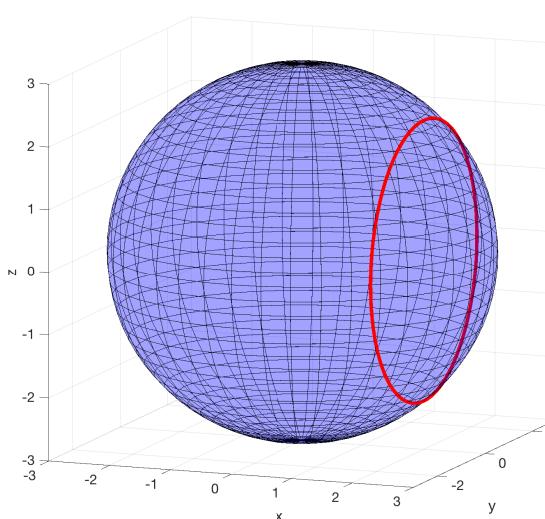
y_array = [fliplr(y1),y2,fliplr(y3),y4]; % plot the circle ...
    anticlockwise
z_array = [fliplr(z1),z2,fliplr(z3),z4];
x_array = ones(1,length(y_array))*x0;
x4 = ones(1,length(y4))*x0;

hold on
plot3(x_array,y_array,z_array,'r','LineWidth',4)

```

Note that the command `sphere` can be used to generate a ball with unit radius, the command `surf` can be used to plot the surface of the ball, and the command `plot3` can be used to plot curves in three-dimensional space. Moreover, the face transparency and the face colour of the ball can be specified when using `surf`, and the width of the curve can be specified when using `plot3`.

The figure plotted is shown below.



Since the cross-section is a circular disk with area $\pi(3^2 - x^2)$, as shown in the figure above, we thus can calculate the volume of the ball by solving the following integral:

$$V = \int_{-3}^3 \pi(3^2 - x^2) dx.$$

It can be implemented in MATLAB as follows:

```
>> syms x
>> f(x) = pi*(3^2 - x^2);
>> V = int(f(x), -3, 3)

V =

36*pi
```

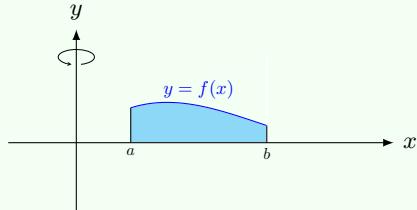
The volume of the ball is thus 36π .

Method of Cylindrical Shells

5.19 Theorem: volume 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 ($0 \leq a < b$), is

$$V = \int_a^b 2\pi x f(x) dx.$$

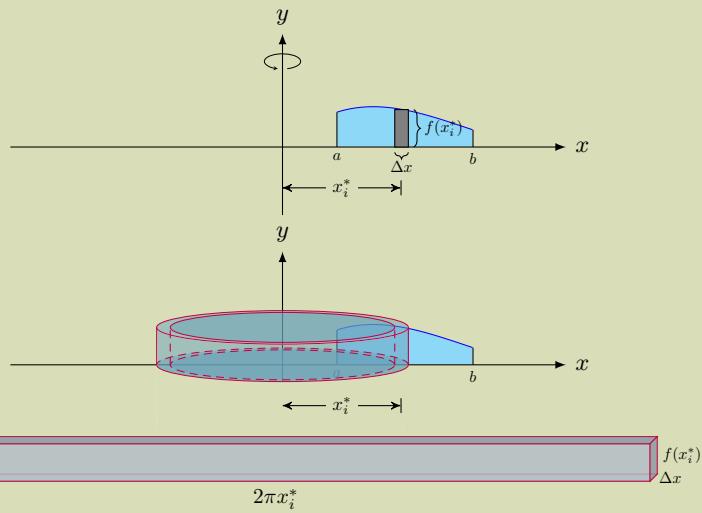


We normally say the above method as finding **volume by using shells**.

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 i th 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

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



Hence 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:

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

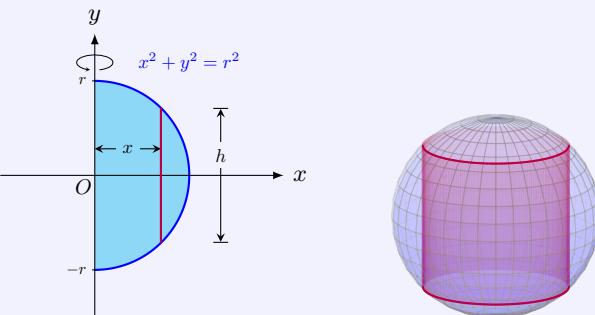
where the terms in the formula can be interpreted as

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

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

5.20 Example: volume of a ball - using shells

Find the volume of the ball of radius r .



The region and a typical shell are shown in the figure. We see that the shell has radius x , circumference $2\pi x$, and height $2\sqrt{r^2 - x^2}$. So the volume is

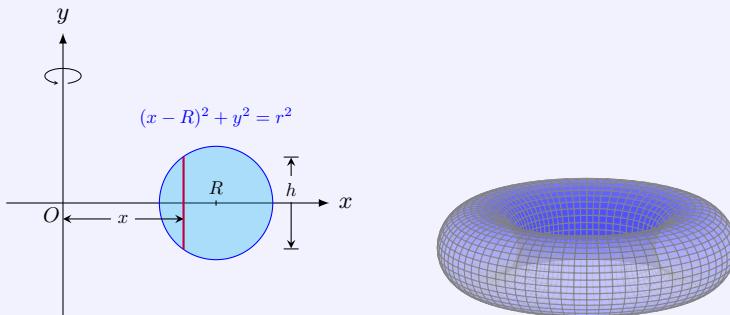
$$\begin{aligned} V &= \int_0^r 2\pi x \cdot 2\sqrt{r^2 - x^2} dx \\ &= (-2\pi) \int_0^r \sqrt{r^2 - x^2} \cdot (r^2 - x^2)' dx \\ &= (-2\pi) \cdot \frac{2}{3}(r^2 - x^2)^{\frac{3}{2}} \Big|_0^r = \frac{4}{3}\pi r^3. \end{aligned}$$

5.21 Example: volume of a solid torus - using shells

Consider the region bounded by the curves

$$(x - R)^2 + y^2 = r^2, \quad 0 < r < R.$$

The solid obtained by rotating the region about the y -axis is the solid torus, as shown in the figure.



We shall use the method of cylindrical shells to compute the volume of the solid.

The region and a typical shell are shown in the figure.

We see that the shell has radius x , circumference $2\pi x$, and height $2\sqrt{r^2 - (x - R)^2}$. So

the volume is

$$\begin{aligned}
 V &= \int_{R-r}^{R+r} 2\pi x \cdot 2\sqrt{r^2 - (x-R)^2} dx \\
 &= \int_{-r}^r 4\pi(u+R)\sqrt{r^2 - u^2} du && u = x - R, du = dx \\
 &= 4\pi \int_{-r}^r (u\sqrt{r^2 - u^2} + R\sqrt{r^2 - u^2}) du \\
 &= 4\pi \int_{-r}^r R\sqrt{r^2 - u^2} du && \text{odd function} \\
 &= 4\pi R \int_{-r}^r \sqrt{r^2 - u^2} du.
 \end{aligned}$$

Since the last integral represents the area of the semicircular region of radius r , we have

$$V = 4\pi R \cdot \frac{1}{2}\pi r^2 = 2\pi^2 Rr^2.$$

5.22 Example: rotating about a horizontal line

Find the volume of the solid obtained by rotating the region bounded by the curves

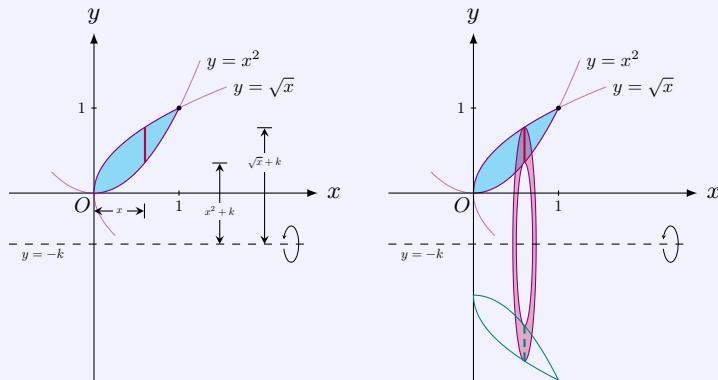
$$y = \sqrt{x} \quad \text{and} \quad y = x^2,$$

about the horizontal line $y = -k$ ($k > 0$).

Solution By solving the simultaneous equations $y = \sqrt{x}$ and $y = x^2$, we see that there are two intersections $(0, 0)$ and $(1, 1)$. So the region bounded by the given curves is $0 \leq x \leq 1$ as well as $0 \leq y \leq 1$.

We compute the volume of the solid by slices and shells.

1. The region and a typical slice are shown in the following figures.

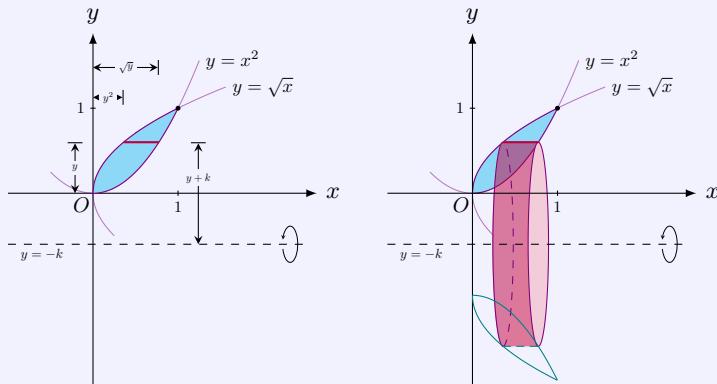


We see that the slice is a washer, with inner radius $x^2 + k$ and outer radius $\sqrt{x} + k$. So

the volume is

$$\begin{aligned} V &= \int_0^1 [\pi(\sqrt{x} + k)^2 - \pi(x^2 + k)^2] \, dx \\ &= \pi \int_0^1 (x + 2k\sqrt{x} - x^4 - 2kx^2) \, dx \\ &= \frac{\pi}{30}(9 + 20k). \end{aligned}$$

2. The region and a typical shell are shown in the following figures.



We see that the shell has radius $y + k$, circumference $2\pi(y + k)$, and height $\sqrt{y} - y^2$. So the volume is

$$\begin{aligned} V &= \int_0^1 2\pi(y + k) \cdot (\sqrt{y} - y^2) \, dy \\ &= 2\pi \int_0^1 \left(y^{\frac{3}{2}} - y^3 + ky^{\frac{1}{2}} - ky^2 \right) \, dx \\ &= \frac{\pi}{30}(9 + 20k). \end{aligned}$$

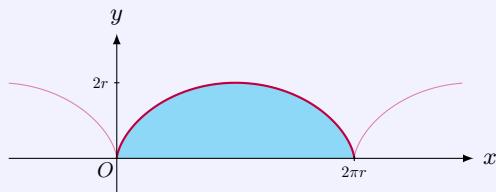
Solid Generated by Parametric Curve

5.23 Example: solid generated by a parametric curve

Rotate the area under one arch of the cycloid

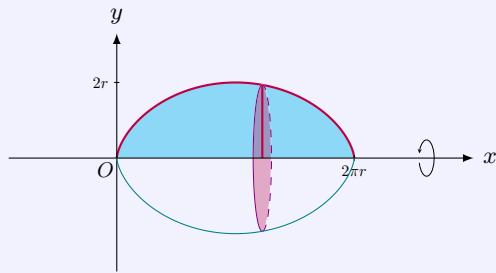
$$x = r(t - \sin t), \quad y = r(1 - \cos t), \quad 0 \leq t \leq 2\pi,$$

about the x -axis and the y -axis, respectively. Find the volumes of the two solids.

**Solution**

1. When rotating the region about the x -axis, we compute the volume of the solid by slices.

A typical slice is shown in the following figure.

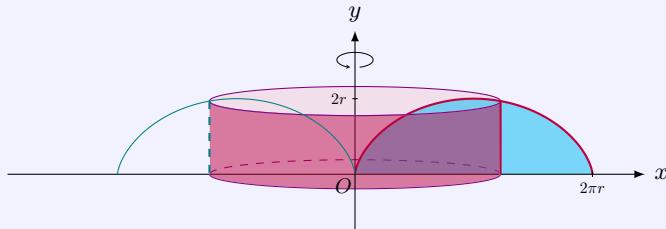


The volume of the solid is $V = \int_0^{2\pi} \pi[y(x)]^2 dx$. We can use the parametric equations for the cycloid as a substitution for the definite integral. Thus, the volume of the solid is

$$\begin{aligned}
 V &= \int_0^{2\pi r} \pi[y(x)]^2 dx \\
 &= \int_0^{2\pi} \pi [r(1 - \cos t)]^2 \cdot [r(t - \sin t)]' dt && \left\{ \begin{array}{l} x = r(t - \sin t) \\ y = r(1 - \cos t) \end{array} \right. \\
 &= \pi r^3 \int_0^{2\pi} (1 - \cos t)^2 dt \\
 &= \pi r^3 \int_0^{2\pi} (1 - 2\cos t + \cos^2 t) dt && \text{expanding} \\
 &= \pi r^3 \int_0^{2\pi} [1 - 2\cos t + \frac{1}{2}(1 + \cos 2t) \\
 &\quad - (1 - \sin^2 t)\cos t] dt \\
 &= \pi r^3 [\frac{5}{2}t - 3\sin t + \frac{3}{4}\sin 2t - \sin t + \frac{1}{3}\sin^3 t]_{t=0}^{2\pi} \\
 &= 5\pi^2 r^3.
 \end{aligned}$$

2. When rotating the region about the y -axis, we compute the volume of the solid by shells.

A typical shell is shown in the following figure.



The volume of the solid is $V = \int_0^{2\pi} 2\pi x \cdot y(x) dx$. We use the parametric equations for the cycloid as a substitution for the definite integral. Thus, the volume of the solid is

$$\begin{aligned} V &= \int_0^{2\pi r} 2\pi x \cdot y(x) dx \\ &= \int_0^{2\pi} 2\pi r(t - \sin t) \cdot [r(1 - \cos t)] \cdot [r(t - \sin t)]' dt \\ &= 2\pi r^3 \int_0^{2\pi} (t - \sin t)(1 - \cos t)^2 dt. \end{aligned}$$

The function $\sin t \cdot (1 - \cos t)^2$ is a 2π periodic odd function, so

$$\int_0^{2\pi} \sin t \cdot (1 - \cos t)^2 dt = \int_{-\pi}^{\pi} \sin t \cdot (1 - \cos t)^2 dt = 0.$$

Hence,

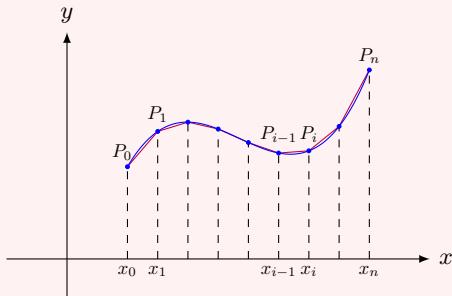
$$\begin{aligned} V &= 2\pi r^3 \int_0^{2\pi} t(1 - \cos t)^2 dt \\ &= 2\pi r^3 \int_0^{2\pi} t(1 - 2\cos t + \cos^2 t) dt \\ &= 2\pi r^3 \int_0^{2\pi} t \left[1 - 2\cos t + \frac{1}{2}(1 + \cos 2t) \right] dt \\ &= 2\pi r^3 \cdot \left[\frac{3}{4}t^2 - 2t \sin t - 2\cos t + \frac{1}{4}t \sin 2t + \frac{1}{8}\cos 2t \right]_{t=0}^{2\pi} \\ &= 6\pi^3 r^3. \end{aligned}$$

5.3 Arc Lengths

5.24 Definition: arc length

Suppose a 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, \dots, x_n and equal width Δx . Denote the endpoints of the line segments to be P_0, P_1, \dots, P_n , where P_i lies on the curve C with coordinates $(x_i, f(x_i))$. Then, the **length** L of the curve C is defined as

$$L = \lim_{n \rightarrow \infty} \sum_{i=1}^n |P_{i-1}P_i|.$$



5.25 Theorem: formulas for arc length of a smooth curve

If f' is continuous on $[a, b]$, then the length of the curve $y = f(x)$, $a \leq x \leq b$, 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 $x = g(y)$, $c \leq y \leq d$, is

$$L = \int_c^d \sqrt{1 + \left(\frac{dx}{dy}\right)^2} dy = \int_c^d \sqrt{1 + [g'(y)]^2} dy.$$

Definition: 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,

$$\begin{aligned} |P_{i-1}P_i| &= \sqrt{(\Delta x)^2 + (\Delta y_i)^2} = \sqrt{(\Delta x)^2 + [f'(x_i^*)\Delta x]^2} \\ &= \sqrt{1 + [f'(x_i^*)]^2}\Delta x. \end{aligned}$$

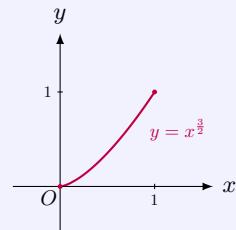
Hence,

$$\begin{aligned} L &= \lim_{n \rightarrow \infty} \sum_{i=1}^n |P_{i-1}P_i| = \lim_{n \rightarrow \infty} \sum_{i=1}^n \sqrt{1 + [f'(x_i^*)]^2}\Delta x \\ &= \int_a^b \sqrt{1 + [f'(x)]^2} dx. \end{aligned}$$

5.26 Example: arc length

Find the arc length of the curve $y = x^{\frac{3}{2}}$ from $(0, 0)$ to $(1, 1)$.

Solution The equation $y = x^{\frac{3}{2}}$ gives $y' = \frac{3}{2}x^{\frac{1}{2}}$. Thus, the arc length of the curve from $(0, 0)$ to $(1, 1)$ is



$$\begin{aligned} L &= \int_0^1 \sqrt{1 + y'^2} dx = \int_0^1 \sqrt{1 + \frac{9}{4}x^2} dx \\ &= \int_1^{\frac{13}{4}} t^{\frac{1}{2}} \cdot \frac{4}{9} dt \quad t = 1 + \frac{9}{4}x \\ &= \frac{4}{9} \left[\frac{2}{3}t^{\frac{3}{2}} \right]_{t=1}^{\frac{13}{4}} = \frac{13\sqrt{13}-8}{27}. \end{aligned}$$

Remark:

The same curve can be expressed as $x = y^{\frac{2}{3}}$, which gives $\frac{dx}{dy} = \frac{2}{3}y^{-\frac{1}{3}}$. Thus, the arc length from $(0, 0)$ to $(1, 1)$ is

$$L = \int_0^1 \sqrt{1 + \left(\frac{dx}{dy}\right)^2} dy = \int_0^1 \sqrt{1 + \frac{4}{9}y^{-\frac{2}{3}}} dy.$$

Evaluation of the last integral (an improper integral!) is technically more complicated.

So, when there are more than one choices in finding a quantity (like area, volume, arc length, etc), one may need to choose a wise setup.

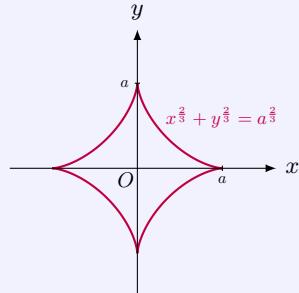
5.27 Example: arc length of astroid

Find the circumference of the astroid $x^{\frac{2}{3}} + y^{\frac{2}{3}} = a^{\frac{2}{3}}$.

Solution The astroid is symmetric about the x -axis and the y -axis. So, its circumference is four times the arc length in the first quadrant. In the first quadrant, $y \geq 0$. In this region, solving the equation of the astroid gives

$$y = \left(a^{\frac{2}{3}} - x^{\frac{2}{3}}\right)^{\frac{3}{2}},$$

where $0 \leq x \leq a$, so that



$$\frac{dy}{dx} = \frac{3}{2} \cdot \left(a^{\frac{2}{3}} - x^{\frac{2}{3}}\right)^{\frac{1}{2}} \cdot (-1)\frac{2}{3}x^{-\frac{1}{3}} = -x^{-\frac{1}{3}} \left(a^{\frac{2}{3}} - x^{\frac{2}{3}}\right)^{\frac{1}{2}}.$$

Thus, the circumference of the astroid is

$$\begin{aligned} L &= 4 \int_0^a \sqrt{1 + \left[x^{-\frac{2}{3}} \left(a^{\frac{2}{3}} - x^{\frac{2}{3}}\right)\right]} dx \\ &= 4 \int_0^a a^{\frac{1}{3}} x^{-\frac{1}{3}} dx \quad \text{an improper integral} \\ &= 4a^{\frac{1}{3}} \lim_{\epsilon \rightarrow 0^+} \int_{\epsilon}^a x^{-\frac{1}{3}} dx \\ &= 4a^{\frac{1}{3}} \lim_{\epsilon \rightarrow 0^+} \left[\frac{3}{2}x^{\frac{2}{3}}\right]_{x=\epsilon}^a = 6a. \end{aligned}$$

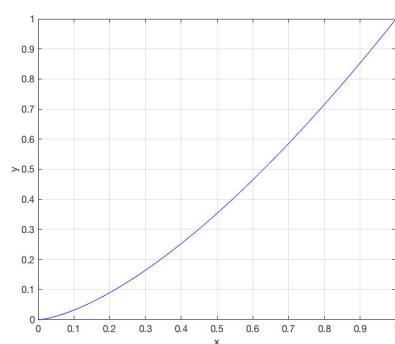
5.28 MATLAB: arc length

Find the arc length of the curve $y = x^{\frac{3}{2}}$ from $(0, 0)$ to $(1, 1)$.

Solution First, we plot the curve of $y = x^{\frac{3}{2}}$ within the interval $[0, 1]$.

```
>> syms x
>> f(x) = x^(3/2);
>> x_array = linspace(0,1,200);
>> figure
>> plot(x_array,f(x_array), 'b')
>> xlabel('x')
>> ylabel('y')
>> grid
```

The figure plotted is shown below.



To calculate the arc length of the curve in MATLAB, we provide two methods.

1. We calculate the arc length of the curve by using the arc length function.

```
>> syms x
>> f(x) = x^(3/2);
>> df(x) = diff(f(x),x)

df(x) =
(3*x^(1/2))/2

>> arc_len = int(sqrt(1 + df(x)^2),0,1)

arc_len =
(13*13^(1/2))/27 - 8/27
```

The result above shows that the arc length is $\frac{13\sqrt{13}}{27} - \frac{8}{27}$.

2. We calculate an approximation of the arc length based on the arc length definition.

```
N = 1000; % number of line segments

x_array = linspace(0,1,N);
y_array = x_array.^(3/2);
arc_len = 0;

for ind = 2:length(x_array)

    x2 = x_array(ind);
    x1 = x_array(ind - 1);
    y2 = y_array(ind);
    y1 = y_array(ind - 1);
    seg_len = sqrt((y2 - y1)^2 + (x2 - x1)^2);
    arc_len = arc_len + seg_len;

end
```

Note that the bigger the number of line segments N is, the more accurate the approximation is.

The result of the arc length can be seen when we type in `arc_len` in the command window, as shown below.

```
>> format long
>> arc_len

arc_len =
1.439709685755622
```

It is shown that when $N = 1000$, the approximation is 1.439709685755622, which is very close to the exact result $\frac{13\sqrt{13}-8}{27} \approx 1.439709873371550$.

Arc Length of Parametric Curve

5.29 Theorem: arc length for parametric curve

Suppose a parametric curve is given by the parametric equations

$$x = f(t), \quad y = g(t), \quad \alpha \leq t \leq \beta,$$

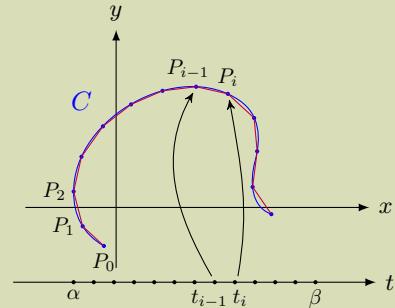
where f' and g' are continuous and the parametric curve is traversed exactly once as t

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 $t_0, t_1, t_2, \dots, t_n$ 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, \dots, P_n approximates C .



The length L of C is defined as the limit of the lengths of these approximating polygons as $n \rightarrow \infty$:

$$L = \lim_{n \rightarrow \infty} \sum_{i=1}^n |P_{i-1}P_i|.$$

Justification: Part 2

By the Mean Value Theorem, there are t_i^* and t_i^{**} in (t_{i-1}, t_i) such that

$$\begin{aligned} \Delta x_i &= f(t_i) - f(t_{i-1}) = f'(t_i^*) \Delta t, \\ \Delta y_i &= g(t_i) - g(t_{i-1}) = g'(t_i^{**}) \Delta t. \end{aligned}$$

Thus,

$$\begin{aligned} |P_{i-1}P_i| &= \sqrt{(\Delta x_i)^2 + (\Delta y_i)^2} = \sqrt{[f'(t_i^*)]^2 \Delta t + [g'(t_i^{**})]^2 \Delta t} \\ &= \sqrt{[f'(t_i^*)]^2 + [g'(t_i^{**})]^2} \Delta t, \end{aligned}$$

and so

$$L = \lim_{n \rightarrow \infty} \sum_{i=1}^n \sqrt{[f'(t_i^*)]^2 + [g'(t_i^{**})]^2} \Delta t.$$

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,

$$L = \int_{\alpha}^{\beta} \sqrt{[f'(t)]^2 + [g'(t)]^2} dt.$$

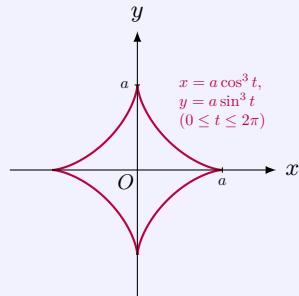
5.30 Example: arc length of a parametric curve

Find the circumference of the astroid, given by the parametric equations

$$x = a \cos^3 t, \quad y = a \sin^3 t,$$

where $0 \leq t \leq 2\pi$.

Solution The astroid is symmetric about the x -axis and the y -axis. So, its circumference is four times the arc length in the first quadrant. Thus,



$$\begin{aligned} L &= 4 \int_0^{\frac{\pi}{2}} \sqrt{[3a \cos^2 t \cdot (-\sin t)]^2 + [3a \sin^2 t \cdot (\cos t)]^2} dt \\ &= 12a \int_0^{\frac{\pi}{2}} \sin t \cos t dt \\ &= 12a \left[\frac{1}{2} \sin^2 t \right]_{t=0}^{\frac{\pi}{2}} = 6a. \end{aligned}$$

5.31 MATLAB: arc length of a parametric curve

Find the circumference of the astroid, given by the parametric equations

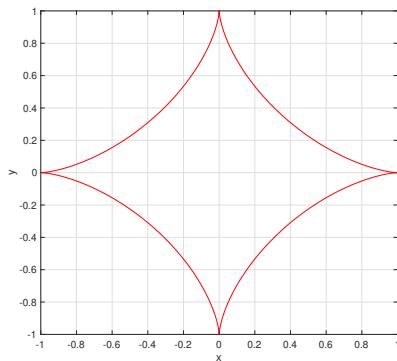
$$x = \cos^3 t, \quad y = \sin^3 t,$$

where $0 \leq t \leq 2\pi$.

Solution First, we plot the curve of $x = \cos^3 t$, $y = \sin^3 t$ where $0 \leq t \leq 2\pi$.

```
>> syms t
>> x(t) = cos(t)^3;
>> y(t) = sin(t)^3;
>> t_array = linspace(0, 2*pi, 200);
>> figure
>> plot(x(t_array), y(t_array), 'r')
>> xlabel('x')
>> ylabel('y')
>> grid
```

The figure plotted is shown below.



We see from the figure above that the plotted astroid is symmetric about the x -axis and the y -axis. So, its circumference is four times the arc length in the first quadrant. We thus first calculate the arc length of the curve in the first quadrant, and then multiply it by 4.

To calculate the arc length of the astroid in MATLAB, we provide two methods.

1. We calculate the arc length of the astroid by using the arc length function.

```
>> syms t
>> x(t) = (cos(t))^3;
>> y(t) = (sin(t))^3;
>> dx(t) = diff(x(t),t)

dx(t) =
-3*cos(t)^2*sin(t)

>> dy(t) = diff(y(t),t)

dy(t) =
3*cos(t)*sin(t)^2

>> arc_len = 4*int(sqrt((dx(t))^2 + (dy(t))^2),0,pi/2)

arc_len =
6
```

The result above shows that the arc length of the astroid is 6.

2. We calculate an approximation of the arc length based on the arc length definition.

```
N = 1000; % number of line segments
```

```

t_array = linspace(0,pi/2,N);
x_array = cos(t_array).^3;
y_array = sin(t_array).^3;
arc_len1 = 0; % arc length in the 1st quadrant

for ind = 2:length(t_array)

    x2 = x_array(ind);
    x1 = x_array(ind - 1);
    y2 = y_array(ind);
    y1 = y_array(ind - 1);
    seg_len = sqrt((x2 - x1)^2 + (y2 - y1)^2);
    arc_len1 = arc_len1 + seg_len;

end

arc_len = 4*arc_len1;

```

Note that the bigger the number of line segments N is, the more accurate the approximation is.

The result of the arc length can be seen when we type in `arc_len` in the command window, as shown below.

```

>> format long
>> arc_len

arc_len =
5.999999381918176

```

It is shown that when $N = 1000$, the approximation is 5.999999381918176, which is very close to the exact result 6.

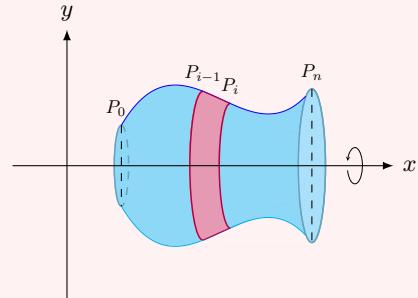
5.32 Definition: surface area of revolution

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, \dots, 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

$$S = \lim_{n \rightarrow \infty} \sum_{i=1}^n S_i,$$

where S_i is the area of the band with slant height $|P_{i-1}P_i|$.



5.33 Theorem: 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

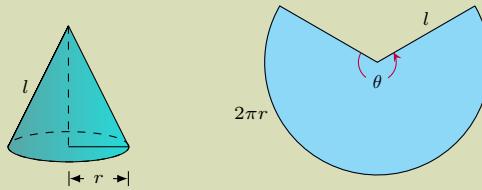
$$S = \int_a^b 2\pi y(x) \sqrt{1 + \left(\frac{dy}{dx}\right)^2} dx = \int_a^b 2\pi f(x) \sqrt{1 + [f'(x)]^2} dx.$$

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

$$A = \frac{1}{2}l^2\theta = \frac{1}{2}l^2 \cdot (2\pi r/l) = \pi rl.$$

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

$$A = \pi r_2(l_1 + l) - \pi r_1 l_1 = \pi [(r_2 - r_1)l_1 + r_2 l].$$

From similar triangles we have

$$\frac{l_1}{r_1} = \frac{l_1 + l}{r_2}$$

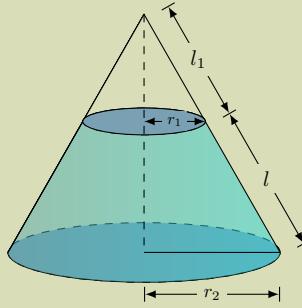
which gives

$$r_2 l_1 = r_1 l_1 + r_1 l \quad \text{or} \quad (r_2 - r_1)l_1 = r_1 l.$$

Hence,

$$A = \pi [r_1 l + r_2 l] = 2\pi r l = (\text{average circumference}) \times (\text{slant height}),$$

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

$$S_i = 2\pi \cdot \frac{1}{2}(f(x_{i-1}) + f(x_i)) \cdot |P_{i-1}P_i|.$$

By the Pythagorean Theorem and the Mean Value Theorem,

$$\begin{aligned} |P_{i-1}P_i| &= \sqrt{(x_i - x_{i-1})^2 + [f(x_i) - f(x_{i-1})]^2} \\ &= \sqrt{(\Delta x)^2 + [f'(x_i^*)\Delta x]^2} = \sqrt{1 + [f'(x_i^*)]^2}\Delta x. \end{aligned}$$

Thus, by approximating $\frac{1}{2}(f(x_{i-1}) + f(x_i))$ using $f(x_i^*)$, we get

$$S_i \approx 2\pi \cdot f(x_i^*) \sqrt{1 + [f'(x_i^*)]^2} \Delta x.$$

Therefore,

$$\begin{aligned} S &= \lim_{n \rightarrow \infty} \sum_{i=1}^n S_i = \lim_{n \rightarrow \infty} \sum_{i=1}^n 2\pi \cdot f(x_i^*) \sqrt{1 + [f'(x_i^*)]^2} \Delta x \\ &= \int_a^b 2\pi f(x) \sqrt{1 + [f'(x)]^2} dx. \end{aligned}$$

5.34 Theorem: general formula for surface area of revolution

For rotations about the x -axis and the y -axis, respectively, the surface area formula becomes

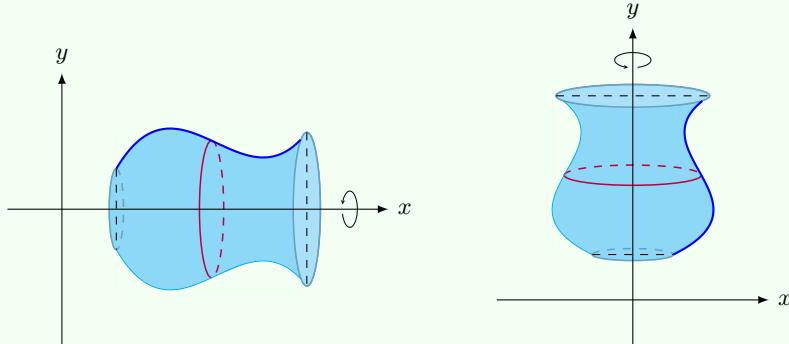
$$S = \int 2\pi y \, ds \quad \text{or} \quad S = \int 2\pi x \, ds,$$

where

$$ds = \sqrt{1 + \left(\frac{dy}{dx}\right)^2} dx \quad \text{or} \quad \sqrt{1 + \left(\frac{dx}{dy}\right)^2} dy.$$

In general, the surface area of rotation is

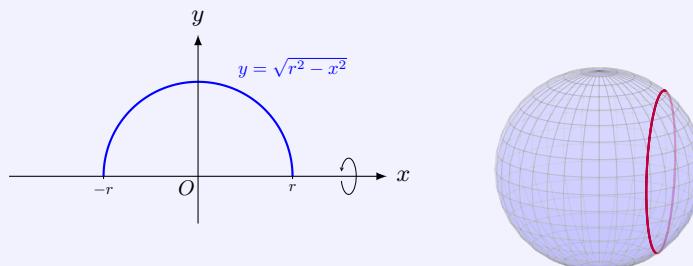
$$\int \underbrace{(2\pi r)}_{\text{circumference}} \underbrace{(ds)}_{\text{arc length}}$$

**5.35 Example:** surface area of a sphere

A sphere can be generated by rotating the semicircle

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

about the x -axis, as shown in the figure. Find its surface area.



Solution It is easy to see that

$$\frac{dy}{dx} = -\frac{x}{\sqrt{r^2 - x^2}}.$$

Thus, the surface area of the sphere is

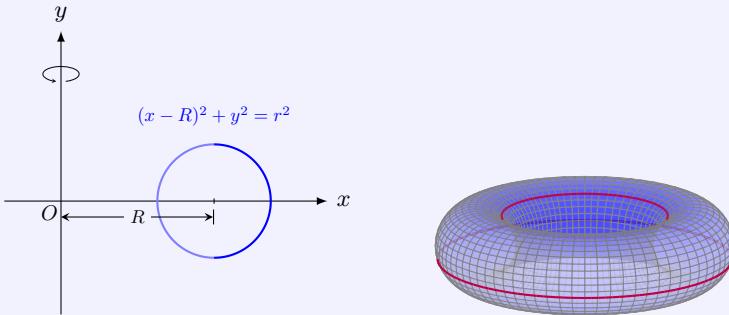
$$\begin{aligned} S &= \int_{-r}^r 2\pi\sqrt{r^2 - x^2} \sqrt{1 + \left(-\frac{x}{\sqrt{r^2 - x^2}}\right)^2} dx \\ &= 2\pi \int_{-r}^r r dx = 4\pi r^2. \end{aligned}$$

5.36 Example: surface area of a torus

A torus can be generated by rotating the circle

$$(x - R)^2 + y^2 = r^2, \quad 0 < r < R,$$

about the y -axis, as shown in the figure.



The torus consists of two portions: the outer portion generated by the rightmost semicircle $x = R + \sqrt{r^2 - y^2}$ ($-r \leq y \leq r$), and the inner portion by the leftmost semicircle $x = R - \sqrt{r^2 - y^2}$ ($-r \leq y \leq r$). It is easy to see that

$$\begin{aligned} \text{along the rightmost semicircle : } \frac{dx}{dy} &= -\frac{y}{\sqrt{r^2 - y^2}}; \\ \text{along the leftmost semicircle : } \frac{dx}{dy} &= \frac{y}{\sqrt{r^2 - y^2}}. \end{aligned}$$

Thus, the surface area of the torus is

$$\begin{aligned}
 S &= \int_{-r}^r 2\pi \left(R + \sqrt{r^2 - y^2} \right) \sqrt{1 + \left(-\frac{y}{\sqrt{r^2 - y^2}} \right)^2} dy \\
 &\quad + \int_{-r}^r 2\pi \left(R - \sqrt{r^2 - y^2} \right) \sqrt{1 + \left(\frac{y}{\sqrt{r^2 - y^2}} \right)^2} dy \\
 &= 4\pi R \int_{-r}^r \frac{r}{\sqrt{r^2 - y^2}} dy \\
 &= 8\pi R \int_0^r \frac{r}{\sqrt{r^2 - y^2}} dy \quad \text{even function} \\
 &= 8\pi R r \int_0^{\frac{\pi}{2}} \frac{r \cos \theta d\theta}{r \cos \theta} \quad y = r \sin \theta \\
 &= 8\pi R r \cdot \frac{\pi}{2} = 4\pi^2 R r.
 \end{aligned}$$

5.37 MATLAB: surface area of a sphere

A sphere can be generated by rotating the semicircle

$$y = \sqrt{3^2 - x^2}, \quad -3 \leq x \leq 3,$$

about the x -axis. Find its surface area.

Solution We calculate the surface area of the sphere by two methods.

1. The surface area of the sphere is calculated by using the formula for surface area of revolution.

```

>> syms x
>> y(x) = sqrt(3^2 - x^2);
>> dy(x) = diff(y(x),x)

dy(x) =
-x/(9 - x^2)^(1/2)

>> int_fun = 2*pi*y(x)*sqrt(1 + dy(x)^2); % integrand
>> S = int(int_fun,-3,3)

S =
36*pi

```

The result above shows that the surface area of the sphere is 36π .

2. An approximation of the surface area is calculated based on the definition.

```
R = 3; % radius of the sphere
N = 1000; % number of line segments

x_array = linspace(-R,R,N);
y_array = sqrt(3^2 - x_array.^2);
S_sum = 0;

for ind = 2:length(x_array)

    x2 = x_array(ind);
    x1 = x_array(ind - 1);
    y2 = y_array(ind);
    y1 = y_array(ind - 1);
    ell = sqrt((x2 - x1)^2 + (y2 - y1)^2);
    S = 2*pi*0.5*(y1 + y2)*ell;
    S_sum = S_sum + S;

end
```

Note that the bigger the number of line segments N is, the more accurate the approximation is.

The result of the surface area can be seen when we type in `S_sum` in the command window, as shown below.

