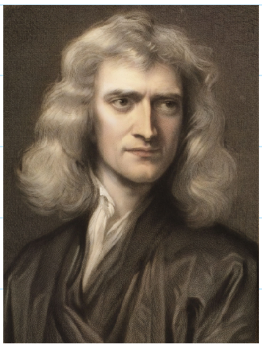
Concise Calculus

with analysis[1]

Jiangang Ying

[1] Copy or distribution is allowed only with agreement of the author.

The great mathematicians who built Calculus



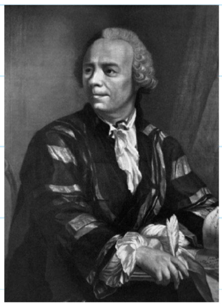
Newton

Leibniz

Fermat

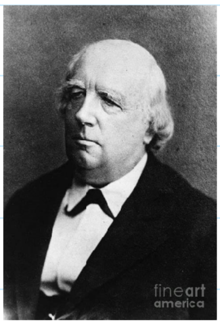
Jakob Bernoulli

Johanne Bernoulli



Euler

Cauchy



Weierstrass

Riemann

PREFACE

What is Calculus? The word Calculus comes from Latin meaning ‘small stone’, because it is like understanding something by looking at small pieces. Calculus includes differential Calculus, which cuts something into small pieces to find how it changes, and integral Calculus, which joins the small pieces together to find how much there is. Calculus is now a basic course for undergraduate students to learn how to analyze the relation between variables. When Calculus first appeared in late 17th century, it was not rigorous at all and more like a magic. After more than 100 years, we could now tell the story with complete logic. The great mathematicians in title page made the major contribution for what Calculus looks like today, though the list may be arguable.

Calculus contains not only a series of techniques how to analyze functions, but also the rigorous theory to explain why. The current textbook is rather compact for a one-semester course of Calculus which I have taught since 2019 for students in UIPE program, Fudan University, and covers usual contents of calculus with the related analysis. Though the analysis is not heavy, it is self-contained and almost all theorems inside are proved carefully, maybe not very rigorously. I hope that every notion in it is necessary and appears naturally. The mathematical rigor is not what it pursues but important for students to understand Calculus. The course is taught in a 16-week semester[2] and each week contains lectures of totally 150 minutes, three 50-minute or two 75-minute lectures, and a 45-minute problem session. The contents with (*) are the reading materials, which may or may not be taught in lectures. I would like to thank Dr Xiaodan Li, my TA for the great help giving to me during this course, and thank all students enrolled in this course for their curiosity and love for learning which push me to complete these notes.

Jiangang Ying
Fall semester, 2021

[2] a normal semester includes 16-week lecture and 2-week examination.

formulae in high school

Newton binomial expansion

$$(a+b)^n = \sum_{j=0}^{n} C_n^j a^j b^{n-j}.$$

sum of geometric series

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

mean-value inequality: for $a, b > 0$,

$$\sqrt{ab} \leqslant \frac{a+b}{2}.$$

triangular inequality: for real a, b,

$$|a+b| \leqslant |a| + |b|.$$

The key trigonometric formulae

$$\begin{aligned}
\sin(x+y) &= \sin x \cos y + \cos x \sin y, \\
\cos(x+y) &= \cos x \cos y - \sin x \sin y,
\end{aligned}$$

and the following follows from above two[3]

$$\begin{aligned}
&\sin^2 x + \cos^2 x = 1,\ \ 1 + \tan^2 x = \frac{1}{\cos^2 x}, \\
&\sin 2x = 2 \sin x \cos x,\ \ \cos 2x = \cos^2 x - \sin^2 x, \\
&\sin x + \sin y = 2 \sin \frac{x+y}{2} \cos \frac{x-y}{2} \\
&\sin x \cos y = \frac{1}{2}(\sin(x+y) + \sin(x-y)), \\
&\sin x \sin y = \frac{1}{2}(\cos(x-y) - \cos(x+y)).
\end{aligned}$$

[3]students should first complete the proof before going further.

Contents

1 numbers and sequences **1**

1.1 numbers and sets . . . 1

1.1.1 numbers . . . 1

1.1.2 definition, statement and proof . . . 2

1.1.3 sets and mappings . . . 4

1.2 functions . . . 6

1.2.1 functions and graphs . . . 6

1.2.2 properties of functions . . . 7

1.2.3 elementary functions . . . 7

1.2.4 inverse functions . . . 9

1.2.5 piece-wise defined functions . . . 10

1.2.6 curves and functions* . . . 11

1.3 sequences and limit . . . 12

1.3.1 sequences . . . 12

1.3.2 sequence limit . . . 13

1.4 properties of limit . . . 15

1.4.1 generic properties . . . 15

1.4.2 subsequences . . . 17

1.5 analysis 1: Euler's number . . . 19

1.5.1 monotone and bounded sequence . . . 19

1.5.2 Euler's number . . . 20

1.5.3 appendix: understanding the real power* . . . 23

2 function limit and continuity **27**

2.1 function limit: definition . . . 27

2.1.1 function limit as $x \to +\infty$ . . . 27

2.1.2 function limit as $x \to a$ . . . 29
2.2 function limit: operations . . . 32
2.2.1 generic properties of limit . . . 32
2.2.2 two important function limits . . . 34
2.2.3 infinitesimals . . . 35
2.2.4 one-side limit . . . 36
2.3 analysis 2: continuous functions . . . 37
2.3.1 continuity . . . 37
2.3.2 two important theorems for continuous functions . . . 38
2.3.3 nested intervals . . . 39
2.3.4 the least upper bound . . . 40
2.3.5 Bolzano-Weierstrass theorem . . . 40
2.3.6 the proof of Theorem 2.3.5 . . . 41

3 derivatives and application **43**
3.1 how to compute derivatives . . . 43
3.1.1 derivatives . . . 43
3.1.2 basic formulae for derivatives . . . 44
3.1.3 derivatives of inverse functions . . . 47
3.1.4 implicit functions and parametrized curves* . . . 49
3.2 application of derivatives 1 . . . 50
3.2.1 mean-value theorem . . . 50
3.2.2 monotonicity . . . 51
3.2.3 L'Hôpital's rule . . . 53
3.3 application of derivatives 2 . . . 54
3.3.1 max/min of a continuous function on a closed interval . . . 54
3.3.2 higher order derivatives and applications . . . 57
3.3.3 differentiation . . . 59

4 integral and application **61**
4.1 indefinite integrals 1 . . . 61
4.1.1 anti-derivatives . . . 61
4.1.2 basic integral formulae . . . 62
4.1.3 linearity of integration . . . 63
4.1.4 integration by parts . . . 64
4.2 indefinite integrals 2 . . . 65

4.2.1 change of variable . . . 65
4.2.2 rational functions . . . 67
4.3 indefinite integrals 3 . . . 70
4.3.1 trigonometric substitution . . . 70
4.3.2 integrating trigonometric functions . . . 72
4.3.3 strategy for integration . . . 72
4.4 Riemann integrals . . . 74
4.4.1 the area under curve and Riemann sums . . . 74
4.4.2 the limit of Riemann sum . . . 78
4.4.3 Newton-Leibniz formula . . . 79
4.4.4 integration by parts and substitution rule . . . 80
4.5 application of integrals . . . 82
4.5.1 area surrounded by curves . . . 82
4.5.2 idea of Riemann sum: cut and add . . . 83
4.5.3 volume of rotating body . . . 86
4.5.4 length of a curve . . . 87
4.6 improper integrals . . . 88
4.6.1 improper integrals . . . 88
4.6.2 absolute/conditional convergence . . . 92
4.7 analysis 3: Riemann integrability . . . 94
4.7.1 uniform continuity . . . 94
4.7.2 integrability of continuous functions . . . 95

5 series and Taylor expansions 98
5.1 series . . . 98
5.1.1 series . . . 98
5.1.2 series with positive terms . . . 99
5.1.3 absolute/conditional convergence . . . 102
5.2 regroup and rearrangement . . . 103
5.2.1 alternating series . . . 103
5.2.2 regroup of series . . . 104
5.2.3 rearrangement of series . . . 105
5.3 analysis 4: rearrangement of series and upper limit . . . 106
5.3.1 Riemann rearrangement theorem . . . 106
5.3.2 the upper limit . . . 108
5.3.3 root test and ratio test . . . 109

5.4 power series 110
5.4.1 power series 111
5.4.2 radius of convergence 112
5.4.3 power series expansion 114
5.5 Taylor expansions 115
5.5.1 term-by-term differentiation and integration 115
5.5.2 Taylor expansions 117
5.5.3 the expression of the remainder 118
5.6 function sequences 121
5.6.1 exchange of limit and integral 122
5.6.2 uniform convergence 123
5.7 analysis 5: uniform convergence 124
5.7.1 uniform convergence of power series 125
5.7.2 Abel's theorem and Leibniz series 126

6 Multi-variate calculus 128
6.1 functions and graphs 128
6.1.1 Euclidean distance on $\mathbf{R}^n$ 128
6.1.2 functions and graphs 130
6.2 limit and derivatives 134
6.2.1 function limit 134
6.2.2 partial derivatives 135
6.2.3 chain rule 137
6.2.4 differentiation* 138
6.3 extremum of multi-variate functions 139
6.3.1 local min/max 139
6.3.2 global max/min 142
6.4 multiple integrals 144
6.4.1 volume of solid under surface 144
6.4.2 iterated integrals 146
6.5 change of variables 151
6.5.1 change of variables 151
6.5.2 polar coordinates 153
6.5.3 Gaussian density 155
References 157

Chapter 1

numbers and sequences

1.1 numbers and sets

1.1.1 numbers

Let us start from numbers. We are all familiar with numbers, but we may not really know numbers. For different people, number is different. For a boy under 10, a number means an integer. For a high school student, a number usually means a fraction. For a math major student in university, a number is usually a real number or complex number or even more.

What is a real number at first? According to what we learned from high school, every real number can be expressed decimally, for example, $0.23, -0.31212\cdots$. In general

$$a = a_0.a_1a_2\cdots a_n\cdots = a_0 + 0.a_1a_2\cdots a_n\cdots ,$$

where $a_0 = [a]$ is the integer part of a (the biggest integer no bigger than a) and any a_j, $j \geqslant 1$, is an integer between 0 and 9, called a digit, the j-th digit after the decimal point. This is different from the usual expression when $a < 0$, for example, $-1.32 = -(1.32) = -1 - 0.32$ usually, while $-1.32 = -2 + 0.68$ here. It is known that this expression is indeed

$$a = a_0 + \frac{a_1}{10} + \frac{a_2}{10^2} + \frac{a_3}{10^3} + \cdots ,$$

which is in fact a series and will be discussed later. The expression is called finite if there exists $j \geqslant 1$ such that $a_k = 0$ for $k \geqslant j$, in which case we write $a = a_0.a_1\cdots a_j$. The decimal expression of a number is not always unique. For example

$$0.1 = 0.0999999\cdots ,$$

which we will prove in the lecture of series. Every number with finite (non-zero) digits has two types of expressions. For uniqueness, we always choose the expression with infinite digits in the sequel.

We should know the relationship between existence and expression. A real number exists intrinsically, not because it has decimal expression. Using expression to denote a number is easier to operate. It is similar to say that one exists intrinsically and is expressed by a name or ID number for convenience.

A number is simply a point in a real line with direction and the points 0,1 labeled, which is called the axis of numbers. Any point in the axis can be represented by the expression. How to position a number with a given expression on the line?

We start with positive integers in school, then zero, then all integers. Their positions on real line are clear. They can be marked by the unit, the distance between 0 and 1. In middle high school, we learned fractions, which are in the form of m/n where n, m are integers, with n being positive. Dividing $[0,1]$ into n pieces equally, $1/n$ is positioned. Then m/n can be easily positioned. A fraction is also called a **rational number**. In particular any integer is rational. All rational number are **dense** on real line namely there are rational numbers between any two different numbers a, b.