```
>> format long
>> S_sum

S_sum =
1.130970267960428e+02
```

It is shown that when $N = 1000$, the approximation is 113.0970267960428 , which is very close to the exact result $36\pi \approx 113.0973355292326$.

Surface Area of Revolution by Parametric Curve

5.38 Theorem:

surface area of revolution by parametric curve

Suppose a parametric curve is given by the parametric equations

$$x = f(t), \quad y = g(t), \quad \alpha \leq t \leq \beta,$$

where f' and g' are continuous and $g(t) \geq 0$. Then the surface area of revolution about the

x -axis is

$$S = \int_{\alpha}^{\beta} 2\pi y \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} dt = \int_{\alpha}^{\beta} 2\pi g(t) \sqrt{[f'(t)]^2 + [g'(t)]^2} dt.$$

Similarly, the surface area of revolution about the y -axis is

$$S = \int_{\alpha}^{\beta} 2\pi x \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} dt = \int_{\alpha}^{\beta} 2\pi f(t) \sqrt{[f'(t)]^2 + [g'(t)]^2} dt.$$

Justification:

The general formula for the surface area 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

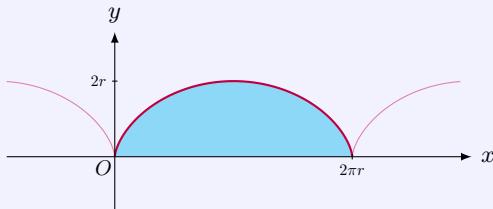
$$S = \int_{\alpha}^{\beta} 2\pi g(t) \sqrt{[f'(t)]^2 + [g'(t)]^2} dt.$$

5.39 Example: solid generated by a parametric curve

Rotate the area under one arch of the cycloid

$$x = r(t - \sin t), \quad y = r(1 - \cos t), \quad 0 \leq t \leq 2\pi,$$

about the x -axis and the y -axis, respectively. Find the surface areas of the two solids.

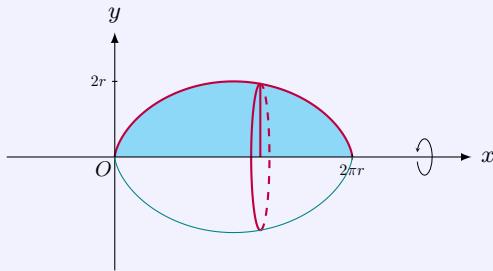


Solution Generally, the surface area of revolution is

$$\int \underbrace{(2\pi r)}_{\text{circumference}} \underbrace{(ds)}_{\text{arc length}} .$$

1. When the area under one arch of the cycloid rotates about the x -axis, a typical circle has radius $y = y(t)$, while the arc length is given by

$$ds = \sqrt{x'(t)^2 + y'(t)^2} dt.$$

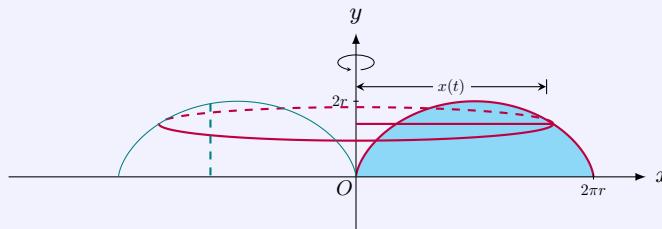


Thus, the surface area of the solid is

$$\begin{aligned} S &= \int_0^{2\pi} 2\pi y(t) \cdot \sqrt{x'(t)^2 + y'(t)^2} dt \\ &= \int_0^{2\pi} 2\pi \cdot r(1 - \cos t) \cdot \sqrt{[r(t - \sin t)]'^2 + [r(1 - \cos t)]'^2} dt \\ &= 2\pi r^2 \int_0^{2\pi} (1 - \cos t) \sqrt{(1 - \cos t)^2 + \sin^2 t} dt \\ &= 2\pi r^2 \int_0^{2\pi} (1 - \cos t) \cdot \sqrt{2 - 2 \cos t} dt \\ &= 2\pi r^2 \int_0^{2\pi} 2 \sin^2(\frac{1}{2}t) \cdot 2 \sin(\frac{1}{2}t) dt && \text{half-angle formula} \\ &= 8\pi r^2 \int_0^\pi (1 - \cos^2 u) \sin u \cdot 2 du && t = 2u \\ &= 16\pi r^2 [-\cos u + \frac{1}{3} \cos^3 u]_{u=0}^\pi = 16\pi r^2 \cdot \frac{4}{3} = \frac{64}{3}\pi r^2. \end{aligned}$$

2. When the area under one arch of the cycloid rotates about the y -axis, a typical circle has radius $x = x(t)$, while the arc length is given by

$$ds = \sqrt{x'(t)^2 + y'(t)^2} dt.$$



Thus, the surface area of the solid is

$$\begin{aligned}
 S &= \int_0^{2\pi} 2\pi x(t) \cdot \sqrt{x'(t)^2 + y'(t)^2} dt \\
 &= \int_0^{2\pi} 2\pi \cdot r(t - \sin t) \cdot \sqrt{[r(t - \sin t)]'^2 + [r(1 - \cos t)]'^2} dt \\
 &= 2\pi r^2 \int_0^{2\pi} (t - \sin t) \sqrt{(1 - \cos t)^2 + \sin^2 t} dt \\
 &= 2\pi r^2 \int_0^{2\pi} (t - \sin t) \cdot \sqrt{2 - 2 \cos t} dt \\
 &= 2\pi r^2 \int_0^{2\pi} (t - \sin t) \cdot 2 \sin(\frac{1}{2}t) dt && \text{half-angle formula} \\
 &= 4\pi r^2 \int_0^{\pi} (2u - \sin 2u) \sin u \cdot 2 du && t = 2u \\
 &= 8\pi r^2 \int_0^{\pi} (2u \sin u - \sin 2u \sin u) du \\
 &= 8\pi r^2 [-2u \cos u + \frac{3}{2} \sin u + \frac{1}{6} \sin 3u]_{u=0}^{\pi} = 8\pi r^2 \cdot 2\pi = 16\pi^2 r^2.
 \end{aligned}$$

5.40 MATLAB: solid generated by a parametric curve

Rotate the area under one arch of the cycloid

$$x = 3(t - \sin t), \quad y = 3(1 - \cos t), \quad 0 \leq t \leq 2\pi,$$

about the x -axis. Find the surface area of the generated solid.

Solution We calculate the surface area of the solid by two methods.

1. The surface area of the solid is calculated by using the formula for surface area of revolution.

```

>> syms t
>> x(t) = 3*(t - sin(t));
>> y(t) = 3*(1 - cos(t));
>> dx(t) = diff(x(t),t);
>> dy(t) = diff(y(t),t);
>> int_fun = 2*pi*y(t)*sqrt(dx(t)^2 + dy(t)^2)

int_fun =
-6*pi*(3*cos(t) - 3)*(sin(t)^2 + (3*cos(t) - 3)^2/9)^(1/2)

>> int_fun2 = matlabFunction(int_fun); % convert to ...
    function handle
>> S = integral(int_fun2,0,2*pi)

```

```
S =
603.1858
```

The result above shows that the surface area of the solid is 603.1858. Note that the command `integral` can only deal with a function handle, so we need to convert the symbolic expression `int_fun` to the function handle `int_fun2` using command `matlabFunction`.

2. An approximation of the surface area is calculated based on the definition.

```
N = 1000; % number of line segments

t_array = linspace(0,2*pi,N);
x_array = 3*(t_array - sin(t_array));
y_array = 3*(1 - cos(t_array));
S_sum = 0;

for ind = 2:length(x_array)

    x2 = x_array(ind);
    x1 = x_array(ind - 1);
    y2 = y_array(ind);
    y1 = y_array(ind - 1);
    ell = sqrt((x2 - x1)^2 + (y2 - y1)^2);
    S = 2*pi*0.5*(y1 + y2)*ell;
    S_sum = S_sum + S;

end
```

Note that the bigger the number of line segments `N` is, the more accurate the approximation is.

The result of the surface area can be seen when we type in `S_sum` in the command window, as shown below.

```
>> S_sum
S_sum =
603.1845
```

It is shown that when $N = 1000$, the approximation is 603.1845.

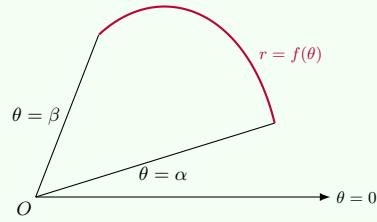
5.5 Lengths and Areas in Polar Coordinates

Arc Length of Polar Curve

5.41 Theorem: 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

$$\begin{aligned} L &= \int_{\alpha}^{\beta} \sqrt{[f(\theta)]^2 + [f'(\theta)]^2} d\theta \\ &= \int_{\alpha}^{\beta} \sqrt{r^2 + (dr/d\theta)^2} d\theta. \end{aligned}$$



Justification:

We write its parametric equations as

$$x = r \cos \theta = f(\theta) \cos \theta, \quad y = r \sin \theta = f(\theta) \sin \theta.$$

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

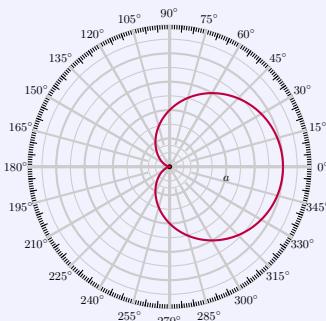
$$L = \int_{\alpha}^{\beta} \sqrt{\left(\frac{dx}{d\theta} \right)^2 + \left(\frac{dy}{d\theta} \right)^2} d\theta = \int_{\alpha}^{\beta} \sqrt{[f(\theta)]^2 + [f'(\theta)]^2} d\theta.$$

5.42 Example: arc length of polar curve

Find the arc length of the cardioid

$$r = a + a \cos \theta,$$

where $a > 0$.



Solution The given polar curve $r = a + a \cos \theta$ is periodic of period 2π . Since $r' = -a \sin \theta$, we get the arc length

$$\begin{aligned} L &= \int_0^{2\pi} \sqrt{(a + a \cos \theta)^2 + (-a \sin \theta)^2} d\theta \\ &= \int_0^{2\pi} a \sqrt{2 + 2 \cos \theta} d\theta \\ &= 2a \int_0^{2\pi} |\cos(\frac{1}{2}\theta)| d\theta \\ &= 2a \int_0^{\pi} \cos(\frac{1}{2}\theta) d\theta + 2a \int_{\pi}^{2\pi} (-1) \cos(\frac{1}{2}\theta) d\theta \\ &= 2a [2 \sin(\frac{1}{2}\theta)]_{\theta=0}^{\pi} - 2a [2 \sin(\frac{1}{2}\theta)]_{\theta=\pi}^{2\pi} = 8a. \end{aligned}$$

5.43 MATLAB: arc length of polar curve

Find the arc length of the cardioid

$$r = 1 + \cos \theta.$$

The arc length of the cardioid can be calculated in MATLAB using the following code:

```
>> syms theta
>> r(theta) = 1 + cos(theta);
>> dr(theta) = diff(r(theta),theta)

dr(theta) =
-sin(theta)

>> int_fun = sqrt(r(theta)^2 + dr(theta)^2);
>> arc_len = int(int_fun,theta,0,2*pi)

arc_len =
```

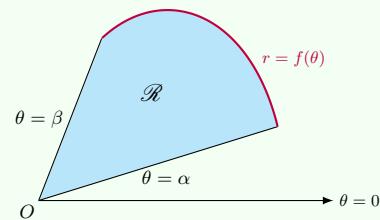
8

The result above shows that the arc length of the cardioid is 8.

5.44 Theorem: area of polar region

Let \mathcal{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 \mathcal{R} is

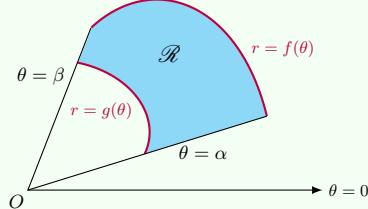
$$A = \int_{\alpha}^{\beta} \frac{1}{2} r^2 d\theta = \int_{\alpha}^{\beta} \frac{1}{2} [f(\theta)]^2 d\theta.$$



Theorem: area between two polar curves

Let \mathcal{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 \mathcal{R} is

$$A = \frac{1}{2} \int_{\alpha}^{\beta} ([f(\theta)]^2 - [g(\theta)]^2) d\theta.$$



Justification:

For the polar curve $r = f(\theta)$, we write its parametric equations as

$$x = r \cos \theta = f(\theta) \cos \theta, \quad y = r \sin \theta = f(\theta) \sin \theta.$$

We divide the interval $[\alpha, \beta]$ into subintervals of equal width $\Delta\theta$ with endpoints $\theta_0, \theta_1, \theta_2, \dots, \theta_n$. If we choose θ_i^* in the i th subinterval $[\theta_{i-1}, \theta_i]$, then the area ΔA_i of the i th 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

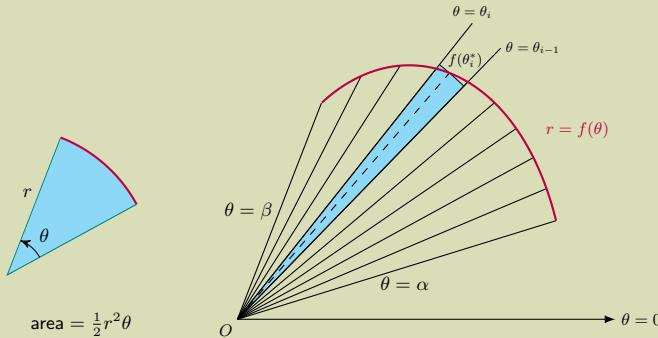
$$\begin{aligned}\Delta A_i &\approx \pi(\text{radius})^2 \cdot \frac{(\text{central angle})}{2\pi} = \frac{1}{2}(\text{radius})^2 \times (\text{central angle}) \\ &= \frac{1}{2} [f(\theta_i^*)]^2 \Delta\theta.\end{aligned}$$

So, an approximation to the total area A of \mathcal{R} is

$$A \approx \sum_{i=1}^n \frac{1}{2} [f(\theta_i^*)]^2 \Delta\theta.$$

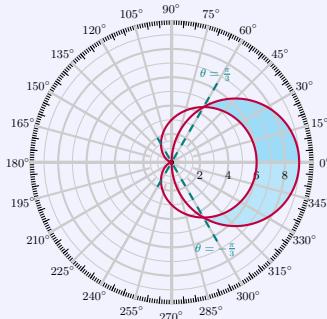
Hence, we obtain the formula for the area A of the polar region \mathcal{R}

$$A = \lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{1}{2} [f(\theta_i^*)]^2 \Delta\theta = \int_{\alpha}^{\beta} \frac{1}{2} [f(\theta)]^2 d\theta.$$



5.45 Example: area between two polar curves

Find the area of the region that lies inside the circle $r = 9 \cos \theta$ outside of the cardioid $r = 3 + 3 \cos \theta$.



Solution The circle $r = 9 \cos \theta$ makes a complete rotation when θ varies from $-\frac{\pi}{2}$ to $\frac{\pi}{2}$;

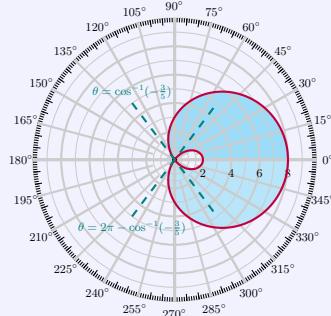
the cardioid $r = 3 + 3 \cos \theta$ make a complete loop when θ varies from $-\pi$ to π . By solving the simultaneous equations $r = 9 \cos \theta$ and $r = 3 + 3 \cos \theta$, we find the intersections at $\theta = -\frac{\pi}{3}$ and $\theta = \frac{\pi}{3}$. Since $r(-\theta) = r(\theta)$ for both polar equations, these curves are symmetric about the polar axis.

Thus, the area of the region that lies inside the circle outside of the cardioid is

$$\begin{aligned} A &= 2 \int_0^{\frac{\pi}{3}} \frac{1}{2} [(9 \cos \theta)^2 - (3 + 3 \cos \theta)^2] d\theta \\ &= \int_0^{\frac{\pi}{3}} (-9 - 18 \cos \theta + 72 \cos^2 \theta) d\theta \\ &= \int_0^{\frac{\pi}{3}} [-9 - 18 \cos \theta + 36(1 + \cos 2\theta)] d\theta \\ &= [27\theta - 18 \sin \theta + 18 \sin 2\theta]_{\theta=0}^{\frac{\pi}{3}} \\ &= 9\pi. \end{aligned}$$

5.46 Example: area between two polar curves

The limaçon $r = 3 + 5 \cos \theta$ consists of an inner loop and an outer loop. Find the area between these loops.



Solution The given polar curve $r = 3 + 5 \cos \theta$ is periodic of period 2π . Since $r(-\theta) = r(\theta)$, the polar graph is symmetric to the polar axis. As θ varies from 0 to $\arccos(-\frac{3}{5})$, the polar equation describes the upper-half outer loop; as θ varies from π to $2\pi - \arccos(-\frac{3}{5})$, it describes the upper-half inner loop. Thus, the area between these loops is

$$A = 2 \int_0^{\arccos(-\frac{3}{5})} \frac{1}{2}(3 + 5 \cos \theta)^2 d\theta - 2 \int_{\pi}^{2\pi - \arccos(-\frac{3}{5})} \frac{1}{2}(3 + 5 \cos \theta)^2 d\theta.$$

We can compute the following indefinite integral:

$$\begin{aligned} \int (3 + 5 \cos \theta)^2 d\theta &= \int (9 + 30 \cos \theta + 25 \cos^2 \theta) d\theta \\ &= \int [9 + 30 \cos \theta + \frac{25}{2}(1 + \cos 2\theta)] d\theta \\ &= \frac{43}{2}\theta + 30 \sin \theta + \frac{25}{4} \sin 2\theta + C. \end{aligned}$$

Hence,

$$\begin{aligned}
 A &= \left[\frac{43}{2}\theta + 30\sin\theta + \frac{25}{4}\sin 2\theta \right]_{\theta=0}^{\arccos(-\frac{3}{5})} \\
 &\quad - \left[\frac{43}{2}\theta + 30\sin\theta + \frac{25}{4}\sin 2\theta \right]_{\theta=\pi}^{2\pi-\arccos(-\frac{3}{5})} \\
 &= \frac{43}{2} \cdot \arccos(-\frac{3}{5}) + 30 \cdot (\frac{4}{5}) + \frac{25}{4} \cdot 2(\frac{4}{5})(-\frac{3}{5}) \\
 &\quad - \frac{43}{2} \cdot [2\pi - \arccos(-\frac{3}{5})] - 30 \cdot (-\frac{4}{5}) - \frac{25}{4} \cdot 2(-\frac{4}{5})(\frac{3}{5}) + \frac{43}{2} \cdot \pi \\
 &= 36 - \frac{43}{2}\pi + 43 \cdot \arccos(-\frac{3}{5}) \\
 &\approx 63.6705.
 \end{aligned}$$

5.47 MATLAB: area between two polar curves

Find the area of the region that lies inside the circle $r = 9\cos\theta$ outside of the cardioid $r = 3 + 3\cos\theta$.

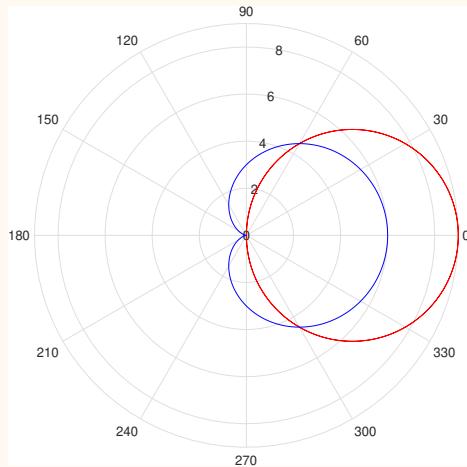
First, we plot these two polar curves in MATLAB.

```

>> syms theta
>> r1(theta) = 9*cos(theta);
>> r2(theta) = 3 + 3*cos(theta);
>> theta_array = linspace(-pi,pi,200);
>> figure
>> polarplot(theta_array,r1(theta_array), 'r')
>> hold on
>> polarplot(theta_array,r2(theta_array), 'b')

```

The figure plotted is shown below.



In order to calculate the area of the specified region, we need to first find the intersections of these two polar curves, which is shown in the following:

```
>> solve(r1(theta) == r2(theta))

ans =

-pi/3
pi/3
```

We see that the circle $r = 9 \cos \theta$ intersects with the cardioid $r = 3 + 3 \cos \theta$ at $\theta = -\frac{\pi}{3}$ and $\theta = \frac{\pi}{3}$. Then, we can calculate the area of the specified region below.

```
>> int_fun = 0.5*(r1(theta)^2 - r2(theta)^2);
>> S = int(int_fun,-pi/3,pi/3)

S =

9*pi
```

The area of the specified region is thus 9π .

5.48 Theorem: moment and center of mass for discrete system

Theorem: moment and center of mass for one-dimensional discrete system

For a system of n particles with masses m_1, m_2, \dots, m_n located at the points x_1, x_2, \dots, x_n on the x -axis, the center of mass of the system is located at

$$\bar{x} = \frac{M}{m} = \frac{\sum_{i=1}^n m_i x_i}{\sum_{i=1}^n m_i},$$

where $M = \sum_{i=1}^n m_i x_i$ is the **moment of the system about the origin**.

Theorem: moments and center of mass for two-dimensional discrete system

For a system of n particles with masses m_1, m_2, \dots, m_n located at the points $(x_1, y_1), (x_2, y_2) \dots, (x_n, y_n)$ in the xy -plane, the center of mass of the system is located at (\bar{x}, \bar{y}) , where

$$\bar{x} = \frac{M_y}{m} = \frac{\sum_{i=1}^n m_i x_i}{\sum_{i=1}^n m_i}, \quad \bar{y} = \frac{M_x}{m} = \frac{\sum_{i=1}^n m_i y_i}{\sum_{i=1}^n m_i}.$$

Here $M_y = \sum_{i=1}^n m_i x_i$ is the **moment of the system about the y -axis** and $M_x = \sum_{i=1}^n m_i y_i$ is the **moment of the system about the x -axis**.

Justification:

Consider a situation with two particles with masses m_1 and m_2 on a rod that lies along x -axis. Suppose the distances from the masses to fulcrum are d_1 and d_2 respectively. The rod will balance if

$$m_1 d_1 = m_2 d_2.$$



Denote the coordinates of m_1 and m_2 to be x_1 and x_2 , and the center of mass at \bar{x} . We see that $d_1 = \bar{x} - x_1$ and $d_2 = x_2 - \bar{x}$. Then we have

$$m_1(\bar{x} - x_1) = m_2(x_2 - \bar{x}) \implies \bar{x} = \frac{m_1 x_1 + m_2 x_2}{m_1 + m_2}.$$

If we have a system of n particles with masses m_1, m_2, \dots, m_n located at the points x_1, x_2, \dots, x_n on the x -axis, and if the rod balances, then the center of mass \bar{x} satisfies:

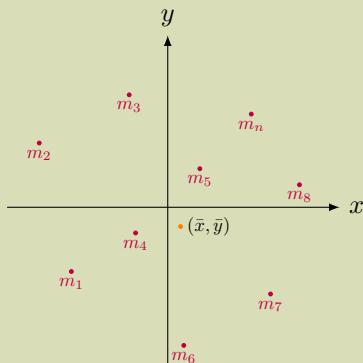
$$\sum_{i=1}^n m_i(x_i - \bar{x}) = 0 \implies \bar{x} = \frac{\sum_{i=1}^n m_i x_i}{\sum_{i=1}^n m_i}.$$

For a system of n particles with masses m_1, m_2, \dots, m_n located at the points $(x_1, y_1), (x_2, y_2) \dots, (x_n, y_n)$ in the xy -plane, we assume that the center of mass of the system is located at (\bar{x}, \bar{y}) . Then, similarly, we have

$$\begin{cases} \sum_{i=1}^n m_i(x_i - \bar{x}) = 0, \\ \sum_{i=1}^n m_i(y_i - \bar{y}) = 0, \end{cases}$$

so that

$$\bar{x} = \frac{M_y}{m} = \frac{\sum_{i=1}^n m_i x_i}{\sum_{i=1}^n m_i}, \quad \bar{y} = \frac{M_x}{m} = \frac{\sum_{i=1}^n m_i y_i}{\sum_{i=1}^n m_i}.$$



Center of Mass Along a Curve

5.49 Theorem: center of mass along a curve

Consider a continuous system of mass distributed along a thin curve \mathcal{C} : $y = f(x)$, $a \leq x \leq b$, with density $\rho(x)$. Then the **center of mass** is (\bar{x}, \bar{y}) , where

$$\bar{x} = \frac{1}{m} \int_a^b \rho(x) \cdot x \sqrt{1 + [f'(x)]^2} dx,$$

$$\bar{y} = \frac{1}{m} \int_a^b \rho(x) \cdot f(x) \sqrt{1 + [f'(x)]^2} dx,$$

and $m = \int_a^b \rho(x) \cdot \sqrt{1 + [f'(x)]^2} dx$ is the total mass along the curve \mathcal{C} .

Theorem: centroid of a curve

In particular, if $\rho(x) = \rho$ is a constant, then the center of mass is the same as the **centroid** of the curve, with

$$\bar{x} = \frac{1}{L} \int_a^b x \sqrt{1 + [f'(x)]^2} dx,$$

$$\bar{y} = \frac{1}{L} \int_a^b f(x) \sqrt{1 + [f'(x)]^2} dx,$$

where $L = \int_a^b \sqrt{1 + [f'(x)]^2} dx$ is the length of the curve \mathcal{C} .

Justification:

We divide the interval $[a, b]$ into n subintervals with endpoints x_0, x_1, \dots, x_n and equal width Δx . Then the i th arc can be approximated by a line segment from $(x_{i-1}, f(x_{i-1}))$ to $(x_i, f(x_i))$, whose approximate length is, by the Mean value Theorem,

$$\begin{aligned}\sqrt{(\Delta x)^2 + (f(x_i) - f(x_{i-1}))^2} &= \sqrt{1 + \left[\frac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}} \right]^2} \Delta x \\ &= \sqrt{1 + [f'(x_i^*)]^2} \Delta x.\end{aligned}$$

The mass along the i th piece of arc is approximately

$$\rho(x_i^*) \sqrt{1 + [f'(x_i^*)]^2} \Delta x.$$

so the moment about the y -axis is

$$\rho(x_i^*) \sqrt{1 + [f'(x_i^*)]^2} \Delta x \cdot x_i^*.$$

Adding these moments, we obtain the approximate moment of \mathcal{C} . By taking the limit as $n \rightarrow \infty$ we obtain the moment of \mathcal{C} about the y -axis:

$$\begin{aligned}M_y &= \lim_{n \rightarrow \infty} \sum_{i=1}^n \rho(x_i^*) x_i^* \sqrt{1 + [f'(x_i^*)]^2} \Delta x \\ &= \int_a^b \rho(x) \cdot x \sqrt{1 + [f'(x)]^2} dx.\end{aligned}$$

In a similar fashion, the moment along the i th piece of arc about the x -axis is

$$\rho(x_i^*) \sqrt{1 + [f'(x_i^*)]^2} \Delta x \cdot f(x_i^*)$$

so we get the moment about the x -axis:

$$\begin{aligned}M_x &= \lim_{n \rightarrow \infty} \sum_{i=1}^n \rho(x_i^*) f(x_i^*) \sqrt{1 + [f'(x_i^*)]^2} \Delta x \\ &= \int_a^b \rho(x) \cdot f(x) \sqrt{1 + [f'(x)]^2} dx.\end{aligned}$$

The mass along the i th piece of arc is

$$\rho(x_i^*) \sqrt{1 + [f'(x_i^*)]^2} \Delta x,$$

so we get the total mass of the curve \mathcal{C} :

$$m = \int_a^b \rho(x) \cdot \sqrt{1 + [f'(x)]^2} dx.$$

Hence, the center of mass of the curve is located at the point $(\bar{x}, \bar{y}) = \left(\frac{M_y}{m}, \frac{M_x}{m}\right)$, that is,

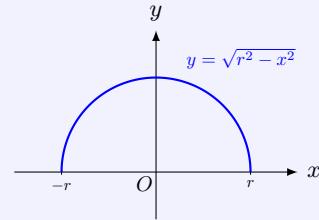
$$\bar{x} = \frac{\int_a^b \rho(x) \cdot x \sqrt{1 + [f'(x)]^2} dx}{\int_a^b \rho(x) \cdot \sqrt{1 + [f'(x)]^2} dx}, \quad \bar{y} = \frac{\int_a^b \rho(x) \cdot f(x) \sqrt{1 + [f'(x)]^2} dx}{\int_a^b \rho(x) \cdot \sqrt{1 + [f'(x)]^2} dx}.$$

It is easy to deduce the centroid from the center of mass when the density is a constant.

5.50 Example: centroid of a semicircle

Find the centroid of the semicircle $y = \sqrt{r^2 - x^2}$ ($-r \leq x \leq r$).

Solution By symmetry, the centroid (\bar{x}, \bar{y}) must lie on the y -axis, so that $\bar{x} = 0$. The length of the semicircle is $L = \pi r$. Since

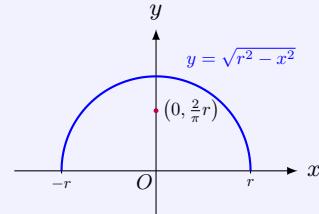


$$y' = \frac{1}{2}(r^2 - x^2)^{-\frac{1}{2}} \cdot (-2x) = -\frac{x}{\sqrt{r^2 - x^2}}.$$

we have

$$\begin{aligned} \bar{y} &= \frac{1}{L} \int_{-r}^r \sqrt{r^2 - x^2} \cdot \sqrt{1 + \left[-\frac{x}{\sqrt{r^2 - x^2}}\right]^2} dx \\ &= \frac{1}{\pi r} \int_{-r}^r r dx = \frac{1}{\pi r} \cdot 2r^2 = \frac{2}{\pi} r. \end{aligned}$$

Hence, the centroid of the semicircle is at $(\bar{x}, \bar{y}) = (0, \frac{2}{\pi} r)$.



5.51 MATLAB: centroid of a semicircle

Find the centroid of the semicircle $y = \sqrt{3^2 - x^2}$ ($-3 \leq x \leq 3$).

Solution First, we notice that the length of the semicircle is $L = 3\pi$. Then, by symmetry, we know that the centroid (\bar{x}, \bar{y}) must lie on the y -axis, so that $\bar{x} = 0$. We will calculate \bar{y} below.

```
>> syms x
```

```

>> f(x) = sqrt(3^2 - x^2);
>> df(x) = diff(f(x),x)

df(x) =

-x/(9 - x^2)^(1/2)

>> L = 3*pi;
>> int_fun_y = f(x)*sqrt(1 + df(x)^2);
>> cen_y = int(int_fun_y,-3,3)/L

cen_y =

6/pi

```

The result above shows that $\bar{y} = \frac{6}{\pi}$.

5.52 Theorem: center of mass along a parametric curve

Consider a continuous system of mass distributed along a parametric curve \mathcal{C} given by the parametric equations

$$x = f(t), \quad y = g(t), \quad \alpha \leq t \leq \beta,$$

with density $\rho(t)$. Suppose f' and g' are continuous. Then the **center of mass** of \mathcal{C} is (\bar{x}, \bar{y}) , where

$$\begin{aligned}\bar{x} &= \frac{1}{L} \int_{\alpha}^{\beta} \rho(t) \cdot f(t) \sqrt{[f'(t)]^2 + [g'(t)]^2} dt, \\ \bar{y} &= \frac{1}{L} \int_{\alpha}^{\beta} \rho(t) \cdot g(t) \sqrt{[f'(t)]^2 + [g'(t)]^2} dt,\end{aligned}$$

and $m = \int_{\alpha}^{\beta} \rho(t) \cdot \sqrt{[f'(t)]^2 + [g'(t)]^2} dt$ is the total mass of \mathcal{C} .

Theorem: centroid of a parametric curve

In particular, if $\rho(t) = \rho$ is a constant, then the center of mass is the same as the **centroid** of the curve, with

$$\begin{aligned}\bar{x} &= \frac{1}{L} \int_{\alpha}^{\beta} f(t) \sqrt{[f'(t)]^2 + [g'(t)]^2} dt, \\ \bar{y} &= \frac{1}{L} \int_{\alpha}^{\beta} g(t) \sqrt{[f'(t)]^2 + [g'(t)]^2} dt,\end{aligned}$$

where $L = \int_{\alpha}^{\beta} \sqrt{[f'(t)]^2 + [g'(t)]^2} dt$ is the length of the curve \mathcal{C} .

Justification:

The moment of \mathcal{C} about the y -axis is

$$M_y = \int \rho x \, ds.$$

For the parametric curve we have

$$ds = \sqrt{[f'(t)]^2 + [g'(t)]^2} \, dt.$$

Thus, by the substitution rule, we get

$$M_y = \int_{\alpha}^{\beta} \rho(t) \cdot f(t) \sqrt{[f'(t)]^2 + [g'(t)]^2} \, dt.$$

Hence, we get

$$\bar{x} = \frac{M_y}{m} = \frac{\int_{\alpha}^{\beta} \rho(t) \cdot f(t) \sqrt{[f'(t)]^2 + [g'(t)]^2} \, dt}{\int_{\alpha}^{\beta} \rho(t) \cdot \sqrt{[f'(t)]^2 + [g'(t)]^2} \, dt}.$$

Similarly, we have

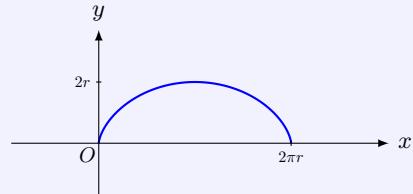
$$\bar{y} = \frac{M_x}{m} = \frac{\int_{\alpha}^{\beta} \rho(t) \cdot g(t) \sqrt{[f'(t)]^2 + [g'(t)]^2} \, dt}{\int_{\alpha}^{\beta} \rho(t) \cdot \sqrt{[f'(t)]^2 + [g'(t)]^2} \, dt}.$$

5.53 Example: centroid of a parametric curve

Find the centroid of one arch of the cycloid

$$x = r(t - \sin t), \quad y = r(1 - \cos t),$$

where $0 \leq t \leq 2\pi$.



Solution The length of one arch of the cycloid is

$$\begin{aligned} L &= \int_0^{2\pi} \sqrt{[r(1 - \cos t)]^2 + (-r \sin t)^2} \, dt \\ &= r \int_0^{2\pi} \sqrt{2 - 2 \cos t} \, dt \\ &= 2r \int_0^{2\pi} \sin(\frac{1}{2}t) \, dt && \text{half-angle formula} \\ &= 2r [-2 \cos(\frac{1}{2}t)]_{t=0}^{2\pi} = 8r. \end{aligned}$$

Without loss of generality, assume the density $\rho = 1$. Then the moment of the curve about the y -axis is

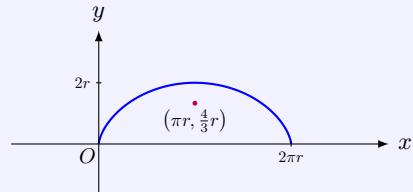
$$\begin{aligned}
 M_y &= \int_0^{2\pi} r(t - \sin t) \cdot \sqrt{[r(t - \sin t)]'^2 + [r(1 - \cos t)]'^2} dt \\
 &= r^2 \int_0^{2\pi} (t - \sin t) \sqrt{(1 - \cos t)^2 + \sin^2 t} dt \\
 &= r^2 \int_0^{2\pi} (t - \sin t) \cdot \sqrt{2 - 2 \cos t} dt \\
 &= r^2 \int_0^{2\pi} (t - \sin t) \cdot 2 \sin(\frac{1}{2}t) dt && \text{half-angle formula} \\
 &= 2r^2 \int_0^\pi (2u - \sin 2u) \sin u \cdot 2 du && t = 2u \\
 &= 4r^2 \int_0^\pi (2u \sin u - \sin 2u \sin u) du \\
 &= 4r^2 [-2u \cos u + \frac{3}{2} \sin u + \frac{1}{6} \sin 3u]_{u=0}^\pi \\
 &= 4r^2 \cdot 2\pi = 8\pi r^2;
 \end{aligned}$$

the moment of the curve about the x -axis is

$$\begin{aligned}
 M_x &= \int_0^{2\pi} r(1 - \cos t) \cdot \sqrt{[r(1 - \cos t)]^2 + (-r \sin t)^2} dt \\
 &= r^2 \int_0^{2\pi} (1 - \cos t) \sqrt{(1 - \cos t)^2 + \sin^2 t} dt \\
 &= r^2 \int_0^{2\pi} (1 - \cos t) \cdot \sqrt{2 - 2 \cos t} dt \\
 &= r^2 \int_0^{2\pi} 2 \sin^2(\frac{1}{2}t) \cdot 2 \sin(\frac{1}{2}t) dt && \text{half-angle formula} \\
 &= 4r^2 \int_0^\pi (1 - \cos^2 u) \sin u \cdot 2 du && t = 2u \\
 &= 8r^2 [-\cos u + \frac{1}{3} \cos^3 u]_{u=0}^\pi \\
 &= 8r^2 \cdot \frac{4}{3} = \frac{32}{3}r^2.
 \end{aligned}$$

Hence the centroid of the curve is (\bar{x}, \bar{y}) , where

$$\begin{aligned}
 \bar{x} &= \frac{M_y}{L} = \frac{8\pi r^2}{8r} = \pi r, \\
 \bar{y} &= \frac{M_x}{L} = \frac{\frac{32}{3}r^2}{8r} = \frac{4}{3}r.
 \end{aligned}$$



5.54 MATLAB: centroid of a parametric curve

Find the centroid of one arch of the cycloid

$$x = 3(t - \sin t), \quad y = 3(1 - \cos t),$$

where $0 \leq t \leq 2\pi$.

Solution For one arch of the cycloid, the centroid is calculated in MATLAB using the following code:

```
>> syms t
>> x(t) = 3*(t - sin(t));
>> y(t) = 3*(1 - cos(t));
>> dx(t) = diff(x(t),t);
>> dy(t) = diff(y(t),t);
>> int_fun_L = sqrt(dx(t)^2 + dy(t)^2);
>> L = int(int_fun_L,0,2*pi)

L =

24

>> int_fun_x = x(t)*sqrt(dx(t)^2 + dy(t)^2);
>> cen_x = int(int_fun_x,0,2*pi)/L

cen_x =

3*pi

>> int_fun_y = y(t)*sqrt(dx(t)^2 + dy(t)^2);
>> int_fun_y2 = matlabFunction(int_fun_y); % convert to ...
    function handle
>> cen_y = integral(int_fun_y2,0,2*pi)/L

cen_y =

4
```

The result above shows that the centroid is $(3\pi, 4)$. Note that the command `int` cannot solve the integral $\int_0^{2\pi} 3(1 - \cos t) \sqrt{(3(t - \sin t))' + (3(1 - \cos t))'} dt$, so we use the command `integral` to numerically evaluate the integral.

Centroid for Planar Region

5.55 Theorem:

centroid for planar region between two curves

Consider a flat plate (called a *lamina*) with uniform density ρ that occupies a region \mathcal{R} . If \mathcal{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 \mathcal{R} is (\bar{x}, \bar{y}) , where

$$\bar{x} = \frac{1}{A} \int_a^b x [f(x) - g(x)] dx, \quad \bar{y} = \frac{1}{A} \int_a^b \frac{1}{2} \{[f(x)]^2 - [g(x)]^2\} dx,$$

where $A = \int_a^b [f(x) - g(x)] dx$ is the area of \mathcal{R} .

Justification: Part 1

First we consider the case that $g(x) \equiv 0$.

We divide the interval $[a, b]$ into n subintervals with endpoints x_0, x_1, \dots, x_n and equal width Δx . Denote $\bar{x}_i = \frac{1}{2}(x_{i-1} + x_i)$ to be the midpoint of the i th subinterval. Then the centroid of the i th 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 i th 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 \mathcal{R} . By taking the limit as $n \rightarrow \infty$ we obtain the moment of \mathcal{R} about the y -axis:

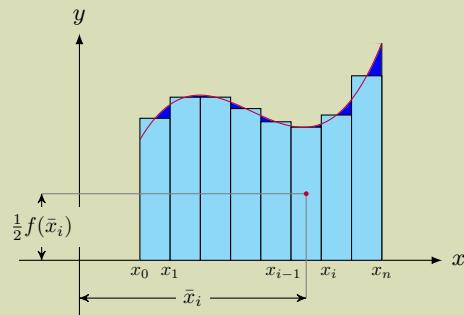
$$M_y = \lim_{n \rightarrow \infty} \sum_{i=1}^n \rho \bar{x}_i f(\bar{x}_i) \Delta x = \rho \int_a^b x f(x) dx.$$

In a similar fashion, the moment of the i th approximating rectangle about the x -axis is

$$\rho f(\bar{x}_i)\Delta x \cdot \frac{1}{2}f(\bar{x}_i)$$

so we get the moment of \mathcal{R} about the x -axis:

$$M_x = \lim_{n \rightarrow \infty} \sum_{i=1}^n \rho \cdot \frac{1}{2} [f(\bar{x}_i)]^2 \Delta x = \rho \int_a^b \frac{1}{2} [f(x)]^2 dx.$$



The mass of the lamina is

$$m = \rho A = \rho \int_a^b f(x) dx.$$

So, the **center of mass** of the plate (or the **centroid** of \mathcal{R} since the density is uniform) is located at the point (\bar{x}, \bar{y}) , where

$$\bar{x} = \frac{M_y}{m} = \frac{\int_a^b x f(x) dx}{\int_a^b f(x) dx}, \quad \bar{y} = \frac{M_x}{m} = \frac{\int_a^b \frac{1}{2} [f(x)]^2 dx}{\int_a^b f(x) dx}.$$

Justification: Part 2

Next, for $g(x) \not\equiv 0$, the mass of the lamina is

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

where A is the area bounded by the functions $y = f(x)$ and $y = g(x)$ over the interval $[a, b]$. The moment of \mathcal{R} about the y -axis and about the x -axis are, respectively,

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

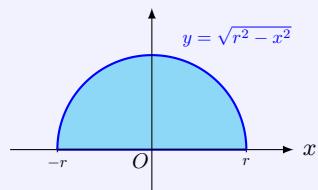
$$M_x = \rho \int_a^b \frac{1}{2} \left\{ [f(x)]^2 - [g(x)]^2 \right\} dx.$$

Hence,

$$\bar{x} = \frac{1}{A} \int_a^b x [f(x) - g(x)] dx, \quad \bar{y} = \frac{1}{A} \int_a^b \frac{1}{2} \left\{ [f(x)]^2 - [g(x)]^2 \right\} dx.$$

5.56 Example: centroid of a semicircular plate

As shown in the figure, we place the plate so that the circular arc is $y = f(x) = \sqrt{r^2 - x^2}$ ($-r \leq x \leq r$). Find the centroid of the plate.

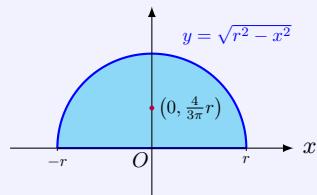


Solution By symmetry, the centroid (\bar{x}, \bar{y}) must lie on the y -axis, so that $\bar{x} = 0$. The area

of the semicircular plate is $A = \frac{1}{2}\pi r^2$. Thus,

$$\begin{aligned}\bar{y} &= \frac{1}{A} \int_{-r}^r \frac{1}{2} [f(x)]^2 dx \\ &= \frac{1}{\frac{1}{2}\pi r^2} \cdot \int_{-r}^r \frac{1}{2} (\sqrt{r^2 - x^2})^2 dx \\ &= \frac{2}{\pi r^2} \int_0^r (r^2 - x^2) dx \quad \text{even function} \\ &= \frac{2}{\pi r^2} [r^2 x - \frac{1}{3}x^3]_{x=0}^r = \frac{4}{3\pi} r.\end{aligned}$$

Hence, the centroid of the semicircular plate is at $(\bar{x}, \bar{y}) = (0, \frac{4}{3\pi} r)$.

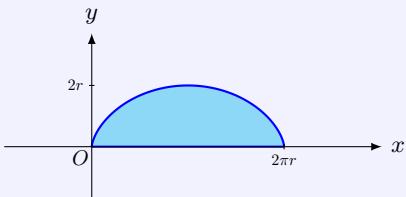


5.57 Example: centroid of a cycloid plate

Find the centroid of the plate between one arch of the cycloid

$$x = r(t - \sin t), \quad y = r(1 - \cos t),$$

where $0 \leq t \leq 2\pi$, and the x -axis.



Solution By applying the substitution rule using the parametric equations, the area of one arch of the cycloid is

$$\begin{aligned}A &= \int_0^{2\pi r} y dx \\ &= \int_0^{2\pi} r(1 - \cos t) \cdot [r(t - \sin t)]' dt \quad \left\{ \begin{array}{l} x = r(t - \sin t) \\ y = r(1 - \cos t) \end{array} \right. \\ &= r^2 \int_0^{2\pi} (1 - 2\cos t + \cos^2 t) dt \\ &= r^2 \int_0^{2\pi} [1 - 2\cos t + \frac{1}{2}(1 + \cos 2t)] dt \quad \text{double-angle formula} \\ &= r^2 [\frac{3}{2}t - 2\sin t + \frac{1}{4}\sin 2t]_{t=0}^{2\pi} = 3\pi r^2.\end{aligned}$$

Without loss of generality, assume the density $\rho = 1$. Then the moment of the plate about

the y -axis is

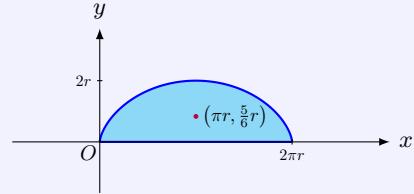
$$\begin{aligned} M_y &= \int_0^{2\pi r} xy \, dx \\ &= \int_0^{2\pi} r(t - \sin t) \cdot r(1 - \cos t) \cdot [r(t - \sin t)]' \, dt & \left\{ \begin{array}{l} x = r(t - \sin t) \\ y = r(1 - \cos t) \end{array} \right. \\ &= r^3 \int_0^{2\pi} (t - \sin t)(1 - \cos t)^2 \, dt = 3\pi^2 r^3; \end{aligned}$$

the moment of the plate about the x -axis is

$$\begin{aligned} M_x &= \int_0^{2\pi r} \frac{1}{2}y^2 \, dx \\ &= \int_0^{2\pi} \frac{1}{2}[r(1 - \cos t)]^2 \cdot [r(t - \sin t)]' \, dt & \left\{ \begin{array}{l} x = r(t - \sin t) \\ y = r(1 - \cos t) \end{array} \right. \\ &= \frac{1}{2}r^3 \int_0^{2\pi} (1 - \cos t)^3 \, dt = \frac{5}{2}\pi r^3. \end{aligned}$$

Hence the centroid of the plate is (\bar{x}, \bar{y}) , where

$$\begin{aligned} \bar{x} &= \frac{M_y}{L} = \frac{3\pi^2 r^3}{3\pi r^2} = \pi r, \\ \bar{y} &= \frac{M_x}{L} = \frac{\frac{5}{2}\pi r^3}{3\pi r^2} = \frac{5}{6}r. \end{aligned}$$



5.58 MATLAB: centroid of a semicircular plate

Find the centroid of the semicircular plate whose circular arc is $y = f(x) = \sqrt{3^2 - x^2}$ ($-3 \leq x \leq 3$).

Solution First, we notice that the area of the semicircular plate is $A = \frac{9\pi}{2}$. Moreover, by symmetry, the centroid (\bar{x}, \bar{y}) must lie on the y -axis, so that $\bar{x} = 0$. Then, we calculate \bar{y} below.

```
>> syms x
>> y(x) = sqrt(3^2 - x^2);
>> A = 3^2*pi/2;
>> int_fun = 0.5*f(x)^2;
>> cen_y = int(int_fun,-3,3)/A

cen_y =
4/pi
```

The result above shows that $\bar{y} = \frac{4}{\pi}$.

5.59 MATLAB: centroid of a cycloid plate

Find the centroid of the plate between one arch of the cycloid

$$x = 3(t - \sin t), \quad y = 3(1 - \cos t),$$

where $0 \leq t \leq 2\pi$, and the x -axis.

Solution The centroid of the plate is calculated in MATLAB using the following code:

```
>> syms t
>> x(t) = 3*(t - sin(t));
>> y(t) = 3*(1 - cos(t));
>> dx(t) = diff(x(t),t);
>> int_fun_A = y(t)*dx(t);
>> A = int(int_fun_A,0,2*pi)

A =

27*pi

>> int_fun_x = x(t)*y(t)*dx(t);
>> cen_x = int(int_fun_x,0,2*pi)/A

cen_x =

3*pi

>> int_fun_y = 0.5*y(t)^2*dx(t);
>> cen_y = int(int_fun_y,0,2*pi)/A

cen_y =

5/2
```

The result above shows that the centroid of the plate is $(3\pi, \frac{5}{2})$.

Pappus's Theorem

5.60 Example: relationship - arc length and surface area of revolution

We have calculated several geometric quantities associated with the cycloid

$$x = r(t - \sin t), \quad y = r(1 - \cos t), \quad 0 \leq t \leq 2\pi.$$

We know that the arc length is $L = 8r$ and the centroid of the arc is $(\bar{x}, \bar{y}) = (\pi r, \frac{4}{3}r)$.

1. For the solid obtained by rotating the region under the curve about the x -axis, we know

that the surface area is $S = \frac{64}{3}\pi r^2$. It is easy to verify that

$$S = \frac{64}{3}\pi r^2 = 2\pi(\frac{4}{3}r) \times (8r) = 2\pi\bar{y} \times L.$$

2. For the solid obtained by rotating the region under the curve about the y -axis, we know that the surface area is $S = 16\pi^2 r^2$. It is easy to verify that

$$S = 16\pi^2 r^2 = 2\pi(\pi r) \times (8r) = 2\pi\bar{x} \times L.$$

5.61 Example: relationship - area and volume of revolution

We have calculated several geometric quantities associated with the area under the cycloid

$$x = r(t - \sin t), \quad y = r(1 - \cos t), \quad 0 \leq t \leq 2\pi.$$

We know that the area under the arc is $A = 3\pi r^2$ and the centroid of region under the arc is $(\bar{x}, \bar{y}) = (\pi r, \frac{5}{6}r)$.

1. For the solid obtained by rotating the region under the curve about the x -axis, we know that the volume is $V = 5\pi^2 r^3$. It is easy to verify that

$$V = 5\pi^2 r^3 = 2\pi(\frac{5}{6}r) \cdot (3\pi r^2) = 2\pi\bar{y} \cdot A.$$

2. For the solid obtained by rotating the region under the curve about the y -axis, we know that the volume is $V = 6\pi^3 r^3$. It is easy to verify that

$$V = 6\pi^3 r^3 = 2\pi(\pi r) \cdot (3\pi r^2) = 2\pi\bar{x} \cdot A.$$

5.62 Theorem: Pappus's Theorem

1. The surface area S of a surface of revolution generated by rotating a plane curve \mathcal{C} about an axis external to \mathcal{C} and on the same plane is equal to the product of the arc length L of \mathcal{C} and the distance d traveled by the centroid of \mathcal{C} .
2. The volume V of a solid of revolution generated by rotating a plane region \mathcal{R} about an external axis is equal to the product of the area A of \mathcal{R} and the distance d traveled by the centroid of \mathcal{R} .

5.7 Work

5.63 Theorem: work done in moving an object along a line

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

$$W = \int_a^b f(x) dx.$$

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, \dots, x_n and equal width Δx . We choose a sample point x_i^* in the i th 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, the work done in moving the object from a to b as the limit of this quantity as $n \rightarrow \infty$. Since the right side is a Riemann sum, we recognize its limit as being a definite integral:

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

5.64 Example: spring stretching

A spring exerts a force of 10 N when stretched 10 cm beyond its natural length. How much work is required to stretch the spring 20 cm beyond its natural length?

Solution According to Hooke's law, the force needed to extend or compress a spring by some distance x meters away from its natural length scales linearly with respect to that distance, that is, $f(x) = kx$. When the spring is stretched 10 cm beyond its natural length, the force is 0.1k. By the hypothesis, $10 = 0.1k$, so that $k = 100$. Thus, $f(x) = 100x$.

Hence the work done in stretching the spring 20 cm beyond its natural length is

$$W = \int_0^{0.2} 100x dx = [50x^2]_{x=0}^{0.2} = 2 \text{ J.}$$

5.65 Example: cable lifting

A cable 60 meters long whose mass is 30 kilograms is hanging over the edge of a tall building and does not touch the ground. How much work is required to lift the top 12 meters of the cable to the top of the building?

Solution To lift the top 12 meters of the cable to the top of the building, the total work

done consists of two parts: (1) W_1 for moving the first top 12 meters of the cable to the top the building; (2) W_2 for moving the bottom $(60 - 12) = 48$ meters of the cable 12 meters higher.

(1) Let us place the origin at the top of the building and the x -axis pointing downward. To compute the work W_1 , we divide the cable into small segments with length Δx . If x_i^* is a point in the i th such segment, then all points in the segment are lifted by approximately the same amount, namely x_i^* . Since the cable weights $\frac{30}{60} = \frac{1}{2}$ kilograms per meter, so the weight of the i th segment is $\frac{1}{2}\Delta x$. Thus, the work done on the i th segment is

$$\left(\frac{1}{2}\Delta x\right) \cdot x_i^* \cdot g = \frac{1}{2}gx_i^*\Delta x,$$

By adding all these approximations and letting the number of segments becomes large (equivalently $\Delta x \rightarrow 0$), we get the work done for moving the first top 12 meters of the cable to the top the building:

$$W_1 = \lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{1}{2}gx_i^*\Delta x = \int_0^{12} \frac{1}{2}gx \, dx = \left[\frac{1}{4}gx^2\right]_{x=0}^{12} = 36g.$$

(2) Similar as above, we divide the second portion of cable into small segments with length Δx . This gives the weight of the i th segment to be $\frac{1}{2}\Delta x$. Thus, the work done on the i th segment is

$$\left(\frac{1}{2}\Delta x\right) \cdot 12 \cdot g = 6g\Delta x,$$

By adding all these approximations and letting the number of segments becomes large (equivalently $\Delta x \rightarrow 0$), we get the work done for moving the bottom 48 meters of the cable to the top the building:

$$W_2 = \lim_{n \rightarrow \infty} \sum_{i=1}^n 6g\Delta x = \int_{12}^{60} 6g \, dx = 6g \times (60 - 12) = 288g.$$

In fact, we can compute W_2 without using integration. Since this portion of cable exerts a force of $\frac{1}{2} \times 48g$ along the direction of motion, the work done is $W_2 = \frac{1}{2} \times 48g \times 12 = 288g$.

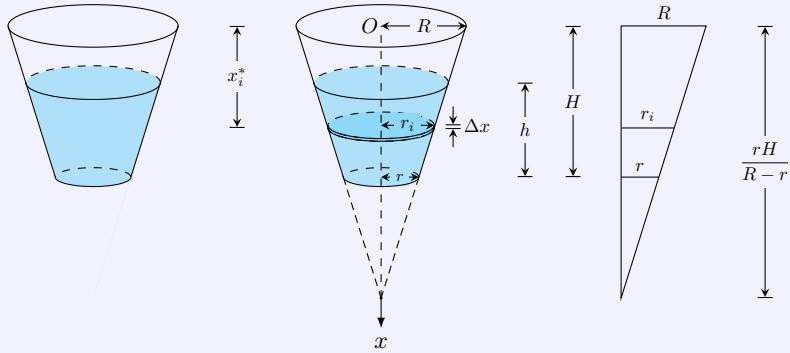
Hence the total work done is

$$W = W_1 + W_2 = 36g + 288g = 324g = 3175.2 \text{ J.}$$

5.66 Example: pumping liquid

A frustum of a cone of height H , top radius R and base radius r contains some liquid of density ρ . Suppose that the depth of the liquid in the frustum is h . How much work does it take to pump the liquid to the top of the container?

Solution Let x be the distance to the top of the container. Then the liquid extends from $x = H - h$ to $x = H$. We sketch a diagram of the situation.



By similar triangles, the radius $r(x)$ of the disk at depth x satisfies

$$\frac{r(x) - r}{R - r} = \frac{H - x}{H},$$

which gives

$$r(x) = (R - r) \frac{H - x}{H} + r.$$

We divide the interval $[H - h, H]$ into subintervals. Then the thin layer corresponding to $[x_{i-1}, x_i]$ has approximate volume

$$\Delta V_i \approx \pi[r(x_i^*)]^2 \Delta x,$$

and the work needed to pump the layer to the top of the container is approximately

$$\Delta W_i \approx (g\rho\Delta V_i)x_i^* = \pi g\rho x_i^*[r(x_i^*)]^2 \Delta x,$$

where g is the gravitational constant. Therefore the total work needed to pump all the liquid to the top of container is approximately

$$\sum_{i=1}^n \Delta W_i = \pi g \rho \sum_{i=1}^n x_i^* [r(x_i^*)]^2 \Delta x.$$

Taking the limit as $\Delta x \rightarrow 0$, we get the total work

$$\begin{aligned} W &= \pi g \rho \int_{H-h}^H x[r(x)]^2 dx = \frac{\pi g \rho}{H^2} \int_{H-h}^H x[(R-r)(H-x) + rH]^2 dx \\ &= \pi g \rho H^2 R^2 \left[a^2 b + \frac{1}{2} a(2a - 3b)b^2 + \frac{1}{3}(1-a)(1-3a)b^3 - \frac{1}{4}(1-a)^2 b^4 \right], \end{aligned}$$

where $a = \frac{r}{R}$ and $b = \frac{h}{H}$.

5.67 MATLAB: spring stretching

A spring exerts a force of 10 N when stretched 10 cm beyond its natural length. How much work is required to stretch the spring 20 cm beyond its natural length?

Solution The problem can be solved in MATLAB as follows:

```
>> syms k
>> g(k) = 0.1*k; % Hooke's law
>> k0 = solve(g(k) == 10) % solve the constant k

k0 =
100

>> syms x
>> int_fun = k0*x;
>> W = int(int_fun,0,0.2)

W =
2
```

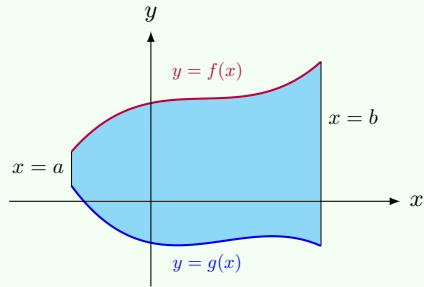
We see that the work required is 2 (J).

Summary

Theorem: area

- If $f(x) \geq g(x)$,

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



- In general,

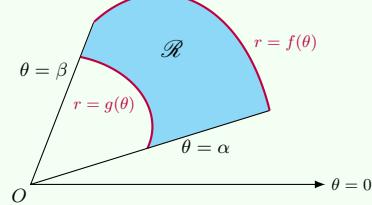
$$A = \int_a^b |f(x) - g(x)| dx$$

- For a region under the parametric curve $x = f(t)$, $y = g(t)$ ($\alpha \leq t \leq \beta$),

$$A = \int_{\alpha}^{\beta} g(t) f'(t) dt$$

- For a polar region bounded by the polar curves $r = f(\theta)$, $r = g(\theta)$, and by the rays $\theta = \alpha$ to $\theta = \beta$,

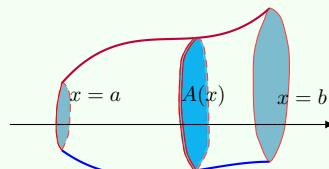
$$A = \frac{1}{2} \int_{\alpha}^{\beta} ([f(\theta)]^2 - [g(\theta)]^2) d\theta$$



Theorem: volume

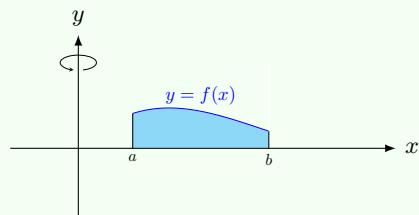
- by slices:

$$V = \int_a^b A(x) dx$$



- by shells:

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



Theorem: arc length

- For $y = f(x)$ ($a \leq x \leq b$),

$$L = \int_a^b \sqrt{1 + [f'(x)]^2} dx.$$

- For parametric curve $x = f(t)$, $y = g(t)$ ($\alpha \leq t \leq \beta$),

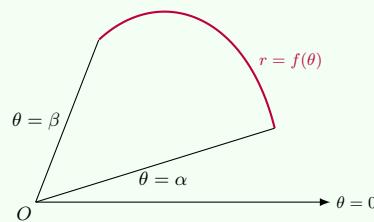
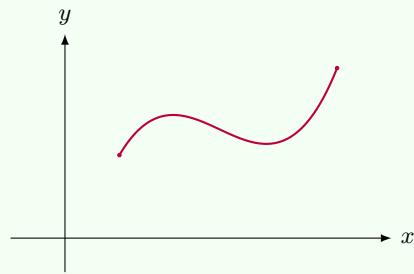
$$L = \int_{\alpha}^{\beta} \sqrt{[f'(t)]^2 + [g'(t)]^2} dt.$$

- In general,

$$L = \int \underbrace{(ds)}_{\text{arc length}}.$$

- For the polar curve $r = f(\theta)$ from $\theta = \alpha$ to $\theta = \beta$,

$$\begin{aligned} L &= \int_{\alpha}^{\beta} \sqrt{[f(\theta)]^2 + [f'(\theta)]^2} d\theta \\ &= \int_{\alpha}^{\beta} \sqrt{r^2 + (dr/d\theta)^2} d\theta \end{aligned}$$

**Theorem:** surface area of revolution

- For a curve $y = f(x)$ ($a \leq x \leq b$),

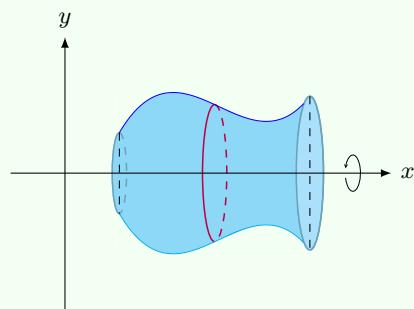
$$S = \int_a^b 2\pi f(x) \sqrt{1 + [f'(x)]^2} dx;$$

- For a parametric curve $x = f(t)$, $y = g(t)$ ($\alpha \leq t \leq \beta$),

$$S = \int_{\alpha}^{\beta} 2\pi g(t) \sqrt{[f'(t)]^2 + [g'(t)]^2} dt.$$

- In general,

$$\int \underbrace{(2\pi r)}_{\text{circumference}} \underbrace{(ds)}_{\text{arc length}}.$$



Theorem: center of mass for discrete system

- For one-dimensional system:

$$\bar{x} = \frac{M}{m} = \frac{\sum_{i=1}^n m_i x_i}{\sum_{i=1}^n m_i}.$$

- For two-dimensional system:

$$\bar{x} = \frac{M_y}{m} = \frac{\sum_{i=1}^n m_i x_i}{\sum_{i=1}^n m_i}, \quad \bar{y} = \frac{M_x}{m} = \frac{\sum_{i=1}^n m_i y_i}{\sum_{i=1}^n m_i}.$$

Theorem: center of mass

- For mass distributed along a curve $y = f(x)$ ($a \leq x \leq b$),

$$\begin{aligned}\bar{x} &= \frac{1}{m} \int_a^b \rho(x) \cdot x \sqrt{1 + [f'(x)]^2} dx, \\ \bar{y} &= \frac{1}{m} \int_a^b \rho(x) \cdot f(x) \sqrt{1 + [f'(x)]^2} dx,\end{aligned}$$

where $m = \int_a^b \rho(x) \cdot \sqrt{1 + [f'(x)]^2} dx$ is the total mass of the curve.

- For mass distributed along a parametric curve $x = f(t)$, $y = g(t)$ ($\alpha \leq t \leq \beta$),

$$\begin{aligned}\bar{x} &= \frac{1}{m} \int_{\alpha}^{\beta} \rho(t) \cdot f(t) \sqrt{[f'(t)]^2 + [g'(t)]^2} dt, \\ \bar{y} &= \frac{1}{m} \int_{\alpha}^{\beta} \rho(t) \cdot g(t) \sqrt{[f'(t)]^2 + [g'(t)]^2} dt,\end{aligned}$$

where $m = \int_{\alpha}^{\beta} \rho(t) \cdot \sqrt{[f'(t)]^2 + [g'(t)]^2} dt$ is the total mass of the curve.

- For a plate occupying the region bounded by two curves $y = f(x)$ and $y = g(x)$ ($a \leq x \leq b$), with $f(x) \geq g(x)$, the centroid is at (\bar{x}, \bar{y}) :

$$\bar{x} = \frac{1}{A} \int_a^b x [f(x) - g(x)] dx, \quad \bar{y} = \frac{1}{A} \int_a^b \frac{1}{2} \left\{ [f(x)]^2 - [g(x)]^2 \right\} dx,$$

where $A = \int_a^b [f(x) - g(x)] dx$ is the area of the region.

Theorem: Pappus's Theorem

1. The surface area S of a surface of revolution generated by rotating a plane curve \mathcal{C} about an axis external to \mathcal{C} and on the same plane is equal to the product of the arc length L of \mathcal{C} and the distance d traveled by the centroid of \mathcal{C} .
2. The volume V of a solid of revolution generated by rotating a plane region \mathcal{R} about an external axis is equal to the product of the area A of \mathcal{R} and the distance d traveled by the centroid of \mathcal{R} .

Theorem: work done by a force

$$W = \int_a^b f(x) dx.$$

Infinite Sequences and Series

6.1 Infinite Sequences

6.1 Definition: infinite sequence

An infinite sequence (or *sequence* for short) is an ordered list of numbers of the form

$$\{a_1, a_2, a_3, \dots, a_n, \dots\}.$$

For each n , the term a_n is called the n th term of the sequence.

For convenience, sometimes we write a sequence as $\{a_n\}$.

6.2 Example: infinite sequences

The following are some examples of sequences.

1. The sequence $\{1, 2, 3, \dots, n, \dots\}$ consists of all the positive integers, with the n th term of the sequence being $a_n = n$.
2. The n th term of the sequence $\{1, -1, 1, -1, 1, -1, \dots\}$ is $a_n = (-1)^{n-1}$.
3. The terms in the sequence

$$\{3, 3.1, 3.14, 3.141, 3.1415, \dots\}$$

are from the truncations of the number π . The n th term of the sequence is the number obtained by truncating π to $n - 1$ decimal place. For this sequence, there is no elementary formula available to express a_n in terms of n .

4. For a given expression of a function f , by evaluating the function at integers, we can obtain a sequence

$$\{a_1, a_2, a_3, \dots, a_n, \dots\},$$

where $a_n = f(n)$ ($n = 1, 2, 3, \dots$). For instance, the infinite sequence

$$\left\{ \cos \frac{\pi}{1}, \cos \frac{\pi}{2}, \cos \frac{\pi}{3}, \dots, \cos \frac{\pi}{n}, \dots \right\}$$

can be generated by the function $f(x) = \cos \frac{\pi}{x}$.

5. The infinite sequence

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

can be defined as

$$a_1 = \sqrt{2}; \quad a_n = \sqrt{2 + a_{n-1}}, \quad \text{for } n \geq 2.$$

By using the formula, we can generate the *whole* sequence *recursively*. In fact, for the first few terms:

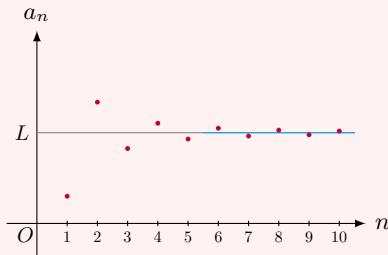
$$\begin{aligned} \text{when } n = 1 : \quad a_1 &= \sqrt{2}; \\ \text{when } n = 2 : \quad a_2 &= \sqrt{2 + a_1} \\ &= \sqrt{2 + \sqrt{2}}; \\ \text{when } n = 3 : \quad a_3 &= \sqrt{2 + a_2} \\ &= \sqrt{2 + \sqrt{2 + \sqrt{2}}}; \\ \text{when } n = 4 : \quad a_4 &= \sqrt{2 + a_3} \\ &= \sqrt{2 + \sqrt{2 + \sqrt{2 + \sqrt{2}}}}. \end{aligned}$$

6.3 Definition: limit of infinite sequence

- An infinite sequence $\{a_n\}$ has the **limit L** , which is a finite number, and we write

$$\lim_{n \rightarrow \infty} a_n = L \quad \text{or} \quad a_n \rightarrow L \text{ as } n \rightarrow \infty$$

if we can make the terms a_n as close to L as we like by taking n sufficiently large.



If $\lim_{n \rightarrow \infty} a_n$ exists, we say the sequence **converges** (or is **convergent**). Otherwise, we say the sequence **diverges** (or is **divergent**).

- The infinite limit $\lim_{n \rightarrow \infty} a_n = \infty$ means that we can make the terms a_n as large as we like by taking n sufficiently large. In this case, we say the sequence **diverges** to infinity. Similarly, we can definite $\lim_{n \rightarrow \infty} a_n = -\infty$.

6.4 Example: convergent/divergent sequences

1. The sequence

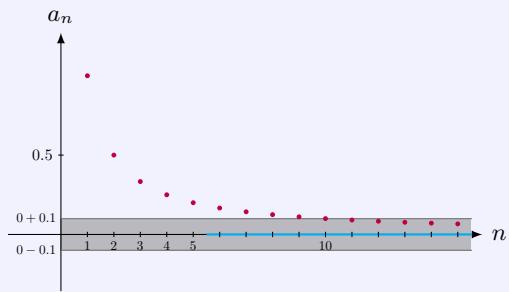
$$\left\{ \frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{n}, \dots \right\}$$

converges to 0, since we can make the terms $a_n = \frac{1}{n}$ as close to 0 as we like by taking n sufficiently large. In fact, *for instance*, to make the terms a_n close to 0 within 0.1, that is,

$$|a_n - 0| < 0.1,$$

we just need to take $n > 10$, as shown in the figure.

n	a_n	n	a_n
1	1.0	11	0.090909
2	0.5	12	0.083333
3	0.333333	13	0.076923
4	0.25	14	0.071429
5	0.2	15	0.066667
6	0.166667	20	0.05
7	0.142857	30	0.033333
8	0.125	40	0.025
9	0.111111	50	0.02
10	0.1	60	0.016667



2. The sequence

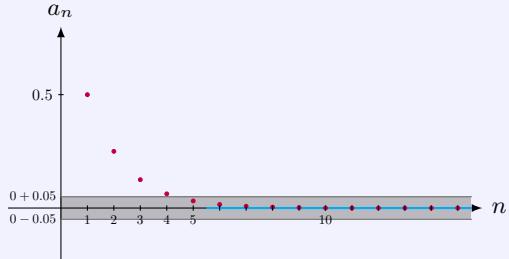
$$\left\{ \frac{1}{2}, \left(\frac{1}{2}\right)^2, \left(\frac{1}{2}\right)^4, \dots, \left(\frac{1}{2}\right)^n, \dots \right\}$$

converges to 0, since we can make the terms $a_n = \left(\frac{1}{2}\right)^n$ as close to 0 as we like by taking n sufficiently large. In fact, *for instance*, to make the terms a_n close to 0 within 0.05, that is,

$$|a_n - 0| < 0.05,$$

we just need to take $n > 4$, as shown in the figure.

n	a_n
1	0.5
2	0.25
3	0.125
4	0.0625
5	0.03125
6	0.015625
7	0.0078125
8	0.00390625
9	0.00195313
10	0.000976563



3. The sequence

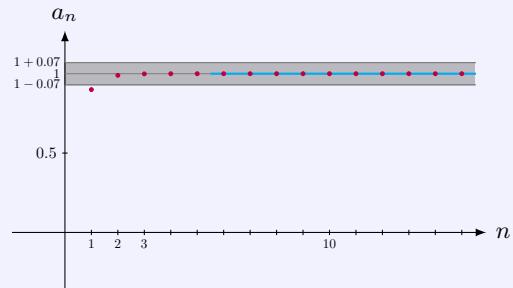
$$\{0.9, 0.99, 0.999, 0.9999, \dots\}$$

converges to 1, since the terms increase in magnitude and approach 1. *For instance*, to make the terms a_n close to 0 within 0.07, that is,

$$|a_n - 0| < 0.07,$$

we just need to take $n > 1$.

n	a_n
1	0.9
2	0.99
3	0.999
4	0.9999
5	0.99999
6	0.999999



4. The sequence

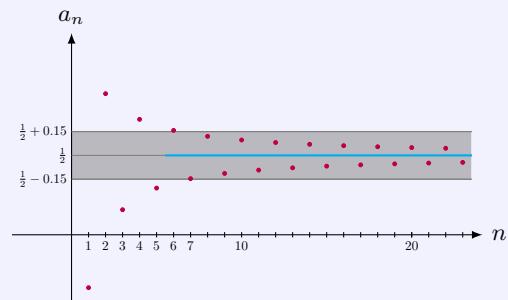
$$\left\{-\frac{1}{3}, \frac{8}{9}, \frac{3}{19}, \dots, \frac{n^2 + 2 \cdot (-1)^n n}{2n^2 + 1}, \dots\right\}$$

converges to $\frac{1}{2}$, since we can make the terms $a_n = \frac{n^2 + 2 \cdot (-1)^n n}{2n^2 + 1}$ as close to $\frac{1}{2}$ as we like by taking n sufficiently large. In fact, *for instance*, to make the terms a_n close to $\frac{1}{2}$ within 0.15, that is,

$$|a_n - \frac{1}{2}| < 0.15,$$

we just need to take $n > 6$, as shown in the figure.

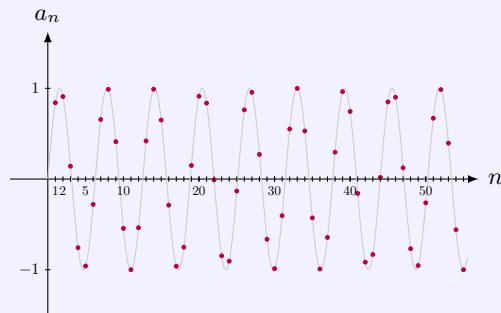
n	a_n	n	a_n
1	-0.333333	11	0.407407
2	0.888889	12	0.581315
3	0.157895	13	0.421829
4	0.727273	14	0.569975
5	0.294118	15	0.432373
6	0.657534	20	0.549313
7	0.353535	30	0.533037
8	0.620155	40	0.524836
9	0.386503	60	0.516595
10	0.597015	80	0.512460



5. The sequence

$$\{\sin 1, \sin 2, \sin 3, \dots, \sin n, \dots\}$$

diverges, since $a_n = \sin n$ does not appear to approach any finite number as $n \rightarrow \infty$.



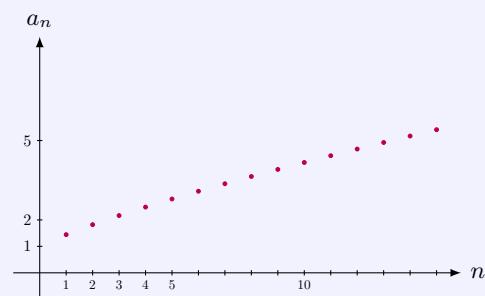
6. The sequence

$$\left\{ \frac{1}{\ln 2}, \frac{2}{\ln 3}, \frac{3}{\ln 4}, \dots, \frac{n}{\ln(n+1)}, \dots \right\}$$

diverges to ∞ . This is because,

$$\begin{aligned} \lim_{x \rightarrow \infty} \frac{x}{\ln(x+1)} &\stackrel{\text{H}}{=} \lim_{x \rightarrow \infty} \frac{1}{\frac{1}{x+1}} \\ &= \lim_{x \rightarrow \infty} (x+1) = \infty, \end{aligned}$$

we have $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} \frac{n}{\ln(n+1)} = \infty$.



6.5 Theorem: limit of a sequence from limit of a function

If $\lim_{x \rightarrow \infty} f(x) = L$ and $f(n) = a_n$ when n is an integer, then $\lim_{n \rightarrow \infty} a_n = L$. The same result holds when $L = \infty$ or $-\infty$.

6.6 Theorem: Limit Laws for sequences

Suppose $\{a_n\}$ and $\{b_n\}$ are convergent sequences and c is a constant. Then the following hold.

1. $\lim_{n \rightarrow \infty} (a_n + b_n) = \lim_{n \rightarrow \infty} a_n + \lim_{n \rightarrow \infty} b_n$.
2. $\lim_{n \rightarrow \infty} (a_n - b_n) = \lim_{n \rightarrow \infty} a_n - \lim_{n \rightarrow \infty} b_n$.
3. $\lim_{n \rightarrow \infty} ca_n = c \lim_{n \rightarrow \infty} a_n$. In particular, $\lim_{n \rightarrow \infty} c = c$.
4. $\lim_{n \rightarrow \infty} (a_n b_n) = \lim_{n \rightarrow \infty} a_n \cdot \lim_{n \rightarrow \infty} b_n$.
5. $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = \frac{\lim_{n \rightarrow \infty} a_n}{\lim_{n \rightarrow \infty} b_n}$ if $\lim_{n \rightarrow \infty} b_n \neq 0$.

6. $\lim_{n \rightarrow \infty} a_n^p = \left[\lim_{n \rightarrow \infty} a_n \right]^p$ if $p > 0$ and $a_n > 0$.

6.7 Example: limits of sequences

1. Show that $\lim_{n \rightarrow \infty} \frac{1}{n^p} = 0$ ($p > 0$).

Solution Because $\lim_{x \rightarrow \infty} \frac{1}{x} = 0$, we know that $\lim_{n \rightarrow \infty} \frac{1}{n} = 0$. Thus, for $p > 0$, by the Limit Laws,

$$\lim_{n \rightarrow \infty} \frac{1}{n^p} = \left(\lim_{n \rightarrow \infty} \frac{1}{n} \right)^p = 0^p = 0.$$

2. Compute $\lim_{n \rightarrow \infty} \frac{n^{\frac{3}{2}} + 2n - 1}{n^{\frac{3}{2}} + 1}$.

Solution Since

$$\frac{n^{\frac{3}{2}} + 2n - 1}{n^{\frac{3}{2}} + 1} = \frac{n^{\frac{3}{2}} \left(1 + 2 \cdot \frac{1}{n^{\frac{1}{2}}} - \frac{1}{n^{\frac{3}{2}}} \right)}{n^{\frac{3}{2}} \left(1 + \frac{1}{n^{\frac{3}{2}}} \right)} = \frac{1 + 2 \cdot \frac{1}{n^{\frac{1}{2}}} - \frac{1}{n^{\frac{3}{2}}}}{1 + \frac{1}{n^{\frac{3}{2}}}},$$

from the Limit Laws, we have

$$\lim_{n \rightarrow \infty} \frac{n^{\frac{3}{2}} + 2n - 1}{n^{\frac{3}{2}} + 1} = \frac{1 + 2 \cdot \lim_{n \rightarrow \infty} \frac{1}{n^{\frac{1}{2}}} - \lim_{n \rightarrow \infty} \frac{1}{n^{\frac{3}{2}}}}{1 + \lim_{n \rightarrow \infty} \frac{1}{n^{\frac{3}{2}}}} = \frac{1 + 2 \cdot 0 - 0}{1 + 0} = 1.$$

6.8 Theorem: Squeeze Theorem

If $a_n \leq b_n \leq c_n$ for $n \geq n_0$ and $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} c_n = L$, then $\lim_{n \rightarrow \infty} b_n = L$.

6.9 Example: limits of sequence - Squeeze Theorem

Show that $\lim_{n \rightarrow \infty} \frac{\cos n}{\sqrt{n}} = 0$.

Solution Since $-1 \leq \cos n \leq 1$ for all n , we have

$$-\frac{1}{\sqrt{n}} \leq \frac{\cos n}{\sqrt{n}} \leq \frac{1}{\sqrt{n}}.$$

Because $\lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} = \lim_{n \rightarrow \infty} \left(-\frac{1}{\sqrt{n}} \right) = 0$, by the Squeeze Theorem, we get that $\lim_{n \rightarrow \infty} \frac{\cos n}{\sqrt{n}} = 0$.

6.10 Example: zero limit and its absolute value

Show that $\lim_{n \rightarrow \infty} |a_n| = 0$ if and only if $\lim_{n \rightarrow \infty} a_n = 0$.

Solution

“ \Rightarrow ”: Suppose $\lim_{n \rightarrow \infty} |a_n| = 0$. By the Limit Laws, $\lim_{n \rightarrow \infty} (-|a_n|) = 0$. Since

$$-|a_n| \leq a_n \leq |a_n|,$$

by the Squeeze Theorem, $\lim_{n \rightarrow \infty} a_n = 0$.

“ \Leftarrow ”: Suppose $\lim_{n \rightarrow \infty} a_n = 0$. By the Limit Laws,

$$\begin{aligned}\lim_{n \rightarrow \infty} \sqrt{a_n^2 + n^{-1}} &= \sqrt{\lim_{n \rightarrow \infty} (a_n^2 + n^{-1})} \\ &= \sqrt{\lim_{n \rightarrow \infty} a_n^2 + \lim_{n \rightarrow \infty} n^{-1}} \\ &= \sqrt{\lim_{n \rightarrow \infty} a_n^2 + 0} = \sqrt{0^2 + 0} = 0.\end{aligned}$$

Since

$$0 \leq |a_n| \leq \sqrt{a_n^2 + n^{-1}},$$

by the Squeeze Theorem, $\lim_{n \rightarrow \infty} |a_n| = 0$.

6.11 Definition: two asymptotic relations

- The notation

$$a_n \ll b_n, \quad \text{as } n \rightarrow \infty,$$

which is read “ a_n is **much smaller** than b_n as n tends to ∞ ,” means

$$\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 0.$$

- The notation

$$a_n \sim b_n, \quad \text{as } n \rightarrow \infty,$$

which is read “ a_n is **asymptotic to** b_n as n tends to ∞ ,” means that

$$a_n - b_n \ll b_n, \quad \text{as } n \rightarrow \infty,$$

or, equivalently,

$$\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 1.$$

6.12 Theorem: growth rates of sequences

For any real numbers $p > 0$, $q > 0$, $r > 0$, and $b > 1$, the following asymptotic relations hold: as $n \rightarrow \infty$,

$$\ln^q n \ll n^p \ll n^p \ln^r n \ll b^n \ll n! \ll n^n.$$

Remark: Stirling's formula

A more precise formula for $n!$ is given by Stirling's formula:

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n, \quad \text{as } n \rightarrow \infty.$$

Justification: $\ln^q n \ll n^p$

Since, as $n \rightarrow \infty$,

$$\ln^q n \ll n^p \iff \frac{\ln^q n}{n^p} \ll 1 \iff \left(\frac{\ln^q n}{n^p}\right)^{1/q} \ll 1 \iff \frac{\ln n}{n^{p/q}} \ll 1 \iff \ln n \ll n^{p/q},$$

we only need to show for any positive real number p ,

$$\ln n \ll n^p, \quad \text{as } n \rightarrow \infty.$$

In fact, by l'Hospital's Rule,

$$\lim_{x \rightarrow \infty} \frac{\ln x}{x^p} \stackrel{\text{H}}{=} \lim_{x \rightarrow \infty} \frac{\frac{1}{x}}{px^{p-1}} = \lim_{x \rightarrow \infty} \frac{1}{px^p} = 0,$$

so that $\lim_{n \rightarrow \infty} \frac{\ln n}{n^p} = 0$. Thus, $\ln n \ll n^p$ as $n \rightarrow \infty$.

Justification: $n^p \ll n^p \ln^r n$

Since, as $n \rightarrow \infty$,

$$n^p \ll n^p \ln^r n \iff 1 \ll \ln^r n \iff 1 \ll \ln n,$$

we only need to show that $\lim_{n \rightarrow \infty} \frac{1}{\ln n} = 0$. The latter is obvious since $\lim_{x \rightarrow \infty} \frac{1}{\ln x} = 0$.

Justification: $n^p \ln^r n \ll b^n$

Since $\ln^q n \ll n^p$ as $n \rightarrow \infty$ for any positive p and q , to show that $n^p \ln^r n \ll b^n$, we only need to show that the asymptotic relation $n^p \ll b^n$ for any positive p and $b > 1$. Furthermore, since

$$n^p \ll b^n \iff n \ll (b^{1/p})^n,$$

it is sufficient to show that $n \ll b^n$ for any $b > 1$.

In fact, we may write $b = 1 + a$, with $a > 0$. Then, by the binomial expansion, we have

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

We get

$$0 < \frac{n}{b^n} < \frac{2}{a^2} \cdot \frac{1}{n - 1}.$$

Because $\lim_{n \rightarrow \infty} \frac{1}{n-1} = 0$, by the Squeeze Theorem, we have $\lim_{n \rightarrow \infty} \frac{n}{b^n} = 0$. That is, $n \ll b^n$.

Justification: $b^n \ll n!$

For the given number b , take a sufficiently large K such that $b < K$. Then, for $n > K$, we have

$$\begin{aligned} 0 &< \frac{b^n}{n!} = \frac{\overbrace{b \cdot b \cdots b \cdot b}^{n \text{ factors}}}{n \cdot (n-1) \cdots 2 \cdot 1} \\ &= \frac{b}{n} \cdot \frac{b}{n-1} \cdots \frac{b}{K} \cdot \frac{b}{K-1} \cdots \frac{b}{2} \cdot \frac{b}{1} \\ &< \frac{b}{n} \cdot \frac{b}{K-1} \cdots \frac{b}{2} \cdot \frac{b}{1}. \end{aligned}$$

Since

$$\lim_{n \rightarrow \infty} \left(\frac{b}{n} \cdot \frac{b}{K-1} \cdots \frac{b}{2} \cdot \frac{b}{1} \right) = \left(b \cdot \frac{b}{K-1} \cdots \frac{b}{2} \cdot \frac{b}{1} \right) \cdot \lim_{n \rightarrow \infty} \frac{1}{n} = 0,$$

by the Squeeze Theorem, we have $\lim_{n \rightarrow \infty} \frac{b^n}{n!} = 0$. That is, $b^n \ll n!$ as $n \rightarrow \infty$.

Justification: $n! \ll n^n$

Since

$$0 < \frac{n!}{n^n} = \frac{n \cdot (n-1) \cdots 2 \cdot 1}{\underbrace{n \cdot n \cdots n \cdot n}_{n \text{ factors}}} = \frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{2}{n} \cdot \frac{1}{n} < \frac{1}{n}.$$

It is known that $\lim_{n \rightarrow \infty} \frac{1}{n} = 0$. By the Squeeze Theorem, we have $\lim_{n \rightarrow \infty} \frac{n!}{n^n} = 0$. That is, $n! \ll n^n$ as $n \rightarrow \infty$.

6.13 Theorem: geometric sequences

Let r be a real number. Then

$$\lim_{n \rightarrow \infty} r^n = \begin{cases} 0, & \text{if } |r| < 1; \\ 1, & \text{if } r = 1; \\ \text{diverges,} & \text{if } r \leq -1 \text{ or } r > 1. \end{cases}$$

Justification:

1. Consider $0 \leq |r| < 1$.

When $r = 0$, it is clear that $\lim_{n \rightarrow \infty} r^n = 0$.

When $0 < r < 1$, we put $b = \frac{1}{r}$ and know that $b > 1$. Since $1 \ll b^n$ as $n \rightarrow \infty$, we have

$$\lim_{n \rightarrow \infty} \frac{1}{b^n} = 0 \implies \lim_{n \rightarrow \infty} r^n = 0.$$

When $-1 < r < 0$, we have $0 < |r| < 1$. Since

$$-|r|^n \leq r^n \leq |r|^n,$$

and $\lim_{n \rightarrow \infty} |r|^n = \lim_{n \rightarrow \infty} (-|r|^n) = 0$, by the Squeeze Theorem, we know that $\lim_{n \rightarrow \infty} r^n = 0$.

2. If $r = 1$, then $r^n \equiv 1$ for all $n \geq 1$. In this case, it is clear that $\lim_{n \rightarrow \infty} r^n = 1$.

3. Suppose $r \leq -1$ or $r > 1$.

When $r > 1$, we know that $1 \ll r^n$, so that $\{r^n\}$ diverges to ∞ .

When $r \leq -1$, $\{|r|^n\}$ diverges to ∞ , so that $\{r^n\}$ oscillates and never approach any finite number. Hence, $\{r^n\}$ diverges in this case.

6.14 Theorem: sequential limit for continuous function

If $\lim_{n \rightarrow \infty} a_n = L$ and the function f is continuous at L , then

$$\lim_{n \rightarrow \infty} f(a_n) = f(L).$$

6.15 Example: limit of sequence

1. Compute the limit $\lim_{n \rightarrow \infty} \sqrt[n]{a}$ for $a > 0$.

Solution Let $a_n = \ln \sqrt[n]{a} = \frac{\ln a}{n}$. Then $\sqrt[n]{a} = e^{a_n}$. We know that $1 \ll n$ as $n \rightarrow \infty$, so that $\lim_{n \rightarrow \infty} a_n = 0$. Because e^x is a continuous function at $x = 0$, we have

$$\lim_{n \rightarrow \infty} \sqrt[n]{a} = \lim_{n \rightarrow \infty} e^{a_n} = e^0 = 1.$$

2. Compute the limit $\lim_{n \rightarrow \infty} \sqrt[n]{n}$.

Solution Let $a_n = \ln \sqrt[n]{n} = \frac{\ln n}{n}$. Then $\sqrt[n]{n} = e^{a_n}$. We know that $\ln n \ll n$ as $n \rightarrow \infty$, so that $\lim_{n \rightarrow \infty} a_n = 0$. Because e^x is a continuous function at $x = 0$, we have

$$\lim_{n \rightarrow \infty} \sqrt[n]{n} = \lim_{n \rightarrow \infty} e^{a_n} = e^0 = 1.$$

3. Compute the limit $\lim_{n \rightarrow \infty} \left(1 - \frac{2}{n}\right)^n$.

Solution Let $a_n = \ln \left(1 - \frac{2}{n}\right)^n = n \cdot \ln \left(1 - \frac{2}{n}\right)$. Then $\left(1 - \frac{2}{n}\right)^n = e^{a_n}$. Replacing n by a continuous variable x^{-1} , we see that $n \rightarrow \infty$ is a special case when $x \rightarrow 0^+$. Applying l'Hospital's Rule gives

$$\lim_{x \rightarrow 0^+} x^{-1} \cdot \ln(1 - 2x) = \lim_{x \rightarrow 0^+} \frac{\ln(1 - 2x)}{x} \stackrel{H}{=} \lim_{x \rightarrow 0^+} \frac{-2}{1} = -2,$$

so we have $\lim_{n \rightarrow \infty} a_n = -2$. Because e^x is a continuous function at $x = -2$, we have

$$\lim_{n \rightarrow \infty} \left(1 - \frac{2}{n}\right)^n = \lim_{n \rightarrow \infty} e^{a_n} = e^{-2}.$$

6.16 Definition: terminology for sequences

- $\{a_n\}$ is **increasing** if $a_{n+1} > a_n$. For example, $\{1, 2, 3, \dots\}$.
- $\{a_n\}$ is **non-decreasing** if $a_{n+1} \geq a_n$. For example, $\{1, 1, 2, 2, 3, 3, \dots\}$.
- $\{a_n\}$ is **decreasing** if $a_{n+1} < a_n$. For example, $\{0, -1, -2, -3, \dots\}$.
- $\{a_n\}$ is **non-increasing** if $a_{n+1} \leq a_n$. For example, $\{0, 0, -1, -1, -2, -2, \dots\}$.
- $\{a_n\}$ is **monotonic** if it is either non-increasing or non-decreasing.
- $\{a_n\}$ is **bounded** if there is number M such that $|a_n| \leq M$ for all n .

6.17 Example: product of sequences

Suppose $\{a_n\}$ is bounded and $\lim_{n \rightarrow \infty} b_n = 0$. Show that $\lim_{n \rightarrow \infty} a_n b_n = 0$.

Solution Since $\{a_n\}$ is bounded, there is a number M such that $|a_n| \leq M$ for all n . Thus, for every n ,

$$-M|b_n| \leq a_n b_n \leq M|b_n|.$$

Since $\lim_{n \rightarrow \infty} b_n = 0$ implies $\lim_{n \rightarrow \infty} |b_n| = 0$, by the Limit Laws, we know that $\lim_{n \rightarrow \infty} M|b_n| = \lim_{n \rightarrow \infty} (-M|b_n|) = 0$. Hence, by the Squeeze Theorem, we get $\lim_{n \rightarrow \infty} a_n b_n = 0$.

6.18 Theorem: Monotonic Sequence Theorem

Every *bounded, monotonic* sequence is convergent.

Remark:

The proof of the theorem is based on a fundamental property of the real numbers that if a sequence (or a nonempty set) is bounded above, then it has a least upper bound. It is an expression of the fact that there is no gap or hole in the real number line. We omit the details of the proof because it is beyond the scope of the text.

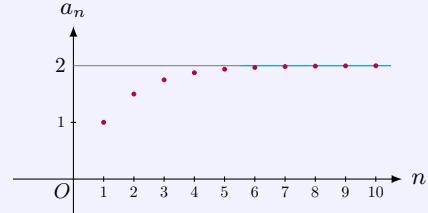
6.19 Example: sequence generated by recurrence relation

Show the sequence $\{a_n\}$, defined by the recurrence relation,

$$a_n = \frac{1}{2}a_{n-1} + 1, \quad (n \geq 2); \quad a_1 = 1,$$

has a limit. What is the limit?

Analysis: By plotting the first few terms of the sequence, we see that the sequence appears increasing and bounded.



Solution We shall prove the following statement by induction: for $n \geq 1$,

$$S_n : \quad \begin{cases} a_{n+1} > a_n, \\ 1 \leq a_n < 2. \end{cases}$$

In fact, for $n = 1$, we have

$$\begin{aligned} a_2 &= \frac{1}{2}a_1 + 1 = \frac{1}{2} \cdot 1 + 1 = \frac{3}{2} > 1 = a_1, \\ 1 &= a_1 < 2. \end{aligned}$$

So S_1 is true. Suppose the statement is true for some k :

$$S_k : \quad \begin{cases} a_{k+1} > a_k, \\ 1 \leq a_k < 2. \end{cases}$$

Then,

$$a_{k+2} = \frac{1}{2}a_{k+1} + 1 > \frac{1}{2}a_k + 1 = a_{k+1},$$

and

$$1 \leq a_k < a_{k+1} = \frac{1}{2}a_k + 1 < \frac{1}{2} \cdot 2 + 1 = 2,$$

that is

$$S_{k+1} : \quad \begin{cases} a_{k+2} > a_{k+1}, \\ 1 \leq a_{k+1} < 2, \end{cases}$$

S_{k+1} is also true. Hence, by the principle of induction, the statement S_n is true for all $n \geq 1$.

Because $\{a_n\}$ is increasing and bounded, by the Monotonic Sequence Theorem, $\lim_{n \rightarrow \infty} a_n$ exists, say L .

To find the limit L , we let n tend to ∞ on the both sides of the recurrence relation $a_n = \frac{1}{2}a_{n-1} + 1$. This yields $L = \frac{1}{2}L + 1$. Solving the equation gives $L = 2$.

6.20 MATLAB: limit of sequence

We will compute the limit of the following sequences in MATLAB.

1. $\lim_{n \rightarrow \infty} \sqrt[n]{a}$ for $a > 0$.

Solution