Are there real numbers other than rational numbers? Yes. We know that the decimal expression of a fraction must be finite or infinite but cyclic, where cyclic means that it will repeat starting from some digit. For example,

$$1/4 = 0.25, \;\; 1/7 = 0.142857142857 \cdots\cdots .$$

A general cyclic decimal expression is written into

$$0.a_1 \cdots a_k \overline{c_1 \cdots c_n},$$

where the digits $c_1 \cdots c_n$ will repeat forever. Conversely any real number with decimal expression finite or infinite but cyclic is a rational number.

1.1.2 definition, statement and proof

Before continuing, we need to say a few words about mathematics. For sciences, the nature determines true or false, but for mathematics, what determines true or false is mathematics itself, precisely the axioms and logic. Therefore we have to be rigorous on every step forward, otherwise mathematics has nothing to trust. Actually we have mentioned many concepts and names above, integers, real numbers, fractions, and rational numbers. Rigorously speaking they should be defined carefully before being used. In mathematics a concept, a terminology or a notation, when it first appears, must be unambiguously defined in a statement called

definition, in which every word must have been clearly defined. The notions in mathematics may take names from the real world, for example, rational, irrational, finite, bounded in the following lectures. A word in the real world may not be used so rigorously, but as long as it is used to name a notion in mathematics, its meaning is unambiguous. However this is not a rigorous textbook and some concepts are just mentioned casually. Students having doubt about these concepts may refer to other more rigorous textbooks.

Definition 1.1.1 Any real number which is not rational is called **irrational**.

Theorem 1.1.2 $\sqrt{2}$ is irrational.

This is a statement (or proposition, claim, assertion, etc), which states the relation between the defined concepts. It could be true or false. When you claim it is true or false, you always need to prove it by the previously proved facts. After it has been proved to be true, we could call it a theorem, if it is important, or a lemma, if it is not so important.

Proof. At first, we need to know the definition of the objects that the statement is concerned. What is $\sqrt{2}$? Actually $\sqrt{2}$ is a number whose square is 2, namely $(\sqrt{2})^2 = 2$. What is an irrational number? An irrational number is a number which is not rational or not a fraction. After we know what they are, we may start to prove. One way to prove it is to prove by contradiction. Suppose that $\sqrt{2}$ is not irrational, or it is rational. Then $\sqrt{2}$ may be written in fraction n/m, where m and n have no common factor. If it leads to a contradiction, it implies that our hypothesis that $\sqrt{2}$ is rational is not true.
Now $\sqrt{2} = n/m$ and we have $2 = n^2/m^2$ or $n^2 = 2m^2$. This implies that n is even or $n = 2k$. This gives $2k^2 = m^2$ which gives back m is even. Hence n and m are even and they have a common factor, which is a contradiction. This contradiction implies that the supposition '$\sqrt{2}$ is not irrational' is wrong, i.e., $\sqrt{2}$ is irrational. □

Hence a number is irrational if and only if its decimal expression is infinite and not cyclic. It is impossible to position [1] such a number in finite time and that's why we can only approach an irrational point but can not catch it precisely. This approaching process is exactly the limit theory we shall learn in this course.

Theorem 1.1.3 Every irrational number can be approximated by a sequence of rational numbers.

The theorem is intuitive but makes sense only after we know what it means by 'approximated by a sequence'. For example the circumference ratio $\pi = 3.1415926\cdots$ is a real number, and was proved to be irrational. People found 62831853071796 digits of π with help of computer

[1] assume that we can only position a multiple and an equal division of a given segment.

in 2021 but never its precise value. Hence π can be approximated by the sequence

$$3,\ 3.1,\ 3.14,\ 3.141,\ 3.1415,\ 3.14159, \cdots, \cdots.$$

1.1.3 sets and mappings

Next we introduce the language of calculus and mathematics: set and mapping, which the readers learned in high school and we briefly review here. A set is to put something specified together. Note that this is not really a definition. The notion of sets is the very base for all notions in mathematics, namely there is nothing more basic to define the notion of sets. However defining a concrete set amounts to telling unambiguously what elements the set contains. We use $A = \{\cdots, \cdots\}$ to list all elements in a set or clearly state what elements a set contains. For example

$$A = \{1, 2, 4, 8\},\ A = \{n : n \text{ is an even integer}\}$$

or A is the set of the real numbers satisfying $x^2 - 2x - 3 = 0$. Every element in a set can be counted only once, for example $A = \{1, 2, 3, 1\} = \{1, 2, 3\}$ and there are 3 elements in A. The way how the elements of a set are listed makes no difference, for example $A = \{1, 2, 3\} = \{3, 2, 1\}$.

Though the notion of sets was not clearly defined, what a set contains exactly must be clearly defined.

Definition 1.1.4 Assume that A is a set. We write $a \in A$ if a is an element of a set A, and $a \notin A$ otherwise. It should be unambiguous that either $a \in A$ or $a \notin A$. For sets A and B, we say that $A \subset B$ if any element in A belongs to B, and $A = B$ if they hold the same elements, namely $A \subset B$ and $B \subset A$.

Set Operations:

1. $A \cup B$, containing all elements either in A or in B;
2. $A \cap B$, containing all elements in A and also in B;
3. $A \setminus B$, containing all elements in A but not in B.

The empty set, denoted by $\emptyset$, is a set which contains nothing. Throughout this notes, let $\mathbf{N}$ be the set of natural numbers, or positive integers, $\mathbf{Z}$ the set of all integers, $\mathbf{Q}$ the set of rational numbers and $\mathbf{R}$ the set of real numbers. Hence $\mathbf{R} \setminus \mathbf{Q}$ is the set of all irrational numbers. For any numbers $a < b$, $(a, b) = \{x \in \mathbf{R} : a < x < b\}$, called open interval, $[a, b] = \{x \in \mathbf{R} : a \leqslant x \leqslant b\}$, called closed interval.

Definition 1.1.5 Assume that X, Y are two sets. A mapping $f : X \mapsto Y$ is a rule according to which any $x \in X$ corresponds a unique $y \in Y$. For $x \in X$, $f(x)$, the element that x corresponds, is called the image of x. The set of all images is denoted by $f(X)$, which is called the range of f and is a subset of Y.

It is very important that any $x \in X$ has a unique image $y \in Y$. For example, if everyone has only one name, the name of a person is a mapping from the set of human beings to the set of names (a name means a combination of letters).. But maybe many persons have the same name and some name may not be the name of any person. This gives another definition.

Definition 1.1.6 Assume $f : X \to Y$ is a mapping. If any two different elements in X have different images, f is called 1-1 or injective. If any element in Y is the image of some element in X, f is called onto or surjective. A 1-1 and onto mapping is called a one-to-one correspondence or bijective.

Example 1.1.1 $y = 2x$*:* $\mathbf{R} \to \mathbf{R}$ *is bijective.* $y = x^2$*:* $\mathbf{R} \to \mathbf{R}$ *is not 1-1 nor onto.* $y = x^2$*:* $\mathbf{R} \to [0, +\infty)$ *is onto but not 1-1.* $y = x^2 :\ [0, +\infty) \to \mathbf{R}$ *is 1-1 but not onto.* $y = x^2\ [0, +\infty) \to [0, +\infty)$ *is one-to-one correspondence.*

Assume that A, B, C are sets, $f : A \to B$ and $g : B \to C$. Define for any $a \in A$,

$$h(a) = g(f(a)) \in C.$$

Then h is a mapping from A to C, which is called the composition of f then g, denoted by $g \circ f$. The composition is an important operation of mappings.

Exercises

1. Write 1/13 into a decimal expression.

2. Prove that the set of rational numbers is dense, namely for any $a < b$, there is a rational number in (a, b).

3. Prove that if $a \in \mathbf{Q}$, $b \in \mathbf{R} \setminus \mathbf{Q}$ and $a \neq 0$, then $ab \in \mathbf{R} \setminus \mathbf{Q}$.

4. Prove that the set of irrational numbers is dense.

5. Prove that $\sqrt{5}$ is irrational.

6. Assume that A, B are two sets. Prove that

$$A \setminus B = A \setminus (A \cap B).$$

1.2 functions

In this lecture we mainly review what we have learned in high schools about functions.

1.2.1 functions and graphs

A function is to express how a variable depends on another variable. We have learned some functions in high school, for example, linear function $y = ax + b$, and quadratic function $y = ax^2 + bx + c$.

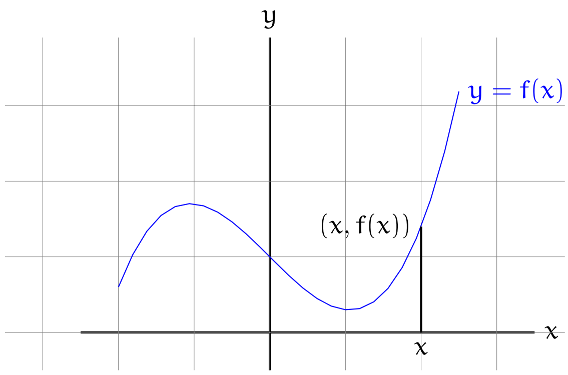
Calculus is mainly about the properties of functions. A function is a particular type of mapping and it can be used to model the relations of variables in real world. For example, the relation between speed and distance, the energy consumption and GDP growth.

Definition 1.2.1 A mapping $f : D \to \mathbf{R}$ is called a function with domain $D \subset \mathbf{R}$.

Roughly speaking a function describes quantitatively how one variable depends on another. A sequence is a special function when $D = \mathbf{N}$. A function is essentially the rule f (and D). For example, $y = f(x)$, $x \in D$, where we should indicate that y, the dependent variable, depends on x, the independent variable or equivalently any $x \in D$ corresponds a unique y. It does not matter which letters are chosen to express the rule. Actually $y = f(x)$, $x \in D$, where y depends on x, and $b = f(a)$, $a \in D$, where b depends on a, express the same function.

For example, the numbers 1,2,3, ..., which we learned in elementary school are abstract, but your math teacher may need to take something concrete to express them. They may use their fingers or use small stones or even draw circles on board.

A function describes the relation between two variables, which can be roughly expressed by its graph. Take a plane coordinate system with a horizontal axis and a vertical axis. A coordinate (a, b) denotes a point on the plane with a, b being horizontal and vertical coordinates. Conventionally, the independent variable varies on horizontal line and the dependent variable varies on vertical line, no matter what letters they use. The set $\{(x, f(x)) : x \in D\}$ is the graph of f, as follows, which may show us how y changes on x intuitively.



1.2.2 properties of functions

Analysing how a function varies is a major task in calculus for example, monotonicity, convexity, the max/min, and etc. We shall introduce boundedness and monotonicity here.

Definition 1.2.2 Given a function $f : D \longrightarrow \mathbf{R}$ and $A \subset D$, f is called bounded on A if there exists $C > 0$ such that for any $x \in A$, $|f(x)| \leqslant C$. If f is bounded on D, we simply say that f is bounded.

If $f(x) \leqslant C$ ($f(x) \geqslant C$) for any $x \in D$, we say f is bounded above (below) by C.

Example 1.2.1 *The boundedness depends on the rule f and also the domain. $y = \sin x$ is bounded. $y = x^2, x \in \mathbf{R}$ is not bounded, but $y = x^2, x \in [0, 1]$ is bounded.*

Definition 1.2.3 A function f on D is called increasing if for any $x_1, x_2 \in D$ with $x_1 < x_2$, $f(x_1) \leqslant f(x_2)$, and called strictly increasing if for any $x_1, x_2 \in D$ with $x_1 < x_2$, $f(x_1) < f(x_2)$.

The notions 'decreasing' and 'strictly decreasing' are defined similarly. A function which is either (resp. strictly) increasing or decreasing is called (resp. strictly) monotone.