```
>> syms a n
>> assume(a, 'positive') % a>0
>> assume(n, 'positive') % n>0
>> assumeAlso(n, 'integer') % n is also an integer
>> limit(a^(1/n), n, Inf)

ans =
1
```

2. $\lim_{n \rightarrow \infty} \sqrt[n]{n}$.

Solution

```
>> syms n
>> assume(n, 'positive') % n>0
>> assumeAlso(n, 'integer') % n is also an integer
>> limit(n^(1/n), n, Inf)

ans =
1
```

3. $\lim_{n \rightarrow \infty} \left(1 - \frac{2}{n}\right)^n$.

Solution

```
>> syms n
>> assume(n, 'positive') % n>0
>> assumeAlso(n, 'integer') % n is also an integer
>> limit((1 - 2/n)^n, n, Inf)
```

```
ans =
exp(-2)
```

6.21 Definition: convergent series

For a given series $\sum_{k=1}^{\infty} a_k = a_1 + a_2 + a_3 + \dots$, let s_n denote its ***n*th partial sum**:

$$s_n = \sum_{k=1}^n a_k = a_1 + a_2 + \dots + a_n.$$

If the sequence $\{s_n\}$ is convergent and $\lim_{n \rightarrow \infty} s_n = s$ exists as a real number, then the series $\sum a_k$ is called **convergent** and we write

$$a_1 + a_2 + a_3 + \dots = s \quad \text{or} \quad \sum_{k=1}^{\infty} a_k = s.$$

The number s is called the **sum** of the series. If the sequence $\{s_n\}$ is divergent, then the series is called **divergent**.

6.22 Theorem: geometric series

Let $a \neq 0$ and r be a real number. Then

$$\sum_{k=1}^{\infty} ar^{k-1} = \begin{cases} \frac{a}{1-r}, & \text{if } |r| < 1; \\ \text{diverges,} & \text{if } |r| \geq 1. \end{cases}$$

Definition: geometric series

The series

$$\sum_{k=1}^{\infty} ar^{k-1} = a + ar + ar^2 + \dots$$

is called a **geometric series**. Each term is obtained from the preceding one by multiplying it by the *common ratio* r .

Justification:

If $r = 1$, the n th partial sum $s_n = a + a + \cdots + a = na$ diverges either to ∞ or $-\infty$, as n tends to ∞ .

If $r \neq 1$, the n th partial sum s_n can be found as follows: by multiplying r to both side of

$$s_n = a + ar + ar^2 + \cdots + ar^{n-1},$$

we get

$$rs_n = ar + ar^2 + \cdots + ar^{n-1} + ar^n.$$

Subtracting these equations, we get

$$s_n - rs_n = a - ar^n,$$

so that

$$s_n = \frac{a - ar^n}{1 - r}.$$

Thus, if $-1 < r < 1$, since $r^n \rightarrow 0$ as $n \rightarrow \infty$, we get

$$\lim_{n \rightarrow \infty} s_n = \lim_{n \rightarrow \infty} \frac{a - ar^n}{1 - r} = \frac{a - a \cdot \lim_{n \rightarrow \infty} r^n}{1 - r} = \frac{a}{1 - r}.$$

If $r \leq -1$ or $r > 1$, the sequence $\{r^n\}$ is divergent, so $\{s_n\}$ is also divergent. Thus, if $r \leq -1$ or $r > 1$, the geometric series diverges.

6.23 Example: repeating decimal

Express $3.\overline{0127}$ as a fraction.

Solution The given number can be written as a geometric series:

$$\begin{aligned} 3.\overline{0127} &= 3 + 0.\overline{0127} = 3 + 0.0127 + 0.0000127 + 0.0000000127 + \cdots \\ &= 3 + \underbrace{0.0127 + 0.0127 \cdot 10^{-3} + 0.0127 \cdot 10^{-6} + \cdots}_{\text{geometric series}} \\ &= 3 + \frac{0.0127}{1 - 10^{-3}} \\ &= 3 + \frac{127 \times 10^{-4}}{10^{-3} \times (10^3 - 1)} \\ &= 3 + \frac{127}{9990} = \frac{30097}{9990}. \end{aligned}$$

Remark:

It is easy to generalize the above argument to show that any repeating decimal can be expressed as a fraction. Hence, any repeating decimal must be a rational number.

6.24 Example: telescoping series

1. Evaluate the series $\sum_{k=1}^{\infty} \left(\frac{1}{2^k} - \frac{1}{2^{k+1}} \right)$.

Solution The partial sum for the given series is

$$\begin{aligned}s_n &= \sum_{k=1}^n \left(\frac{1}{2^k} - \frac{1}{2^{k+1}} \right) \\&= \left(\frac{1}{2^1} - \frac{1}{2^2} \right) + \left(\frac{1}{2^2} - \frac{1}{2^3} \right) + \left(\frac{1}{2^3} - \frac{1}{2^4} \right) + \cdots + \left(\frac{1}{2^n} - \frac{1}{2^{n+1}} \right) \\&= \frac{1}{2^1} - \frac{1}{2^{n+1}}.\end{aligned}$$

Thus,

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

Hence,

$$\sum_{k=1}^{\infty} \left(\frac{1}{2^k} - \frac{1}{2^{k+1}} \right) = \frac{1}{2}.$$

This is called *telescoping series* since the partial sum collapses (like a pirate's collapsing telescope) into just two terms.

2. Evaluate the series $\sum_{k=1}^{\infty} \frac{1}{k(k+1)}$.

Solution The partial sum for the given series can be evaluated by writing it as a telescoping sum:

$$\begin{aligned}s_n &= \sum_{k=1}^n \frac{1}{k(k+1)} = \sum_{k=1}^n \left(\frac{1}{k} - \frac{1}{k+1} \right) \\&= \left(\frac{1}{1} - \frac{1}{2} \right) + \left(\frac{1}{2} - \frac{1}{3} \right) + \left(\frac{1}{3} - \frac{1}{4} \right) + \cdots + \left(\frac{1}{n} - \frac{1}{n+1} \right) \\&= \frac{1}{1} - \frac{1}{n+1}.\end{aligned}$$

Thus,

$$\lim_{n \rightarrow \infty} s_n = \lim_{n \rightarrow \infty} \left(\frac{1}{1} - \frac{1}{n+1} \right) = 1.$$

Hence,

$$\sum_{k=1}^{\infty} \frac{1}{k(k+1)} = 1.$$

6.25 Theorem: properties of convergent series

1. If $\sum a_k$ converges and c is a real number, then the series $\sum ca_k$ converges and

$$\sum ca_k = c \sum a_k.$$

2. If $\sum a_k$ and $\sum b_k$ converge, then the series $\sum(a_k \pm b_k)$ converges and

$$\sum(a_k \pm b_k) = \sum a_k \pm \sum b_k.$$

3. If N is a positive integer, then $\sum_{k=1}^{\infty} a_k$ and $\sum_{k=N}^{\infty} a_k$ either both converge or both diverge.

In general, changing a *finite* number of terms in a convergent series does not change its convergence, although it does change the value of the series.

Justification:

1. Let $s_n = \sum_{k=1}^n a_k$ and $\lim_{n \rightarrow \infty} s_n = s$. Then the n th partial sum for the series $\sum_{k=1}^{\infty} ca_k$ is

$$u_n = \sum_{k=1}^n ca_k = c \sum_{k=1}^n a_k = cs_n,$$

and

$$\lim_{n \rightarrow \infty} u_n = \lim_{n \rightarrow \infty} cs_n = c \cdot \lim_{n \rightarrow \infty} s_n = c \cdot s.$$

That is, $\sum ca_k$ is convergent and

$$\sum_{k=1}^{\infty} ca_k = c \sum_{k=1}^{\infty} a_k.$$

2. Let

$$s_n = \sum_{k=1}^n a_k, \quad t_n = \sum_{k=1}^n b_k,$$

and $\lim_{n \rightarrow \infty} s_n = s$, $\lim_{n \rightarrow \infty} t_n = t$. Then the n th partial sum for the series $\sum_{k=1}^{\infty} (a_k + b_k)$ is

$$u_n = \sum_{k=1}^n (a_k + b_k) = \sum_{k=1}^n a_k + \sum_{k=1}^n b_k = s_n + t_n,$$

and

$$\lim_{n \rightarrow \infty} u_n = \lim_{n \rightarrow \infty} (s_n + t_n) = \lim_{n \rightarrow \infty} s_n + \lim_{n \rightarrow \infty} t_n = s + t.$$

That is, $\sum(a_k + b_k)$ is convergent and

$$\sum_{k=1}^{\infty}(a_k + b_k) = \sum_{k=1}^{\infty} a_k + \sum_{k=1}^{\infty} b_k.$$

Similarly we can show that $\sum(a_k - b_k)$ is convergent and

$$\sum_{k=1}^{\infty}(a_k - b_k) = \sum_{k=1}^{\infty} a_k - \sum_{k=1}^{\infty} b_k.$$

3. For fixed integer N with $1 < N < n$,

$$\sum_{k=N}^n a_k = \sum_{k=1}^n a_k - \sum_{k=1}^{N-1} a_k.$$

If $\sum_{k=1}^{\infty} a_k$ converges, then the right side has a finite value as n tends to ∞ , so $\sum_{k=N}^{\infty} a_k$ converges. Similarly, it is easy to see that, from the equality

$$\sum_{k=1}^n a_k = \sum_{k=1}^{N-1} a_k + \sum_{k=N}^n a_k,$$

if $\sum_{k=N}^{\infty} a_k$ converges, then $\sum_{k=1}^{\infty} a_k$ converges.

If $\sum b_k$ is generated by changing a finite number of terms in a convergent series $\sum a_k$, then there exists a sufficiently large N , such that, when $n > N$,

$$\sum_{k=N}^n b_k = \sum_{k=N}^n a_k.$$

Thus, when $n > N$, the equality holds:

$$\sum_{k=1}^n b_k = \sum_{k=1}^n a_k + K,$$

where K is a fixed constant. It is easy to see from the last equality that $\sum_{k=1}^{\infty} b_k$ converges,

while its value may differ from the value of $\sum_{k=1}^{\infty} a_k$.

6.26 Remark: finding the sums

Besides the geometric series and telescoping series, generally, it is not easy to find the sum of a series. However, it is possible to study convergence of a series without knowing its sum.

6.27 MATLAB: telescoping series

In MATLAB, we can use the command `symsum` to evaluate the sum of a series. In the following, we will show two examples.

$$1. \sum_{k=1}^{\infty} \left(\frac{1}{2^k} - \frac{1}{2^{k+1}} \right).$$

Solution

```
>> syms k
>> f = 1/2^k - 1/2^(k + 1);
>> symsum(f, k, 1, Inf)

ans =
1/2
```

$$2. \sum_{k=1}^{\infty} \frac{1}{k(k+1)}.$$

Solution

```
>> syms k
>> f = 1/(k*(k + 1));
>> symsum(f, k, 1, Inf)

ans =
1
```

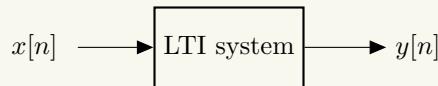
6.28 Engineering Example: discrete-time convolution

The convolution is an operation on two functions to produce a third function that shows how the shape of one of the original functions is modified by the other one. For two functions $f[n]$ and $g[n]$ defined on the set \mathbb{Z} of integers, the discrete-time convolution of $f[n]$ and $g[n]$ is given by

$$f[n] \star g[n] = \sum_{m=-\infty}^{\infty} f[m]g[n-m].$$

The discrete-time convolution has wide applications in engineering fields, e.g., computer vision, natural language processing and digital signal processing (DSP) systems.

For example, the block diagram for a linear time-invariant (LTI) system is shown below, where $x[n]$ is the input and $y[n]$ is the output.



The output of the system $y[n]$ is the **convolution** of the input $x[n]$ with the system's response to an impulse, i.e., the impulse response. Mathematically, it can be written as

$$\begin{aligned} y[n] &= x[n] \star h[n] \\ &= \sum_{m=-\infty}^{\infty} x[m]h[n-m] \end{aligned}$$

where $h[n]$ is the impulse response of the LTI system.

6.29 Engineering Example: discrete-time correlation

Correlation is a way to compare signals with each other and search for similarities between signals. The signal correlation can be performed either between two different signals or with one signal.

Specifically, the correlation between two *different* signals is called **cross-correlation**. Given two signals $x[n]$ and $y[n]$, the cross-correlation between $x[n]$ and $y[n]$ is defined as

$$R_{xy}[k] = \sum_{m=-\infty}^{\infty} x[m]y[m-k].$$

There are many applications of signal cross-correlation in signal processing systems. For example, when a signal is corrupted by another undesirable signal, e.g., noise, during transmission, the cross-correlation is then used to estimate the signal at the receiver side.

The correlation with *one* signal is called **auto-correlation**. For the signal $x[n]$, its auto-correlation is defined as

$$R_{xx}[k] = \sum_{m=-\infty}^{\infty} x[m]x[m-k].$$

The auto-correlation is widely applied in radar, sonar, satellite, wireless communication systems, etc.

6.30 Engineering Example: geometric distribution

Assume a programmer has 85% chance of finding a bug every time he compiles his code. Then, the probability that he finds the bug at the k th ($k \geq 1$) trial can be calculated as

$$p(k) = (1 - 0.85)^{k-1} 0.85.$$

Note that the probability that the programmer finds the bug at the k th trial, i.e., $p(k)$, obeys the geometric distribution.

Suppose that the total number of chances that the programmer can try to find the bug every workday is n . Then, the probability that the programmer can find the bug before he

gets off work is

$$s_n = \sum_{k=1}^n p(k) = \sum_{k=1}^n (1 - 0.85)^{k-1} 0.85.$$

In particular, when $n \rightarrow \infty$, it can be proved that

$$s = \sum_{k=1}^{\infty} (1 - 0.85)^{k-1} 0.85 = 1.$$

The result above shows that the series $\sum_{k=1}^{\infty} (1 - 0.85)^{k-1} 0.85$ converges to 1. Note that this result holds for any bug finding probability $q \in (0, 1]$, i.e., $\sum_{k=1}^{\infty} (1 - q)^{k-1} q = 1$ for any $0 < q \leq 1$. It means if given enough times of trial, no matter how proficient the programmer is, he can find the bug for sure.

6.31 Engineering Example: Poisson distribution

The Poisson distribution is a discrete probability distribution that describes the likelihood of a given number of events occurring within a fixed interval of time (or space), if these events occur 1) with a known constant average rate, and 2) independently of the time since the last event occurred. The Poisson distribution can be used to model quite a large number of processes, e.g., the number of births expected during the night in a hospital, the number of arrivals in any service facility such as at an ATM, bank, bus station and petrol pump, and the number of clicks on any website.

Let the probability of k arrivals at a bus station during a fixed time interval be denoted as $p(k)$. Since $p(k)$ obeys the Poisson distribution, we have

$$p(k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

where λ is the average number of arrivals at the bus station during the fixed time interval. Furthermore, it can be proved that

$$s = \sum_{k=0}^{\infty} p(k) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = 1.$$

The convergent series above means there is either no arrival or at least one arrival at the bus station during the fixed time interval, which is consistent with the intuition.

Divergence Test

6.32 Theorem: Divergence Test

If $\sum a_k$ converges, then $\lim_{k \rightarrow \infty} a_k = 0$. Hence, if $\lim_{k \rightarrow \infty} a_k \neq 0$, then the series diverges.

Justification:

If $\sum a_k$ is convergent, by the definition, $\lim_{n \rightarrow \infty} s_n = s$ exists, where $s_n = \sum_{k=1}^n a_k$. Thus,

$$\lim_{k \rightarrow \infty} a_k = \lim_{k \rightarrow \infty} (s_k - s_{k-1}) = \lim_{k \rightarrow \infty} s_k - \lim_{k \rightarrow \infty} s_{k-1} = s - s = 0.$$

6.33 Example: divergent series

Show that the series $\sum_{k=1}^{\infty} \frac{(-1)^k k^2}{k^2 + 1}$ diverges.

Solution Since, as $k \rightarrow \infty$,

$$\left| \frac{(-1)^k k^2}{k^2 + 1} \right| = \frac{k^2}{k^2 + 1} = \frac{k^2}{k^2 \left(1 + \frac{1}{k^2}\right)} = \frac{1}{1 + \frac{1}{k^2}} \rightarrow 1 \neq 0,$$

we know that $\lim_{k \rightarrow \infty} \frac{(-1)^k k^2}{k^2 + 1} \neq 0$. Hence, by the Divergence Test, the given series is divergent.

6.34 Remark: Divergence Test

1. If $\lim_{k \rightarrow \infty} a_k \neq 0$, then either the limit does not exist, or the limit exists but does not equal to 0.
2. If $\lim_{k \rightarrow \infty} a_k = 0$, the Divergence Test is inconclusive. In other words, the zero limit of the sequence $\{a_n\}$ is not sufficient for the convergence of the series $\sum a_k$. An example of such kind of series is the so-called *harmonic series* $\sum \frac{1}{k}$.

6.35 Example: harmonic series

Show that the harmonic series $\sum_{k=1}^{\infty} \frac{1}{k}$ diverges.

Solution We can show that the partial sums $s_2, s_4, s_8, s_{16}, s_{32}, \dots$ become unbounded, so

the sequence $\{s_n\}$ does not have a finite limit. For the first few terms, we see that

$$\begin{aligned}s_2 &= 1 + \frac{1}{2}, \\ s_4 &= 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) > 1 + \frac{1}{2} + \left(\frac{1}{4} + \frac{1}{4}\right) = 1 + \frac{2}{2}, \\ s_8 &= 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) \\ &> 1 + \frac{1}{2} + \left(\frac{1}{4} + \frac{1}{4}\right) + \left(\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}\right) = 1 + \frac{3}{2}, \\ s_{16} &= 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) + \left(\frac{1}{9} + \frac{1}{10} + \cdots + \frac{1}{16}\right) \\ &> 1 + \frac{1}{2} + \left(\frac{1}{4} + \frac{1}{4}\right) + \left(\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}\right) + \left(\frac{1}{16} + \frac{1}{16} + \cdots + \frac{1}{16}\right) = 1 + \frac{4}{2}.\end{aligned}$$

In fact, we see generally that

$$s_{2^n} > 1 + \frac{n}{2}.$$

Thus, $\{s_n\}$ does not have a finite limit. That is, the harmonic series diverges.

6.36 MATLAB: divergent series

Show the convergence/divergence of the series $\sum_{k=1}^{\infty} \frac{(-1)^k k^2}{k^2 + 1}$ in MATLAB.

Solution We will use the Divergence Test to verify the convergence/divergence of the series, as shown in the following:

```
>> syms k
>> f = (-1)^k*k^2/(k^2 + 1);
>> limit(abs(f), k, Inf)

ans =
1
```

Note that in MATLAB, the command `abs(x)` can calculate the absolute value of x . From the result above, we see that $\lim_{k \rightarrow \infty} \left| \frac{(-1)^k k^2}{k^2 + 1} \right| = 1 \neq 0$, which means $\lim_{k \rightarrow \infty} \frac{(-1)^k k^2}{k^2 + 1} \neq 0$.

Hence, according to the Divergence Test, we know that the series $\sum_{k=1}^{\infty} \frac{(-1)^k k^2}{k^2 + 1}$ is divergent.

6.3 Integral Test

6.37 Theorem: Integral Test

Suppose f is a *continuous, positive, decreasing* function on $[1, \infty)$ and let $a_k = f(k)$. Then the series $\sum_{k=1}^{\infty} a_k$ is convergent if and only if the improper integral $\int_1^{\infty} f(x) dx$ is convergent.

In other words:

1. If $\int_1^\infty f(x) dx$ is convergent, then $\sum_{k=1}^\infty a_k$ is convergent;
2. If $\int_1^\infty f(x) dx$ is divergent, then $\sum_{k=1}^\infty a_k$ is divergent.

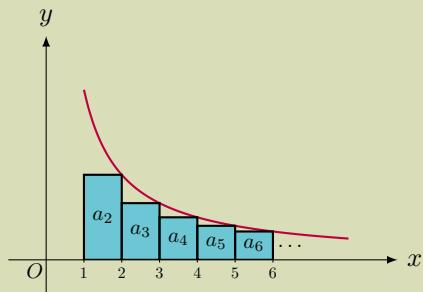
Justification:

1. Suppose $\int_1^\infty f(x) dx$ converges. To show that $\sum_{k=1}^\infty f(k)$ converges, since $f \geq 0$, we only need to show that $\sum_{k=1}^n f(k)$ is bounded with respect to n .

For each positive integer k , put $h(x) = f(k+1)$ for $x \in [k, k+1]$. Then h is a function defined on $[1, \infty)$. Since f is decreasing monotonically, we know that $h \leq f$. Hence

$$\begin{aligned}\sum_{k=1}^n f(k) &= f(1) + \sum_{k=2}^n f(k) \\ &\leq f(1) + \int_1^n h(x) dx \\ &\leq f(1) + \int_1^\infty f(x) dx \leq f(1) + \int_1^\infty f(x) dx,\end{aligned}$$

so that $\sum_{k=1}^n f(k)$ is bounded.

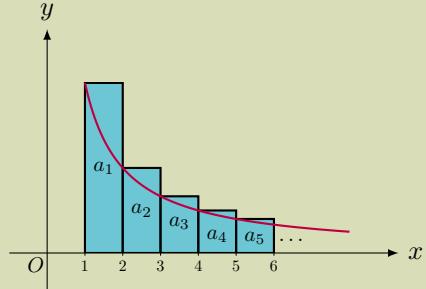


2. Suppose $\int_1^\infty f(x) dx$ is divergent. We show that $\sum_{k=1}^\infty f(k)$ converges by contradiction.

For each positive integer k , put $g(x) = f(k)$ for $x \in [k, k+1]$. Then g is a function defined on $[1, \infty)$. Since f is decreasing monotonically, it is clear that $g \geq f$. For any b , we choose an integer n such that $n \geq b$. Then

$$\begin{aligned} \int_1^b f(x) dx &\leq \int_1^n f(x) dx \leq \int_1^n g(x) dx \\ &= \sum_{k=1}^{n-1} f(k) \leq \sum_{k=1}^{\infty} f(k), \end{aligned}$$

so that $\int_1^b f(x) dx$ is bounded, which implies that the improper integral is convergent. This contradicts to the hypothesis that $\int_1^{\infty} f(x) dx$ is divergent.



6.38 Remark: Integral Test

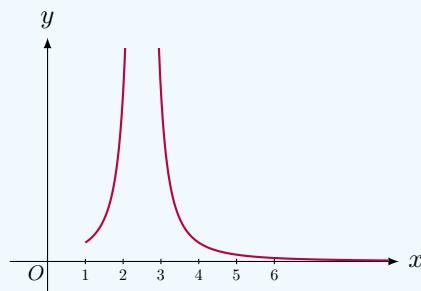
The hypotheses in the Integral Test say that f is a continuous, positive, decreasing function on $[1, \infty)$. However, it is not necessary to start the series or the integral at $n = 1$. In fact, if f is decreasing for x larger than some number N , then the same conclusion holds. The reason is that changing a finite number of terms in a convergent series does not change its convergence.

As an example, consider the infinite series $\sum_{k=1}^{\infty} \frac{1}{(2k-5)^2}$. As usual, if one takes $f(x) = \frac{1}{(2x-5)^2}$, then f is not decreasing on $[1, \infty)$ (also not continuous in this interval). However, we can consider the interval $[3, \infty)$ on which f is continuous, positive, and decreasing. We can similarly have the conclusion that $\sum_{k=3}^{\infty} \frac{1}{(2k-5)^2}$ is convergent since $\int_3^{\infty} \frac{1}{(2x-5)^2} dx$ is convergent.

Since

$$\sum_{k=1}^{\infty} \frac{1}{(2k-5)^2} = \sum_{k=1}^2 \frac{1}{(2k-5)^2} + \sum_{k=3}^{\infty} \frac{1}{(2k-5)^2},$$

we know that $\sum_{k=1}^{\infty} \frac{1}{(2k-5)^2}$ is indeed convergent.



6.39 Example: harmonic series - Integral Test

Apply the Integral Test to show that the harmonic series $\sum_{k=1}^{\infty} \frac{1}{k}$ diverges.

Solution Consider the function $f(x) = \frac{1}{x}$. It is clear that f is a continuous, positive, decreasing function on $[1, \infty)$. The improper integral $\int_1^{\infty} \frac{1}{x} dx$ is divergent:

$$\int_1^{\infty} \frac{1}{x} dx = \lim_{b \rightarrow \infty} \int_1^b \frac{1}{x} dx = \lim_{b \rightarrow \infty} \ln x \Big|_{x=1}^b = \lim_{b \rightarrow \infty} \ln b = \infty.$$

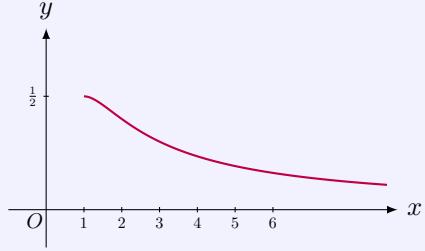
Hence, by the Integral Test, the harmonic series $\sum_{k=1}^{\infty} \frac{1}{k}$ is divergent.

6.40 Example: Integral Test

1. Determine whether the series $\sum_{k=1}^{\infty} \frac{k}{k^2 + 1}$ converges.

Solution Consider the function $f(x) = \frac{x}{x^2 + 1}$. It is clear that f is a continuous, positive function on $[1, \infty)$. It is decreasing on $(1, \infty)$, since

$$\begin{aligned} f'(x) &= \left(\frac{x}{x^2 + 1} \right)' = \frac{1 \cdot (x^2 + 1) - x \cdot 2x}{(x^2 + 1)^2} \\ &= -\frac{x^2 - 1}{(x^2 + 1)^2} < 0. \end{aligned}$$



The improper integral $\int_1^{\infty} \frac{x}{x^2 + 1} dx$ is divergent:

$$\begin{aligned} \int_1^{\infty} \frac{x}{x^2 + 1} dx &= \lim_{b \rightarrow \infty} \int_1^b \frac{x}{x^2 + 1} dx = \lim_{b \rightarrow \infty} \frac{1}{2} \ln(x^2 + 1) \Big|_{x=1}^b \\ &= \lim_{b \rightarrow \infty} \frac{1}{2} [\ln(b^2 + 1) - \ln 2] = \infty. \end{aligned}$$

Hence, by the Integral Test, the series $\sum_{k=1}^{\infty} \frac{k}{k^2 + 1}$ is divergent.

2. Determine whether the series $\sum_{k=1}^{\infty} \frac{1}{k^2 + 2}$ converges.

Solution Consider the function $f(x) = \frac{1}{x^2 + 2}$. It is clear that f is a continuous, posi-

tive, decreasing function on $[1, \infty)$. The improper integral $\int_1^\infty \frac{1}{x^2+2} dx$ is convergent:

$$\begin{aligned}\int_1^\infty \frac{1}{x^2+2} dx &= \lim_{b \rightarrow \infty} \int_1^b \frac{1}{x^2+2} dx \\ &= \lim_{b \rightarrow \infty} \frac{1}{\sqrt{2}} \tan^{-1} \left(\frac{1}{\sqrt{2}} x \right) \Big|_{x=1}^b \\ &= \lim_{b \rightarrow \infty} \frac{1}{\sqrt{2}} \left[\tan^{-1} \left(\frac{1}{\sqrt{2}} b \right) - \tan^{-1} \frac{1}{\sqrt{2}} \right] \\ &= \frac{1}{\sqrt{2}} \left(\frac{\pi}{2} - \tan^{-1} \frac{1}{\sqrt{2}} \right).\end{aligned}$$

Hence, by the Integral Test, the series $\sum_{k=1}^\infty \frac{1}{k^2+2}$ is convergent.

3. Determine whether the series $\sum_{k=1}^\infty \frac{1}{\sqrt{2k-1}}$ converges.

Solution Consider the function $f(x) = \frac{1}{\sqrt{2x-1}}$. It is clear that f is a continuous, positive, decreasing function on $[1, \infty)$. The improper integral $\int_1^\infty \frac{1}{\sqrt{2x-1}} dx$ is divergent:

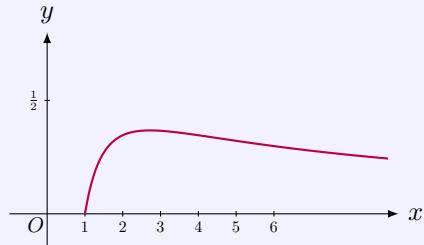
$$\begin{aligned}\int_1^\infty \frac{1}{\sqrt{2x-1}} dx &= \lim_{b \rightarrow \infty} \int_1^b \frac{1}{\sqrt{2x-1}} dx \\ &= \lim_{b \rightarrow \infty} \sqrt{2x-1} \Big|_{x=1}^b = \lim_{b \rightarrow \infty} (\sqrt{2b-1} - 1) = \infty.\end{aligned}$$

Hence, by the Integral Test, the series $\sum_{k=1}^\infty \frac{1}{\sqrt{2k-1}}$ is divergent.

4. Determine whether the series $\sum_{k=1}^\infty \frac{\ln k}{k}$ converges.

Solution Consider the function $f(x) = \frac{\ln x}{x}$. It is clear that f is a continuous, positive function on $[1, \infty)$. It is decreasing on $(3, \infty)$, since

$$\begin{aligned}f'(x) &= \left(\frac{\ln x}{x} \right)' = \frac{\frac{1}{x} \cdot x - \ln x \cdot 1}{x^2} \\ &= -\frac{\ln x - 1}{(x^2 + 1)^2} < 0.\end{aligned}$$



The improper integral $\int_3^\infty \frac{\ln x}{x} dx$ is divergent:

$$\begin{aligned}\int_3^\infty \frac{\ln x}{x} dx &= \lim_{b \rightarrow \infty} \int_3^b \frac{\ln x}{x} dx \\ &= \lim_{b \rightarrow \infty} \frac{1}{2} \ln^2 x \Big|_{x=3}^b = \lim_{b \rightarrow \infty} \frac{1}{2} (\ln^2 b - \ln^2 3) = \infty.\end{aligned}$$

Thus, by the Integral Test, the series $\sum_{k=3}^\infty \frac{\ln k}{k}$ is divergent. Hence $\sum_{k=1}^\infty \frac{\ln k}{k}$ is divergent.

5. Show that the series $\sum_{k=1}^\infty \frac{1}{\sqrt{k^3}}$ converges.

Solution Consider the function $f(x) = \frac{1}{\sqrt{x^3}}$. It is clear that f is a continuous, positive, decreasing function on $[1, \infty)$. The improper integral $\int_1^\infty \frac{1}{\sqrt{x^3}} dx$ is convergent:

$$\begin{aligned}\int_1^\infty \frac{1}{\sqrt{x^3}} dx &= \lim_{b \rightarrow \infty} \int_1^b \frac{1}{\sqrt{x^3}} dx \\ &= \lim_{b \rightarrow \infty} -\frac{2}{\sqrt{x}} \Big|_{x=1}^b = \lim_{b \rightarrow \infty} 2 \left(1 - \frac{1}{\sqrt{b}}\right) = 2.\end{aligned}$$

Hence, by the Integral Test, the series $\sum_{k=1}^\infty \frac{1}{\sqrt{k^3}}$ is convergent. This is a special case of the p -series.

6.41 Theorem: convergence of the p -series

The **p -series** $\sum_{k=1}^\infty \frac{1}{k^p}$ converges for $p > 1$ and diverges for $p \leq 1$.

Justification:

If $p < 0$, then $\lim_{k \rightarrow \infty} \frac{1}{k^p} = \infty$. If $p = 0$, then $\lim_{k \rightarrow \infty} \frac{1}{k^p} = 1$. In either case $\lim_{k \rightarrow \infty} \frac{1}{k^p} \neq 0$, so the p -series diverges by the Divergence Test.

If $p > 0$, then the function $f(x) = \frac{1}{x^p}$ is clearly continuous, positive, and decreasing on

$[1, \infty)$. For the improper integral $\int_1^\infty \frac{1}{x^p} dx$,

$$\begin{aligned}\text{when } p \neq 1 : \quad \int_1^\infty \frac{1}{x^p} dx &= \lim_{b \rightarrow \infty} \int_1^b \frac{1}{x^p} dx = \lim_{b \rightarrow \infty} \left. \frac{x^{-p+1}}{-p+1} \right|_{x=1}^b \\ &= \lim_{b \rightarrow \infty} \left(\frac{b^{-p+1}}{-p+1} - \frac{1}{-p+1} \right) \\ &= \begin{cases} \frac{1}{p-1}, & \text{if } p > 1; \\ \infty, & \text{if } p < 1; \end{cases}\end{aligned}$$

$$\begin{aligned}\text{when } p = 1 : \quad \int_1^\infty \frac{1}{x^p} dx &= \lim_{b \rightarrow \infty} \int_1^b \frac{1}{x} dx = \lim_{b \rightarrow \infty} \ln x|_{x=1}^b \\ &= \lim_{b \rightarrow \infty} \ln b = \infty,\end{aligned}$$

so that it converges if $p > 1$ and diverges if $p \leq 1$. It follows from the Integral Test that the p -series $\sum_{k=1}^\infty \frac{1}{k^p}$ converges if $p > 1$ and diverges if $0 < p \leq 1$.

6.42 MATLAB: Integral Test

We will determine whether the following series are convergent by applying the Integral Test.

$$1. \sum_{k=1}^\infty \frac{k}{k^2 + 1}.$$

Solution