Example 1.2.2 *$y = ax + b$ is increasing on $\mathbf{R}$ when $a > 0$. $y = x^2$ is not monotone for $x \in \mathbf{R}$, but it is increasing on $[0, +\infty)$ and decreasing on $(-\infty, 0]$. $y = \sin x$ is not monotone for $x \in \mathbf{R}$, but increasing on $[-\pi/2, \pi/2]$ and decreasing on $[\pi/2, 3\pi/2]$.*

The notion of even and odd function is known for us. A function $y = f(x)$, $x \in D = (-a, a)$ is even if $f(-x) = f(x)$ for any $x \in D$, and is odd if $f(-x) = -f(x)$ for any $x \in D$.
The domain of the functions above can be closed interval. Intuitively, the graph of an even function is symmetric about y-axis and that of an odd function is symmetric about the origin.

1.2.3 elementary functions

Recall 4 types of functions we have learned in high school:

1. power function $y = x^a$, where a is a constant.

2. exponential function $y = a^x$, where $a > 0, a \neq 1$.

3. logarithmic function $y = \log_a x$, where $a > 0, a \neq 1$.

4. trigonometric function $y = \sin x$, $y = \cos x$, $\cdots$.

We assume that students know their properties and graphs. These functions with their natural domain, where the function can be defined unambiguously, belong to a function family, called basic elementary functions. Be careful that the real power a^x, $x \in \mathbf{R}$ has not precisely defined in high school, so we shall discuss it later.

They are not enough for our modeling. How to obtain new functions? Usually by the four operations $(+, -, \times, \div)$ and compositions. The operation is known, for example, $y = x^{-1} + x$, $y = \frac{2^x}{\log x}$.

What is composition? Simply speaking, if y depends on u and u depends on x, then y depends on x, by composing them together. The composition of two functions $y = f(x)$, $x \in D_f$, and $y = g(x)$, $x \in D_g$, g then f, is a new function

$$y = f(g(x)),$$

with its natural domain. The rule of composition is denoted by $f \circ g$, i.e., $(f \circ g)(x) = f(g(x))$. This means that $x \in D_g$ corresponds a unique $g(x)$ by the rule g which in turn corresponds a unique $f(g(x))$ by the rule f, when $g(x) \in D_f$. The domain of $f \circ g$ is $\{x \in D_g : g(x) \in D_f\}$. Similarly, the composition, f then g, is $y = g(f(x))$ or $g \circ f$. It is known that $f \circ g$ differs from $g \circ f$, for example, $f(x) = \sin x$ and $g(x) = x^2$. The composition is a useful way to get a new function. We may even compose many times.

A function, which is obtained by a finite times of arithmetic operations, $+, -, \times, /$, called the four operations, and compositions, of basic elementary functions, is called **an elementary function**. Every elementary function has its natural domain. It is seen that as long as the class of basic elementary functions is determined, so is the class of elementary functions. In calculus, the family of elementary functions are mainly concerned.

Example 1.2.3 *We may write as many as you like.*

1. *A polynomial is a linear combination of integer power functions,*

$$y = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0,$$

which is called of degree n if the leading coefficient $a_n \neq 0$. The case $n = 1$ and $n = 2$ are linear and quadratic functions respectively.

2. $y = \sqrt{x + \sin x}$

3. $y = 2^{\sin x + \tan x / \ln x}$.

4. $y = \sqrt{1 + \sqrt{1 + \sin\sqrt{x}}}$.

5. $y = |x|$ *is also an elementary function, because* $|x| = \sqrt{x^2}$.

Remark. The function f is called a **linear combination** of $f_1, f_2, \cdots, f_n$ if there exist real numbers $a_1, a_2, \cdots, a_n$ such that for any $x \in D$, the common domain,

$$f(x) = a_1 f_1(x) + a_2 f_2(x) + \cdots + a_n f_n(x).$$

The operations of addition and scalar product (a constant multiple) are called **linear operation**.

1.2.4 inverse functions

We know in high school that $y = \sqrt{x}$ is the inverse of $y = x^2$, and logarithmic function $y = \log_2 x$ is the inverse of exponential function $y = 2^x$.

How to define an inverse function in general? A function $y = f(x)$ describes the rule f how y depends on x. When it is injective, x actually depends on y too. This inverse dependence defines inverse function.

Definition 1.2.4 If every $y \in E \subset \mathbf{R}$ determines a unique $x \in D \subset \mathbf{R}$, denoted by $x = f^{-1}(y)$, through the equation $y = f(x)$, then $x = f^{-1}(y),\ x \in E$ describes the rule f^{-1} how x depends on y, and is called the inverse function or simply inverse of f.

A function $y = f(x), x \in D$ is injective if and only if every horizontal line intersects the graph of f at most once. In particular, a strictly monotone function is injective. The expression $y = f(x)$ describes two different rules: (1) how y depends on x; (2) how x depends on y. The graph of $y = f^{-1}(x)$, $x \in f(D)$ is obtained by flipping the graph of $y = f(x)$ according to the line $y = x$, namely the graphs $y = f(x)$ and $y = f^{-1}(x)$ are symmetric with respect to the line $y = x$. It is easy to verify that f^{-1} and f have the same monotonicity when they are monotone.

Example 1.2.4 1. *The inverse function is obtained by solving equation* $y = f(x)$. *Solving* $y = 2x + 3$ *we get* $x = (y - 3)/2$.

2. $y = x^2$. *It is not a one-to-one correspondence from* $\mathbf{R}$ *to* $\mathbf{R}$*, but it is from* $[0, \infty)$ *to* $[0, +\infty)$ *or from* $(-\infty, 0]$ *to* $[0, +\infty)$. *Hence* $y = x^2$ *has no inverse function, but as a function on* $[0, +\infty)$ *its inverse is* $x = \sqrt{y}$, $y > 0$*, or as a function on* $(-\infty, 0]$ *its inverse is* $x = -\sqrt{y}$, $y > 0$.

3. *Solving* $y = a^x$ *we get* $x = \log_a y$*, which is a name for the unique solution. This actually means that we know a unique solution exists but we do not know what it is, so we name it as* $\log_a y$*.*

4. *The function* $f(x) = \begin{cases} x, & x \leqslant 0, \\ x^{-1}, & x > 0 \end{cases}$ *is injective but not monotone and* $f^{-1} = f$*.*

Exponential functions and logarithmic functions are mutually inverse. The inverse of a power function is still a power function. To inverse trigonometric functions, we will obtain the inverse functions $y = \arcsin x,\ x \in [-1, 1],\ y = \arccos x,\ x \in [-1, 1],\ y = \arctan x,\ x \in \mathbf{R}$. We shall put them into the class of basic elementary functions to make our class of elementary functions richer. Roughly speaking in our course we have the following classes of basic elementary functions

1. power functions $y = x^a$;
2. exponential functions $y = a^x$;
3. logarithmic functions $y = \log_a x$;
4. trigonometric functions $y = \sin x, \cos x, \tan x$;
5. inverse trigonometric functions $y = \arcsin x$, $\arctan x$.

Since $\mathrm{ctan} x = (\tan x)^{-1}$ and $\arccos x = \pi/2 - \arcsin x,\ x \in [-1, 1]$, we avoid using $\mathrm{ctan} x$ and $\arccos x$ to save notations.

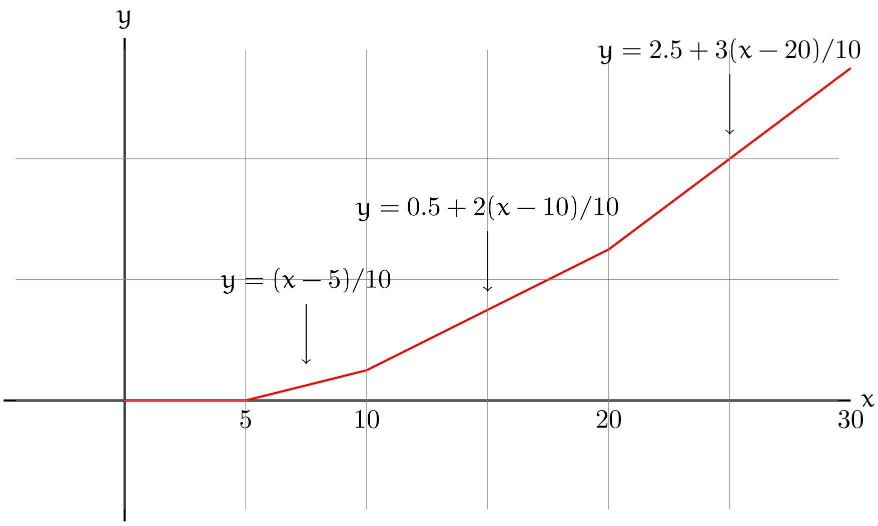
1.2.5 piece-wise defined functions

The class of elementary functions is very important for us, but we may construct functions ourselves. For example

$$f(x) = \begin{cases} x^2, & x > 0, \\ 1, & x = 0, \\ x - 1, & x < 0. \end{cases}$$

The function is obtained by putting different functions with disjoint domains together so it is called a piece-wise defined function. This approach gives us a lot more new functions which are no longer elementary functions.

Example 1.2.5 *The rate of income tax depends on the income. For example, the graph shows the dependence of* y*, the tax levied, on* x*, the annual income.*



Classification of functions is for convenience only and not based on rigorous definition. For example, the function

$$f(x) = \begin{cases} 3x, & x \geqslant 0; \\ x, & x < 0, \end{cases}$$

looks like a piece-wise defined function, but it can be expressed into $f(x) = |x| + 2x$, an elementary function. We do not care too much about if a function is elementary or not.

1.2.6 curves and functions*

The graph of a function on a plane with xy-coordinate system looks like a curve. Actually the notion 'function' is rigorously defined, while 'curve' is not. Roughly speaking, a curve is a set of points on plane which looks like a rope. In this sense, the graph of a function may be a broken rope and a 'continuous' function looks like a curve.

There are different ways to characterize a curve. An equation of x, y, for example,

$$x^2 + y^2 = 1$$

characterizes a circle, $x^2 - y^2 = 1$ characterizes a hyperbola. In general all points satisfying an equation, written as $F(x, y) = 0$, determines a curve.

A curve is not necessarily a graph of a function. It is easy to know that a curve is the graph of a function when every vertical line intersects the curve at most one point, which is called the vertical line test. If it is the case, then for any x on the x-axis, $x \in D$ when the vertical line and the curve have one point in common and the y-value of this point is defined to be $f(x)$. Hence this gives a function $y = f(x)$, $x \in D$. It is also seen that a piece on a curve which satisfies the vertical line test determines a function.

For example $x^2 + y^2 = 1$ is a circle, which is not a function because it does not satisfies the vertical line test. However if we cut the circle into two parts with x-axis, then both the upper and lower half, $y = \sqrt{1 - x^2}$ and $y = -\sqrt{1 - x^2}$ respectively, are functions. A function given this way is called an implicit function, which may or may not be able to write as $y = f(x)$ explicitly.

There are other ways to determine a curve. For example, parametrized curve

$$\begin{cases} x = f(t) \\ y = g(t) \end{cases}$$

also determines a curve, because each t in domain determines a point (x, y) in the plane. For example, $x = \cos t,\ y = \sin t$ denotes the unit circle.

Remark. Some particular functions appeared in history occasionally, but the general notion of functions was formulated first time in 18th century by L. Euler. In modern mathematics, a function is just a particular mapping.

Exercises

1. find the natural domain of following functions

$$(a)\ y = \sqrt{\sin x - 1}; \qquad (b)\ y = \sqrt{x^2 - 1};$$

$$(c)\ y = (x^2 - 1)^{-1}; \qquad (d)\ y = \frac{1}{1 + x^2}.$$

2. are the functions above bounded?

3. Using the monotonicity of $y = x^2$ to find the intervals where the functions in b,c,d of the problem 1 are monotone.

4. Assume that $y = f(x)$ and $y = g(x)$ are odd on $[-1, 1]$. Prove that $y = f(x)g(x)$ is even.

5. Define a function $y = f(x)$ on $\mathbf{R}$ such that it is even and $f(x) = (x - 1)^2$ for $x \geqslant 0$.

6. Find the inverse function for $y = e^x - 1$.

1.3 sequences and limit

1.3.1 sequences

Let's come to the sequence limit first. The major difference between high school math and college math is the difference between finite and infinite. We start now from the simplest infinite object: sequences.

Definition 1.3.1 A sequence is a mapping $a : \mathbf{N} \to \mathbf{R}$.

Conventionally write $a_n = a(n)$ for $n \in \mathbf{N}$ and to denote the sequence by (a_n), or simply a_n if no confusion is caused. We could indicate where a sequence starts if it is important, for example, $a_n, n \geqslant 0$. We have learned sequences in high school, for example, we know the arithmetic sequence $a_n = dn + c$, $n \geqslant 1$, where d, c are constant, and also the geometric sequence $a_n = ar^n$, $n \geqslant 1$, where a, r are constant. Here we are interested in how the sequence changes when n is large. Let us define two important properties for sequences.

Definition 1.3.2

1. (monotonicity) (a_n) is called increasing(or decreasing), if for any $n \geqslant 1$, $a_{n+1} \geqslant (\leqslant) a_n$. A sequence which is either increasing or decreasing is called monotone. If $\geqslant$ is changed to $>$ ($\leqslant$ to $<$), then we say (a_n) is strictly increasing (strictly decreasing).

2. (bounded) (a_n) is called bounded if there exists a constant C such that for any $n \geqslant 1$, $|a_n| \leqslant C$.

Read the definition of boundedness above, and it means that a constant C bounds all a_n. Image that if we change the order of words, say for any n there exists C such that $|a_n| \leqslant C$, then it means that one C bounds only one a_n, and is totally different from the original definition. Notice the difference between finite and bounded. The word 'finite' is applied to a number, for example, n is finite, while the word 'bounded' is applied to a set of numbers, for example, a sequence is bounded.

1.3.2 sequence limit

Definition 1.3.3 The sequence (a_n) converges to L, $a_n \to L$, or $\lim_n a_n = L$ if for any $\varepsilon > 0$, there exists N such that for any $n \geqslant N$, $|a_n - L| < \varepsilon$. Otherwise we say that (a_n) does not converge to L.

It is extremely important to read this definition 100 times carefully and to really understand every words. The sentence like "for any ..., there exists ... such that for any ..., something holds." will be used often later.

1. 'there exists N such that for $n > N$ ...' means that 'for n large enough' or 'starting from some n'. The existence of N, not the exact value, is important.

2. 'for any $\varepsilon > 0$, (important: it could be arbitrarily small), $|a_n - L| < \varepsilon$ for n large enough' means that when n is large enough, a_n will be fallen into the small interval $(L-\varepsilon, L+\varepsilon)$, which is called a neighborhood of L. It amounts to say that there are only finitely many a_n out of $(L - \varepsilon, L + \varepsilon)$ for any $\varepsilon > 0$.

3. for the inequality sign used here, ε has to be positive, but $n \geqslant N$ can be replaced with $n > N$, and $|a_n - L| < \varepsilon$ with $|a_n - L| \leqslant \varepsilon$.

4. It is seen that $a_n \to L$ is equivalent to $a_n - L \to 0$.

5. It is extremely important to negate a notion, for example, how to say that a sequence (a_n) does not converge to L. That (a_n) does not converges to L means that $\exists \varepsilon > 0$ so that for any N, there exists $n > N$ it holds that $|a_n - L| > \varepsilon$, or $\exists \varepsilon > 0$, so that $|a_n - L| > \varepsilon$ for infinitely many n.

Look at the definition of limit, we see that the inequality plays a major role and hence it is important to learn the techniques of inequality.

The limit of a sequence is a property of the sequence when n goes to infinity. It is nothing to do with the first finite many terms. It means that even if we throw away or change values of the first finite many a_n, it will not change the limit.

As seen in this lecture, we have defined several concepts: sequence, monotonicity, boundedness and convergence. It needs not to say that the only way to understand a concept is to read definition carefully. Actually I think that a major thing you should know about how to learn math is to read definition carefully.

Example 1.3.1 *1. Prove that* $\lim_{n\to+\infty} \frac{1}{n} = 0$. *For* $\varepsilon > 0$, *we may find an integer* $N \in \mathbf{N}$ *such that* $1/N < \varepsilon$, *for example,* $N > \varepsilon^{-1}$. *This simple fact is actually a fundamental result, called the axiom of Archimedes. When* $n > N$, *we have* $\frac{1}{n} < \frac{1}{N} < \varepsilon$. *By the definition,* $1/n$ *converges to* 0.

2. When $|r| < 1$, $\lim_n r^n = 0$. *In order to have* $|r^n| < \varepsilon$, *it is enough that*

$$n > \frac{\log_2 \varepsilon}{\log_2 |r|}.$$

We then take N *to be the least integer bigger than* $\log_2 \varepsilon / \log_2 |r|$.

3. Any real number can be approximated by a sequence of rational numbers.

The following theorem provides a useful way to prove the limit.

Theorem 1.3.4 If $|a_n - L| \leqslant b_n$ when n is large enough and $\lim_n b_n = 0$, then $\lim_n a_n = L$.

We are now in a position to define convergence.

Definition 1.3.5 The sequence (a_n) converges if there exists a number L such that a_n converges to L, and diverges otherwise.

We should also notice the difference between '(a_n) converges to L' and '(a_n) converges'. The former means we know the limit and the latter means there exists such a limit which we may not know.

Remark. People usually think that in the history of Calculus, the notion of limit was presented about 150 years after Calculus was presented. Actually in his book, the art of conjecturing, published in 1813, Jakob Bernoulli proved the famous law of large numbers rigorously by a rigorous ε-N approach. However no one in his era realized the importance of this notion so that the language of limit was reinvented more than 100 years later by Cauchy, Weierstrass and etc.

1.4 properties of limit

1.4.1 generic properties

The following theorem is important and shall be proved carefully. The similar result will appear whenever a notion of limit is defined.

Theorem 1.4.1 Assume that (a_n), (b_n) and (c_n) are sequences.

1. Uniqueness: The limit of a sequence is unique if exists.

2. Comparison: If $a_n \geqslant 0$ for large n and $\lim_n a_n = L$, then $L \geqslant 0$.

3. Sandwich: If (b_n) and (c_n) converge to the same limit L and for large n,

$$b_n \leqslant a_n \leqslant c_n,$$

then $\lim_n a_n = L$.

4. Boundedness: A convergent sequence is bounded.

5. four Operations: If the limits in the right side exist and the one appeared in denominator is not zero, then the limit on the left also exists and the identity holds.

   (a) $\lim_n(a_n + b_n) = \lim_n a_n + \lim_n b_n$;

   (b) $\lim_n(a_n b_n) = \lim_n a_n \cdot \lim_n b_n$;

   (c) $\lim_n(a_n/b_n) = \lim_n a_n / \lim_n b_n$.

Proof. 1. Suppose that (a_n) converges to L_1 and L_2. We need to verify they are the same. According to the definition of convergence, for any $\varepsilon > 0$, there exists N such that when $n > N$,

$$|a_n - L_1| < \varepsilon \text{ and } |a_n - L_2| < \varepsilon.$$

It concludes that picking such an n,

$$|L_1 - L_2| \leqslant |L_1 - a_n| + |a_n - L_2| < 2\varepsilon.$$

But $|L_1 - L_2|$ is a fixed number and 2ε can be arbitrarily small. This forces that $|L_1 - L_2| = 0$.
2. This is called the comparison theorem. As $L < 0$, $-L/2 > 0$. Taking this as ε, there exists N such that for $n > N$, $|a_n - L| < -L/2$, i.e., $3L/2 < a_n < L/2 < 0$. What we proved is more than the statement.
3. This result follows from the following inequality

$$\begin{aligned}|a_n - L| &\leqslant |a_n - b_n| + |b_n - L| \\ &\leqslant |c_n - b_n| + |b_n - L| \leqslant |c_n - L| + 2|b_n - L|.\end{aligned}$$

4. Assume that $\lim_n a_n = L$. There exists N such that $|a_n - L| < 1$. It follows that

$$|a_n| \leqslant |a_n - L| + |L| \leqslant 1 + |L|.$$

Then we may find a constant C which bounds (a_n).
5. Assume $\lim_n a_n = A$ and $\lim_n b_n = B$. We use Theorem 1.3.4.
(a) It is easily seen that

$$0 \leqslant |a_n + b_n - (A + B)| \leqslant |a_n - A| + |b_n - B|.$$

(b) It is seen that for n large

$$|a_n b_n - AB| \leqslant |a_n||b_n - B| + |B||a_n - A| \leqslant (|A| + 1)|b_n - B| + |B||a_n - A|.$$

(c) Assume $B > 0$. Similar to the proof of 2, we know that $b_n > B/2 > 0$ for large n and hence

$$\begin{aligned}&\left|\frac{a_n}{b_n} - \frac{A}{B}\right| = \left|\frac{a_n B - b_n A}{b_n B}\right| \\ &\leqslant \frac{|a_n - A||B| + |b_n - B||A|}{|b_n|B} \\ &\leqslant \frac{|a_n - A||B| + |b_n - B||A|}{B^2/2}.\end{aligned}$$

□

Example 1.4.1 *1. For $a > 0$, prove that $\lim_{n\to\infty} \sqrt[n]{a} = 1$. Assume that $a > 1$. Let $a_n = \sqrt[n]{a} - 1$. Then it suffices to show a_n converges to 0. In fact $a_n > 0$ and by Newton's binomial formula,*

$$a = (1 + a_n)^n = 1 + na_n + \cdots > na_n.$$

Hence

$$0 < a_n < \frac{a}{n}$$

and by the sandwich theorem, $\lim a_n = 0$.

2. *Find limit* $\lim_n \frac{n^2-1}{2n^2+n-1}$. *Divide both numerator and denominator by* n^2,
$$\frac{n^2-1}{2n^2+n-1} = \frac{1-\frac{1}{n^2}}{2+\frac{1}{n}-\frac{1}{n^2}},$$
and it is seen that the limit of numerator and denominator are 1 and 2 respectively. Hence the limit is $1/2$.

3. *Prove that* $\lim_n \frac{n}{a^n} = 0$ *for* $a > 1$. *Write* $a = 1 + b$ *with* $b > 0$. *Then*
$$a^n = (1+b)^n > C_n^2 b^2 = \frac{n^2-n}{2} b^2.$$
Hence
$$0 < \frac{n}{a^n} < \frac{n}{n(n-1)b^2/2} = \frac{2}{(n-1)b^2},$$
which converges to zero.

Remark. How to understand a statement?

1. Many statements above are in the form of 'if A, then B.', where A and B are statements. This statement means that A implies B, or A is a sufficient condition for B and B is a necessary condition for A. Its inverse statement is 'if B, then A', negative statement is 'if not A, then not B' and inverse-negative statement is 'if not B, then not A'. The truth of statement 'if A then B' is equivalent to the truth of its inverse-negative statement 'if not B then not A'. Similarly the truth of inverse statement 'if B then A' is equivalent to the truth of negative statement 'if not A then not B', because it is the inverse-negative of the inverse.

2. The comparison theorem says that when $a_n \geqslant 0$ for large n, $\lim a_n \geqslant 0$ if exists. Even when $a_n > 0$ for large n, the conclusion that $\lim a_n \geqslant 0$ is still true, but not necessarily $\lim a_n > 0$. For example $a_n = 1/n$. The example $a_n = -1/n$ shows that the inverse of the comparison theorem is not true.