```
>> syms x
>> f(x) = x/(x^2 + 1);
>> int(f(x), 1, Inf)

ans =
Inf
```

Since $\int_1^\infty \frac{x}{x^2 + 1} dx = \infty$, we thus know that $\sum_{k=1}^\infty \frac{k}{k^2 + 1}$ is divergent.

$$2. \sum_{k=1}^\infty \frac{1}{k^2 + 2}.$$

Solution

```
>> syms x
```

```
>> f(x) = 1/(x^2 + 2);
>> int(f(x),1,Inf)

ans =

(pi*2^(1/2))/4 - (2^(1/2)*atan(2^(1/2)/2))/2
```

We see from the result above that $\int_1^\infty \frac{1}{x^2+2} dx < \infty$, we thus know that $\sum_{k=1}^{\infty} \frac{1}{k^2+2}$ is convergent.

3. $\sum_{k=1}^{\infty} \frac{1}{\sqrt{2k-1}}.$

Solution

```
>> syms x
>> f(x) = 1/sqrt(2*x - 1);
>> int(f(x),1,Inf)

ans =

Inf
```

Since $\int_1^\infty \frac{1}{\sqrt{2x-1}} dx = \infty$, the series $\sum_{k=1}^{\infty} \frac{1}{\sqrt{2k-1}}$ is thus divergent.

4. $\sum_{k=1}^{\infty} \frac{\ln k}{k}.$

Solution

```
>> syms x
>> f(x) = log(x)/x;
>> int(f(x),1,Inf)

ans =

Inf
```

Since $\int_1^\infty \frac{\ln x}{x} dx = \infty$, the series $\sum_{k=1}^{\infty} \frac{\ln k}{k}$ is thus divergent.

5. $\sum_{k=1}^{\infty} \frac{1}{\sqrt{k^3}}.$

```
>> syms x
```

```

>> f(x) = 1/sqrt(x^3);
>> int(f(x),1,Inf)

ans =

2

```

We see from the result above that $\int_1^\infty \frac{1}{\sqrt{x^3}} dx < \infty$, the series $\sum_{k=1}^\infty \frac{1}{\sqrt{k^3}}$ is thus convergent.

6.43 Theorem: remainder estimate for the Integral Test

Suppose f is a *continuous, positive, decreasing* function on $[1, \infty)$. Let $a_k = f(k)$ and suppose $\sum_{k=1}^\infty a_k = s$ is *convergent*. Denote $R_n = s - s_n = \sum_{k=n+1}^\infty a_k$. Then

1. $\int_{n+1}^\infty f(x) dx \leq R_n \leq \int_n^\infty f(x) dx;$
2. $s_n + \int_{n+1}^\infty f(x) dx \leq s \leq s_n + \int_n^\infty f(x) dx.$

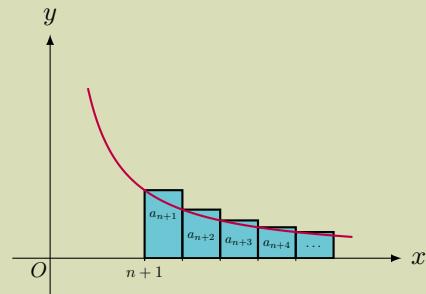
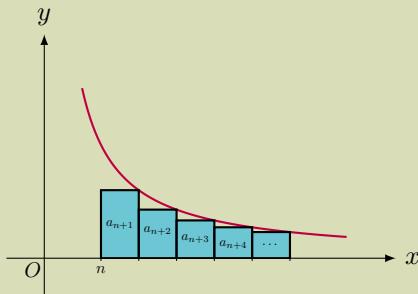
Justification:

1. We apply the same ideas as in the Integral Test, assuming that f is decreasing on $[n, \infty)$. Comparing the areas of the rectangles with the area under $y = f(x)$ for $x > n$, we see that

$$R_n = a_{n+1} + a_{n+2} + \dots \leq \int_n^\infty f(x) dx.$$

Similarly, we have

$$R_n = a_{n+1} + a_{n+2} + \dots \geq \int_{n+1}^\infty f(x) dx.$$



2. Adding s_n to each side of the inequalities in part 1, we get the desired inequalities in part 2.

6.44 Example:

In 1734, Euler derived the equality

$$\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \dots = \frac{\pi^2}{6}.$$

1. How many terms of the series must be used to obtain an approximation that is within 10^{-4} of the exact value of the series?
2. Find an approximation to the series using 50 terms of the series.

Solution The function associated with this series is $f(x) = \frac{1}{x^2}$. Clearly it is continuous, positive, and decreasing on $[1, \infty)$.

1. Using the bound on the remainder in the Integral Test, we have

$$R_n < \int_n^{\infty} f(x) dx = \int_n^{\infty} \frac{dx}{x^2} = \frac{1}{n}.$$

Thus, to ensure that $R_n < 10^{-4}$, we need to choose n so that $\frac{1}{n} < 10^{-4}$, or $n > 10,000$.

Hence, the finite series $\sum_{k=1}^{10,001} \frac{1}{k^2}$ gives an approximation that is within 10^{-4} of the exact value of the series.

2. Using the bounds on the series in the Integral Test, we have

$$s_{50} + \int_{51}^{\infty} \frac{dx}{x^2} < s < s_{50} + \int_{50}^{\infty} \frac{dx}{x^2},$$

so that

$$s_{50} + \frac{1}{51} < s < s_{50} + \frac{1}{50},$$

where $s_{50} = \sum_{k=1}^{50} \frac{1}{k^2}$. By using a calculator, we have $s_{50} \approx 1.62513273$, which has 8th decimal place accuracy. Thus,

$$1.64474057 < s < 1.64513273.$$

Taking the average of these two bounds as our approximation of s , we find that $s \approx 1.64493665$.

For comparison, $\frac{\pi^2}{6} = 1.64493406\cdots$. We see that our approximation matches first five decimal places.

6.4 Convergence Tests for Nonnegative Series

Comparison Test

6.45 Theorem: Comparison Test

Suppose that $\sum a_k$ and $\sum b_k$ are series with *nonnegative* terms.

1. If $\sum b_k$ is *convergent* and $a_k \leq b_k$ for all k , then $\sum a_k$ is also convergent;
2. If $\sum b_k$ is *divergent* and $a_k \geq b_k$ for all k , then $\sum a_k$ is also divergent.

Justification:

1. Suppose $\sum b_k$ is convergent. Let

$$s_n = \sum_{k=1}^n a_k, \quad t_n = \sum_{k=1}^n b_k, \quad t = \sum_{k=1}^{\infty} b_k$$

Since both series have positive terms, the sequences $\{s_n\}$ and $\{t_n\}$ are increasing. Because $t_n \rightarrow t$, we have $t_n \leq t$ for all n . Since $a_k \leq b_k$, 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 $\sum a_k$ converges.

2. Suppose $\sum b_k$ is divergent. We show that $\sum a_k$ is divergent by contradiction.

In fact, if $\sum a_k$ is convergent, because $b_k \leq a_k$, from part 1, we know that $\sum b_k$ is convergent, a contradiction.

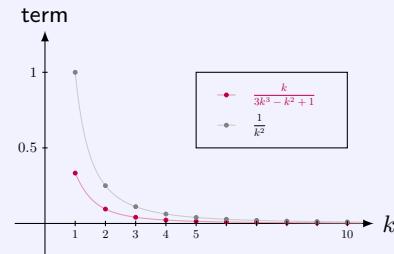
6.46 Example: Comparison Test

1. Determine whether the series $\sum_{k=1}^{\infty} \frac{k}{3k^3 - k^2 + 1}$ converges.

Solution The given series has all positive terms. For any integer $k \geq 1$,

$$\begin{aligned} \frac{k}{3k^3 - 2k + 1} &= \frac{k}{2k^3 + (k^3 - k^2) + 1} \\ &\leq \frac{k}{2k^3 + 1} < \frac{k}{2k^3} \\ &< \frac{1}{k^2}. \end{aligned}$$

Since the p -series $\sum_{k=1}^{\infty} \frac{1}{k^2}$ is convergent ($p = 2$), by the Comparison Test, the series $\sum_{k=1}^{\infty} \frac{k}{3k^3 - k^2 + 1}$ is convergent.

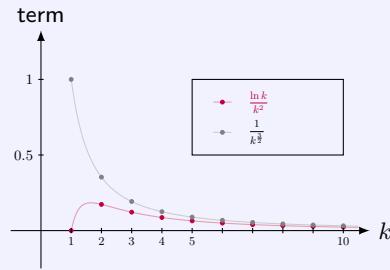


2. Determine whether the series $\sum_{k=1}^{\infty} \frac{\ln k}{k^2}$ converges.

Solution The given series has nonnegative terms. For any integer $k \geq 1$, since $\ln k < \sqrt{k}$, we know that

$$\frac{\ln k}{k^2} < \frac{\sqrt{k}}{k^2} = \frac{1}{k^{3/2}}.$$

Since the p -series $\sum_{k=1}^{\infty} \frac{1}{k^{3/2}}$ is convergent ($p = \frac{3}{2}$), by the Comparison Test, the series $\sum_{k=1}^{\infty} \frac{\ln k}{k^2}$ is convergent.



6.47 Example: Comparison Test - decimal series

Show that any decimal series of the form

$$0.a_1a_2a_3\cdots,$$

where a_i is an integer satisfying $0 \leq a_i \leq 9$, is convergent. In other words, it always represents a real number.

Solution The decimal series can be rewritten as a series

$$0.a_1a_2a_3\cdots = a_1 \cdot 10^{-1} + a_2 \cdot 10^{-2} + a_3 \cdot 10^{-3} + \cdots = \sum_{k=1}^{\infty} a_k \cdot 10^{-k}.$$

It is clear that it is a series with nonnegative terms. Since $0 \leq a_k \leq 9$, we know that

$$0 \leq a_k \cdot 10^{-k} \leq 9 \cdot 10^{-k} \quad \text{for all } k \geq 1.$$

The geometric series $\sum_{k=1}^{\infty} 9 \cdot 10^{-k}$ has common ratio 10^{-1} , so it converges. Hence, by the Comparison Test, the decimal series always converges.

6.48 Theorem: Limit Comparison Test

Suppose that $\sum a_k$ and $\sum b_k$ are series with $a_k \geq 0$ and $b_k > 0$. Assume

$$\lim_{k \rightarrow \infty} \frac{a_k}{b_k} = c.$$

1. If $0 < c < \infty$ (that is, c is a finite positive number), then $\sum a_k$ and $\sum b_k$ either both converge or both diverge.
2. If $c = 0$ and $\sum b_k$ converges, then $\sum a_k$ converges.
3. If $c = \infty$ and $\sum b_k$ diverges, then $\sum a_k$ diverges.

Justification:

1. If $0 < c < \infty$, let m and M be positive numbers such that $m < c < M$. Because $\frac{a_k}{b_k} \rightarrow c$ as $k \rightarrow \infty$, there is an integer K such that

$$m < \frac{a_k}{b_k} < M \quad \text{for } k > K,$$

or equivalently,

$$mb_k < a_k < Mb_k \quad \text{for } k > K.$$

If $\sum b_k$ converges, so does $\sum Mb_k$. Thus $\sum a_k$ converges by part 1 of the Comparison Test. If $\sum b_k$ diverges, so does $\sum mb_k$ and part 2 of the Comparison Test shows that $\sum a_n$ diverges.

2. If $c = 0$, then, similarly as in Case 1, there is an integer K such that

$$a_k < Mb_k \quad \text{for } k > K.$$

It follows from part 1 of the Comparison Test that if $\sum b_k$ converges, then $\sum a_k$ converges.

3. If $c = \infty$, similarly as in Case 1, there is an integer K such that

$$mb_k < a_k \quad \text{for } k > K.$$

It follows from part 2 of the Comparison Test that if $\sum b_k$ diverges, then $\sum a_k$ diverges.

6.49 Example: Limit Comparison Test

1. Determine whether the series $\sum_{k=1}^{\infty} \frac{k}{3k^3 - k^2 + 1}$ converges.

Solution The given series has all positive terms. We calculate the limit of the ratio of the k th terms of the series and the p -series $\sum_{k=1}^{\infty} \frac{1}{k^2}$ ($p = 2$):

$$\lim_{k \rightarrow \infty} \frac{\frac{k}{3k^3 - 2k + 1}}{\frac{1}{k^2}} = \lim_{k \rightarrow \infty} \frac{k^3}{3k^3 - 2k + 1} = \lim_{k \rightarrow \infty} \frac{1}{3 - \frac{2}{k^2} + \frac{1}{k^3}} = \frac{1}{3}.$$

Since the p -series ($p = 2$) is convergent, by the Limit Comparison Test, the series $\sum_{k=1}^{\infty} \frac{k}{3k^3 - k^2 + 1}$ is convergent.

2. Determine whether the series $\sum_{k=1}^{\infty} \frac{k^2}{3k^3 - k^2 + 1}$ converges.

Solution The given series has all positive terms. We calculate the limit of the ratio of

the k th terms of the series and the harmonic series $\sum_{k=1}^{\infty} \frac{1}{k}$:

$$\lim_{k \rightarrow \infty} \frac{\frac{k^2}{3k^3 - 2k + 1}}{\frac{1}{k}} = \lim_{k \rightarrow \infty} \frac{k^3}{3k^3 - 2k + 1} = \lim_{k \rightarrow \infty} \frac{1}{3 - \frac{2}{k^2} + \frac{1}{k^3}} = \frac{1}{3}.$$

Since the harmonic series is divergent, by the Limit Comparison Test, the series $\sum_{k=1}^{\infty} \frac{k^2}{3k^3 - k^2 + 1}$ is divergent.

3. Determine whether the series $\sum_{k=1}^{\infty} \frac{\ln k}{k^2}$ converges.

Solution The given series has nonnegative terms. Since $\ln k \ll k^{\frac{3}{2}}$ as $k \rightarrow \infty$, we have

$$\lim_{k \rightarrow \infty} \frac{\frac{\ln k}{k^2}}{\frac{1}{k^{\frac{3}{2}}}} = \lim_{k \rightarrow \infty} \frac{\ln k}{k^{\frac{1}{2}}} = 0.$$

It is known that the p -series $\sum_{k=1}^{\infty} \frac{1}{k^{\frac{3}{2}}}$ is convergent ($p = \frac{3}{2}$), by the Limit Comparison Test, the series $\sum_{k=1}^{\infty} \frac{\ln k}{k^2}$ is convergent.

4. Determine whether the series $\sum_{k=1}^{\infty} \frac{\ln k}{k}$ converges.

Solution The given series has nonnegative terms. Since $1 \ll \ln k$ as $k \rightarrow \infty$, we have

$$\lim_{k \rightarrow \infty} \frac{\frac{\ln k}{k}}{\frac{1}{k}} = \lim_{k \rightarrow \infty} \ln k = \infty.$$

It is known that the harmonic series $\sum_{k=1}^{\infty} \frac{1}{k}$ is divergent, by the Limit Comparison Test, the series $\sum_{k=1}^{\infty} \frac{\ln k}{k}$ is divergent.

Ratio Test and Root Test

6.50 Theorem: Ratio Test

Suppose that $\sum a_k$ is a series with *positive* terms.

1. If $\lim_{k \rightarrow \infty} \frac{a_{k+1}}{a_k} = L < 1$, then the series $\sum a_k$ converges.
2. If $\lim_{k \rightarrow \infty} \frac{a_{k+1}}{a_k} = L > 1$ or $\lim_{k \rightarrow \infty} \frac{a_{k+1}}{a_k} = \infty$, then the series $\sum a_k$ diverges.
3. If $\lim_{k \rightarrow \infty} \frac{a_{k+1}}{a_k} = 1$, the test is inconclusive.

Justification:

1. If $\lim_{k \rightarrow \infty} \frac{a_{k+1}}{a_k} = L < 1$, we take any number r such that $L < r < 1$. Thus, there is K such that when $k \geq K$,

$$\begin{aligned} \left| \frac{a_{k+1}}{a_k} - L \right| &< r - L \implies -(r - L) < \frac{a_{k+1}}{a_k} - L < r - L \\ &\implies \frac{a_{k+1}}{a_k} < r \\ &\implies a_{k+1} < r \cdot a_k. \end{aligned}$$

Putting k successively equal to $K, K + 1, K + 2, \dots$ in the inequality, we obtain

$$\begin{aligned} a_{K+1} &< r \cdot a_K, \\ a_{K+2} &< r \cdot a_{K+1} < r^2 \cdot a_K, \\ a_{K+3} &< r \cdot a_{K+2} < r^3 \cdot a_K, \\ &\vdots \end{aligned}$$

so that

$$a_{K+k} < r^k \cdot a_K, \quad \text{for all } k \geq 1.$$

Since $|r| = r < 1$, the geometric series $\sum_{k=1}^{\infty} a_K r^k$ converges. By the Comparison Test, the series $\sum_{k=K+1}^{\infty} a_k$ converges. Hence, $\sum_{k=1}^{\infty} a_k$ converges.

2. To show $\sum a_k$ diverges, by the Divergence Test, it is sufficient to show that $\lim_{k \rightarrow \infty} a_k \neq 0$.

Since $\lim_{k \rightarrow \infty} \frac{a_{k+1}}{a_k} = L > 1$, there is K such that when $k \geq K$,

$$\begin{aligned} \left| \frac{a_{k+1}}{a_k} - L \right| &< (L - 1) \implies -(L - 1) < \frac{a_{k+1}}{a_k} - L \\ &\implies \frac{a_{k+1}}{a_k} > 1 \\ &\implies a_{k+1} > a_k > \dots > a_K. \end{aligned}$$

Hence, $a_k \geq a_K$ for $k \geq K$. This implies that $\lim_{k \rightarrow \infty} a_k \neq 0$.

Similarly as above, if $\lim_{k \rightarrow \infty} \frac{a_{k+1}}{a_k} = \infty$, we can also show that $\lim_{k \rightarrow \infty} a_k \neq 0$.

3. Consider the p -series $\sum_{k=1}^{\infty} \frac{1}{k}$ with $p = 1$ and $\sum_{k=1}^{\infty} \frac{1}{k^2}$ with $p = 2 > 1$. The former is divergent and the latter is convergent. For both series, we have

$$\lim_{k \rightarrow \infty} \frac{a_{k+1}}{a_k} = 1.$$

Hence, the Ratio Test is inconclusive if $\lim_{k \rightarrow \infty} \frac{a_{k+1}}{a_k} = 1$.

6.51 Example: Ratio Test

1. Determine whether the series $\sum_{k=1}^{\infty} \frac{b^k}{k!}$ converges, where $b > 0$.

Solution The given series has all positive terms. Since, for any $b > 0$,

$$\lim_{k \rightarrow \infty} \frac{\frac{b^{k+1}}{(k+1)!}}{\frac{b^k}{k!}} = \lim_{k \rightarrow \infty} \frac{b}{k+1} = 0,$$

by the Ratio Test, the series $\sum_{k=1}^{\infty} \frac{b^k}{k!}$ is convergent for any positive b .

2. Determine whether the series $\sum_{k=1}^{\infty} \frac{k!}{k^k}$ converges.

Solution The given series has all positive terms. The ratio of two consecutive terms is

$$\frac{\frac{(k+1)!}{(k+1)^{k+1}}}{\frac{k!}{k^k}} = \frac{1}{\frac{(k+1)^k}{k^k}} = \frac{1}{\left(1 + \frac{1}{k}\right)^k}.$$

Since $e = \lim_{x \rightarrow \infty} \left(1 + \frac{1}{x}\right)^x$, we have $e = \lim_{k \rightarrow \infty} \left(1 + \frac{1}{k}\right)^k$, so that

$$\lim_{k \rightarrow \infty} \frac{\frac{(k+1)!}{(k+1)^{k+1}}}{\frac{k!}{k^k}} = \frac{1}{\lim_{k \rightarrow \infty} \left(1 + \frac{1}{k}\right)^k} = \frac{1}{e} < 1.$$

By the Ratio Test, the series $\sum_{k=1}^{\infty} \frac{k!}{k^k}$ is convergent.

6.52 MATLAB: Ratio Test

We will determine whether the following two series converge by applying the Ratio Test.

$$1. \sum_{k=1}^{\infty} \frac{b^k}{k!} \text{ where } b > 0.$$

Solution

```
>> syms b k
>> assume(b, 'positive') % b > 0
>> assume(k, 'positive') % k > 0
>> assumeAlso(k, 'integer') % k is also an integer
>> a(k) = b^k/factorial(k);
>> ratio_a = simplify(a(k + 1)/a(k));
>> limit(ratio_a, k, inf)

ans =
0
```

Note that we can set assumptions on symbolic objects using commands `assume` and `assumeAlso`. To calculate factorial in MATLAB, the command `factorial` is used. Moreover, the command `simplify` is used for symbolic simplification, so that the calculation of the limit for the factorial ratio becomes feasible.

The result above shows that $\lim_{k \rightarrow \infty} \frac{a_{k+1}}{a_k} = 0 < 1$, then the series $\sum_{k=1}^{\infty} \frac{b^k}{k!}$ is convergent.

$$2. \sum_{k=1}^{\infty} \frac{k!}{k^k}.$$

Solution

```
>> syms k
>> assume(k, 'positive') % k > 0
>> assumeAlso(k, 'integer') % k is also an integer
>> a(k) = factorial(k)/k^k;
>> ratio_a = simplify(a(k + 1)/a(k));
>> limit(ratio_a, k, inf)

ans =
exp(-1)
```

From the result above, we see that $\lim_{k \rightarrow \infty} \frac{a_{k+1}}{a_k} = e^{-1} < 1$, so that the series $\sum_{k=1}^{\infty} \frac{k!}{k^k}$ is convergent.

6.53 Theorem: Root Test

Suppose that $\sum a_k$ is a series with *nonnegative* terms.

1. If $\lim_{k \rightarrow \infty} \sqrt[k]{a_k} = L < 1$, then the series $\sum a_k$ converges.
2. If $\lim_{k \rightarrow \infty} \sqrt[k]{a_k} = L > 1$ or $\lim_{k \rightarrow \infty} \sqrt[k]{a_k} = \infty$, then the series $\sum a_k$ diverges.
3. If $\lim_{k \rightarrow \infty} \sqrt[k]{a_k} = 1$, the test is inconclusive.

Justification:

1. If $\lim_{k \rightarrow \infty} \sqrt[k]{a_k} = L < 1$, we take any number r such that $L < r < 1$. Thus, there is K such that when $k \geq K$,

$$\begin{aligned} |\sqrt[k]{a_k} - L| &< r - L \implies -(r - L) &< \sqrt[k]{a_k} - L < r - L \\ &\implies \sqrt[k]{a_k} &< r \\ &\implies a_k &< r^k. \end{aligned}$$

When $|r| < 1$, the geometric series $\sum r^k$ converges. By the Comparison Test, the series $\sum_{k=K}^{\infty} a_k$ converges. Hence, $\sum_{k=1}^{\infty} a_k$ converges.

2. To show $\sum a_k$ diverges, by the Divergence Test, it is sufficient to show that $\lim_{k \rightarrow \infty} a_k \neq 0$.

Since $\lim_{k \rightarrow \infty} \sqrt[k]{a_k} = L > 1$, there is K such that when $k \geq K$,

$$\begin{aligned} |\sqrt[k]{a_k} - L| &< (L - 1) \implies -(L - 1) &< \sqrt[k]{a_k} - L \\ &\implies \sqrt[k]{a_k} &> 1 \\ &\implies a_k &> 1. \end{aligned}$$

This implies that $\lim_{k \rightarrow \infty} a_k \neq 0$.

Similarly as above, if $\lim_{k \rightarrow \infty} \sqrt[k]{a_k} = \infty$, we can also show that $\lim_{k \rightarrow \infty} a_k \neq 0$.

3. Consider the p -series $\sum_{k=1}^{\infty} \frac{1}{k^p}$ with $p = 1$ and $\sum_{k=1}^{\infty} \frac{1}{k^2}$ with $p = 2 > 1$. The former is divergent and the latter is convergent. For both series, we have

$$\lim_{k \rightarrow \infty} \sqrt[k]{\frac{1}{k^p}} = 1.$$

Hence, the Root Test is inconclusive if $\lim_{k \rightarrow \infty} \sqrt[k]{a_k} = 1$.

6.54 Example: Root Test

1. Determine whether the series $\sum_{k=1}^{\infty} \left(\frac{3k-1}{4k+1}\right)^k$ converges.

Solution The given series has all positive terms. It is easy to have

$$\lim_{k \rightarrow \infty} \sqrt[k]{\left(\frac{3k-1}{4k+1}\right)^k} = \lim_{k \rightarrow \infty} \frac{3k-1}{4k+1} = \lim_{k \rightarrow \infty} \frac{3 - \frac{1}{k}}{4 + \frac{1}{k}} = \frac{3}{4} < 1.$$

By the Root Test, the series $\sum_{k=1}^{\infty} \left(\frac{3k-1}{4k+1}\right)^k$ is convergent.

2. Determine whether the series $\sum_{k=1}^{\infty} \left(\frac{k}{k+1}\right)^{-k^2}$ converges.

Solution The given series has all positive terms. Since $e = \lim_{x \rightarrow \infty} \left(1 + \frac{1}{x}\right)^x$, we have

$$e = \lim_{k \rightarrow \infty} \left(1 + \frac{1}{k}\right)^k, \text{ so that}$$

$$\lim_{k \rightarrow \infty} \sqrt[k]{\left(\frac{k}{k+1}\right)^{-k^2}} = \lim_{k \rightarrow \infty} \left(\frac{k+1}{k}\right)^k = \lim_{k \rightarrow \infty} \left(1 + \frac{1}{k}\right)^k = e > 1.$$

By the Root Test, the series $\sum_{k=1}^{\infty} \left(\frac{k}{k+1}\right)^{-k^2}$ is divergent.

6.55 MATLAB: Root Test

We will determine whether the following two series converge by applying the Root Test.

1. $\sum_{k=1}^{\infty} \left(\frac{3k-1}{4k+1}\right)^k$.

Solution

```
>> syms k
>> assume(k, 'positive') % k > 0
>> assumeAlso(k, 'integer') % k is also an integer
>> a(k) = ((3*k - 1)/(4*k + 1))^k;
>> root_a = (a(k))^(1/k);
>> limit(root_a, k, inf)

ans =
```

3/4

The result above shows that $\lim_{k \rightarrow \infty} \frac{a_{k+1}}{a_k} = \frac{3}{4} < 1$, then the series $\sum_{k=1}^{\infty} \left(\frac{3k-1}{4k+1} \right)^k$ is convergent.

$$2. \sum_{k=1}^{\infty} \left(\frac{k}{k+1} \right)^{-k^2}.$$

Solution

```
>> syms k
>> assume(k, 'positive') % k > 0
>> assumeAlso(k, 'integer') % k is also an integer
>> a(k) = (k/(k + 1))^-(-k^2);
>> root_a = (a(k))^(1/k);
>> limit(root_a, k, inf)

ans =

exp(1)
```

We see from the result above that $\lim_{k \rightarrow \infty} \frac{a_{k+1}}{a_k} = e > 1$, so that the series $\sum_{k=1}^{\infty} \left(\frac{k}{k+1} \right)^{-k^2}$ is divergent.

6.56 Example: alternating series

An alternating series is a series whose terms are alternately positive and negative. The following are two examples of alternating series:

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \dots = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k},$$

$$-1 + \sqrt[3]{2} - \sqrt[3]{3} + \sqrt[4]{4} - \sqrt[5]{5} + \sqrt[6]{6} - \dots = \sum_{k=1}^{\infty} (-1)^k \sqrt[k]{k}.$$

In general, an alternating series $\sum_{k=1}^{\infty} a_k$ has its k th term in the form

$$a_k = (-1)^{k-1} b_k \quad \text{or} \quad a_k = (-1)^k b_k,$$

where $b_k > 0$ for all k .

6.57 Theorem: Alternating Series Test

The alternating series

$$\sum_{k=1}^{\infty} (-1)^{k-1} b_k = b_1 - b_2 + b_3 - b_4 + b_5 - b_6 + \dots, \quad b_k > 0$$

converges if

1. if $\{b_k\}$ is a *positive non-increasing* sequence, $b_k \geq b_{k+1} > 0$, and
2. $\lim_{k \rightarrow \infty} b_k = 0$.

Justification:

First consider the even partial sums:

$$\begin{aligned} s_2 &= b_1 - b_2 \geq 0, && \text{since } b_1 \geq b_2 \\ s_4 &= s_2 + (b_3 - b_4) \geq s_2, && \text{since } b_3 \geq b_4 \\ &\vdots \\ s_{2n} &= s_{2n-2} + (b_{2n-1} - b_{2n}) \geq s_{2n-2} && \text{since } b_{2n-1} \geq b_{2n} \end{aligned}$$

Also,

$$s_{2n} = b_1 - (b_2 - b_3) - (b_4 - b_5) - \dots - (b_{2n-2} - b_{2n-1}) - b_{2n} \leq b_1.$$

So, $\{s_{2n}\}$ is increasing and bounded above. Hence, it is convergent by the Monotonic Sequence Theorem. Denote $s = \lim_{n \rightarrow \infty} s_{2n}$. Now we compute the limit of the odd partial sums:

$$\lim_{n \rightarrow \infty} s_{2n+1} = \lim_{n \rightarrow \infty} (s_{2n} + b_{2n+1}) = \lim_{n \rightarrow \infty} s_{2n} + \lim_{n \rightarrow \infty} b_{2n+1} = s + 0 = s.$$

Since both the even and odd partial sums converge to s , we have $\lim_{n \rightarrow \infty} s_n = s$ and so the series is convergent.

6.58 Example: alternating harmonic series

The alternating harmonic series

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \dots = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} = \sum_{k=1}^{\infty} (-1)^{k-1} b_k,$$

where $b_k = \frac{1}{k}$ satisfies

$$b_k \geq b_{k+1} > 0 \quad \text{and} \quad \lim_{k \rightarrow \infty} b_k = 0.$$

Hence, by the Alternating Series Test, the series converges.

6.59 MATLAB: Alternating Series Test

Determine whether the series $\sum_{k=1}^{\infty} (-1)^{k-1} b_k$ where $b_k = \frac{1}{k}$ converges.

Solution Clearly, we know that $b_k = \frac{1}{k} > 0$ when $k \geq 1$. Hence, we will check whether $\lim b_k = 0$ and $b_k \geq b_{k+1}$ in the following:

```
>> syms k
>> assume(k, 'positive') % k > 0
>> assumeAlso(k, 'integer') % k is also an integer
>> b(k) = 1/k;
>> limit(b(k), k, inf)

ans =
0

>> Delta_b = b(k) - b(k + 1);
>> isAlways(Delta_b > 0)

ans =
logical

1
```

Note that the command `isAlways` can be used to test whether a mathematical statement is true or not. If the mathematical statement is true, then the result will be 1; if it is false, then the result will be 0.

From the above result, we see that $\lim b_k = 0$ and $b_k > b_{k+1}$. Therefore, according to the Alternating Series Test, the series $\sum_{k=1}^{\infty} (-1)^{k-1} b_k$ where $b_k = \frac{1}{k}$ is convergent.

6.60 Theorem: remainder estimate for alternating series

If $s = \sum_{k=1}^{\infty} (-1)^k b_k$ is the sum of an alternating series that satisfies

1. $b_{k+1} \leq b_k$ for all k ;

2. $\lim_{k \rightarrow \infty} b_k = 0$

then

$$|R_n| = |s - s_n| \leq b_{n+1}.$$

Justification:

The even partial sums $\{s_{2n}\}$ form an increasing sequence:

$$s_2 = b_1 - b_2 \geq 0,$$

since $b_1 \geq b_2$

$$s_4 = s_2 + (b_3 - b_4) \geq s_2,$$

since $b_3 \geq b_4$

\vdots

$$s_{2n} = s_{2n-2} + (b_{2n-1} - b_{2n}) \geq s_{2n-2}$$

since $b_{2n-1} \geq b_{2n}$

Similarly, the odd partial sums $\{s_{2n+1}\}$ form a decreasing sequence. Because we know that both $\{s_{2n}\}$ and $\{s_{2n+1}\}$ converge to s , we get

$$s_{2n} \leq s \leq s_{2n+1}, \quad \text{for all } n.$$

It follows that

$$|R_n| = |s - s_n| \leq |s_{n+1} - s_n| = b_{n+1}.$$

6.61 Example:

Consider the series

$$\sum_{k=0}^{\infty} \frac{(-1)^k}{k!} = 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \dots$$

1. How many terms of the series must be used to obtain an approximation that is within 10^{-4} of the exact value of the series?
2. What is the approximation value?

Solution Write

$$\sum_{k=0}^{\infty} \frac{(-1)^k}{k!} = \sum_{k=0}^n \frac{(-1)^k}{k!} + R_n.$$

1. Using the bound on the remainder in the Alternating Series Test, we have

$$|R_n| < \frac{1}{(n+1)!}.$$

Thus, to ensure that $|R_n| < 10^{-4}$, we need to choose n so that $\frac{1}{(n+1)!} < 10^{-4}$, or $n > 7$. Hence, the finite series $\sum_{k=0}^7 \frac{(-1)^k}{k!}$ gives an approximation that is within 10^{-4} of the exact value of the series.

2. Using the bounds on the series in the Alternating Series Test, we have

$$s_7 - \frac{1}{8!} < s < s_7 + \frac{1}{8!},$$

where $s_7 = \sum_{k=0}^7 \frac{(-1)^k}{k!}$. By using a calculator, we have $s_7 \approx 0.36786$, which has 5th decimal place accuracy. Thus,

$$0.36784 < s < 0.36789.$$

Taking the average of these two bounds as our approximation of s , we find that $s \approx 0.36787$.

For comparison, $\sum_{k=0}^{\infty} \frac{(-1)^k}{k!} = e^{-1} = 0.3678794412\cdots$. We see that our approximation indeed is within 10^{-4} of the exact value of the series.

Absolute Convergence and Conditional Convergence

6.62 Definition:

absolute convergece and conditional convergence

- A series $\sum a_k$ is called **absolutely convergent** if the series of absolute values $\sum |a_k|$ is convergent.
- A series $\sum a_k$ is called **conditionally convergent** if it is convergent but *not* absolutely convergent.

6.63 Theorem:

convergence of absolutely convergent series

If a series $\sum a_k$ is *absolutely convergent*, then it is convergent.

Justification:

It is clear that

$$0 \leq a_k + |a_k| \leq 2|a_k|, \quad \text{for all } k.$$

If $\sum a_k$ is absolutely convergent, then $\sum 2|a_k|$ is convergent. By the Comparison Test, $\sum(a_k + |a_k|)$ is convergent. Thus,

$$\sum a_k = \sum(a_k + |a_k|) - \sum |a_k|$$

is convergent.

6.64 Example: absolutely/conditionally convergent series

1. The p -series $\sum_{k=1}^{\infty} \frac{1}{k^2}$, with $p = 2$, is convergent, so the alternating series

$$\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k^2} = 1 - \frac{1}{4} + \frac{1}{9} - \frac{1}{16} + \dots$$

is *absolutely convergent*.

2. The alternating harmonic series $\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k}$ is convergent by the Alternating Series Test, but the harmonic series $\sum_{k=1}^{\infty} \frac{1}{k} = \sum_{k=1}^{\infty} \left| \frac{(-1)^{k-1}}{k} \right|$ is divergent, so the alternating harmonic series is *conditionally convergent*.

6.65 Remark: differences between absolutely and conditionally convergent series

There are some fundamental differences between absolutely convergent series and conditionally convergent series.

Riemann discovered a striking result, known as the *Riemann Rearrangement Theorem*. It says that for any given number α , a conditionally convergent series may be rearranged in its terms so that the rearranged series converges to α .

Summary

Definition: limit of infinite sequence

- $\lim_{n \rightarrow \infty} a_n = L \iff$ we can make the terms a_n as close to L as we like by taking n sufficiently large.
 - $\lim_{n \rightarrow \infty} a_n$ exists $\iff \{a_n\}$ **converges** (or is **convergent**).
 - $\lim_{n \rightarrow \infty} a_n$ does not exist $\iff \{a_n\}$ **diverges** (or is **divergent**).
- $\lim_{n \rightarrow \infty} a_n = \infty \iff$ we can make the terms a_n as large as we like by taking n sufficiently large.
 - $\lim_{n \rightarrow \infty} a_n = \infty \iff \{a_n\}$ **diverges to infinity**.
 - $\lim_{n \rightarrow \infty} a_n = -\infty \iff \{a_n\}$ **diverges to negative infinity**.

Definition: terminology for sequences

- $\{a_n\}$ is **increasing** if $a_{n+1} > a_n$. For example, $\{1, 2, 3, \dots\}$.
- $\{a_n\}$ is **non-decreasing** if $a_{n+1} \geq a_n$. For example, $\{1, 1, 2, 2, 3, 3, \dots\}$.
- $\{a_n\}$ is **decreasing** if $a_{n+1} < a_n$. For example, $\{0, -1, -2, -3, \dots\}$.
- $\{a_n\}$ is **non-increasing** if $a_{n+1} \leq a_n$. For example, $\{0, 0, -1, -1, -2, -2, \dots\}$.
- $\{a_n\}$ is **monotonic** if it is either non-increasing or non-decreasing.
- $\{a_n\}$ is **bounded** if there is number M such that $|a_n| \leq M$ for all n .

6.66 Definition: two asymptotic relations

- $a_n \ll b_n$ (as $n \rightarrow \infty$) $\iff \lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 0$.
- $a_n \sim b_n$ (as $n \rightarrow \infty$) $\iff a_n - b_n \ll b_n$ (as $n \rightarrow \infty$) $\iff \lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 1$.

Theorem: limit of a sequence from limit of a function

If $\lim_{x \rightarrow \infty} f(x) = L$ and $f(n) = a_n$ when n is an integer, then $\lim_{n \rightarrow \infty} a_n = L$. The same result holds when $L = \infty$ or $-\infty$.

Theorem: Limit Laws for sequences

Suppose $\{a_n\}$ and $\{b_n\}$ are convergent sequences and c is a constant. Then the following hold.

1. $\lim_{n \rightarrow \infty} (a_n + b_n) = \lim_{n \rightarrow \infty} a_n + \lim_{n \rightarrow \infty} b_n.$
2. $\lim_{n \rightarrow \infty} (a_n - b_n) = \lim_{n \rightarrow \infty} a_n - \lim_{n \rightarrow \infty} b_n.$
3. $\lim_{n \rightarrow \infty} ca_n = c \lim_{n \rightarrow \infty} a_n.$ In particular, $\lim_{n \rightarrow \infty} c = c.$
4. $\lim_{n \rightarrow \infty} (a_n b_n) = \lim_{n \rightarrow \infty} a_n \cdot \lim_{n \rightarrow \infty} b_n.$
5. $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = \frac{\lim_{n \rightarrow \infty} a_n}{\lim_{n \rightarrow \infty} b_n}$ if $\lim_{n \rightarrow \infty} b_n \neq 0.$
6. $\lim_{n \rightarrow \infty} a_n^p = \left[\lim_{n \rightarrow \infty} a_n \right]^p$ if $p > 0$ and $a_n > 0.$

Theorem: Squeeze Theorem

If $a_n \leq b_n \leq c_n$ for $n \geq n_0$ and $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} c_n = L$, then $\lim_{n \rightarrow \infty} b_n = L.$

Theorem: growth rates of sequences

For any real numbers $p > 0$, $q > 0$, $r > 0$, and $b > 1$, the following asymptotic relations hold: as $n \rightarrow \infty$,

$$\ln^q n \ll n^p \ll n^p \ln^r n \ll b^n \ll n! \ll n^n.$$

Theorem: geometric sequences

Let r be a real number. Then

$$\lim_{n \rightarrow \infty} r^n = \begin{cases} 0, & \text{if } |r| < 1; \\ 1, & \text{if } r = 1; \\ \text{diverges}, & \text{if } r \leq -1 \text{ or } r > 1. \end{cases}$$

Theorem: sequential limit for continuous function

If $\lim_{n \rightarrow \infty} a_n = L$ and the function f is continuous at L , then

$$\lim_{n \rightarrow \infty} f(a_n) = f(L).$$

Theorem: Monotonic Sequence Theorem

Every *bounded, monotonic* sequence is convergent.

Definition: convergent series

$$\sum_{k=1}^{\infty} a_k \text{ converges to } s \iff \lim_{n \rightarrow \infty} s_n = s, \text{ where } s_n = \sum_{k=1}^n a_k.$$

Theorem: geometric series

Let $a \neq 0$ and r be a real number. Then

$$\sum_{k=1}^{\infty} ar^{k-1} = \begin{cases} \frac{a}{1-r}, & \text{if } |r| < 1; \\ \text{diverges,} & \text{if } |r| \geq 1. \end{cases}$$

Theorem: properties of convergent series

1. If $\sum a_k$ converges and c is a real number, then the series $\sum ca_k$ converges and

$$\sum ca_k = c \sum a_k.$$

2. If $\sum a_k$ and $\sum b_k$ converge, then the series $\sum(a_k \pm b_k)$ converges and

$$\sum(a_k \pm b_k) = \sum a_k \pm \sum b_k.$$

3. If N is a positive integer, then $\sum_{k=1}^{\infty} a_k$ and $\sum_{k=N}^{\infty} a_k$ either both converge or both diverge.

In general, changing a *finite* number of terms in a convergent series does not change its convergence, although it does change the value of the series.

Theorem: Divergence Test

If $\sum a_k$ converges, then $\lim_{k \rightarrow \infty} a_k = 0$. Hence, if $\lim_{k \rightarrow \infty} a_k \neq 0$, then the series diverges.

Theorem: Integral Test

Suppose f is a *continuous, positive, decreasing* function on $[1, \infty)$ and let $a_k = f(k)$. Then the series $\sum_{k=1}^{\infty} a_k$ is convergent if and only if the improper integral $\int_1^{\infty} f(x) dx$ is convergent. In other words:

1. If $\int_1^{\infty} f(x) dx$ is convergent, then $\sum_{k=1}^{\infty} a_k$ is convergent;
2. If $\int_1^{\infty} f(x) dx$ is divergent, then $\sum_{k=1}^{\infty} a_k$ is divergent.

Theorem: remainder estimate for the Integral Test

Suppose f is a *continuous, positive, decreasing* function on $[1, \infty)$. Let $a_k = f(k)$ and suppose $\sum_{k=1}^{\infty} a_k = s$ is *convergent*. Denote $R_n = s - s_n = \sum_{k=n+1}^{\infty} a_k$. Then

1. $\int_{n+1}^{\infty} f(x) dx \leq R_n \leq \int_n^{\infty} f(x) dx;$
2. $s_n + \int_{n+1}^{\infty} f(x) dx \leq s \leq s_n + \int_n^{\infty} f(x) dx.$

Theorem: convergence of the p -series

The p -series $\sum_{k=1}^{\infty} \frac{1}{k^p}$ converges for $p > 1$ and diverges for $p \leq 1$.

Theorem: Comparison Test

Suppose that $\sum a_k$ and $\sum b_k$ are series with *nonnegative* terms.

1. If $\sum b_k$ is *convergent* and $a_k \leq b_k$ for all k , then $\sum a_k$ is also *convergent*;
2. If $\sum b_k$ is *divergent* and $a_k \geq b_k$ for all k , then $\sum a_k$ is also *divergent*.

Theorem: Limit Comparison Test

Suppose that $\sum a_k$ and $\sum b_k$ are series with $a_k \geq 0$ and $b_k > 0$. Assume

$$\lim_{k \rightarrow \infty} \frac{a_k}{b_k} = c.$$

1. If $0 < c < \infty$ (that is, c is a finite positive number), then $\sum a_k$ and $\sum b_k$ either both converge or both diverge.
2. If $c = 0$ and $\sum b_k$ *converges*, then $\sum a_k$ converges.
3. If $c = \infty$ and $\sum b_k$ *diverges*, then $\sum a_k$ diverges.

Theorem: Ratio Test

Suppose that $\sum a_k$ is a series with *positive* terms.

1. If $\lim_{k \rightarrow \infty} \frac{a_{k+1}}{a_k} = L < 1$, then the series $\sum a_k$ converges.
2. If $\lim_{k \rightarrow \infty} \frac{a_{k+1}}{a_k} = L > 1$ or $\lim_{k \rightarrow \infty} \frac{a_{k+1}}{a_k} = \infty$, then the series $\sum a_k$ diverges.

- 3.** If $\lim_{k \rightarrow \infty} \frac{a_{k+1}}{a_k} = 1$, the test is inconclusive.

Theorem: Root Test

Suppose that $\sum a_k$ is a series with *nonnegative* terms.

- 1.** If $\lim_{k \rightarrow \infty} \sqrt[k]{a_k} = L < 1$, then the series $\sum a_k$ converges.
- 2.** If $\lim_{k \rightarrow \infty} \sqrt[k]{a_k} = L > 1$ or $\lim_{k \rightarrow \infty} \sqrt[k]{a_k} = \infty$, then the series $\sum a_k$ diverges.
- 3.** If $\lim_{k \rightarrow \infty} \sqrt[k]{a_k} = 1$, the test is inconclusive.

Theorem: Alternating Series Test

The alternating series

$$\sum_{k=1}^{\infty} (-1)^{k-1} b_k = b_1 - b_2 + b_3 - b_4 + b_5 - b_6 + \dots, \quad b_k > 0$$

converges if

- 1.** if $\{b_k\}$ is a *positive non-increasing* sequence, $b_k \geq b_{k+1} > 0$, and
- 2.** $\lim_{k \rightarrow \infty} b_k = 0$.

Theorem: remainder estimate for alternating series

If $s = \sum_{k=1}^{\infty} (-1)^k b_k$ is the sum of an alternating series that satisfies

1. $b_{k+1} \leq b_k$ for all k ;
2. $\lim_{k \rightarrow \infty} b_k = 0$

then

$$|R_n| = |s - s_n| \leq b_{n+1}.$$

Definition: absolute convergece and conditional convergence

- A series $\sum a_k$ is called **absolutely convergent** if the series of absolute values $\sum |a_k|$ is convergent.
- A series $\sum a_k$ is called **conditionally convergent** if it is convergent but *not* absolutely convergent.

Theorem: convergence of absolutely convergent series

If a series $\sum a_k$ is *absolutely convergent*, then it is convergent.

Series of Functions

7.1 Approximating Functions with Polynomials

Taylor Polynomials

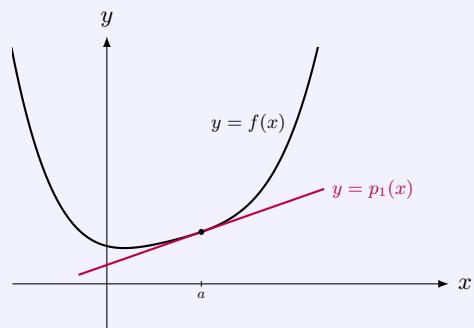
7.1 Example: linear approximation and quadratic approximation

1. We know that if a function f is differentiable at a point a , then it can be approximated near a by its tangent line $y = f(a) + f'(a)(x - a)$ (or “linearization”). This is a linear approximation to f at the point a . Since it is a first-degree polynomial, it is convenient to write

$$p_1(x) = f(a) + f'(a)(x - a).$$

Clearly, this polynomial p_1 matches f in *value* and in *slope* at a . That is,

$$p_1(a) = f(a) \quad \text{and} \quad p'_1(a) = f'(a).$$



2. If f has a fairly constant slope near a , the tangent line provides a good approximation. However, if the graph of f near a has a larger curvature, then the linear approximation may not be good enough. To improve accuracy, we consider a quadratic approximating polynomial p_2 by adding one new term to the linear polynomial:

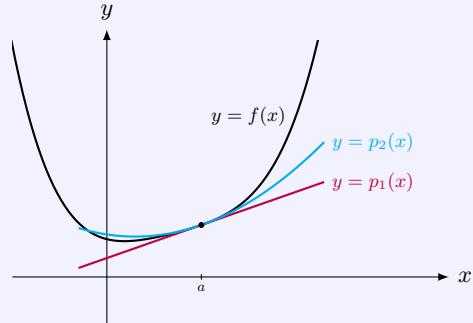
$$p_2(x) = f(a) + f'(a)(x - a) + c_2(x - a)^2.$$

The constant c_2 should be chosen so that p_2 is a good approximation to f near a . To do so, we require that p_2 agree with f in *value, slope, and concavity* at a :

$$p_2(a) = f(a), \quad p'_2(a) = f'(a), \quad \text{and} \quad p''_2(a) = f''(a),$$

where we assume that $f(a)$, $f'(a)$, and $f''(a)$ exist. In fact, we can easily verify that indeed p_2 satisfies the first and the second equations. The third equation holds if c_2 satisfies $2c_2 = f''(a)$, or $c_2 = \frac{1}{2}f''(a)$. Hence, the quadratic approximating polynomial is

$$p_2(x) = f(a) + f'(a)(x - a) + \frac{f''(a)}{2}(x - a)^2.$$



7.2 Example: linear approximation and quadratic approximation for \sqrt{x}

Find the linear approximation and the quadratic approximation for \sqrt{x} at $x = 1$. Then use them to approximate the value of $\sqrt{1.03}$.

Solution For the function $f(x) = \sqrt{x}$, we have

$$f'(x) = \frac{1}{2}x^{-\frac{1}{2}}, \quad f''(x) = -\frac{1}{4}x^{-\frac{3}{2}}.$$

Thus, the linear approximation and the quadratic approximation at $x = 1$ are, respectively,

$$\begin{aligned} p_1(x) &= f(1) + f'(1)(x - 1) \\ &= 1 + \frac{1}{2}(x - 1), \\ p_2(x) &= f(1) + f'(1)(x - 1) + \frac{f''(1)}{2}(x - 1)^2 \\ &= 1 + \frac{1}{2}(x - 1) - \frac{1}{8}(x - 1)^2. \end{aligned}$$

Thus, by using these formulas, we obtain the approximating values of $\sqrt{1.03}$:

$$\begin{aligned} p_1(1.03) &= 1 + \frac{1}{2}(1.03 - 1) = 1.015, \\ p_2(1.03) &= 1 + \frac{1}{2}(1.03 - 1) - \frac{1}{8}(x - 1)^2 = 1.0148875. \end{aligned}$$

For comparison, the value of $\sqrt{1.03}$ is $1.014889156509222\dots$. We see that the quadratic approximation indeed yields better accuracy.

7.3 Definition: Taylor polynomials

Let f be a function with f', f'', \dots , and $f^{(n)}$ defined at a . The **n th-order Taylor polynomial** for f with its **center** at a is

$$T_n(x) = f(a) + f'(a)(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x - a)^n.$$

7.4 Theorem: uniqueness of Taylor polynomials

Suppose f is a function with f', f'', \dots , and $f^{(n)}$ defined at a . If P_n is a polynomial of degree n satisfying

$$P_n(a) = f(a), \quad P'_n(a) = f'(a), \quad P''_n(a) = f''(a), \quad \dots, \quad P_n^{(n)}(a) = f^{(n)}(a),$$

Then

$$P_n(x) \equiv T_n(x) = f(a) + f'(a)(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x - a)^n.$$

Justification:

Write P_n as

$$P_n(x) = c_0 + c_1(x - a) + \cdots + c_n(x - a)^n.$$

The equation $P_n(a) = f(a)$ gives $c_0 = f(a)$. Differentiating the expression of P_n and evaluating at $x = a$ gives $P'_n(a) = c_1$, so that $c_1 = f'(a)$. In general, differentiating the expression of P_n for k times and evaluating at $x = a$ gives $P_n^{(k)}(a) = c_k \cdot k!$, so that $c_k = \frac{f^{(k)}(a)}{k!}$. Hence,

$$P_n(x) = f(a) + f'(a)(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x - a)^n = T_n(x).$$

7.5 Example: Taylor polynomials for e^x

Find the Taylor polynomials of order $n = 0, 1, 2, 3$, and 4 for $f(x) = e^x$ centered at 0 . Graph f and the polynomials. Use $T_4(x)$ to approximate the value of $e^{0.1}$. What is the n th-order Taylor polynomial centered at 0 for general n ?

Solution For any integer $k \geq 0$, we have

$$f^{(k)}(x) = e^x \implies f^{(k)}(0) = 1.$$

Thus, the first few Taylor polynomials for $f(x) = e^x$ centered at 0 are:

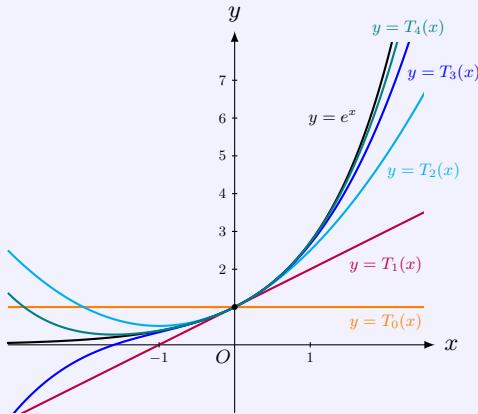
$$T_0(x) = f(0) = 1,$$

$$T_1(x) = f(0) + f'(0)x = 1 + x,$$

$$T_2(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 = 1 + x + \frac{x^2}{2!},$$

$$T_3(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!},$$

$$T_4(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \frac{f^{(4)}(0)}{4!}x^4 = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!}.$$



Using the Taylor polynomial T_4 , we get the approximation to $e^{0.1}$ (to 8th decimal places):

$$T_4(0.1) \approx 1.10517083.$$

The actual value of $e^{0.1}$ to 8th decimal places is 1.10517092.

In general, we have the n th-order Taylor polynomial centered at 0:

$$T_n(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} = \sum_{k=0}^n \frac{x^k}{k!}.$$

7.6 Theorem: asymptotic approximation of Taylor polynomials

If $f^{(n)}(a)$ exists, then $\lim_{x \rightarrow a} \frac{f(x) - T_n(x)}{(x - a)^n} = 0$, or equivalently, by using little- o notation,

$$f(x) = T_n(x) + o(|x - a|^n) = \sum_{k=0}^n \frac{f^{(k)}(a)}{k!} (x - a)^k + o(|x - a|^n), \quad \text{as } x \rightarrow a.$$

Justification:

By applying l'Hospital's Rule, we have

$$\begin{aligned}
 \lim_{x \rightarrow a} \frac{f(x) - T_n(x)}{(x - a)^n} &= \lim_{x \rightarrow a} \frac{f(x) - T_{n-1}(x) - \frac{f^{(n)}(a)}{n!}(x - a)^n}{(x - a)^n} \\
 &= \lim_{x \rightarrow a} \frac{f(x) - T_{n-1}(x)}{(x - a)^n} - \frac{f^{(n)}(a)}{n!} \\
 &\stackrel{\text{H}}{=} \lim_{x \rightarrow a} \frac{f'(x) - T'_{n-1}(x)}{n(x - a)^{n-1}} - \frac{f^{(n)}(a)}{n!} \stackrel{\text{H}}{=} \dots \\
 &\stackrel{\text{H}}{=} \lim_{x \rightarrow a} \frac{f^{(n-1)}(x) - T^{(n-1)}_{n-1}(x)}{n(n-1) \cdots 2(x - a)} - \frac{f^{(n)}(a)}{n!} \\
 &= \lim_{x \rightarrow a} \frac{f^{(n-1)}(x) - f^{(n-1)}(a)}{n(n-1) \cdots 2(x - a)} - \frac{f^{(n)}(a)}{n!} \\
 &= \frac{1}{n!} \cdot \lim_{x \rightarrow a} \frac{f^{(n-1)}(x) - f^{(n-1)}(a)}{x - a} - \frac{f^{(n)}(a)}{n!} \\
 &= \frac{1}{n!} \cdot f^{(n)}(a) - \frac{f^{(n)}(a)}{n!} = 0.
 \end{aligned}$$

7.7 Pitfall: Taylor approximation

From the conclusion that $f(x) = T_n(x) + o(|x - a|^n)$ (as $x \rightarrow a$), we know that $T_n(x)$ approximates $f(x)$ as $x \rightarrow a$ and the accuracy improves as n increases. However, this does not imply that the sequence $\{T_n(x)\}$ approaches $f(x)$ as $n \rightarrow \infty$. In fact, generally $f(x) \neq \lim_{n \rightarrow \infty} T_n(x)$.

As an example, consider the function

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

We can show that $f^{(n)}(0) = 0$ ($n \geq 0$) by induction. So, the n -th order Taylor polynomial for f at 0 is

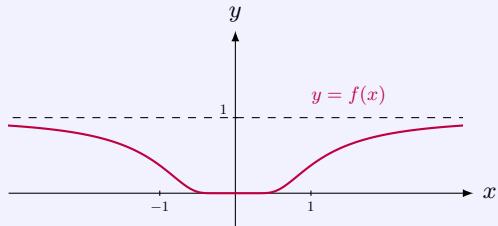
$$T_n(x) = 0 + 0 \cdot x + \cdots + 0 \cdot x^n = 0.$$

By Applying l'Hospital's Rule, we can see that as $x \rightarrow 0$,

$$f(x) = T_n(x) + o(|x|^n) = o(|x|^n), \quad \text{for all } n \geq 0.$$

However, for any $x \neq 0$,

$$f(x) \neq 0 = 0 + 0 \cdot x + \cdots + 0 \cdot x^n + \cdots = \lim_{n \rightarrow \infty} T_n(x).$$



7.8 MATLAB: Taylor polynomials

Find the Taylor polynomial of order $n = 4$ for $f(x) = e^x$ centered at 0.

Solution In MATLAB, we can use the command `taylor` to calculate the Taylor polynomial approximation for a function. The following MATLAB code shows how to find the Taylor polynomial of order 4 for $f(x) = e^x$ centered at 0.

```
>> syms x
>> f(x) = exp(x);
>> taylor(f(x),x,0,'Order',4 + 1) % centered at 0, order 4 ...
    plus one constant term

ans =
x^4/24 + x^3/6 + x^2/2 + x + 1
```

Since the first term of the Taylor polynomial is a constant, the Taylor polynomial of order 4 for $f(x) = e^x$ has 5 terms. In other words, the value of '`Order`' in the command `taylor` is the required order of the Taylor polynomial plus 1.

7.2 Represent Functions by Power Series

7.9 Definition: remainder of Taylor series

Let T_n be the Taylor polynomial of order n for f at a :

$$T_n(x) = f(a) + f'(a)(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x - a)^n.$$

The term $R_n(x) = f(x) - T_n(x)$ is called the **remainder** of the **Taylor series**. The latter refers to the following “formal” power series of f :

$$f(a) + f'(a)(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x - a)^n + \cdots.$$

This power series is also called **Taylor expansion** of f at a . A Taylor series centered at 0 is called a **Maclaurin series**.

7.10 Definition: power series

A **power series** is a series of the form

$$\sum_{k=0}^{\infty} c_k x^k = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + \cdots,$$

where x is a variable and the c_k ’s are constants called the **coefficients** of the series.

More generally, a series of the form

$$\sum_{k=0}^{\infty} c_k(x-a)^k = c_0 + c_1(x-a) + c_2(x-a)^2 + \cdots,$$

is called a **power series in** $(x-a)$ or a **power series centered at** a or a **power series about** a .

7.11 Pitfall: power series

For any given function f , if $f^{(n)}(a)$ ($n \geq 0$) all exist, then the Taylor series for f at a is defined as

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x-a)^k.$$

The series is defined *formally*, since the series may not converge for some x at which f is defined.

As an example, consider the function $f(x) = \frac{1}{1+x}$. It has the following power series (centered at 0):

$$1 - x + x^2 - x^3 + x^4 - x^5 + \cdots.$$

In fact, we can have by induction that

$$f^{(n)}(x) = \frac{(-1)^n n!}{(1+x)^{n+1}},$$

so that $f^{(n)}(0) = (-1)^n n!$. Thus, the Taylor series for f at 0 is

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k = \sum_{k=0}^{\infty} \frac{(-1)^k k!}{k!} x^k = \sum_{k=0}^{\infty} (-1)^k x^k.$$

It is clear that the function f is defined everywhere except at $x = -1$, but the series only converges for $|x| < 1$.

7.12 Theorem: convergence theorem of Taylor series

If $\lim_{n \rightarrow \infty} R_n(x) = 0$ in an open interval I that contains a , then

$$f(x) = \lim_{n \rightarrow \infty} T_n(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x-a)^k, \quad x \in I.$$

Justification:

If $\lim_{n \rightarrow \infty} R_n(x) = 0$, then

$$\lim_{n \rightarrow \infty} T_n(x) = \lim_{n \rightarrow \infty} [f(x) - R_n(x)] = f(x) - \lim_{n \rightarrow \infty} R_n(x) = f(x).$$

It is clear that $\{T_n(x)\}$ are the partial sums of the series

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!} (x-a)^k = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(k)}(a)}{k!}(x-a)^k + \cdots.$$

Thus, we have

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!} (x-a)^k.$$

7.13 Engineering Example: Z-transform

The Z-transform converts a sequence of discrete-time signals into a complex frequency-domain representation for easy interpretations. The Z-transform can be defined as either a two-sided or one-sided transform.

Specifically, given a discrete-time signal $x[n]$, the *two-sided* Z-transform is defined as

$$X(z) = \sum_{n=-\infty}^{\infty} x[n]z^{-n}.$$

where z is, in general, a complex number. Alternatively, if the signal $x[n]$ is defined only for $n \geq 0$, the *one-sided* Z-transform is defined as

$$X(z) = \sum_{n=0}^{\infty} x[n]z^{-n}.$$

The Z-transform has been applied widely in science and engineering fields, e.g., probability theory, digital signal processing (DSP) and control systems.

7.14 Engineering Example: continuous-time Fourier series

The Fourier series can represent periodic signals as sums of complex exponentials (or, sine functions and cosine functions). Given a continuous-time periodic signal $x(t)$, its continuous-time Fourier series is defined as

$$x(t) = \sum_{k=-\infty}^{\infty} a_k e^{j\frac{2\pi}{T}kt}$$

where a_k is the Fourier series coefficient determined by $x(t)$, and T is the period of $x(t)$. From the above definition, we see that the Fourier series can represent signals by their harmonics $e^{j\frac{2\pi}{T}kt}$.

The Fourier series has many applications in electrical engineering, signal processing, image processing, acoustics, optics, etc. For example, since the Fourier series can represent signals by their harmonics, it is thus very useful for the signals which can be naturally described by harmonics, e.g., music signals.

Moreover, for a linear time-invariant (LTI) system with frequency response $H(j\omega)$, if the

input is given by $x(t) = \sum_{k=-\infty}^{\infty} a_k e^{j\frac{2\pi}{T}kt}$, then the output is

$$y(t) = \sum_{k=-\infty}^{\infty} a_k H\left(j\frac{2\pi}{T}k\right) e^{j\frac{2\pi}{T}kt}.$$

Since the amplitude of each component is scaled by $H(j\frac{2\pi}{T}k)$, as seen above, the output of the LTI system is thus a “filtered” version of the input. In other words, the Fourier series motivates a new representation of a system as a filter. This is useful for many systems, such as the speech synthesis system.

Remainder Formulas

7.15 Theorem: integral form of the remainder

If $f^{(n+1)}$ is continuous in an open interval I that contains a , then

$$R_n(x) = \frac{1}{n!} \int_a^x (x-t)^n f^{(n+1)}(t) dt, \quad x \in I.$$

Justification:

We prove the remainder formula by induction.

For $n = 0$, the formula gives

$$\frac{1}{0!} \int_a^x (x-t)^0 f'(t) dt = f(x) - f(a) = f(x) - T_0(x) = R_0(x).$$

So, the formula is true for $n = 0$.

Suppose the formula holds for $n = k$:

$$R_k(x) = \frac{1}{k!} \int_a^x (x-t)^k f^{(k+1)}(t) dt, \quad x \in I.$$

Then, by integration by parts, we get

$$\begin{aligned} & \frac{1}{(k+1)!} \int_a^x (x-t)^{k+1} f^{(k+2)}(t) dt \\ &= \left[\frac{1}{(k+1)!} (x-t)^{k+1} f^{(k+1)}(t) \right]_{t=a}^x + \frac{1}{k!} \int_a^x (x-t)^k f^{(k+1)}(t) dt \\ &= R_k(x) - \frac{f^{(k+1)}(a)}{(k+1)} (x-a)^{k+1} \\ &= f(x) - T_k(x) - \frac{f^{(k+1)}(a)}{(k+1)} (x-a)^{k+1} \\ &= f(x) - T_{k+1}(x) = R_{k+1}(x). \end{aligned}$$

So, the formula is also true for $n = k + 1$.

Hence, the remainder formula holds for all $n \geq 0$.

7.16 Theorem: Lagrange remainder

If $f^{(n+1)}$ is continuous in an open interval I that contains a , then there is c between a and x such that

$$R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}.$$

This formula is called the **Lagrange remainder** of Taylor series.

Justification:

Let $x \in I$ be a fixed number.

If $x > a$, on the interval $[a, x]$, the function $f^{(n+1)}$ has its minimum and maximum at s_1 and s_2 :

$$m = f^{(n+1)}(s_1), \quad M = f^{(n+1)}(s_2),$$

respectively. Since $m \leq f^{(n+1)}(t) \leq M$ on $[a, x]$, we have

$$\frac{m}{n!} \int_a^x (x-t)^n dt \leq \frac{1}{n!} \int_a^x (x-t)^n f^{(n+1)}(t) dt \leq \frac{M}{n!} \int_a^x (x-t)^n dt,$$

so that

$$m(x-a)^{n+1} \leq (n+1) \int_a^x (x-t)^n f^{(n+1)}(t) dt \leq M(x-a)^{n+1},$$

or equivalently,

$$m \leq \frac{(n+1)}{(x-a)^{n+1}} \int_a^x (x-t)^n f^{(n+1)}(t) dt \leq M.$$

By the Intermediate Value Theorem, there is $c \in [a, x]$, such that

$$f^{(n+1)}(c) = \frac{(n+1)}{(x-a)^{n+1}} \int_a^x (x-t)^n f^{(n+1)}(t) dt.$$

Hence, from the integral form of the remainder,

$$R_n(x) = \frac{1}{n!} \int_a^x (x-t)^n f^{(n+1)}(t) dt = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}.$$

If $x < a$, the argument is similar.

7.17 Theorem: remainder estimate

If $|f^{(n+1)}(x)| \leq M$ in an open interval I that contains a , then

$$|R_n(x)| \leq \frac{M}{(n+1)!}|x-a|^{n+1}, \quad x \in I.$$

Hence, if $|f^{(n+1)}(x)| \leq M$ for every n in I , then

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!} (x-a)^k, \quad x \in I.$$

Justification:

If $|f^{(n+1)}(x)| \leq M$ in I , by the Lagrange remainder,

$$|R_n(x)| = \left| \frac{f^{(n+1)}(c)}{(n+1)!} (x-a)^{n+1} \right| \leq \frac{M}{(n+1)!} |x-a|^{n+1}, \quad x \in I.$$

We know that $|x-a|^{n+1} \ll (n+1)!$ as $n \rightarrow \infty$. Thus, if $|f^{(n+1)}(x)| \leq M$ for every n in I , then $\lim_{n \rightarrow \infty} R_n(x) = 0$. Hence, by the convergence theorem of the Taylor series,

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!} (x-a)^k, \quad x \in I,$$

7.18 Example: convergent Taylor series

1. Show that for all $x \in \mathbb{R}$,

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots$$

Solution Let $f(x) = e^x$. For all n , since $f^{(n)}(x) = e^x$, we have $f^{(n)}(0) = e^0 = 1$. Thus,

$$T_n(x) = \sum_{k=0}^n \frac{f^{(k)}(0)}{k!} x^k = \sum_{k=0}^n \frac{x^k}{k!}$$

is well-defined for every $n \geq 0$. Its Lagrange remainder is

$$R_n(x) = f(x) - T_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!} x^{n+1},$$

where c is between 0 and x . For any fixed x ,

$$|f^{(n+1)}(c)| = e^c \leq e^{|x|},$$

so that

$$-\frac{e^{|x|}}{(n+1)!} |x|^{n+1} \leq R_n(x) \leq \frac{e^{|x|}}{(n+1)!} |x|^{n+1}.$$

Since $|x|^{n+1} \ll (n+1)!$ as $n \rightarrow \infty$, by the Squeeze Theorem, $\lim_{n \rightarrow \infty} R_n(x) = 0$. Hence, by the convergence theorem of Taylor series, we have

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}, \quad x \in \mathbb{R}.$$

2. Show that for all $x \in \mathbb{R}$,

$$\sin x = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} x^{2k+1} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots,$$

$$\cos x = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} x^{2k} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots.$$

Solution Let $f(x) = \sin x$. Directly computation gives, for $k = 0, 1, 2, \dots$,

$$f^{(4k)}(0) = 0, \quad f^{(4k+1)}(0) = 1, \quad f^{(4k+2)}(0) = 0, \quad f^{(4k+3)}(0) = -1.$$

Thus, for $n \geq 1$,

$$T_{2n+1}(x) = T_{2n+2}(x) = \sum_{k=0}^n \frac{(-1)^k}{(2k+1)!} x^{2k+1}.$$

The Lagrange remainder is

$$R_n(x) = f(x) - T_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!} x^{n+1},$$

where c is between 0 and x . For any fixed x , it is easy to see that $f^{(n)}(x)$ is either $\pm \sin x$ or $\pm \cos x$. Thus, we have

$$|f^{(n+1)}(c)| \leq 1,$$

so that

$$-\frac{1}{(n+1)!} |x|^{n+1} \leq R_n(x) \leq \frac{1}{(n+1)!} |x|^{n+1}.$$

Since $|x|^{n+1} \ll (n+1)!$ as $n \rightarrow \infty$, by the Squeeze Theorem, $\lim_{n \rightarrow \infty} R_n(x) = 0$. Hence, by the convergence theorem of Taylor series, we have

$$\sin x = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} x^{2k+1} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots, \quad x \in \mathbb{R}.$$

In a similar manner, we can have

$$\cos x = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} x^{2k} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots, \quad x \in \mathbb{R}.$$

7.3 Properties of Power Series

Radius of Convergence and Interval of Convergence

7.19 Example: power series

1. The polynomial $P_n(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$ is a power series centered at 0. It has finite number of terms and is meaningful (convergent) for any $x \in \mathbb{R}$.

2. The power series $\sum_{k=0}^{\infty} k!x^k$ converges only when $x = 0$ (it is a convention that $0! = 1$).

In fact, for any $x \neq 0$, we know that, as $k \rightarrow \infty$, $\left(\frac{1}{|x|}\right)^k \ll k!$, so that $1 \ll k!|x|^k$.

Thus, $\lim_{k \rightarrow \infty} k!x^k \neq 0$ for any $x \neq 0$. Hence, if $x \neq 0$, the series $\sum_{k=0}^{\infty} k!x^k$ always diverges by the Divergence Test.

3. The power series $\sum_{k=0}^{\infty} \frac{x^k}{\sqrt{k!}}$ converges for any $x \in \mathbb{R}$.

In fact, for any $x \in \mathbb{R}$,

$$\lim_{k \rightarrow \infty} \frac{\left| \frac{x^{k+1}}{\sqrt{(k+1)!}} \right|}{\left| \frac{x^k}{\sqrt{k!}} \right|} = \lim_{k \rightarrow \infty} \frac{|x|}{\sqrt{k+1}} = 0.$$

By the Ratio Test, the series $\sum_{k=0}^{\infty} \frac{x^k}{\sqrt{k!}}$ is absolutely convergent, so it is convergent.

4. The power series $\sum_{k=0}^{\infty} x^k$ converges for $x \in (-1, 1)$ and diverges for $x \in (-\infty, -1] \cup [1, \infty)$.

In fact, for $x \in (-1, 1)$, the power series is a convergent geometric series. For $x \in (-\infty, -1] \cup [1, \infty)$, the limit of k th term x^k is not zero: $\lim_{k \rightarrow \infty} x^k \neq 0$. By the Divergence Test, the power series is divergent.

5. The power series $\sum_{k=0}^{\infty} \frac{(-1)^k x^k}{k+1}$ converges for $x \in (-1, 1]$ and diverges for $x \in (-\infty, -1] \cup (1, \infty)$.

In fact, for any fixed $x \in \mathbb{R}$,

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

By the Ratio Test, the series is absolutely convergent for $|x| < 1$ and divergent for $|x| > 1$. Hence, the series is convergent for $x \in (-1, 1)$ and divergent for $x \in (-\infty, -1) \cup (1, \infty)$.

For $x = -1$, the power series is the divergent harmonic series $\sum_{k=1}^{\infty} \frac{1}{k}$.

For $x = 1$, the power series is the convergent alternating harmonic series $\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k}$.

6. The power series $\sum_{k=0}^{\infty} \left(\frac{k}{k+1}\right)^k (x+1)^k$ converges for $x \in (-2, 0)$ and diverges for $x \in (-\infty, -2] \cup [0, \infty)$.

In fact, for any fixed $x \in \mathbb{R}$,

$$\lim_{k \rightarrow \infty} \sqrt[k]{\left(\frac{k}{k+1}\right)^k (x+1)^k} = \lim_{k \rightarrow \infty} \frac{k \cdot |x+1|}{k+1} = |x+1|.$$

By the Root Test, the series is absolutely convergent for $|x+1| < 1$ and divergent for $|x+1| > 1$. Hence, the series is convergent for $x \in (-2, 0)$ and divergent for $x \in (-\infty, -2) \cup (0, \infty)$.

For $x = -2$ or $x = 0$, since

$$\lim_{k \rightarrow \infty} \left| \left(\frac{k}{k+1}\right)^k (x+1)^k \right| = \lim_{k \rightarrow \infty} \frac{1}{\left(1 + \frac{1}{k}\right)^k} = \frac{1}{e} \neq 0,$$

by the Divergence Test, the series diverges.

7.20 Theorem: convergence of power series

For a given power series $\sum_{k=0}^{\infty} c_k(x-a)^k$ there are only three possibilities:

1. The series converges only when $x = a$.
2. The series converges for all x .
3. There is a positive number R such that the series converges if $|x-a| < R$ and diverges if $|x-a| > R$.

Definition: radius of convergence

The number R is called the **radius of convergence** of the power series.

By convention, the radius of convergence is $R = 0$ when the power series converges only at the center a and $R = \infty$ when the power series converges for all x . The **interval of convergence** of a power series is the interval that consists of all values of x for which the series converges.

Remark: interval of convergence

When $0 < R < \infty$, the theorem says nothing about the convergence of the series at the endpoint points $|x - a| = R$ (*i.e.*, $x = a \pm R$) of the interval. Generally, the interval of convergence *cannot* be obtained from the radius of convergence.

Justification:

The theorem is a consequence of the following claim: if $\sum a_k r^k$ converges, then $\sum a_k x^k$ absolutely converges for $|x| < |r|$.

In fact, for convergent series $\sum a_k r^k$, we have $\lim_{k \rightarrow \infty} a_k r^k = 0$. Thus, there is K such that when $k > K$,

$$|a_k x^k| < 1.$$

For $|x| < |r|$, this implies that when $k > K$,

$$|a_k x^k| = |a_k r^k| \cdot \left| \frac{x}{r} \right|^k \leq \left| \frac{x}{r} \right|^k.$$

Since $\left| \frac{x}{r} \right| < 1$, the geometric series $\sum \left| \frac{x}{r} \right|^k$ converges. Then, by the comparison test, the series $\sum_{k=K}^{\infty} a_k x^k$ absolutely converges. Hence, $\sum_{k=1}^{\infty} a_k x^k$ absolutely converges.

7.21 Method: determine the interval of convergence

Let $\sum_{k=0}^{\infty} c_k (x - a)^k$ be a given power series.

1. Compute the radius of convergence R by using the Ratio Test (or sometimes the Root Test) by

$$R = \frac{1}{\lim_{k \rightarrow \infty} \left| \frac{c_{k+1}}{c_k} \right|} = \lim_{k \rightarrow \infty} \left| \frac{c_k}{c_{k+1}} \right| \quad \text{or} \quad R = \frac{1}{\lim_{k \rightarrow \infty} \sqrt[k]{|c_k|}}.$$

2. The Ratio and Root Tests are inconclusive 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.

7.22 Example: radius of convergence and interval of convergence

- Determine the radius of convergence and the interval of convergence of the power series

$$\sum_{k=0}^{\infty} \frac{x^{2k}}{(k+1)^k}.$$

Solution Since

$$\lim_{k \rightarrow \infty} \sqrt[k]{\left| \frac{x^{2k}}{(k+1)^k} \right|} = |x|^2 \lim_{k \rightarrow \infty} \frac{1}{k+1} = 0,$$

by the Root Test, for all x , the power series absolutely converges, so it converges. Thus, the radius of convergence is $R = \infty$ and the interval of convergence is $(-\infty, \infty)$.

- Determine the radius of convergence and the interval of convergence of the power series

$$\sum_{k=0}^{\infty} \frac{k(x+1)^k}{k+1}.$$

Solution

- (a) Compute the radius of convergence by the Ratio Test.

Since

$$\lim_{k \rightarrow \infty} \frac{\left| \frac{(k+1)(x+1)^{k+1}}{k+2} \right|}{\left| \frac{k(x+1)^k}{k+1} \right|} = |x+1| \cdot \lim_{k \rightarrow \infty} \frac{(k+1)^2}{k(k+2)} = |x+1|,$$

by the Ratio Test, when $|x+1| < 1$, the power series converges; when $|x+1| > 1$, it diverges. Thus, the radius of convergence is $R = 1$.

So, the series converges if $-2 < x < 0$; diverges if $x < -2$ or $x > 0$.

- (b) Determine whether the power series is convergent at the endpoints.

At the left endpoint $x = -2$, the series is $\sum_{k=0}^{\infty} \frac{(-1)^k k}{k+1}$. Since

$$\lim_{k \rightarrow \infty} \left| \frac{(-1)^k k}{k+1} \right| = \lim_{k \rightarrow \infty} \frac{k}{k+1} = 1 \neq 0,$$

so that $\lim_{k \rightarrow \infty} \frac{(-1)^k k}{k+1} \neq 0$. From the Divergence Test, the series is divergent.

At the right endpoint $x = 0$, the series is $\sum_{k=0}^{\infty} \frac{k}{k+1}$. Again from the Divergence Test, it is divergent from the Divergence Test.

Hence, for the power series $\sum_{k=0}^{\infty} \frac{k(x+1)^k}{k+1}$, the radius of convergence $R = 1$ and the interval of convergence is $(-2, 0)$.

3. Determine the radius of convergence and the interval of convergence of the power series $\sum_{k=0}^{\infty} \frac{(x-2)^k}{\sqrt{k+1}}$.

Solution

- (a) Compute the radius of convergence by the Ratio Test.

Since

$$\lim_{k \rightarrow \infty} \frac{\left| \frac{(x-2)^{k+1}}{\sqrt{k+1}+1} \right|}{\left| \frac{(x-2)^k}{\sqrt{k+1}} \right|} = |x-2| \cdot \lim_{k \rightarrow \infty} \frac{\sqrt{k+1}}{\sqrt{k+1}+1} = |x-2|,$$

by the Ratio Test, when $|x-2| < 1$, the power series converges; when $|x-2| > 1$, it diverges. Thus, the radius of convergence is $R = 1$.

So, the series converges if $1 < x < 3$; diverges if $x < 1$ or $x > 3$.

- (b) Determine whether the power series is convergent at the endpoints.

At the left endpoint $x = 1$, the series is $\sum_{k=0}^{\infty} \frac{(-1)^k}{\sqrt{k+1}}$, which is an alternating series.

Let $b_k = \frac{1}{\sqrt{k+1}}$. Then, $\{b_k\}$ is a positive decreasing sequence and $\lim_{k \rightarrow \infty} b_k = 0$. By the Alternating Series Test, this series is convergent.

At the right endpoint $x = 3$, the series is $\sum_{k=0}^{\infty} \frac{1}{\sqrt{k+1}}$. Since

$$\lim_{k \rightarrow \infty} \frac{1}{\sqrt{k+1}} = 1,$$

and the p -series $\sum_{k=0}^{\infty} \frac{1}{\sqrt{k}}$ is divergent ($p = \frac{1}{2}$), by the Limit Comparison Test, the series $\sum_{k=0}^{\infty} \frac{1}{\sqrt{k+1}}$ is divergent.

Hence, for the power series $\sum_{k=0}^{\infty} \frac{(x-2)^k}{\sqrt{k+1}}$, the radius of convergence $R = 1$ and the interval of convergence is $[1, 3)$.

7.23 MATLAB: radius of convergence and interval of convergence

Determine the radius of convergence and the interval of convergence of the power series $\sum_{k=0}^{\infty} \frac{k(x+1)^k}{k+1}$.

Solution

First, we compute the radius of convergence by the Ratio Test.