3. Intuitively 'if A then B' is denoted by $A \Rightarrow B$, and we say that A is stronger than B. More intuitively it can be compare to the inequality $A \geqslant B$. In this sense, the inverse is $B \geqslant A$, the negative is $-A \geqslant -B$, and the inverse-negative is $-B \geqslant -A$. If $A \geqslant B$ and $B \geqslant A$, then they are equivalent: $A = B$. Transitivity: if $A \geqslant B$ and $B \geqslant C$, then $A \geqslant C$.

1.4.2 subsequences

Definition 1.4.2 Assume that a sequence (a_n) is the function $f : \mathbf{N} \to \mathbf{R}$. If D is an infinite subset of $\mathbf{N}$, then a subsequence of (a_n) is the function $f : D \to \mathbf{R}$.

A subsequence is still a sequence. To pick a subsequence from a sequence (a_n), we usually pick a subsequence (k_n) from $\mathbf{N}$, or an infinite subset of $\mathbf{N}$, and obtain a subsequence (a_{k_n}). For example, given a sequence $a_n = (-1)^n$, all even terms form a subsequence $a_{2n} = (-1)^{2n} = 1$ and all odd terms forms a subsequence too. You could have any way you like to pick a subsequence, but must be infinitely many.

The following theorem is obvious by reading the definition of convergence.

Theorem 1.4.3 If the sequence (a_n) converges to L, then every subsequence converges to L.

This theorem can be used to prove a sequence does not converge, for example, $a_n = (-1)^n$, which contains two subsequences converging to different limits.

If the sequence (a_n) converges to L and there are infinitely many a_n which are non-negative, then they form a non-negative subsequence which converges to L and it follows that $L \geqslant 0$. Note the difference between '$a_n \geqslant 0$ for large n' and 'there are infinitely many $a_n \geqslant 0$'.

A sequence could go to infinity, for example, $a_n = n$ or $a_n = (-1)^n n$.

Definition 1.4.4 A sequence a_n goes to $+\infty$ if for any $C > 0$ there exists N such that $a_n > C$ for $n > N$. A sequence a_n goes to $-\infty$ if $-a_n$ goes to $+\infty$. A sequence a_n goes to ∞ if $|a_n|$ goes to $+\infty$.

Remark. Read the definition of convergence and going to infinity carefully. The first sentence of convergence is for any $\varepsilon > 0$ and the first sentence of going to infinity is for any $C > 0$. They look the same, but different very much. For convergence, the important point is small ε, because it illustrates that a_n is near the limit, while for going to infinity, the important point is big C, because it illustrates that a_n could be arbitrarily large.

Clearly $a_n = n$ goes to $+\infty$, $a_n = (-1)^n n$ goes to ∞. The following theorem is obvious.

Theorem 1.4.5 A sequence a_n goes to ∞ if and only if a_n^{-1} converges to 0.

According to the definition of boundedness, that a sequence (a_n) is unbounded means that for any C there exists n such that $|a_n| > C$. Therefore (a_n) is unbounded if and only if a subsequence of (a_n) goes to ∞. For example, the sequence $a_n = n^{(-1)^n}$ is unbounded but does not go to ∞.

Lemma 1.4.6 A sequence (a_n) is unbounded if and only if there is a sub-sequence (a_{k_n}) which goes to infinity.

Proof. The 'if' part is easy. We prove the 'only if' part. Assume (a_n) is unbounded, i.e., (a_n) is not bounded. Read the definition. This means that any $C > 0$ could not bound (a_n). Any integer n can't bound (a_n), i.e., $\exists k_n$ so that $|a_{k_n}| > n$. This gives a subsequence (a_{k_n}) which goes to infinity. $\square$

Exercises

1. Prove that the arithmetic sequence $a_n = an+b$ increases when $a > 0$ and the geometric sequence $a_n = ar^n$ increases when $a > 0$ and $r > 1$.
2. Prove that the sequence $a_n = an + b$ is bounded only if $a = 0$ and the sequence $a_n = ar^n$ is bounded if and only if $a = 0$ or $|r| \leqslant 1$.
3. Prove that if a_n converges then it is bounded.
4. Does the sequence $a_n = \sqrt[n]{n}$ converge? If so, what is the limit?
5. Prove that $\lim r^n = 0$ when $|r| < 1$.
6. How to state that (a_n) does not converge by definition?
7. How to state that (a_n) is not bounded?
8. Assume (a_n) is an integer sequence and converges to L. Prove that L is an integer and $a_n = L$ for large n.

1.5 analysis 1: Euler's number

1.5.1 monotone and bounded sequence

Here comes the first legendary theorem.

Theorem 1.5.1 A monotone and bounded sequence converges.

To prove this, we need to prepare a lemma.

Lemma 1.5.2 An increasing and bounded sequence of integers (a_n) no longer increases for large n.

For any $n \geqslant 1$, $a_n \leqslant a_{n+1}$. If $a_n < a_{n+1}$ some n, we say the sequence has an increment. That (a_n) is an integer sequence implies that that each increment $a_{n+1} - a_n \geqslant 1$. Since (a_n) is bounded, it has at most finitely many increments and this means that it no longer increases for large n.

Proof. Assume that (a_n) is a bounded and increasing sequence. Then they can be decimally expressed as

$$\begin{aligned}
a_1 &= x_0^{(1)} + 0.x_1^{(1)}x_2^{(1)}\cdots;\\
a_2 &= x_0^{(2)} + 0.x_1^{(2)}x_2^{(2)}\cdots;\\
&\cdots\cdots\cdots\\
a_n &= x_0^{(n)} + 0.x_1^{(n)}x_2^{(n)}\cdots;\\
&\cdots\cdots\cdots.
\end{aligned}$$

Intuitively the limit should be the 'biggest' number, which may not be in the sequence. How to find it? The integer sequence $(x_0^{(n)})$ is increasing and bounded. $x_0^{(n)}$ must be equal to some integer x_0 from some n, say n_0, and never increases anymore. Then we write down x, put a point on the right as above and throw away the terms before n_0. Observe now $x_1^{(n)}$. This sequence must also increase and by the same reason must be equal to some integer x_1 from some n, say n_1, and never increases anymore. Then we write down x_1 as $x_0 + 0.x_1$ and throw away the terms before n_1. Go on and on. In this way, we will pick out $x_2, x_3, \cdots$ and form a real number

$$L := x_0 + 0.x_1x_2x_3\cdots.$$

Finally it is easy to verify that a_n converges to L. □

Note that L is obtained digit by digit. A simpler method to obtain the first k digits $x_0.x_1\cdots x_k$ is to realize that the limit of the sequence $[10^k \cdot a_n]$, $n \geqslant 1$, is $10^k(x_0 + 0.x_1\cdots x_k)$.

1.5.2 Euler's number

All of previous preparation are used to analyse and prove the existence of the most important limit.

Theorem 1.5.3 Compound interest formula

$$a_n = \left(1 + \frac{1}{n}\right)^n.$$

This sequence is increasing and bounded. It converges and the limit is a real number, denoted by e, called Euler's number.

We all know interest. When I save money in bank, I get interest. If I save \$1 in bank and get \$$1 + r$ back in one year, the value r is usually positive and called annual interest rate. But to maximize my interest, I can compound the interest. Precisely take it out in half a year, get half year interest \$$1 + r/2$ back and put it in bank. Then in the end of the year, I will get totally $(1+r/2)^2$ back. People intuitively believe that $(1+r/2)^2$ is more than $1+r$. Yes it is true because

$$(1+r/2)^2 = 1+r+r^2/4 > 1+r.$$

This gives us reason to 'take out and save again' money more often, called compound interest. If I do it every 3 months, I will get totally $(1+r/4)^4$ back. If I do it monthly, I will get totally $(1+r/12)^{12}$ back. If I am more extreme and do it daily, I will get totally

$$(1+r/365)^{365}$$

back. Theoretically If I do it n times in a year, I will get totally

$$(1+r/n)^n.$$

A question may arise naturally. Will this action increase the interest a lot more? We shall tell you that this action will increase the interest but not a lot more.

Proof. Using the Newton's binomial formula, we have

$$a_n = \left(1+\frac{1}{n}\right)^n = 1+1+\cdots+C_n^k(1/n)^k+\cdots+(1/n)^n.$$

The right side has totally $n+1$ terms, and write

$$b_{k,n} := C_n^k(1/n)^k,\ \ k=0,1,\cdots,n.$$

Then $b_{0,n}=b_{1,n}=1$ and when $k>1$,

$$\begin{aligned} b_{k,n} &= \frac{n!}{k!(n-k)!}\frac{1}{n^k} = \frac{n(n-1)\cdots(n-k+1)}{k!\cdot n^k} \\ &= \frac{1}{k!}\frac{n-1}{n}\frac{n-2}{n}\cdots\frac{n-k+1}{n}. \end{aligned}$$

Now

$$a_{n+1} = \left(1+\frac{1}{n+1}\right)^{n+1} = \sum_{k=0}^{n+1} b_{k,n+1}$$

has $n+2$ terms. When $k>1$,

$$\begin{aligned} b_{k,n+1} :&= \frac{1}{k!}\frac{n}{n+1}\frac{n-1}{n+1}\cdots\frac{n+1-k+1}{n+1} \\ &> b_{k,n} = \frac{1}{k!}\frac{n-1}{n}\frac{n-2}{n}\cdots\frac{n-k+1}{n}. \end{aligned}$$

Comparing then a_n with a_{n+1} term by term

$$\begin{array}{llllllll} a_n = & b_{0,n} & +b_{1,n} & +\cdots & +b_{k,n} & +\cdots & +b_{n,n} & \\ a_{n+1} = & b_{0,n+1} & +b_{1,n+1} & +\cdots & +b_{k,n+1} & +\cdots & +b_{n,n+1} & +b_{n+1,n+1} \end{array}$$

we conclude that $a_{n+1} > a_n$, namely the sequence increases. This means that the more often it is compounded, the more interest we would get.
However it is evident that for any n, since $\frac{n-k+1}{n} < 1$, when $k \leqslant n$,

$$b_{k,n} \leqslant \frac{1}{k!} = \frac{1}{1 \cdot 2 \cdot 3 \cdots k} \leqslant \frac{1}{2^{k-1}}.$$

Hence

$$\left(1+\frac{1}{n}\right)^n \leqslant 1+1+\cdots+\frac{1}{2^{k-1}}+\cdots < 1+1+1 = 3.$$

This means that I will not get a lot more. When $r = 1$ (the real world interest rate is about 5%) and I save \$1, I will get \$2 back without compound and will get no more than \$3 no matter how many times of compounds. □

Remark. The number e is called the Euler's number which is the most important number in Calculus. However, before Euler, some mathematicians, such as J. Napier, G. Leibniz, and Jakob Bernoulli, also made their contribution for the appearance of this number.

Exercises

1. Find

$$\lim\left(1+\frac{1}{n}\right)^{n+1}, \ \lim\left(1-\frac{1}{n}\right)^n.$$

2. Prove that for any $n > 1$,

$$\left(1+\frac{1}{n}\right)^n < \sum_{j=0}^{n} \frac{1}{j!} < e,$$

where $0! = 1$.

3. Please give an algorithm to approximate e so that the first 4 digits after the decimal point are correct.

4. Prove that the sequence $a_n = \sum_{j=1}^{n} \frac{1}{j^2}$ converges.

5. Prove

$$a_n = \left(1+\frac{1}{n^2}\right)^{n^2} \text{ and } b_n = \left(1+\frac{1}{n^2}\right)^n$$

converge and find their limits.

hint: Since for $0 \leqslant j \leqslant n^2$,

$$\left(1+\frac{1}{n^2}\right)^j \leqslant \left(1+\frac{1}{n^2}\right)^{n^2} < 3,$$

we have

$$0 < b_n - 1 < \frac{1}{n^2} \sum_{j=1}^{n} \left(1 + \frac{1}{n^2}\right)^j < \frac{3n}{n^2}.$$

6. Logic deduction is not all about mathematics. Observation is important as well. The Fibonacci's sequence $(a_n : n \geqslant 1)$ is defined recursively by

$$a_{n+2} = a_{n+1} + a_n, \ n \geqslant 1,$$

with the initial values $a_2 = a_1 = 1$. Observe the sequence $b_n = a_{n+1}/a_n$ for $n = 1, 2, 3, 4, 5, 6, 7, \cdots$. Write down and prove your observations. Finally prove that b_n converges and find the limit.