```
>> syms k x
>> assume(k,'positive') % k>0
>> assumeAlso(k,'integer') % k is also an integer
>> a(k) = k*(x + 1)^k/(k + 1);
>> ratio_a = simplify(abs(a(k + 1)/a(k)));
>> L = limit(ratio_a,k,inf)

L =

abs(x + 1)
```

According to the Ratio Test, when $|x + 1| < 1$, the power series is convergent; when $|x + 1| > 1$, it is divergent. Thus, the radius of convergence is $R = 1$.

Moreover, from the result above, we can know that the series converges if $-2 < x < 0$, and diverges if $x < -2$ or $x > 0$. To determine the interval of convergence, we need to determine whether the power series is convergent at the endpoints $x = -2$ and $x = 0$, which is shown in the following:

```
>> a_left = subs(a(k),x,-2) % when x = -2
a_left =
((-1)^k*k)/(k + 1)
>> lim_a_left = limit(abs(a_left),k,inf)
lim_a_left =
1
>> a_right = subs(a(k),x,0) % when x = 0
a_right =
k/(k + 1)
>> lim_a_right = limit(a_right,k,inf)
lim_a_right =
1
```

We see from the results above that $\lim_{k \rightarrow \infty} \frac{k(-2+1)^k}{k+1} = 1 \neq 0$ and $\lim_{k \rightarrow \infty} \frac{k(0+1)^k}{k+1} = 1 \neq 0$. So, according to the Divergence Test, the power series is divergent at both endpoints $x = -2$ and $x = 0$. Therefore, the interval of convergence of this power series is $(-2, 0)$.

Operations on Power Series

7.24 Theorem: combining power series

Suppose $\sum c_k x^k$ and $\sum d_k x^k$ converge to $f(x)$ and $g(x)$, respectively, in an open interval I containing 0. Then

- Sum and Difference:** $\sum (c_k \pm d_k) x^k$ converges to $f(x) \pm g(x)$ in I , that is,

$$\sum_{k=0}^{\infty} (c_k \pm d_k) x^k = \sum_{k=0}^{\infty} c_k x^k \pm \sum_{k=0}^{\infty} d_k x^k, \quad x \in I.$$

- Product:** $\sum (c_0 d_k + c_1 d_{k-1} + \cdots + c_{k-1} d_1 + c_k d_0) x^k$ converges to $f(x) \cdot g(x)$ in I , that is,

$$\sum_{k=0}^{\infty} \left(\sum_{j=0}^k c_j d_{k-j} \right) x^k = \left(\sum_{k=0}^{\infty} c_k x^k \right) \left(\sum_{k=0}^{\infty} d_k x^k \right), \quad x \in I.$$

- Quotient:** If $d_0 \neq 0$, then in a neighborhood of $x = 0$, the power series $\sum a_k x^k$ converges to $\frac{f(x)}{g(x)}$, that is,

$$\sum_{k=0}^{\infty} a_k x^k = \frac{\sum_{k=0}^{\infty} c_k x^k}{\sum_{k=0}^{\infty} d_k x^k},$$

where

$$a_0 = \frac{c_0}{d_0}, \quad a_k = \frac{1}{d_0} \left[c_k - \sum_{j=1}^k d_j a_{k-j} \right] \quad \text{for } k \geq 1.$$

- Composition:** If $h(x) = bx^m$, where m is a positive integer and $b \neq 0$ a real number, the power series $\sum c_k (h(x))^k$ converges to $f(h(x))$ for x such that $h(x) \in I$.

7.25 Example: product and quotient of power series

- The product can be found by a vertical set-up. For instance, for the power series representations

$$e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \cdots, \quad x \in \mathbb{R},$$

$$\cos x = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \cdots, \quad x \in \mathbb{R},$$

to compute the Maclaurin series for $e^x \cos x$, we collect like terms just as for polynomials:

$$\begin{array}{r}
 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \dots \\
 \times \quad 1 \qquad - \frac{1}{2}x^2 \qquad + \dots \\
 \hline
 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \dots \\
 \qquad - \frac{1}{2}x^2 - \frac{1}{2}x^3 - \dots \\
 \\[1em]
 + \qquad \qquad \qquad \ddots
 \end{array}$$

Thus,

$$e^x \cos x = 1 + x - \frac{1}{3}x^3 - \dots, \quad x \in \mathbb{R}.$$

2. The quotient can be found by a vertical set-up. For instance, for the power series representations

$$\cos x = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \dots, \quad x \in \mathbb{R},$$

to compute the Maclaurin series for $\tan x = \frac{\sin x}{\cos x}$, we use a procedure like long division:

$1 - \frac{1}{2}x^2 + \frac{1}{24}x^4 - \dots$	$x + \frac{1}{3}x^3 + \frac{2}{15}x^5 + \dots$
	$x - \frac{1}{6}x^3 + \frac{1}{120}x^5 - \dots$
	$x - \frac{1}{2}x^3 + \frac{1}{24}x^5 - \dots$
	$\frac{1}{3}x^3 - \frac{1}{30}x^5 + \dots$
	$\frac{1}{3}x^3 - \frac{1}{6}x^5 + \dots$
	$\frac{2}{15}x^5 + \dots$
	$\frac{2}{15}x^5 + \dots$
	\dots

Thus,

$$\tan x = x + \frac{1}{3}x^3 + \frac{2}{15}x^5 + \cdots, \quad x \in (-\frac{\pi}{2}, \frac{\pi}{2}).$$

The Maclaurin series for the tangent function $\tan x$ has a smaller interval of convergence due to the fact that the denominator $\cos x$ becomes zero at $x = \pm\frac{\pi}{2}$.

7.26 MATLAB: product and quotient of power series

1. Given two power series $f(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!}$ ($x \in (-\infty, \infty)$) and $g(x) =$

$\sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!}$ ($x \in (-\infty, \infty)$), find the power series for the product of these two power series, i.e., $f(x)g(x)$.

Solution

```

>> syms x k
>> assume(k, 'integer') % k is an integer
>> a(k) = x^k/factorial(k);
>> b(k) = (-1)^k*x^(2*k)/factorial(2*k);
>> f(x) = symsum(a(k), k, 0, inf)

f(x) =

exp(x)

>> g(x) = symsum(b(k), k, 0, inf)

g(x) =

cos(x)

>> prod_term = f(x)*g(x)

prod_term =

exp(x)*cos(x)

>> prod_series = ...
    taylor(prod_term, x, 'ExpansionPoint', 0, 'Order', 6)

prod_series =
- x^5/30 - x^4/6 - x^3/3 + x + 1

```

It is shown that, up to the order of 5, the product of these two series can be written as

$$f(x)g(x) = 1 + x - \frac{1}{3}x^3 - \frac{1}{6}x^4 - \frac{1}{30}x^5 + \dots, \quad x \in \mathbb{R}.$$

2. Given two power series $f(x) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!}$ ($x \in (-\infty, \infty)$) and $g(x) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!}$ ($x \in (-\infty, \infty)$), find the power series for the quotient of these two power series, i.e., $\frac{f(x)}{g(x)}$.

Solution

```

>> syms x k
>> assume(k, 'integer') % k is an integer
>> a(k) = (-1)^k*x^(2*k + 1)/factorial(2*k + 1);
>> b(k) = (-1)^k*x^(2*k)/factorial(2*k);
>> f(x) = symsum(a(k), k, 0, inf)

f(x) =

sin(x)

>> g(x) = symsum(b(k), k, 0, inf)

g(x) =

cos(x)

>> quot_term = f(x)/g(x)

quot_term =

sin(x)/cos(x)

>> quot_series = ...
    taylor(quot_term, x, 'ExpansionPoint', 0, 'Order', 6)

quot_series =

(2*x^5)/15 + x^3/3 + x

```

It is shown that, up to the order of 5, the quotient of these two series can be written as

$$\frac{f(x)}{g(x)} = x + \frac{1}{3}x^3 + \frac{2}{15}x^5 + \dots$$

Note that the denominator $g(x) = \cos x$ becomes 0 when $x = \pm\frac{\pi}{2}$. Hence, the power series for the quotient is convergent when $x \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$.

7.27 Theorem: differentiation and integration of power series

If the power series $\sum_{k=0}^{\infty} c_k(x - a)^k$ has radius of convergence $R > 0$, then the function f defined by

$$f(x) = c_0 + c_1(x - a) + c_2(x - a)^2 + \dots = \sum_{k=0}^{\infty} c_k(x - a)^k$$

is differentiable (and therefore continuous) on the interval $(a - R, a + R)$ and

$$1. f'(x) = c_1 + 2c_2(x-a) + 3c_3(x-a)^2 + \cdots = \sum_{k=1}^{\infty} kc_k(x-a)^{k-1}.$$

$$2. \int f(x) dx = C + c_0(x-a) + c_1 \frac{(x-a)^2}{2} + c_2 \frac{(x-a)^3}{3} + \cdots = C + \sum_{k=0}^{\infty} c_k \frac{(x-a)^{k+1}}{k+1}.$$

The radii of convergence of these power series are both still R .

7.28 Example: differentiation of power series

Find $1 + 3x^2 + 5x^4 + 7x^6 + \cdots$.

Solution Finite geometric series gives

$$\sum_{k=0}^n x^k = 1 + x + x^2 + x^3 + \cdots + x^n = \frac{1 - x^{n+1}}{1 - x},$$

so that

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} (-1)^k x^k = 1 + x + x^2 + x^3 + \cdots, \quad x \in (-1, 1).$$

Substituting x by $-x$ (or making a composition) gives

$$\frac{1}{1+x} = \sum_{k=0}^{\infty} (-1)^k x^k = 1 - x + x^2 - x^3 + \cdots, \quad x \in (-1, 1).$$

Thus, differentiating these two equalities, respectively, gives

$$\left(\frac{1}{1-x} \right)' = 1 + 2x + 3x^2 + 4x^3 + \cdots, \quad x \in (-1, 1);$$

$$\left(\frac{1}{1+x} \right)' = -1 + 2x - 3x^2 + 4x^3 - \cdots, \quad x \in (-1, 1).$$

The difference of the last two equalities is

$$\begin{aligned} 2(1 + 3x^2 + 5x^4 + 7x^6 + \cdots) &= \left(\frac{1}{1-x} \right)' - \left(\frac{1}{1+x} \right)' \\ &= \frac{1}{(1-x)^2} + \frac{1}{(1+x)^2} = \frac{2(1+x^2)}{(1-x^2)^2}, \end{aligned}$$

so that

$$1 + 3x^2 + 5x^4 + 7x^6 + \cdots = \frac{1+x^2}{(1-x^2)^2}, \quad x \in (-1, 1).$$

7.29 Example: integration of power series

1. Finite geometric series gives

$$\sum_{k=0}^n (-1)^k x^k = 1 - x + x^2 - x^3 + \cdots + (-1)^n x^n = \frac{1 - (-x)^{n+1}}{1 + x},$$

so that

$$\frac{1}{1+x} = \sum_{k=0}^{\infty} (-1)^k x^k = 1 - x + x^2 - x^3 + \dots, \quad \text{for } x \in (-1, 1).$$

Integrating the equality gives

$$\ln(1+x) = C + x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots, \quad \text{for } x \in (-1, 1).$$

By putting $x = 0$ on both of the last equality, we get $C = 0$. Hence, for $x \in (-1, 1)$, we have

$$\ln(1+x) = \sum_{k=1}^{\infty} (-1)^{k-1} \frac{x^k}{k} = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots.$$

At the endpoints $x = -1$, the Maclaurin series is the divergent harmonic series. At the other endpoint $x = 1$, the Maclaurin series is the convergent alternating harmonic series. We can get the Maclaurin series for the function $\ln(1+x)$:

$$\ln(1+x) = \sum_{k=1}^{\infty} (-1)^{k-1} \frac{x^k}{k} = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots, \quad \text{for } x \in (-1, 1].$$

The following result can be applied to justify the equality at $x = 1$:

If $\sum c_k$ converges and $f(x) = \sum_{k=0}^{\infty} c_k x^k$ ($-1 < x < 1$), then $\lim_{x \rightarrow 1^-} f(x) = \sum_{k=0}^{\infty} c_k$.

2. Notice that $\int \frac{dx}{1+x^2} = \tan^{-1} x + C$. So, to find the Maclaurin series for the function $\tan^{-1} x$, we may start with the Maclaurin series for the function $\frac{1}{1+x}$:

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

Substituting x by x^2 gives

$$\frac{1}{1+x^2} = 1 - x^2 + x^4 - x^6 + \dots, \quad x \in (-1, 1).$$

Integrating the equality gives

$$\tan^{-1} x = C + x - \frac{1}{3}x^3 + \frac{1}{5}x^5 - \frac{1}{7}x^7 + \dots, \quad x \in (-1, 1).$$

By putting $x = 0$ on both of the last equality, we get $C = 0$. Hence, for $x \in (-1, 1)$, we have

$$\tan^{-1} x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{2k+1} = x - \frac{1}{3}x^3 + \frac{1}{5}x^5 - \frac{1}{7}x^7 + \dots.$$

At each of the endpoints $x = \pm 1$, the Maclaurin series is a convergent alternating series, by Alternating Series Test. We can get the Maclaurin series for the function $\tan^{-1} x$:

$$\tan^{-1} x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{2k+1} = x - \frac{1}{3}x^3 + \frac{1}{5}x^5 - \frac{1}{7}x^7 + \dots, \quad \text{for } x \in [-1, 1].$$

Again, the equality holds at both endpoints due to the same result stated above.

7.30 MATLAB: differentiation of power series

Find the power series for $\left(\sum_{k=0}^{\infty} x^k\right)' - \left(\sum_{k=0}^{\infty} (-1)^k x^k\right)', x \in (-1, 1)$.

Solution

```
>> syms x k
>> assume(k, 'integer') % k is an integer
>> a(k) = x^k;
>> b(k) = (-1)^k * x^k;
>> term1 = symsum(a(k), k, 0, inf);
>> term2 = symsum(b(k), k, 0, inf);
>> term1_diff = diff(term1, x)

term1_diff =
piecewise(1 < x, 0, abs(x) < 1, 1/(x - 1)^2)

>> term2_diff = diff(term2, x)

term2_diff =
piecewise(x < -1, 0, abs(x) < 1, -1/(x + 1)^2)

>> Delta_term_diff = term1_diff - term2_diff;
>> Delta_term_diff_series = ...
    taylor(Delta_term_diff, x, 'ExpansionPoint', 0, 'Order', 7)

Delta_term_diff_series =
14*x^6 + 10*x^4 + 6*x^2 + 2
```

It is shown that, up to the order of 6, the power series for $\left(\sum_{k=0}^{\infty} x^k\right)' - \left(\sum_{k=0}^{\infty} (-1)^k x^k\right)'$ can

be written as

$$\left(\sum_{k=0}^{\infty} x^k \right)' - \left(\sum_{k=0}^{\infty} (-1)^k x^k \right)' = 2 + 6x^2 + 10x^4 + 14x^6 + \dots, \quad x \in (-1, 1).$$

7.31 MATLAB: integration of power series

Find the power series for $\int \sum_{k=0}^{\infty} (-1)^k x^k dx, \quad x \in (-1, 1)$.

Solution

```
>> syms x k
>> assume(k, 'integer') % k is an integer
>> a(k) = (-1)^k * x^k;
>> int_fun = symsum(a(k), k, 0, inf);
>> int_term = int(int_fun, x);
>> int_series = taylor(int_term, x, 'ExpansionPoint', 0, 'Order', 5)

int_series =
- x^4/4 + x^3/3 - x^2/2 + x
```

It is shown that, up to the order of 4, the power series for $\int \sum_{k=0}^{\infty} (-1)^k x^k dx, \quad x \in (-1, 1)$, can be written as

$$\int \sum_{k=0}^{\infty} (-1)^k x^k dx = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + \dots.$$

7.32 Theorem: common Maclaurin series and their radii of convergence

$$1. \quad \frac{1}{1-x} = \sum_{k=0}^{\infty} x^k = 1 + x + x^2 + x^3 + \dots, \quad x \in (-1, 1), \quad R = 1.$$

$$2. \quad e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots, \quad x \in (-\infty, \infty), \quad R = \infty.$$

$$3. \quad \sin x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots, \quad x \in (-\infty, \infty), \quad R = \infty.$$

$$4. \quad \cos x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots, \quad x \in (-\infty, \infty), \quad R = \infty.$$

$$5. \quad \tan^{-1} x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{2k+1} = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \dots, \quad x \in [-1, 1], \quad R = 1.$$

6. $\ln(1+x) = \sum_{k=1}^{\infty} (-1)^{k-1} \frac{x^k}{k} = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots, \quad x \in (-1, 1], \quad R = 1.$

7. $(1+x)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k = 1 + \alpha x + \frac{\alpha(\alpha-1)}{2!} x^2 + \frac{\alpha(\alpha-1)(\alpha-2)}{3!} x^3 + \dots$, at least for
 $x \in (-1, 1), \quad R = 1.$

7.33 Example: binomial expansion

Expand the function $y = \sqrt{a^2 + x^2}$ ($a > 0$) as its Maclaurin series.

Solution Write the function as

$$y = a [1 + (x/a)^2]^{\frac{1}{2}}.$$

By the binomial expansion

$$(1+x)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k, \quad x \in (-1, 1).$$

we get, for $-a < x < a$,

$$\begin{aligned} \sqrt{a^2 + x^2} &= a \sum_{k=0}^{\infty} \binom{\frac{1}{2}}{k} (x/a)^{2k} \\ &= a \left[1 + \frac{\frac{1}{2}}{1!} \cdot \left(\frac{x}{a}\right)^2 + \frac{\frac{1}{2}(\frac{1}{2}-1)}{2!} \cdot \left(\frac{x}{a}\right)^4 + \frac{\frac{1}{2}(\frac{1}{2}-1)(\frac{1}{2}-2)}{3!} \cdot \left(\frac{x}{a}\right)^6 + \dots \right] \\ &= a + \frac{\frac{1}{2}}{1!} \cdot a^{-1} x^2 + \frac{\frac{1}{2}(\frac{1}{2}-1)}{2!} \cdot a^{-3} x^4 + \frac{\frac{1}{2}(\frac{1}{2}-1)(\frac{1}{2}-2)}{3!} \cdot a^{-5} x^6 + \dots . \end{aligned}$$

7.34 Theorem: theorem of analytic function

Suppose f has a power series representation (expansion) at a :

$$f(x) = \sum_{k=0}^{\infty} c_k (x-a)^k, \quad |x-a| < R,$$

where $R > 0$ or $R = \infty$. Then

1. The coefficients are given by the formula

$$c_k = \frac{f^{(k)}(a)}{k!}, \quad k \geq 0.$$

2. $\lim_{x \rightarrow a} \frac{f(x) - T_n(x)}{(x-a)^n} = 0$, or equivalently,

$$f(x) = T_n(x) + o(|x-a|^n) = \sum_{k=0}^n \frac{f^{(k)}(a)}{k!} (x-a)^k + o(|x-a|^n), \quad \text{as } x \rightarrow a.$$

3. $f(x) = \lim_{n \rightarrow \infty} T_n(x)$ for $|x - a| < R$.

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 k , $k = 0, 1, 2, \dots$, we differentiate and substitute $x = a$ to have

$$f^{(k)}(a) = 2 \cdot 3 \cdot 4 \cdots k \cdot c_k = k! \cdot c_k,$$

so that

$$c_k = \frac{f^{(k)}(a)}{k!}, \quad k \geq 0.$$

- By part 1, we get $T_n(x) = \sum_{k=0}^n c_k(x - a)^k$. Thus,

$$\lim_{x \rightarrow a} \frac{f(x) - T_n(x)}{(x - a)^n} = \lim_{x \rightarrow a} \frac{(x - a)^{n+1} \cdot \sum_{k=0}^{\infty} c_{n+k}(x - a)^k}{(x - a)^n} = 0.$$

- By part 1, we get $T_n(x) = \sum_{k=0}^n c_k(x - a)^k$. Thus,

$$f(x) = \sum_{k=0}^{\infty} c_k(x - a)^k = \lim_{n \rightarrow \infty} T_n(x).$$

7.35 Example: find higher derivative

For $f(x) = \tan^{-1} x$, find $f^{(2019)}(0)$.

Solution We know that the function $f(x) = \tan^{-1} x$ has the following Maclaurin series:

$$\tan^{-1} x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{2k+1}, \quad x \in [-1, 1].$$

By the uniqueness of Taylor expansion, we have

$$\frac{f^{(2k+1)}(0)}{(2k+1)!} = \frac{(-1)^k}{2k+1}, \quad k \geq 0.$$

Thus,

$$f^{(2019)}(0) = (-1)^{1009} \cdot 2018! = -2018!.$$

7.36 MATLAB: find higher derivative

For $f(x) = \tan^{-1} x$, find $f^{(2019)}(0)$.

Solution First, the Maclaurin series of $f(x) = \tan^{-1} x$ can be written as

$$f(x) = \tan^{-1} x = \sum_{k=0}^{\infty} c_k x^k, \quad x < R.$$

According to the theorem of analytic function, the coefficient of the term with order k is given by

$$c_k = \frac{f^{(k)}(0)}{k!}, \quad k \geq 0.$$

Hence, we have $f^{(2019)}(0) = 2019! c_{2019}$. To calculate $f^{(2019)}(0)$, we thus only need to calculate c_{2019} , which is shown as follows:

```
>> syms x
>> f(x) = atan(x);
>> f_series = taylor(f(x),x,'ExpansionPoint',0,'Order',2019 + ...
1); % expand f(x) up to order 2019
>> c_array = coeffs(f_series); % find all the coefficients of ...
the series, the coefficient of the term with the highest ...
order is the last element of the array
>> ind = length(c_array); % the location of the coefficient ...
for x^2019 in the array
>> c = c_array(ind) % the value of the last coefficient, ...
i.e., the one for x^2019

c =
-1/2019
```

In the MATLAB code above, the command `coeffs` can return all the coefficients of a polynomial, and the command `length` can return the length of a vector.

We see from the result above that $c_{2019} = -\frac{1}{2019}$. Therefore, we have

$$f^{(2019)}(0) = 2019! c_{2019} = -2018!.$$

7.4 Applications of Taylor Series

Limits by Taylor Series

7.37 Example: limit by Taylor series

Find $\lim_{x \rightarrow 0} \frac{x^3 - 6x \cos(x^2) + 6 \sin x}{x^5}$.

Solution The limit has the indeterminate form $\frac{0}{0}$. Although we may apply l'Hospital's Rule, alternatively we use the technique of Taylor series. In fact, near $x = 0$, we have

$$\begin{aligned}\sin x &= x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 + \dots, \\ \cos(x^2) &= 1 - \frac{1}{2!}(x^2)^2 + \dots,\end{aligned}$$

so that, as $x \rightarrow 0$,

$$\begin{aligned}\sin x &= x - \frac{1}{6}x^3 + \frac{1}{120}x^5 + o(|x|^5), \\ \cos(x^2) &= 1 - \frac{1}{2}x^4 + o(|x|^4),\end{aligned}$$

Hence,

$$\begin{aligned}&\lim_{x \rightarrow 0} \frac{x^3 - 6x \cos(x^2) + 6 \sin x}{x^5} \\ &= \lim_{x \rightarrow 0} \frac{x^3 - 6x \left[1 - \frac{1}{2}x^4 + o(|x|^4)\right] + 6 \left[x - \frac{1}{6}x^3 + \frac{1}{120}x^5 + o(|x|^5)\right]}{x^5} \\ &= \lim_{x \rightarrow 0} \frac{\frac{61}{20}x^5 + o(|x|^5)}{x^5} = \frac{61}{20}.\end{aligned}$$

7.38 Example: limit by Taylor series

Find $\lim_{x \rightarrow \infty} [x^4 \sin(x^{-4}) - x^2 e^{-1/x^2} + x^2]$.

Solution If making the substitution $x = t^{-1}$, we see that $x \rightarrow \infty \iff t \rightarrow 0^+$. Thus,

$$\begin{aligned}\lim_{x \rightarrow \infty} [x^4 \sin(x^{-4}) - x^2 e^{-1/x^2} + x^2] &= \lim_{t \rightarrow 0^+} [t^{-4} \sin(t^4) - t^{-2} e^{-t^2} + t^{-2}] \\ &= \lim_{t \rightarrow 0^+} \frac{\sin(t^4) - t^2 e^{-t^2} + t^2}{t^4}.\end{aligned}$$

The last limit has the indeterminate form $\frac{0}{0}$. We use the technique of Taylor series. In fact, near $t = 0$, we have

$$\begin{aligned}\sin(t^4) &= t^4 + \dots, \\ e^{-t^2} &= 1 + \frac{1}{1!}(-t^2) + \dots,\end{aligned}$$

so that, as $t \rightarrow 0^+$,

$$\begin{aligned}\sin(t^4) &= t^4 + o(|t|^4), \\ e^{-t^2} &= 1 - t^2 + o(|t|^2),\end{aligned}$$

Hence,

$$\begin{aligned}&\lim_{t \rightarrow 0^+} \frac{\sin(t^4) - t^2 e^{-t^2} + t^2}{t^4} \\ &= \lim_{t \rightarrow 0^+} \frac{[t^4 + o(|t|^4)] - t^2 [1 - t^2 + o(|t|^2)] + t^2}{t^4} \\ &= \lim_{t \rightarrow 0^+} \frac{2t^4 + o(|t|^4)}{t^4} = 2.\end{aligned}$$

Therefore,

$$\lim_{x \rightarrow \infty} [x^4 \sin(x^{-4}) - x^2 e^{-1/x^2} + x^2] = 2.$$

Approximation of Definite Integrals

7.39 Example: approximation of a definite integral

Approximate $\int_0^1 e^{-x^2} dx$ within 10^{-4} of the exact value.

Solution We begin with the Maclaurin series

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots, \quad x \in (-\infty, \infty).$$

By replacing x with $-x^2$, we have

$$e^{-x^2} = \sum_{k=0}^{\infty} \frac{(-x^2)^k}{k!} = 1 - \frac{x^2}{1!} + \frac{x^4}{2!} - \frac{x^6}{3!} + \cdots, \quad x \in (-\infty, \infty).$$

Thus,

$$\begin{aligned}\int_0^1 e^{-x^2} dx &= \left(x - \frac{x^3}{3 \cdot 1!} + \frac{x^5}{5 \cdot 2!} - \frac{x^7}{7 \cdot 3!} + \cdots \right) \Big|_{x=0}^1 \\ &= 1 - \frac{1}{3 \cdot 1!} + \frac{1}{5 \cdot 2!} - \frac{1}{7 \cdot 3!} + \cdots = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1) \cdot k!}.\end{aligned}$$

This shows that the definite integral can be expressed as a convergent alternating series, with $b_k = \frac{1}{(2k+1) \cdot k!}$ satisfying:

$$0 < b_{k+1} < b_k, \quad \lim_{k \rightarrow \infty} b_k = 0.$$

Using the bound on the remainder in the Alternating Series Test, we have

$$|R_n| = \left| \int_0^1 e^{-x^2} dx - \sum_{k=0}^n \frac{(-1)^k}{(2k+1) \cdot k!} \right| < \frac{1}{(2n+3) \cdot (n+1)!}.$$

Thus, to ensure that $|R_n| < 10^{-4}$, we need to choose n so that $\frac{1}{(2n+3) \cdot (n+1)!} < 10^{-4}$, or $n > 6$. Hence, the finite series $\sum_{k=0}^6 \frac{(-1)^k}{(2k+1) \cdot k!}$ gives an approximation that is within 10^{-4} of the exact value of the series.

Using the bounds on the series in the Alternating Series Test, we have

$$s_6 - \frac{1}{15 \cdot 7!} < \int_0^1 e^{-x^2} dx < s_6 + \frac{1}{15 \cdot 7!},$$

where $s_6 = \sum_{k=0}^6 \frac{(-1)^k}{(2k+1) \cdot k!}$. By using a calculator, we have $s_6 \approx 0.74684$, which has 5th decimal place accuracy. Thus,

$$0.74683 < s < 0.74685.$$

Taking the average of these two bounds as our approximation of $\int_0^1 e^{-x^2} dx$, we find that $\int_0^1 e^{-x^2} dx \approx 0.74684$.

For comparison, $\int_0^1 e^{-x^2} dx = 0.7468241324\ldots$. We see that our approximation indeed is within 10^{-4} of the exact value of the series.

Solving Differential Equations

7.40 Example: solving differential equation by power series

Solve the differential equation $y'(t) = y(t) - 1$, subject to the initial value condition $y(0) = 3$.

Solution We assume the solution has a power series solution of the form

$$y(t) = \sum_{k=0}^{\infty} c_k t^k = c_0 + c_1 t + c_2 t^2 + c_3 t^3 + \cdots,$$

where the coefficients $\{c_k\}$ will be determined.

Since $y(0) = 3$, we get $c_0 = 3$.

The other coefficients can be determined one by one. In fact, substituting the series solution into the differential equation gives

$$c_1 + 2c_2 t + 3c_3 t^2 + 4c_4 t^3 + \cdots = (c_0 + c_1 t + c_2 t^2 + c_3 t^3 + \cdots) - 1.$$

Putting $t = 0$ on both sides of the equation gives

$$c_1 = c_0 - 1 \implies c_1 = 2. \quad \text{since } c_0 = 3$$

We differentiate both sides of the above equation to have

$$2c_2 + 3 \cdot 2c_3t + 4 \cdot 3c_4t^2 + \dots = (c_1 + 2c_2t + 3c_3t^2 + \dots) - 0.$$

Again, putting $t = 0$ on both sides of the last equation gives

$$2c_2 = c_1 \implies c_2 = 1. \quad \text{since } c_1 = 2$$

We repeat the above process by differentiating the equation once more then putting $t = 0$ on the resulted equation. Recursively, we get:

$$c_3 = \frac{1}{3}, \quad c_4 = \frac{1}{4 \cdot 3}, \quad c_5 = \frac{1}{5 \cdot 4 \cdot 3}, \quad \dots$$

Thus, we obtain an “formal” solution:

$$y(t) = 3 + 2t + t^2 + \frac{t^3}{3} + \frac{t^4}{4 \cdot 3} + \frac{t^5}{5 \cdot 4 \cdot 3} + \dots$$

The formal solution can be re-written as (formally!)

$$y(t) = 1 + 2 \left(1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \frac{t^4}{4!} + \frac{t^5}{5!} + \dots \right).$$

We recognize that the right-hand side is actually the Maclaurin series of the function $1 + 2e^t$.

Finally, we can verify that the formally obtained solution is indeed the true solution of the original problem, since

$$\begin{aligned} y'(t) &= (1 + 2e^t)' = 2e^t = (1 + 2e^t) - 1 = y(t) - 1; \\ y(0) &= (1 + 2e^t)|_{t=0} = 3. \end{aligned}$$

Summary

Definition: basic concepts

- Let f be a function with f', f'', \dots , and $f^{(n)}$ defined at a . The **n th-order Taylor polynomial** for f with its **center** at a is

$$T_n(x) = f(a) + f'(a)(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x - a)^n.$$

- Let T_n be the Taylor polynomial of order n for f at a . The term $R_n(x) = f(x) - T_n(x)$ is called the **remainder** of the **Taylor series**. The latter refers to the following “formal” power series of f :

$$f(a) + f'(a)(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x - a)^n + \cdots.$$

This power series is also called **Taylor expansion** of f at a . A Taylor series centered at 0 is called a **Maclaurin series**.

- A **power series in $(x - a)$** or a **power series centered at a** or a **power series about a** is a series of the form

$$\sum_{k=0}^{\infty} c_k (x - a)^k = c_0 + c_1(x - a) + c_2(x - a)^2 + \cdots.$$

Theorem: uniqueness of Taylor polynomials

Suppose f is a function with f', f'', \dots , and $f^{(n)}$ defined at a . If P_n is a polynomial of degree n satisfying

$$P_n(a) = f(a), \quad P'_n(a) = f'(a), \quad P''_n(a) = f''(a), \quad \dots, \quad P_n^{(n)}(a) = f^{(n)}(a),$$

Then

$$P_n(x) \equiv T_n(x) = f(a) + f'(a)(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x - a)^n.$$

Theorem: asymptotic approximation of Taylor polynomials

If $f^{(n)}(a)$ exists, then $\lim_{x \rightarrow a} \frac{f(x) - T_n(x)}{(x - a)^n} = 0$, or equivalently, by using little- o notation,

$$f(x) = T_n(x) + o(|x - a|^n) = \sum_{k=0}^n \frac{f^{(k)}(a)}{k!}(x - a)^k + o(|x - a|^n), \quad \text{as } x \rightarrow a.$$

Theorem: convergence theorem of Taylor series

If $\lim_{n \rightarrow \infty} R_n(x) = 0$ in an open interval I that contains a , then

$$f(x) = \lim_{n \rightarrow \infty} T_n(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!} (x-a)^k, \quad x \in I.$$

Theorem: integral form of the remainder

If $f^{(n+1)}$ is continuous in an open interval I that contains a , then

$$R_n(x) = \frac{1}{n!} \int_a^x (x-t)^n f^{(n+1)}(t) dt, \quad x \in I.$$

Theorem: Lagrange remainder

If $f^{(n+1)}$ is continuous in an open interval I that contains a , then there is c between a and x such that

$$R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!} (x-a)^{n+1}.$$

This formula is called the **Lagrange remainder** of Taylor series.

Theorem: remainder estimate

If $|f^{(n+1)}(x)| \leq M$ in an open interval I that contains a , then

$$|R_n(x)| \leq \frac{M}{(n+1)!} |x-a|^{n+1}, \quad x \in I.$$

Hence, if $|f^{(n+1)}(x)| \leq M$ for every n in I , then

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!} (x-a)^k, \quad x \in I.$$

Theorem: convergence of power series

For a given power series $\sum_{k=0}^{\infty} c_k (x-a)^k$ there are only three possibilities:

1. The series converges only when $x = a$.
2. The series converges for all x .
3. There is a positive number R such that the series converges if $|x-a| < R$ and diverges if $|x-a| > R$.

Definition: radius of convergence

The number R is called the **radius of convergence** of the power series.

By convention, the radius of convergence is $R = 0$ when the power series converges only at the center a and $R = \infty$ when the power series converges for all x . The **interval of convergence** of a power series is the interval that consists of all values of x for which the series converges.

Method: determine the interval of convergence

Let $\sum_{k=0}^{\infty} c_k(x - a)^k$ be a given power series.

1. Compute the radius of convergence R by using the Ratio Test (or sometimes the Root Test) by

$$R = \frac{1}{\lim_{k \rightarrow \infty} \left| \frac{c_{k+1}}{c_k} \right|} = \lim_{k \rightarrow \infty} \left| \frac{c_k}{c_{k+1}} \right| \quad \text{or} \quad R = \frac{1}{\lim_{k \rightarrow \infty} \sqrt[k]{|c_k|}}.$$

2. The Ratio and Root Tests are inconclusive 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.

Theorem: combining power series

Suppose $\sum c_k x^k$ and $\sum d_k x^k$ converge to $f(x)$ and $g(x)$, respectively, in an open interval I containing 0. Then

1. **Sum and Difference:** $\sum(c_k \pm d_k)x^k$ converges to $f(x) \pm g(x)$ in I , that is,

$$\sum_{k=0}^{\infty} (c_k \pm d_k)x^k = \sum_{k=0}^{\infty} c_k x^k \pm \sum_{k=0}^{\infty} d_k x^k, \quad x \in I.$$

2. **Product:** $\sum (c_0 d_k + c_1 d_{k-1} + \cdots + c_{k-1} d_1 + c_k d_0)x^k$ converges to $f(x) \cdot g(x)$ in I , that is,

$$\sum_{k=0}^{\infty} \left(\sum_{j=0}^k c_j d_{k-j} \right) x^k = \left(\sum_{k=0}^{\infty} c_k x^k \right) \left(\sum_{k=0}^{\infty} d_k x^k \right), \quad x \in I.$$

3. **Quotient:** If $d_0 \neq 0$, then in a neighborhood of $x = 0$, the power series $\sum a_k x^k$

converges to $\frac{f(x)}{g(x)}$, that is,

$$\sum_{k=0}^{\infty} a_k x^k = \frac{\sum_{k=0}^{\infty} c_k x^k}{\sum_{k=0}^{\infty} d_k x^k},$$

where

$$a_0 = \frac{c_0}{d_0}, \quad a_k = \frac{1}{d_0} \left[c_k - \sum_{j=1}^k d_j a_{k-j} \right] \quad \text{for } k \geq 1.$$

- 4. Composition:** If $h(x) = bx^m$, where m is a positive integer and $b \neq 0$ a real number, the power series $\sum c_k (h(x))^k$ converges to $f(h(x))$ for x such that $h(x) \in I$.

Theorem: differentiation and integration of power series

If the power series $\sum_{k=0}^{\infty} c_k (x-a)^k$ has radius of convergence $R > 0$, then the function f defined by

$$f(x) = c_0 + c_1(x-a) + c_2(x-a)^2 + \cdots = \sum_{k=0}^{\infty} c_k (x-a)^k$$

is differentiable (and therefore continuous) on the interval $(a-R, a+R)$ and

1. $f'(x) = c_1 + 2c_2(x-a) + 3c_3(x-a)^2 + \cdots = \sum_{k=1}^{\infty} kc_k (x-a)^{k-1}$.
2. $\int f(x) dx = C + c_0(x-a) + c_1 \frac{(x-a)^2}{2} + c_2 \frac{(x-a)^3}{3} + \cdots = C + \sum_{k=0}^{\infty} c_k \frac{(x-a)^{k+1}}{k+1}$.

The radii of convergence of these power series are both still R .

Theorem: common Maclaurin series and their radii of convergence

1. $\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k = 1 + x + x^2 + x^3 + \cdots, \quad x \in (-1, 1), \quad R = 1.$
2. $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots, \quad x \in (-\infty, \infty), \quad R = \infty.$
3. $\sin x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots, \quad x \in (-\infty, \infty), \quad R = \infty.$
4. $\cos x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots, \quad x \in (-\infty, \infty), \quad R = \infty.$
5. $\tan^{-1} x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{2k+1} = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots, \quad x \in [-1, 1], \quad R = 1.$

6. $\ln(1+x) = \sum_{k=1}^{\infty} (-1)^{k-1} \frac{x^k}{k} = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots, \quad x \in (-1, 1], \quad R = 1.$

7. $(1+x)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k = 1 + \alpha x + \frac{\alpha(\alpha-1)}{2!} x^2 + \frac{\alpha(\alpha-1)(\alpha-2)}{3!} x^3 + \dots$, at least for
 $x \in (-1, 1), \quad R = 1.$

Theorem: theorem of analytic function

Suppose f has a power series representation (expansion) at a :

$$f(x) = \sum_{k=0}^{\infty} c_k (x-a)^k, \quad |x-a| < R,$$

where $R > 0$ or $R = \infty$. Then

1. The coefficients are given by the formula

$$c_k = \frac{f^{(k)}(a)}{k!}, \quad k \geq 0.$$

2. $\lim_{x \rightarrow a} \frac{f(x) - T_n(x)}{(x-a)^n} = 0$, or equivalently,

$$f(x) = T_n(x) + o(|x-a|^n) = \sum_{k=0}^n \frac{f^{(k)}(a)}{k!} (x-a)^k + o(|x-a|^n), \quad \text{as } x \rightarrow a.$$

3. $f(x) = \lim_{n \rightarrow \infty} T_n(x)$ for $|x-a| < R$.

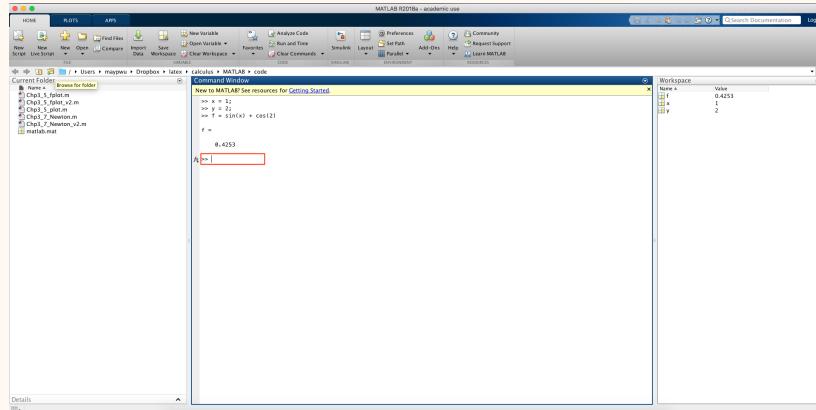
Appendix: Introduction to MATLAB

7.41 MATLAB: What is MATLAB?

MATLAB is short for “Matrix Laboratory”. It is a multi-paradigm numerical computing environment and proprietary programming language developed by MathWorks. MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages, including C, C++, C#, Java, Fortran and Python. Although MATLAB is intended primarily for numerical computing, an optional toolbox uses the MuPAD symbolic engine, allowing access to symbolic computing abilities [Wikipedia].

7.42 MATLAB: Command Window

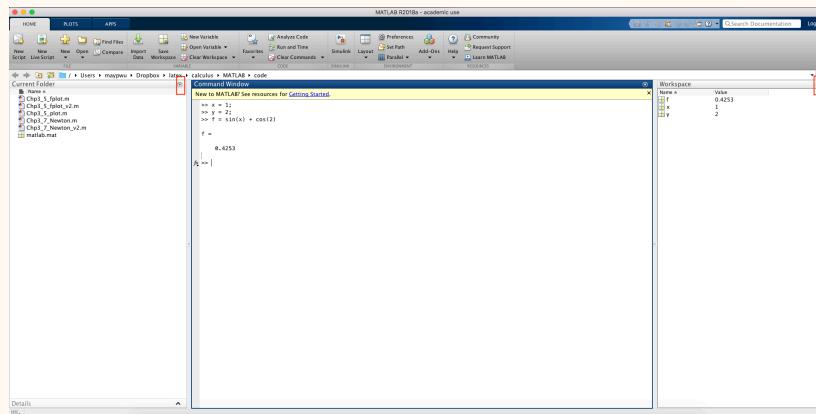
The command window is where the commands of MATLAB are typed in and calculations are carried out. In particular, in the command window, the MATLAB commands are typed in after two arrows `>>`, as shown in the following figure.



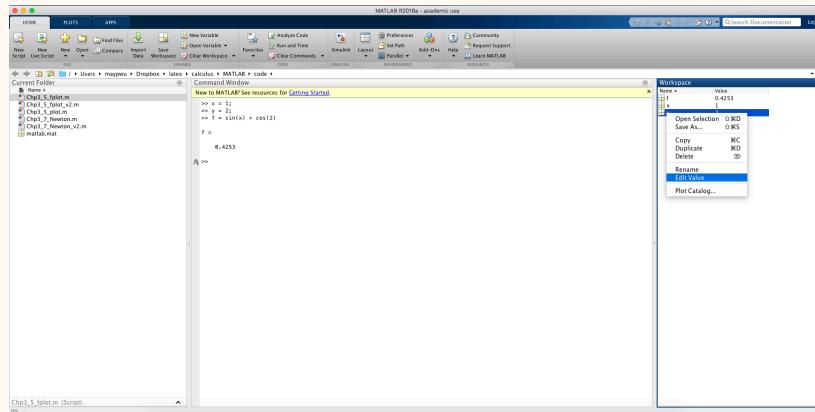
After typing in the command and pressing “Enter” (in Windows) or “return” (in Mac), the calculation will be carried out.

7.43 MATLAB: Current Folder and Workspace

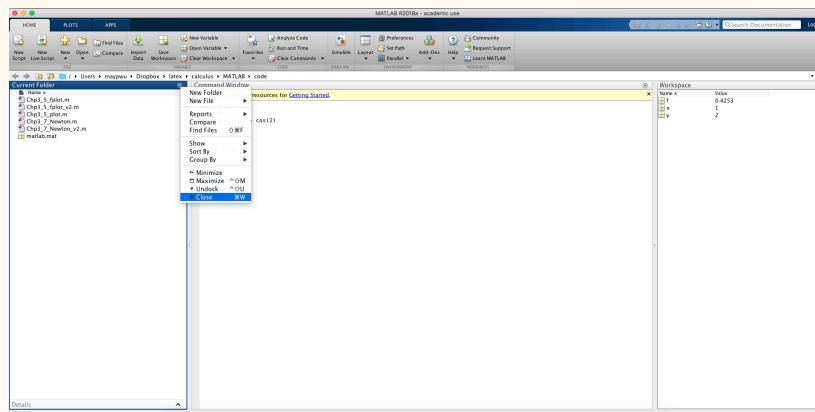
Next to the command window, there are two other windows, namely, “Current Folder” window and “Workspace” window, as shown in the following figure.



In particular, the “Current Folder” window on the left enables us to view, create, open, move, and rename files and folders in the current folder. For the “Workspace” window on the right, it enables us to view and interactively manage each variable or object used in MATLAB. For example, we can edit the content a variable directly in the workspace by right clicking the scalar and selecting the “Edit Value”, as shown in the following figure.



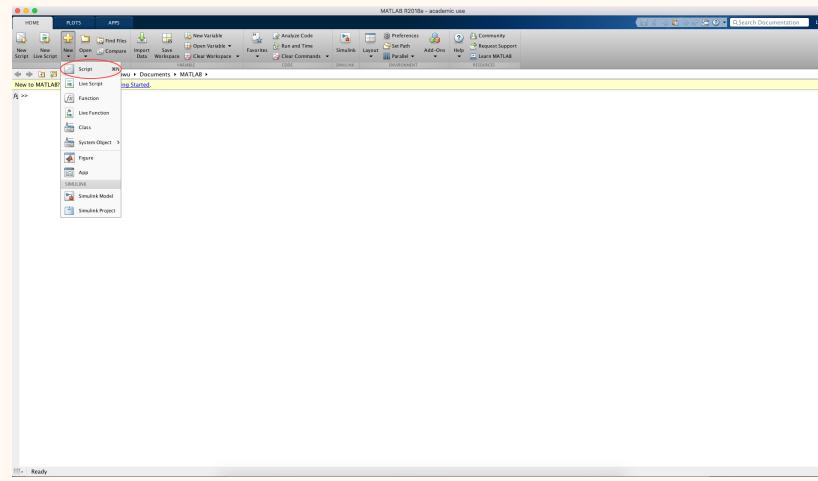
Note that if we click the “Close” option under the triangle menu, the “Current Folder” or “Workspace” window will be closed.



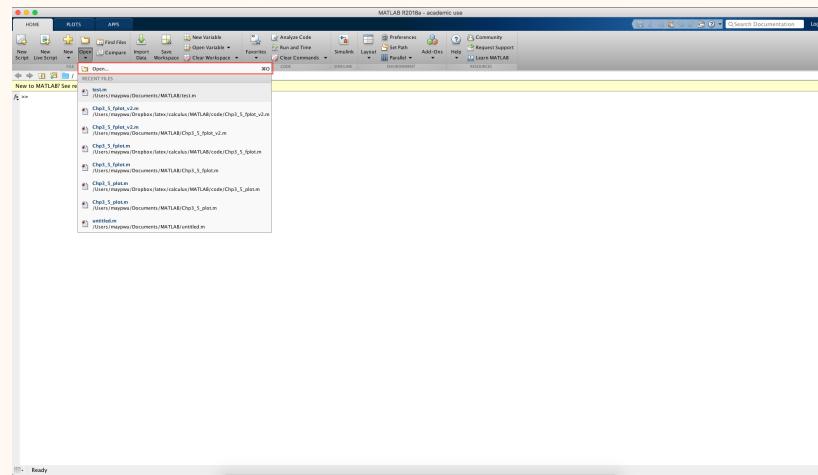
7.44 MATLAB: M-File

The MATLAB commands can be put into one file called m-file (or script file), and then we can instruct MATLAB to execute these commands.

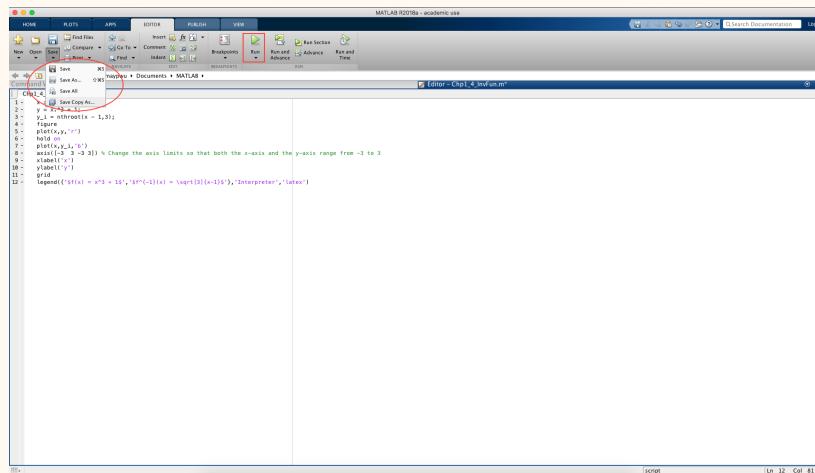
To create a m-file, we simply need to click the “New” option from the “HOME” menu and select “Script”, as shown in the following figure.



To open an existing m-file, we simply need to click the “Open” option from the “HOME” menu and select the file required from the directory, as shown in the following figure.



To save an m-file, we can click the “Save” option from the “EDITOR” menu, as shown in the following figure, and choose a suitable destination from the directory for the m-file. We can also save an existing m-file to another destination by clicking the “Save As” option, and then choose a suitable destination from the directory. Note that the m-file should be saved with a “.m” extension.



In the m-file, unlike in the command window, there are no double arrows after which the command should be typed in. Moreover, all the commands in the m-file will not be executed before clicking the “Run” button in the “EDITOR” menu which is shown in the figure above. After clicking the “Run” button, the result will be shown if there is no error; otherwise, the error information will be displayed in the command window.

7.45 MATLAB: Assignment of A Value and Semicolon

To explain how to assign a value and when to use the semicolon in MATLAB, we first provide some examples as follows. Specifically, the following examples show the calculations of $\cos(1) + 1$ in the command window.