7. prove for any positive integer n and real number $0 < x < 1$, that $(1-x)^n \geqslant 1 - nx$. Now there is a clever idea to prove $a_n \leqslant a_{n+1}$,

$$\left(1 + \frac{1}{n}\right)^{-1} = 1 - \frac{1}{n+1} \leqslant \left(1 - \frac{1}{(n+1)^2}\right)^{n+1} = \left(\left(1 - \frac{1}{n+1}\right)\left(1 + \frac{1}{n+1}\right)\right)^{n+1}.$$

1.5.3 appendix: understanding the real power*

It is important to understand the relation $c = a^b$ for $a, b, c \in \mathbf{R}$, which defines 3 classes of functions: (1) power function, c depends on a; (2) exponential function, c depends on b; (3) logarithmic function, b depends on c.

The real power a^x, a raised to a real number x, when well-defined, should obey the following operational rules

1. $a^{x_1+x_2} = a^{x_1} a^{x_2}$;

2. $(a^{x_1})^{x_2} = a^{x_1 x_2}$;

3. $(ab)^x = a^x b^x$.

How to understand a^b? We define it step-by-step. When b is a positive integer, a^b is the product of a number of a's, for example, $a^2 = a \cdot a$. Then $a^0 = 1$ when b is zero and $a^b = (a^{-1})^{-b}$ when b is a negative integer. Later in middle high school we learned square roots, $\sqrt{a}$, and also $a^{1/b}$. However it could not be very clearly explained in high school how to define a^b when b is not a rational. Let's review this procedure.

When b is an integer and $a \neq 0$, a^b is clear. We recall the properties of power operation:

$$a^{b+c} = a^b \cdot a^c, \ a^0 = 1, \tag{1.1}$$

where b, c are integers.

Next we need to define a^b when b is a rational. This can be done as long as $a^{1/b}$ is defined, when $b \in \mathbf{N}$. It is not so trivial.

Lemma 1.5.4 When $a > 0$ and $b \in \mathbf{N}$, there is a unique positive number x satisfying the equation $x^b = a$.

The solution x is called the b-th root of a, and denoted by $a^{1/b}$. This lemma could't be proved in high school, not even now, but will be proved later by the continuity of function. From this example, we see that a notion can be defined as long as it is uniquely determined. Then we define

$$a^{m/n} := (a^{1/n})^m = (a^m)^{1/n},$$

and this defines a^x for $x \in \mathbf{Q}$, which satisfies the power property (1.1). Let's prepare a lemma for later use.

Lemma 1.5.5 Assume $a > 1$.

1. $a^x > 1$ for $x > 0$ and $x \in \mathbf{Q}$.

2. If $(x_n) \subset \mathbf{Q}$ and $\lim x_n = 0$, then $\lim a^{x_n} = 1$.

Proof. 1. It is easy to see that $a^n > 1$ for any $n \in \mathbf{N}$. It follows that $a^{1/n} > 1$ for any $n \in \mathbf{N}$. Hence for $x = m/n$ with $m, n \in \mathbf{N}$,

$$a^{m/n} = (a^m)^{1/n} > 1.$$

2. It is easy to verify that for any $z \in \mathbf{Q}$,

$$|a^z - 1| \leqslant a^{|z|} - 1.$$

Since $\lim_n a^{1/n} = 1$, for any $\varepsilon > 0$, $\exists N \in \mathbf{N}$ such that $0 < a^{1/N} - 1 < \varepsilon$. When n is large, $|x_n| < 1/N$ and then $|a^{x_n} - 1| \leqslant |a^{|x_n|} - 1| < |a^{1/N} - 1 < \varepsilon$. □

The last step is to define a^x for $a > 0$ and $x \in \mathbf{R}$, which still satisfies the power property (1.1).

Theorem 1.5.6 Assume $a > 1$ without loss of generality.

1. for any $x \in \mathbf{R}$, and any sequence $(x_n) \subset \mathbf{Q}$ increasing and converging to x, $\lim_n a^{x_n}$ does not depend on the choice of (x_n). Define then $a^x = \lim_n a^{x_n}$.

2. $a^x > 1$ when $x > 0$.

3. $y = a^x$ is strictly increasing.

Proof. 1. Therefore, by Lemma 1.5.5-1, (a^{x_n}) is increasing and bounded. This implies that it converges to a limit L. To prove that L does not depend on the choice of the sequence

(x_n), only on a and x, assume that (y_n) is another rational sequence increasing to x. Then $(x_n - y_n)$ is a rational sequence having limit zero, and it follows that

$$\lim_n a^{x_n - y_n} = 1 \text{ or } \lim_n a^{x_n} = \lim_n a^{y_n}.$$

Why do we need to prove this? in order to make $a^x = \lim_n a^{x_n}$ well-defined or uniquely defined.

2. When x is a positive real number, there exists $N \in \mathbf{N}$ such $1/N < x$. Then for n large, $x_n > 1/N$ by the comparison theorem. This implies that $a^x = \lim_n a^{x_n} \geqslant a^{1/N} > 1$.

3. for any $x_1, x_2 \in \mathbf{R}$ with $x_2 > x_1$, it follows from 3 that

$$a^{x_2} - a^{x_1} = a^{x_1}(a^{x_2 - x_1} - 1) > 0.$$

   It follows that $y = a^x$, $x \in \mathbf{R}$, is strictly increasing.

This completes the proof. □

When $0 < a < 1$, $y = a^x = (a^{-1})^{-x}$ is strictly decreasing. A function f has the same monotonicity as its inverse function and hence $y = \log_a x$ is strictly increasing (decreasing) when $a > 1$ $(0 < a < 1)$.

From the beginning to now, the readers may ask why we keep proving some results which look so obvious. I would say that every step in mathematics should be rigorously based on reason, and on the proven facts, not on the facts in our minds, where there might be many facts we know but do not know why. This makes mathematics convincing.

We may now consider the relation $x^y = z$. This relation gives birth to three types of basic elementary functions: exponential, power and logarithmic functions. When x is fixed, z can be viewed as a function of y, which is an exponential function, conversely y can be viewed as a function of z, which is a logarithmic function. When y is fixed, z can be viewed as a function of x, which is a power function.

When $a \in \mathbf{R}$, $y = x^a$ is a power function. Its domain depends on a. For example, the domain of $y = x^2$ is $\mathbf{R}$ and the domain $y = \sqrt{x} = x^{1/2}$ is $x \geqslant 0$. No matter what a is, the domain always contain $(0, +\infty)$.

Lemma 1.5.7 For $b > 0$, the power function $y = x^b$, $x > 0$ is strictly increasing.

It is easy to prove by Theorem 1.5.6-3. In fact, take $x_2 > x_1 > 0$, we have

$$\frac{x_2^b}{x_1^b} = \left(\frac{x_2}{x_1}\right)^b > 1,$$

and hence $x_2^b > x_1^b$.

It can be seen from this lemma that the solution $x^b = a$ is unique if exists, when $a > 0$, and $b \in \mathbf{N}$.

Exercises

1. Assume that $a > 1$. Prove that for any $z \in \mathbf{R}$,
$$|a^z - 1| \leqslant a^{|z|} - 1.$$

2. Assume that $y = f(x)$ is a polynomial and a is a root, i.e., $f(a) = 0$. Prove that there exists a polynomial $y = g(x)$ such that $f(x) = (x - a)g(x)$.

3. How to get a symmetric curve? (1) Find the coordinate of the point symmetric to (x_0, y_0) about the straight line $y = kx + a$. (2) Find the equation describing the curve symmetric to $y = x^2$ about the same line.

4. Try to draw roughly the curve $x^2 + y^3 = 1$. Is y a function of x? How about the curve $x^3 + y^2 = 1$?

Chapter 2

function limit and continuity

2.1 function limit: definition

2.1.1 function limit as $x \to +\infty$

Let us define the function limit when $x \to +\infty$. What is limit? Limit is like the tendency that a function changes. We may recall the tendency of basic elementary functions when $x \to +\infty$. For example, when $a > 0$, x^a increases to infinity, and when $a < 0$, x^a decreases to 0; when $a > 1$, a^x increases to infinity, and when $a < 1$, a^x decreases to 0; and $\sin x$ oscillates.

Definition 2.1.1 Assume that the function $y = f(x)$ is defined for large x, or with domain $(A, +\infty)$ for some A. We say that $f(x)$ converges to L, $f(x) \to L$ as $x \to +\infty$, or $\lim_{x\to+\infty} f(x) = L$, if for any $\varepsilon > 0$, there exists $N > 0$ such that for any $x \geqslant N$, $|f(x) - L| < \varepsilon$.

This is similar to the limit of sequence. The difference is that for function limit $|f(x) - L| < \varepsilon$ holds for all $x > N$ but for sequence limit this holds only for $x > N$, being integers. Since x is always between two integers n and $n + 1$, $f(x)$ may compare to the value $f(n)$ when f is monotone. This is the first trick how to compute the function limit.

Example 2.1.1

1. *The function $y = x^{-1}$ converges to 0 when $x \to +\infty$. In fact when $n \leqslant x < n + 1$,*

$$(n + 1)^{-1} < x^{-1} \leqslant n^{-1}.$$

   Because the sequences on both sides converge to zero, x^{-1} converges to zero when $x \to +\infty$.

2. *Assume that $a > 1$. Find*

$$\lim_{x\to+\infty} \frac{x}{a^x}.$$

Actually we know that a^x will increase faster and faster when x increases, and will eventually succeed any power function. Hence we guess the limit is zero. To prove it, when $x \in [n, n+1)$, $a^n \leqslant a^x < a^{n+1}$. Hence

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

From Example1.4.1, it follows that the sequences on both sides converge to zero.

3. *Here gives an example the trick does not work. The sine function $y = \sin x$ does not converge when $x \to +\infty$, because it is impossible to have $L \in \mathbf{R}$, so that $|\sin x - L| < 1/2$ for large x.*

The properties listed in Theorem 1.4.1 for sequence limit also hold for function limit.

Theorem 2.1.2 Assume that f, g and h are functions on $[A, \infty)$.

1. Uniqueness: The limit of $f(x)$ as $x \to \infty$ is unique if exists.

2. Comparison: If $f(x) \geqslant 0$ for large x and $\lim_{x\to+\infty} f(x) = L$, then $L \geqslant 0$. It is also true when both '$\geqslant$' are changed to '$\leqslant$'.

3. Sandwich: If $g(x)$ and $h(x)$ converge to the same limit L as $x \to +\infty$ and for large x,

$$g(x) \leqslant f(x) \leqslant h(x),$$

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

4. Boundedness: If $f(x)$ converges as $x \to +\infty$, then f is bounded near $+\infty$.

5. Operations: If each limit in the right side exists and the one appeared in denominator is not zero, then we have the following results ($\lim = \lim\limits_{x\to+\infty}$).

   (a) $\lim(f(x) + g(x)) = \lim f(x) + \lim g(x)$.

   (b) $\lim f(x)g(x) = \lim f(x) \cdot \lim g(x)$;

   (c) $\lim \dfrac{f(x)}{g(x)} = \dfrac{\lim f(x)}{\lim g(x)}$.

We will carefully discuss this when we talk about function limit when $x \to a$ later. The following is an example we use this theorem.