```

>> cos(1) + 1

ans =
1.5403

>> r1 = cos(1) + 1

r1 =
1.5403

>> r2 = cos(1) + 1;
>> r2

r2 =

```

1.5403

We note that “`=`” means “assign a value” in MATLAB (The symbol “`==`” represents “equal” in MATLAB). If no variable is specified, then the answer of `cos(1)+1` is assigned to the default variable `ans`, as can be seen from the first command of the MATLAB code above. Note that in MATLAB, the number “1” in `cos(1)` is in radians (not degrees). If a certain variable is specified, e.g., `r1` in the second command, then the value of `cos(1)+1` is assigned to `r1` after executing this command.

Next, we explain the use of the semicolon (i.e., “`;`”). Specifically, if no semicolon is used after a command, then the result of this command will be displayed directly after executing this command; while if a semicolon is used after a command, then the result of this command will not be displayed, as can be seen from the last two commands of the MATLAB code above.

7.46 MATLAB: Comments

In MATLAB, comments are written after `%`, and MATLAB does not attempt to compute the comments. For example, in the MATLAB code below, “`Change the axis limits so that both the x-axis and the y-axis range from -3 to 3`” are the comments.

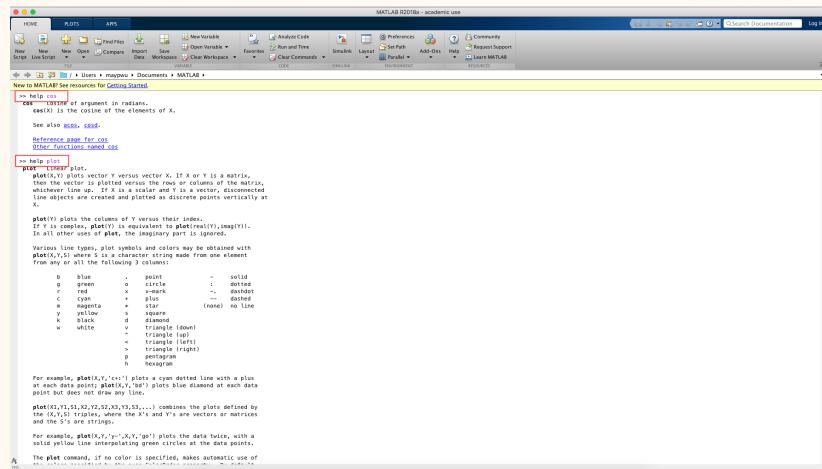
```

1 Ch7_4.m
2 x = -pi:0.02:pi;
3 y = sin(x);
4 z = x.*exp(10*x)^(1/3);
5 plot(x,y,'r')
6 hold on
7 plot(x,z,'b')
8 axis([-3 3 -3 3]) % Change the axis limits so that both the x-axis and the y-axis range from -3 to 3
9 xlabel('x')
10 ylabel('y')
11 title('Plot of sin(x) and x^(1/3)')
12 legend('sin(x) = x^3 + 1x^2 + (-1)x + 1sqrt(3)x(-1)x^2y','Interpreter','latex')

```

7.47 MATLAB: Help

In MATLAB, there is extensive documentation explaining the use of all its built-in functions. To check the use of a function, we simply need to type in `help function name` after `>>` in the command window. For example, in the following figure, two commands, namely, `help cos` and `help plot`, are typed in to check the use of `cos` and `plot`, respectively.



7.48 MATLAB: Vector

The most common variable type in MATLAB is matrix. In Calculus course, we will need to frequently use the one-dimensional matrices, i.e., vectors, in MATLAB. In the following, we will introduce two ways to define a vector in MATLAB.

1. initial:increment:terminator

The first way to define a vector is by specifying the initial value, the increment (or step), and the terminator of the vector.

```
>> vec1 = 1:2:9

vec1 =

    1     3     5     7     9

>> vec2 = 1:2:10

vec2 =

    1     3     5     7     9

>> vec = 1:5

vec =

    1     2     3     4     5
```

The MATLAB code above defines two vectors named `vec1` and `vec2`, respectively. There are 5 elements in both of the vectors, namely, 1, 3, 5, 7 and 9. In particular, `vec1` starts at 1 (i.e., the initial value is 1), the increment from the previous value (i.e., the step) is 2, and it stops at 9.

However, sometimes the vector does not stop at the `terminator` specified in the command. With the value of each element changing from the value of the previous element by adding an increment, the vector stops at a value which is not exceeding the `terminator` but closest to the terminator. For example, `vec2` stops at 9 instead of 10.

Note that if the `increment` is not specified in the command, the default increment is 1, as can be seen from the third command above.

2. `linspace`

Using the command `linspace` is another way of defining a vector.

```
>> vec3 = linspace(1,9,5)

vec3 =
    1     3     5     7     9

>> vec4 = linspace(1,10,5)

vec4 =
    1.0000    3.2500    5.5000    7.7500   10.0000
```

From the MATLAB code above, we see that the command `linspace(x1,x2,N)` defines an N -element vector starting at `x1` and stopping at `x2`. In this way, the increment equals to $\frac{x_2 - x_1}{N - 1}$.

Note that if we use the command `linspace(x1,x2)`, then a 100-element vector with initial value `x1` and ending value `x2` will be generated.

7.49 MATLAB: Basic Computations

Some basic matrix operators are listed in the following.

Operator	Meaning
+	Matrix addition
-	Matrix subtraction
*	Matrix multiplication
.*	Element-by-element multiplication
./	Element-by-element division
.^	Matrix exponentiation
.^	Element-by-element exponentiation

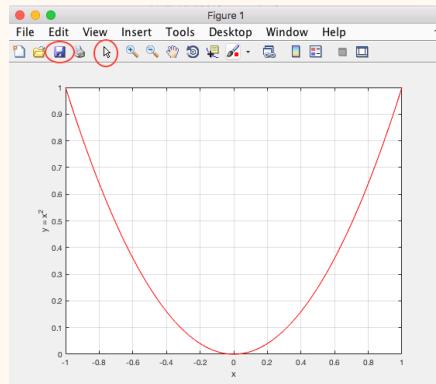
We see from the above table that some operators for matrices are the same with the ones for scalars, but some are different and require an additional . before the scalar operator.

7.50 MATLAB: Plot

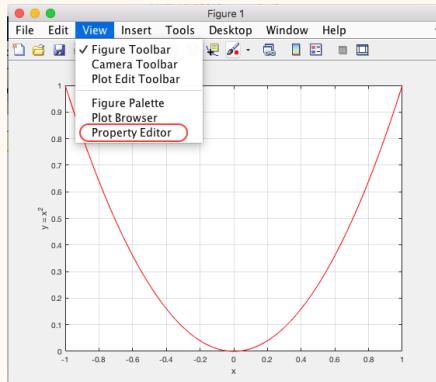
MATLAB has powerful graph-plotting features. For example, the command `plot` can be used to generate a graph based on two vectors. The following MATLAB code generates the figure of $y = x^2$ by using the command `plot`.

```
>> x = -1:0.01:1;
>> y = x.^2;
>> figure % create a figure window
>> plot(x,y,'r') % graph is in red
>> xlabel('x') % X-axis label
>> ylabel('y = x^2') % Y-axis label
>> grid % add grid lines
```

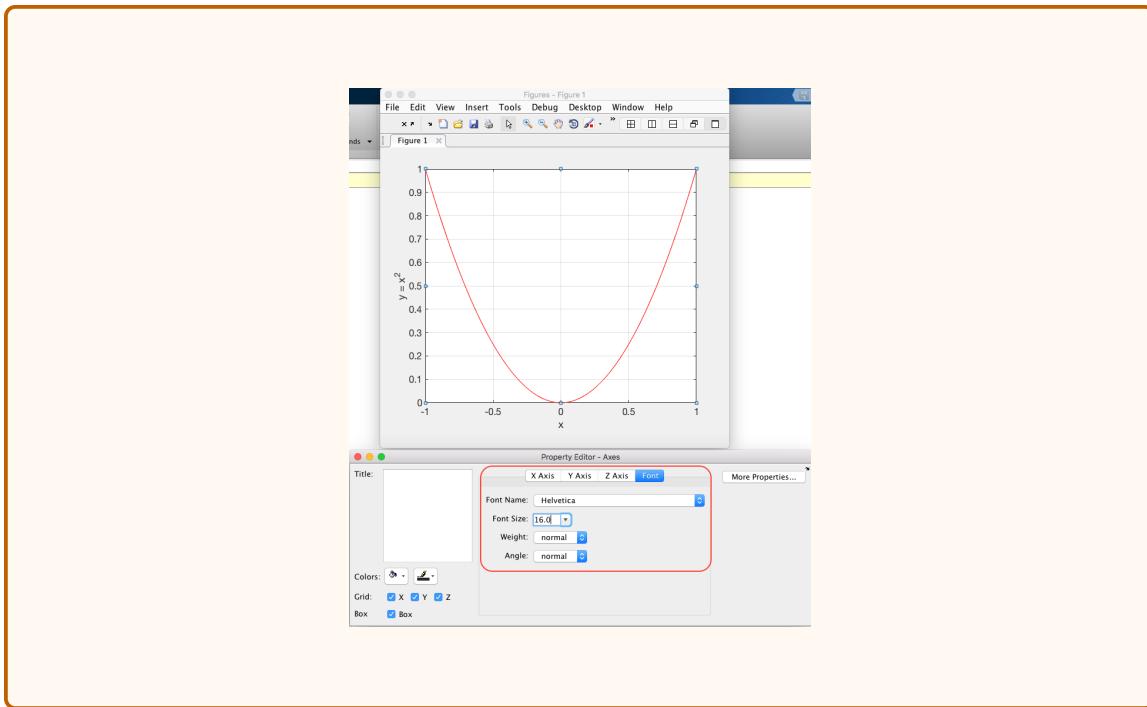
The figure generated is shown in the following:



Note that by clicking the third icon, we can save the figure. Moreover, after clicking the fifth icon, i.e., the arrow, we can choose the “Property Editor” option from the “View” menu, as highlighted in the following figure.



Then, a “Property Editor” window will appear as follows, and we can edit the figure, e.g., change the font, etc.



Index

- e , 34
 - p -series, 362
- absolute convergence, 380
 - absolute extreme, 97
 - absolute maximum, 97
 - absolute minimum
 - MATLAB, 105
 - absolution minimum, 97
 - alternating harmonic series, 377
 - MATLAB, 378
 - Alternating Series Test
 - MATLAB, 378
 - remainder estimate, 378
 - antiderivative, 205
 - arc length, 286
 - MATLAB, 289
 - parametric curve
 - MATLAB, 292
 - polar curve
 - MATLAB, 306
 - arc length function, 286
 - area
 - between two curves
 - general case, 265
 - bounded by two curves, 263
 - along the y -axis, 266
 - polar curve, 307
 - MATLAB, 310
 - polar region, 307
 - cancellation equations, 38
 - center of mass, 321
 - along a curve, 313
 - discrete system
 - 1d, 311
 - 2d, 312
 - parametric curve, 316
 - centroid, 320
 - of a curve
 - MATLAB, 315
 - parametric curve, 316
 - MATLAB, 319
 - planar region between two curves, 320
 - MATLAB, 323
 - chain rule, 68
 - Comparison Test, 367
 - for improper integrals, 253
 - composite function, 29
 - MATLAB, 29
 - concave down, 121
 - concave up, 121
 - concavity, 121

- concavity test, 121
- conditional convergence, 380
- constant function, 112
- continuity
 - composite functions, 30
 - elementary functions, 6, 46
 - inverse function, 39
- continuous function, 27
 - in an interval, 28
- continuous functions
 - extreme value theorem, 102
- convergence
 - improper integral
 - over infinite interval, 248
 - unbounded integrand, 250
 - indefinite integral, 205
 - infinite limit, 25
 - integral, 187
 - interior point, 99, 101
 - inverse function, 37
 - inverse trigonometric functions, 42
 - limit, 10
 - limit at infinities, 19
 - local extreme, 97
 - local maximum, 97
 - local minimum, 97
 - logarithmic function, 40
 - monotone, 107
 - odd function, 141
 - one-sided limits, 16
 - one-to-one, 37
 - periodic function, 141
 - polar coordinates, 155
 - polar curve, 155
 - series
 - absolute convergence, 380
 - conditional convergence, 380
 - convergence, 348
 - divergence, 348
 - the number e , 34
 - trigonometric functions, 5
 - vector
 - MATLAB, 433
- definite integral, 187
 - MATLAB, 196
 - polynomial
 - MATLAB, 197
- definition
 - absolute extreme, 97
 - absolute maximum, 97
 - absolute minimum, 97
 - composite function, 29
 - concave down, 121
 - concave up, 121
 - continuous function, 27
 - critical point, 100
 - definite integral, 187
 - derivative, 56
 - differential, 87
 - even function, 141
 - exponential function, 4
 - function, 1
 - higher derivatives, 90
 - improper integral
 - over infinite interval, 248
 - unbounded integrand, 250
 - indefinite integral, 205
 - infinite limit, 25
 - integral, 187
 - interior point, 99, 101
 - inverse function, 37
 - inverse trigonometric functions, 42
 - limit, 10
 - limit at infinities, 19
 - local extreme, 97
 - local maximum, 97
 - local minimum, 97
 - logarithmic function, 40
 - monotone, 107
 - odd function, 141
 - one-sided limits, 16
 - one-to-one, 37
 - periodic function, 141
 - polar coordinates, 155
 - polar curve, 155
 - series
 - absolute convergence, 380
 - conditional convergence, 380
 - convergence, 348
 - divergence, 348
 - the number e , 34
 - trigonometric functions, 5
 - vector
 - MATLAB, 433
- exponentiation, 4
- delta function, 12
- derivative, 56
 - arithmetic rules, 61
 - chain rule, 68
 - exponential functions, 75
 - Fermat's theorem, 99, 101
 - higher, 90
 - inverse function, 71
 - inverse trigonometric functions, 73
 - logarithmic functions, 75
 - MATLAB, 58
 - monotone test, 111
 - notations, 57

- trigonometric functions, 65
- derivative of inverse function, 71
- derivatives of exponential functions, 75
- derivatives of inverse trigonometric functions, 73
- derivatives of logarithmic functions, 75
- derivatives of trigonometric functions, 65
- determine the interval of convergence
 - MATLAB, 405
- determine the radius of convergence
 - MATLAB, 405
- differentiable, 56
- differential, 87
- differentiation of power series
 - MATLAB, 413
- Dirac delta function, 12
- discontinuous, 28
- divergent series, 348
- domain, 1
- error bound
 - midpoint rule, 239
 - Simpson's rule, 241
 - trapezoidal rule, 239
- Euler's formula, 7
- even function, 141
- exponential function
 - MATLAB, 34
- extreme point, 97
- extreme value theorem, 102
- Fermat's Theorem, 99, 101
- first derivative test, 114
 - MATLAB, 119
- formula
 - arc length, 286
 - parametric curve, 290
 - polar curve, 307
- area
 - bounded by two curves, 263
 - bounded by two polar curves, 307
- center of mass
 - along a curve, 313
 - parametric curve, 316
- centroid
 - parametric curve, 316
 - planar region between two curves, 320
- surface area of revolution, 295
- parametric curve, 300
- volume by cylindrical shells, 279
- function, 1
 - absolute extreme, 97
 - absolute maximum, 97
 - absolute minimum, 97
 - arc length, 286
 - composite
 - MATLAB, 29
 - composition, 29
 - concave down, 121
 - concave up, 121
 - continuous, 27
 - critical point, 100
 - decreasing, 107
 - differentiable, 56
 - Dirac delta, 12
 - discontinuous, 28
 - domain, 1
 - even, 141
 - exponential
 - MATLAB, 34
 - increasing, 107
 - inverse, 37
 - MATLAB, 39
 - inverse trigonometric, 42
 - local extreme, 97
 - local maximum, 97
 - local minimum, 97
 - logarithmic, 40
 - MATLAB, 42
 - monotone, 107
 - natural domain, 2
 - odd, 141
 - periodic, 141
 - range, 1
 - transcendental, 4
 - trigonometric, 5
 - MATLAB, 44
- Fundamental Theorem of Calculus, 200
- geometric sequences, 343
- geometric series, 348
- graph of the inverse function, 39
- growth rates of sequences, 341
- Guidelines for Curve Sketching, 142
- harmonic series

- alternating, 377
 - MATLAB, 378
- divergence, 356
 - by Integral Test, 360
- higher derivatives, 90
- horizontal asymptote, 20
- horizontal line test, 37
- implicit differentiation, 81
 - MATLAB, 82
- improper integral
 - over infinite interval, 248
 - convergence, 248
 - MATLAB, 253
 - unbounded integrand, 250
 - convergence, 250
 - MATLAB, 254
- increasing/decreasing test, 111
- indefinite integral, 205
 - linearity, 205
 - MATLAB, 209
 - some elementary functions, 206
- infinite limit, 25
- integrable, 187
- integral
 - lower limit, 187
 - numerical approximation
 - left point rule, 236
 - midpoint rule, 237
 - right point rule, 237
 - Simpson's rule, 240
 - trapezoidal rule, 237
 - upper limit, 187
- integral form of the remainder, 397
- Integral Test, 357
 - MATLAB, 363
 - remainder estimate, 365
- integrand, 187
- integrating products of $\sin mx$ and $\cos nx$, 223
- integrating products of powers of $\sin x$ and $\cos x$, 225
- integrating rational functions by partial fractions, 229
- integration by parts, 218, 226
- integration of power series
 - MATLAB, 414
- interior point, 99, 101
- Intermediate Value Theorem, 30
- inverse function, 37
 - cancellation equations, 38
 - derivative, 71
 - find, 38
 - graph, 39
 - MATLAB, 39
 - trigonometric
 - MATLAB, 44
- inverse trigonometric functions, 42
 - MATLAB, 44
- l'Hospital's Rule, 134
- Lagrange remainder, 398
- laws of exponents, 4
- laws of logarithms, 41
- left point rule, 236
 - MATLAB, 242
- left-hand limit, 16
 - MATLAB, 17
- Leibniz Formula, 93
- Leibniz Rule, 93
- limit, 10
 - $\lim_{x \rightarrow +\infty} \left(1 + \frac{1}{x}\right)^x = e$, 34
 - $\lim_{\theta \rightarrow 0} \frac{\sin \theta}{\theta} = 1$, 18
- arithmetic rules, 13
- at infinity, 20
- at negative infinity, 20
- at positive infinity, 19
- finite, 10
- infinite, 25
- one-sided, 16
- sequence, 336
 - convergence, 336
 - divergence, 336
 - from a function, 339
 - MATLAB, 347
- Limit Comparison Test, 368
- limit laws for sequences, 339
- limit of sequence, 336
- linearity
 - indefinite integral, 205
- linearization, 86
- local extreme, 97
- local maximum, 97
 - MATLAB, 103

local minimum, 97
MATLAB, 103
logarithmic differentiation, 77
logarithmic function, 40
MATLAB, 42
lower limit, 187

Maclaurin series, 394
common functions, 414
MATLAB
 `Inf`, 255
 `NaN`, 255
 `abs`, 357
 `assumeAlso`, 373
 `assume`, 373
 `cla`, 269
 `coeffs`, 417
 `collect`, 210
 `compose`, 30, 69
 `diff`, 58, 197
 `exp`, 34
 `factorial`, 373
 `finverse`, 39
 `format long`, 172
 `fplot`, 149
 `fzero`, 104
 `help`, 432
 `inf`, 21
 `integral`, 196, 253, 254, 304
 `int`, 196, 209, 253, 254
 `isAlways`, 378
 `length`, 417
 `limit`, 12
 `linspace`, 149, 434
 `log2`, `log10`, `log`, 42
 `matlabFunction`, 104, 304
 `partfrac`, 235
 `plot`, 9, 153, 267, 435
 `polarplot`, 160
 `polyint`, 197
 `pretty`, 210
 `quadgk`, 253, 254
 `rsums`, 195
 `simplify`, 373
 `solve`, 82
 `subplot`, 149
 `subs`, 58
 `symsum`, 353
 `taylor`, 394
 `trapz`, 244
 `viscircles`, 269
 `vpa`, 66
 absolute minimum, 105
 alternating harmonic series, 378
 Alternating Series Test, 378
 arc length, 289
 parametric curve, 292
 polar curve, 306
 area
 polar curve, 310
 under a parametric curve, 272
 centroid
 of a curve, 315
 parametric curve, 319
 planar region between two curves, 323
 chain rule, 69
 command help, 432
 command window, 427
 comment, 432
 composite function, 29
 concavity test, 122
 current folder, 428
 definite integral, 196
 polynomial, 197
 derivative, 58
 determine the interval of convergence,
 405
 determine the radius of convergence, 405
 differentiation of power series, 413
 exponential function, 34
 figure plot, 435
 harmonic series
 alternating, 378
 implicit differentiation, 82
 improper integral
 over infinite interval, 253
 unbounded integrand, 254
 increasing/decreasing test, 113
 indefinite integral, 209
 Integral Test, 363
 integration of power series, 414
 inverse function, 39
 inverse trigonometric functions, 44
 left point rule, 242
 left-hand limit, 17
 limit

- sequence, 347
- limits at infinities, 21
- local maximum, 103
- local minimum, 103
- logarithmic function, 42
- m-file, 429
- method
 - Simpson's rule, 244
 - midpoint rule, 243
 - monotone test, 113
 - Newton's iteration, 171
 - Newton's method, 171
 - one-sided limit, 17
 - parametric curves, 153
 - partial fractions for rational functions, 234
 - polar curves, 160
 - product of power series, 408
 - quotient of power series, 408
 - Ratio Test, 373
 - Riemann sum, 195
 - right point rule, 242
 - right-hand limit, 17
 - Root Test, 375
 - semicolon, 431
 - series
 - divergence, 357
 - single quote, 113
 - substitution, 58
 - surface area of revolution, 299
 - parametric curve, 303
 - symbolic function, 8
 - Taylor polynomials, 394
 - telescoping series, 353
 - trapezoidal rule, 244
 - trigonometric functions, 44
 - value assignment, 431
 - vector, 149
 - vector definition, 433
 - volume, 277
 - work, 329
 - workspace, 428
 - Mean Value Theorem, 108
 - method
 - determine the interval of convergence, 403
 - find the absolute extrema, 103
 - find the inverse function, 38
 - graph of the inverse function, 39
 - implicit differentiation, 81
 - integrating products of $\sin mx$ and $\cos nx$, 223
 - integrating products of powers of $\sin x$ and $\cos x$, 225
 - integrating rational functions by partial fractions, 229
 - logarithmic differentiation, 77
 - numerical integration, 236
 - Simpson's approximation, 239
 - Simpson's rule
 - MATLAB, 244
 - solving inequality, 112
 - trigonometric substitutions, 215
 - midpoint rule, 237
 - error bound, 239
 - MATLAB, 243
 - moment, 311, 312
 - about the x -axis, 312
 - about the y -axis, 312
 - about the origin, 311
 - monotone, 107
 - monotone test, 111
 - Monotonic Sequence Theorem, 345
 - natural domain, 2
 - Newton's iteration, 168
 - MATLAB, 171
 - Newton's Law of Cooling, 4
 - Newton's method, 168
 - MATLAB, 171
 - notations of derivative, 57
 - numerical integration
 - left point rule, 236
 - midpoint rule, 237
 - right point rule, 237
 - Simpson's rule, 240
 - trapezoidal rule, 237
 - odd function, 141
 - one-sided limit
 - MATLAB, 17
 - one-sided limits, 16
 - one-to-one, 37
 - parametric curve, 151
 - parametric curves
 - MATLAB, 153

- parametric equations, 151
partial fractions for rational functions
 MATLAB, 234
partial sum, 348
period, 141
periodic function, 141
pitfall
 arithmetic rules fail, 14
 continuous but not differentiable, 60
 critical points, 101
 improper integral, 252
 zero of the denominator \Leftrightarrow vertical asymptote, 26
polar coordinates, 155
polar curve, 155
 slope of the tangent, 158
polar curves
 MATLAB, 160
power series, 394
 centered at a , 394
product of power series
 MATLAB, 408
properties
 exponential functions, 33
 logarithmic functions, 41
quotient of power series
 MATLAB, 408
range, 1
Ratio Test, 370
 MATLAB, 373
reciprocal rule, 64
recursive formula, 220
related rates, 84
relationship between differentiability and continuity, 59
relationship between the limit and the one-sided limits, 16
remainder
 Taylor series, 394
 integral form, 397
 Lagrange, 397
remainder estimate
 Alternating Series Test, 378
 Integral Test, 365
 Taylor series, 398
Riemann integral, 187
Riemann sum, 187
 MATLAB, 195
right point rule, 237
 MATLAB, 242
right-hand limit, 16
 MATLAB, 17
rigorous definition of limit, 12
rigorousdefinitionoflimit, 12
Root Test, 374
 MATLAB, 375
rules of differentiation, 61
second derivative test, 117
 MATLAB, 120
sequential limit, 344
series
 absolute convergence, 380
 conditional convergence, 380
 convergence, 348
 divergence, 348
 MATLAB, 357
 partial sum, 348
sigma notation, 186
Simpson's rule, 240
 error bound, 241
slant asymptote, 22, 147
Squeeze Theorem, 14
 for sequences, 340
Stirling's formula, 342
substitution
 definite integral, 212
 indefinite integral, 210
summation notation, 186
surface area of revolution, 295
 general formula, 297
 MATLAB, 299
 parametric curve, 300
 MATLAB, 303
table
 indefinite integral, 206
Taylor expansion, 394
Taylor polynomials, 391
 asymptotic approximation, 392
 MATLAB, 394
 uniqueness, 391
Taylor series, 394
 convergence theorem, 395

- integral form of the remainder, 397
- Lagrange remainder, 398
- remainder, 394
- telescoping series, 350
 - MATLAB, 353
- test
 - concavity, 121
 - MATLAB, 122
 - first derivative, 114
 - MATLAB, 119
 - increasing/decreasing, 111
 - MATLAB, 113
 - monotone, 111
 - MATLAB, 113
 - second derivative, 117
 - MATLAB, 120
- the number e , 34
- theorem
 - Alternating Series Test
 - remainder estimate, 378
 - analytic function, 415
 - arithmetic rules of limits, 13
 - chain rule, 68
 - Comparison Test, 367
 - concavity test, 121
 - constant function, 112
 - continuity of basic elementary functions, 6, 46
 - continuity of composition of continuous functions, 30
 - continuity of the inverse function, 39
 - continuity with arithmetic operations, 28
 - derivatives of exponential functions, 75
 - derivatives of inverse trigonometric functions, 73
 - derivatives of logarithmic functions, 75
 - derivatives of trigonometric functions, 65
 - differentiation of inverse function, 71
 - extreme value theorem, 102
 - Fermat's Theorem, 99, 101
 - first derivative test, 114
 - Fundamental Theorem of Calculus, 200
 - geometric sequences, 343
 - growth rates of sequences, 341
 - horizontal line test, 37
 - integrable functions, 187
 - Integral Test, 357
 - remainder estimate, 365
 - integration by parts, 218, 226
 - intermediate value, 30
 - l'Hospital's Rule, 134
 - Leibniz Formula, 93
 - Limit Comparison Test, 368
 - Limit Laws for sequences, 339
 - linear approximation, 86
 - Mean Value, 108
 - monotone test, 111
 - monotonic sequence, 345
 - properties of exponential functions, 33
 - properties of logarithmic functions, 41
 - Ratio Test, 370
 - relationship between differentiability and continuity, 59
 - relationship between the limit and the one-sided limits, 16
 - Root Test, 374
 - rules of differentiation, 61
 - second derivative test, 117
 - sequential limit of continuous function, 344
 - squeeze, 14
 - Taylor series
 - convergence theorem, 395
 - integral form of the remainder, 397
 - Lagrange remainder, 398
 - theorem of analytic function, 415
 - transcendental function, 4
 - trapezoidal rule, 237
 - error bound, 239
 - MATLAB, 244
 - trigonometric functions, 5
 - MATLAB, 44
 - trigonometric substitutions, 215
 - unit-impulse signal, 12
 - upper limit, 187
 - vector
 - MATLAB, 149
 - vertical asymptote, 26
 - volume, 273
 - by cross sections, 273
 - by cylindrical shells, 279
 - MATLAB, 277
 - using shells, 279
 - using slices, 274

work, 325
MATLAB, 329

zero derivative
constant function, 112