Example 2.1.2 *In general, for any given integer $j \in \mathbf{N}$,*

$$\lim_{x\to+\infty} \frac{x^j}{a^x} = \lim_{x\to+\infty} \left(\frac{x}{(a^{1/j})^x}\right)^j = 0,$$

because $a^{1/j} > 1$ for any positive integer j.

Application to proving inequality: Assume $a > 1$. *When* x *is large enough,*

$$x^n + a_1 x^{n-1} + \cdots + a_{n-1}x + a_n \leqslant a^x.$$

Since

$$\lim_{x \to +\infty} \frac{x^n + a_1 x^{n-1} + \cdots + a_{n-1}x + a_n}{a^x} = 0,$$

by the comparison theorem, when x *is large,*

$$\frac{x^n + a_1 x^{n-1} + \cdots + a_{n-1}x + a_n}{a^x} < 1.$$

This means that any increasing exponential function increases faster than any polynomial eventually.

The limits of $f(x)$ when $x \to +\infty$, $x \to -\infty$ and $|x| \to \infty$, even $x \to a$, can be defined similarly. The students who are interested should learn to write down the definitions. We may talk about sequence limit without mentioning where n goes, since n can only go one direction to infinity. However we can not talk about the limit of $f(x)$ without specifying where x goes, since x can go to any point.

2.1.2 function limit as $x \to a$

When we talk about the limit of function $y = f(x)$, $x \in D$ as $x \to a$, we assume that f is defined near a i.e., on a neighborhood of a, or an interval $(a - \delta_0, a + \delta_0)$ for some $\delta_0 > 0$, but f may not de defined at a.

Definition 2.1.3 We say $f(x)$ converges to L when $x \to a$, denoted by

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

if for any $\varepsilon > 0$, there exists $\delta > 0$ such that

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

for any $x \in (a - \delta, a + \delta) \setminus \{a\}$. We say $f(x)$ converges when $x \to a$ if there exists $L \in \mathbf{R}$ such that

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

The function limit $\lim_{x \to a} f(x)$ is more precisely

$$\lim_{x \to a, x \neq a, x \in D} f(x).$$

It is obvious that $\lim_{x \to a} f(x) = L$ is equivalent to $\lim_{x \to a}(f(x) - L) = 0$ and also to $\lim_{x \to a} |f(x) - L| = 0$. The definition above is called the ε-δ definition, which describes rigorously how a function approaches its limit when x approaches a by two inequalities. Note that $|f(x) - L| < \varepsilon$ is required for $x \in D$ and $0 < |x - a| < \delta$. Thus the limit describes the function behavior around a but not at a.

Let's see how we use the definition to prove function limits. We shall prove that for any basic elementary function f, if a is in the domain of f, then

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

Example 2.1.3 *1. For any $a \in \mathbf{R}$, find $\lim_{x\to a} x^2$ and prove it.*

First guessing the limit is easy. When x goes to a, x^2 goes to a^2 quite intuitively. Now we need to prove it. For any $\varepsilon > 0$, how to find $\delta > 0$, such that for any $x \in (a - \delta, a + \delta) \setminus \{a\}$, $|x^2 - a^2| < \varepsilon$? It is not easy to solve this inequality directly. We should make it bigger and simpler but still very small. At first, we may only need to consider x is in a neighborhood of a, say $|x - a| < 1$. Then $|x| \leqslant |x - a| + |a| < 1 + |a|$ and

$$|x^2 - a^2| = |x + a||x - a| \leqslant (|x| + |a|)|x - a| \leqslant (1 + 2|a|)|x - a|.$$

The righ side is simple and still small and we may easily see that as long as $\delta = \dfrac{\varepsilon}{1 + 2|a|}$, when $|x - a| < \delta$, it holds that

$$|x^2 - a^2| \leqslant (1 + 2|a|)|x - a| < \varepsilon.$$

It is seen that δ depends on a and ε.

2. *Prove $\lim_{x\to 0} a^x = 1$, for $a > 0, a \neq 1$. When $a > 1$, for any $\varepsilon > 0$, there is N such that $0 < a^{1/N} - 1 < \varepsilon$. Hence as long as $|x| < 1/N$, $|a^x - 1| < a^{1/N} - 1 < \varepsilon$, i.e., $\lim_{x\to 0} a^x = 1$. Hence*

$$\lim_{x\to b} a^x = a^b \lim_{x \to b} a^{x-b} = a^b.$$

3. *when $a > 0$, $\lim_{x\to a} \ln x = \ln a$. Assume $x > a$. To require*

$$0 < \ln x - \ln a = \ln \frac{x}{a} < \varepsilon,$$

it suffices to have

$$0 < x - a < a(e^{\varepsilon} - 1).$$

When $0 < \delta < a(e^{\varepsilon} - 1)$, we will have

$$0 < \ln x - \ln a < \varepsilon.$$

4. *$\lim_{x\to a} \sin x = \sin a$. It can be shown by the following inequality*

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

The following theorem describes function limits by sequence limits.

Theorem 2.1.4 $\lim_{x\to a} f(x) = L$ if and only if for any sequence $(x_n) \subset I \setminus \{a\}$ convergent to a, it holds that $\lim_{n\to\infty} f(x_n) = L$.

If $\lim_{x\to a} f(x) = L$, it is clear that $f(x_n) \to L$ as any sequence $x_n \to a$. If it is not true that $\lim_{x\to a} f(x) = L$, according to the definition, $\exists \varepsilon > 0$ so that $\forall \delta > 0$, $\exists x_\delta \in (a-\delta, a+\delta) \setminus \{a\}$, $|f(x_\delta) - L| > \varepsilon$. For $\delta = 1/n$, we have a sequence $x_n \to a$ so that $|f(x_n) - L| > \varepsilon$.

Example 2.1.4 *1. The function* $y = \sin \frac{1}{x}$ *does not converge. Two sequences* $x_n = \frac{1}{2n\pi}$ *and* $y_n = \frac{1}{2n\pi+\pi/2}$ *make* $f(x_n)$ *and* $f(y_n)$ *have different limits*

$$\lim_n \sin \frac{1}{x_n} = \sin(2n\pi) = 0, \ \lim_n \sin \frac{1}{y_n} = \sin(2n\pi + \pi/2) = 1.$$

2. $\lim_{x\to a} \arcsin x = \arcsin a$. *Assume* $b = \arcsin a$ *or* $\sin b = a$. *Set* $x_n := \sin(b - 1/n)$ *and* $x_n' = \sin(b + 1/n)$. *Then* $x_n \uparrow a$, $x_n' \downarrow a$ *and* $x_n < a < x_n'$. *It follows from the strict monotonicity that when* $x \in (x_n, x_n')$, *we have*

$$b - 1/n < \arcsin x < b + 1/n,$$

i.e., $|\arcsin x - \arcsin a| < 1/n$.

Exercises

1. Write down the proof for Theorem 2.2.1 by mimicking the proof of Theorem 1.4.1.
2. Find $\lim \frac{x}{1+x^2}$ when $x \to +\infty$ and when $x \to 0$ and then prove.
3. Prove that $|\sin x| \leqslant |x|$ for any $x \in \mathbf{R}$.
4. Prove that $\lim_{x\to a} \cos x = \cos a$.
5. Prove that $\lim_{x\to 0} x \sin(1/x) = 0$.
6. Prove that if f is bounded near a, $\lim_{x\to a}(x-a)f(x) = 0$.
7. Assume that f^{-1} with the domain D is the inverse function of f, and $a \in D$, $a = f(b)$. Write down a rigorous argument to prove that

$$\lim_{x\to a} f^{-1}(x) = b = f^{-1}(a)$$

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

2.2 function limit: operations

2.2.1 generic properties of limit

From these examples we see that it is usually difficult to prove a limit by ε-δ language, so we would like to avoid it if possible. We need to develop techniques to compute limits. The limit properties listed below are similar to the properties we proved for sequence limit. They are called the generic properties of limit because they are actually true for all types of limits. The function f is said to be bounded near a if there exist $K > 0$ and $\delta > 0$ such that $|f(x)| \leqslant K$ for any $x \in (a-\delta, a+\delta) \setminus \{a\}$.

Theorem 2.2.1 1. Uniqueness: The limit is unique if exists.

2. Comparison: If f is non-negative near a and $\lim_{x\to a} f(x) = L$, then $L \geqslant 0$.

3. Sandwich: If $g(x) \leqslant f(x) \leqslant h(x)$ holds true near a and

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

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

4. Boundedness: If $\lim_{x\to a} f(x)$ exists, then f is bounded near a.

5. Operations: If the functions below are defined on a same domain, the limits on the right side exist and the denominator is non zero, then we have the following results ($\lim = \lim\limits_{x\to a}$).

$$\begin{aligned}
&\lim[f(x) + g(x)] = \lim f(x) + \lim g(x),\\
&\lim f(x)g(x) = \lim f(x) \cdot \lim g(x),\\
&\lim \frac{f(x)}{g(x)} = \frac{\lim f(x)}{\lim g(x)}.
\end{aligned}$$

Let us prove the multiplication formula. In fact f is bounded near a, i.e., there exists $K > 0$, $\varepsilon > 0$ such that $|f(x)| \leqslant K$ when $0 < |x-a| < \delta$. Hence

$$\begin{aligned}
|f(x)g(x) - LM| &= |f(x)g(x) - f(x)M + f(x)M - LM|\\
&\leqslant |f(x)| \cdot |g(x) - M| + |M| \cdot |f(x) - L||M|\\
&\leqslant K \cdot |g(x) - M| + |M| \cdot |f(x) - L|.
\end{aligned}$$

This gives the conclusion and the other statements are left for students to prove. We stated this theorem intentionally three times from sequence limit to function limit and the purpose is to make a deep impression on the readers.

By the generic properties of limit, the following limits can be obtained without turning to ε-δ.

Example 2.2.1 1. *When* $a > 0$*, for any integer* b*,* $\lim_{x\to a} x^b = a^b$*. Since* $\lim_{x\to a} x = a$*, we have*

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

When b *is a positive integer,*

$$\lim_{x\to a} x^b = \lim_{x\to a} x \cdots x = a \cdots a = a^b.$$

When b *is a negative integer,*

$$\lim_{x\to a} x^b = \lim_{x\to a} \frac{1}{x^{-b}} = \frac{1}{a^{-b}} = a^b.$$

2. $\lim_{x\to a} \tan x = \tan a$*, where* $\cos a \neq 0$*.*

$$\lim_{x\to a} \frac{\sin x}{\cos x} = \frac{\lim_{x\to a} \sin x}{\lim_{x\to a} \cos x} = \frac{\sin a}{\cos a} = \tan a.$$

Actually it is not easy to prove $\lim_{x\to a} x^b = a^b$ by the definition when b is not an integer, especially when $b \notin \mathbf{Q}$. When $a > 0$, it can be written as $x^b = e^{b\ln x}$ and we know that $\ln x \to \ln a$. Hence this can be done by a powerful property, the composition of limit.

Theorem 2.2.2 Assume that $\lim_{x\to a} g(x) = b$ and $\lim_{y\to b} f(y) = L$. Then

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

This theorem is quite intuitive because when x is near a, $g(x)$ is near b and then $f(g(x))$ approaches L. The proof can be written this way easily.

Example 2.2.2 *Assume that* $a > 0$*.*

1. $\lim_{x\to b} a^x = \lim_{x\to b} a^b \cdot a^{x-b} = a^b$.
2. $\lim_{x\to a} x^b = \lim_{x\to a} e^{b\ln x} = e^{b\ln a} = a^b$.

From this theorem, we know that it is not important where x goes when we talk about a limit. If necessary, every limit $\lim_{x\to a} f(x)$ can be transferred to a limit as $x \to \infty$, or $x \to 0$,

$$\lim_{x\to a} f(x) = \lim_{x\to\infty} f(a + 1/x) = \lim_{x\to 0} f(x + a).$$

In summary, We have the following theorem.

Theorem 2.2.3 For any elementary function f with domain D, if $a \in D$, then

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

2.2.2 two important function limits

We now introduce two fundamental limits in calculus. The first limit

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

For any $x > 0$, let $n = [x]$, the integer part of x. Then we have

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

Since $x \to +\infty$ if and only if $n \to \infty$, it follows that the two ends of the inequality above converge to e and

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

To prove

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

we may take $y = -x \to +\infty$, and then

$$\begin{aligned}\lim_{x\to-\infty}\left(1+\frac{1}{x}\right)^x &= \lim_{y\to+\infty}\left(1+\frac{1}{-y}\right)^{-y} = \lim_{y\to+\infty}\left(1+\frac{1}{y-1}\right)^y \\ &= \lim_{y\to+\infty}\left(1+\frac{1}{y-1}\right)^{y-1}(1+\frac{1}{y-1}) = e\cdot 1 = e.\end{aligned}$$

This proves that

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

The second limit

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

This limit may be proved by the inequality that for $x \in (0, \pi/2)$,

$$\sin x < x < \tan x,$$

from which, it follows that

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

But $\lim_{x\to 0}\cos x = 1$ and hence the limit 2 follows.

Example 2.2.3 *The variants of the limit*

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

The 1st variant:

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

The 2nd variant: taking logarithm by theorem above

$$\lim_{y\to 0}\frac{\log(1+y)}{y}=1.$$

The 3rd variant: taking $z=\log(1+y)$ *and* $y=e^z-1$,

$$\lim_{z\to 0}\frac{e^z-1}{z}=1.$$

All these limits are important.

Example 2.2.4 *Other limits can be calculated by four operations,*

$$\begin{aligned}
\lim_{x\to 0}(1+2x)^{1/x} &= \lim_{y\to\infty}\left(1+\frac{2}{y}\right)^{y}\\
&= \lim_{y\to\infty}\left(1+\frac{2}{y}\right)^{y/2}\cdot\left(1+\frac{2}{y}\right)^{y/2}=e^2.\\
\lim_{x\to 0}\frac{\tan x}{x} &= \lim_{x\to 0}\frac{\sin x}{x}\frac{1}{\cos x}=1\cdot\frac{1}{1}=1.\\
\lim_{x\to+\infty}x^{-1}\log x &= \lim_{y\to+\infty}\frac{y}{e^y}=0.\\
\lim_{x\to+\infty}x^{1/x} &= \lim_{x\to+\infty}e^{x^{-1}\log x}\\
&= e^{\lim\limits_{x\to+\infty}x^{-1}\log x}\\
&= e^0=1,
\end{aligned}$$

2.2.3 infinitesimals

If $\lim_{x\to a}f(x)=0$, we say $f(x)$ is an infinitesimal (when $x\to a$). Hence an infinitesimal is a limit, not a number. The different infinitesimals can be compared. Assume that both $f(x)$ and $g(x)$ are infinitesimals, and

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

If $L=0$, we say $f(x)$ is a higher order infinitesimal than $g(x)$, denoted by $f(x)=o(g(x))$. If $L\neq 0$, we say they are the same order. More precisely if $L=1$, we say they are equivalent infinitesimals, denoted by $f(x)\sim g(x)$ as $x\to a$. Of course if $L\neq 0$, $f(x)\sim Lg(x)$. For example as $x\to 0$ we have

$$\begin{aligned}
&\sin x\sim x;\\
&\tan x\sim x;\\
&\ln(1+x)\sim x;\\
&e^x-1\sim x.
\end{aligned}$$

When we compute limit in a product, we may use infinitesimal substitution.

Theorem 2.2.4 If $f(x) \sim g(x)$ as $x \to a$, then

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

Example 2.2.5

$$\lim_{x\to 0} \frac{\ln(1+x^2)}{x} = \lim_{x\to 0} \frac{x^2}{x} = 0.$$
$$\lim_{x\to+\infty} \frac{2^{-1/x}-1}{\sim (1/x)} = \lim_{x\to+\infty} \frac{-(1/x)\ln 2}{1/x} = -\ln 2.$$

When $x \to a$, $f(x)$ is called an infinity, if its inverse $1/f(x)$ is an infinitesimal. Similarly the infinities are 'higher order', 'same order' and 'equivalent' if their inverses are.

2.2.4 one-side limit

The types of limits are rich. As long as x approaches to somewhere, we may talk about where $f(x)$ approaches. For example, we may talk about one-side limit of a function. We say $f(x)$ has right limit L when $x \to a$, denoted by

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

if for any $\varepsilon > 0$, there exists $\delta > 0$ such that when $x \in (a, a+\delta)$, we have

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

The left limit $\lim_{x\to a-} f(x)$ is defined similarly.

Example 2.2.6 *Assume that*

$$f(x) = \begin{cases} 1+x, & x > 0, \\ 100, & x = 0, \\ -1+x^2, & x < 0. \end{cases}$$

Then

$$\lim_{x\to 0+} f(x) = 1, \quad \lim_{x\to 0-} f(x) = -1.$$

Theorem 2.2.5 $\lim_{x\to a} f(x) = L$ if and only if

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

Exercises

1. Find the following limits when $x \to +\infty$.

$$(1)\ \lim \frac{x}{1+x^2}; \qquad (2)\ \lim \frac{x^2-1}{x^2+x-2};$$
$$(3)\ \lim \frac{\ln x}{\sqrt{x}}; \qquad (4)\ \lim \frac{\sin x}{x}.$$

2. Find the following limits.

$$(1)\ \lim_{x\to 0} \frac{x}{1+x^2}; \qquad (2)\ \lim_{x\to 1} \frac{x^2-1}{x^2+x-2};$$
$$(3)\ \lim_{x\to 0} \frac{\tan x}{x}; \qquad (4)\ \lim_{x\to 0} \frac{1-\cos x}{x^2};$$
$$(5)\ \lim_{x\to 0} \frac{\sqrt{x+4}-2}{x}; \qquad (6)\ \lim_{x\to 1} \frac{x^6-1}{x^9-1}.$$

2.3 analysis 2: continuous functions

2.3.1 continuity

The function limit equals the function value. This is a very important property called continuity.

Definition 2.3.1 Assume that $y = f(x),\ x \in D$, and now $a \in D$. We say that f is continuous at a, if for any $\varepsilon > 0$ there exists $\delta > 0$ such that $|f(x)-f(a)| < \varepsilon$ holds for $x \in D \cap (a-\delta, a+\delta)$. If f is not continuous at a, we say that f is discontinuous at a.

The continuity means that the value of f near a, if any, is near $f(a)$. Note that we only talk about the continuity of f at a point a in its domain. When a is a limit point of D, the continuity amounts to say that

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

Read the definition carefully. If $a \in D$ is not a limit point of D (called an isolated point, namely there is no point of D other than a in a neighborhood of a), for example $y = \sqrt{\sin x - 1}$, then f is continuous at a automatically because in this case 'there exists $\delta > 0$ such that $|f(x) - f(a)| < \varepsilon$ holds for $x \in D \cap (a - \delta, a + \delta)$' holds automatically.

The following theorem is a direct consequence of generic properties of limit.

Theorem 2.3.2
1. Assume that f, g are two functions defined near a. If f, g are continuous at a, then $f+g$, fg are continuous at a and $1/g$ is continuous at a, if in addition $g(a) \neq 0$.
2. If g is continuous at a and f is continuous at $g(a)$, then $f(g(x))$ is continuous at a.

An easy corollary of the theorem above is that an elementary function is continuous at any point in its domain.

Example 2.3.1 *1.* $y = x^2$ *is continuous at any point.*

2. The function

$$y = \begin{cases} 1, & x > 0, \\ 2x, & x \leqslant 0 \end{cases}$$

is continuous except $x = 0$*, at which, the function limit does not exist.*

3. The function $y = \frac{\sin x}{x}$ *is elementary, so it is continuous when* $x \neq 0$*. It has no definition at* $x = 0$*, but has limit* 1 *when* $x \to 0$*. Hence if we define*

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

then f *is no longer an elementary function but continuous everywhere. If we define* $f(0)$ *to be other value, it is not continuous at* $x = 0$*.*

In the case where $\lim_{x \to a} f(x) = L$ but f is not continuous at a, we say a is a (point with) removable discontinuity, because we may redefine $f(a) = L$ to make f continuous at a.

2.3.2 two important theorems for continuous functions

A function $y = f(x),\ x \in D$ is called continuous on D if it is continuous at every point of D. A continuous function on a closed interval $[a, b]$ has two important theorems. The first one is called the intermediate-value theorem, which says that a continuous function on a closed interval reaches any value between two function values.

Theorem 2.3.3 A continuous function f on $[a, b]$ reaches any value m between $f(a)$ and $f(b)$, i.e., there exists x_0 such that $f(x_0) = m$,

As we said before, a continuous (never broken) function is like a rope. A rope with two ends being in different sides of a line must cross the line somewhere.

Example 2.3.2 *About the* n*-th root* $a^{1/n}$*. Assume that* $a > 0$*,* $n \in \mathbf{N}$*. Then* $y = x^n$ *is continuous, strictly increasing on* $[0, +\infty)$ *and* $x^n \uparrow +\infty$ *as* $x \uparrow +\infty$*. Hence the equation* $x^n = a$ *has a unique positive solution, which is denoted by* $x = a^{1/n}$*.*

Definition 2.3.4 Assume that $y = f(x)$, $x \in D$. If there exists $x_0 \in D$ such that $f(x_0) \geqslant f(x)$ for any $x \in D$, we say that f reaches its maximum at x_0. If there exists $x_0 \in D$ such that $f(x_0) \leqslant f(x)$ for any $x \in D$, we say that f reaches its minimum at x_0. These are the global extremum for f.

The function $y = f(x),\ x \in (0, 1]$ can reach maximum but not minimum. The second one is that a continuous function on a closed interval reaches its maximum and minimum.

Theorem 2.3.5 Assume that $y = f(x)$ is continuous on $[a, b]$. Then f can reach both its maximum and minimum.

Two results look very intuitive but not so easy to prove rigorously. First we understand that the continuity is essential by finding counter-examples when the function f has discontinuity inside $[a, b]$.

2.3.3 nested intervals

The intermediate-value theorem is equivalent to the Bolzano's zero theorem below.

Theorem 2.3.6 (Bolzano) Assume that $y = f(x)$ is continuous on an interval I. If $a, b \in I$ and $f(a) \cdot f(b) < 0$, then there exists $x_0 \in (a, b)$ such that $f(x_0) = 0$, i.e., the equation $f(x) = 0$ has a solution on (a, b).

How to find the root x_0? Here we give a routine which always works. Without loss of generality, assume $f(a) < 0$ and $f(b) > 0$. Take the middle point $(a + b)/2$. If it is a root, then $x_0 = (a+b)/2$, otherwise, either $f((a+b)/2) > 0$ or $f((a+b)/2) < 0$, say the former, we denote $a_1 = a$ and $b_1 = (a + b)/2$ and look for the root on $[a_1, b_1]$. Continue this procedure, either some middle point is a root and then we stop, or we get an interval sequence $[a_n, b_n]$. Every time, the interval will be half of previous one. Finally we will track down a point x_0 which will certainly satisfy $f(x_0) = 0$ by continuity. In this section we would like to illustrate the proof of Theorem 2.3.5: a continuous function on a bounded closed interval reaches its maximum and minimum. Look at the idea of the proof of Theorem 2.3.6 and there we use so-called cut-half technique: cut the interval into two parts, keep the part where the target is and then cut again. Finally we obtain a sequence of intervals $[a_n, b_n]$ satisfying

1. $a_n \nearrow$ and $b_n \searrow$;
2. $b_n - a_n \to 0$.

Then (a_n) and (b_n) converges to the same number $x_0 \in [a, b]$, and it follows that $f(x_0) = 0$ because 0 is between $f(a_n)$ and $f(b_n)$. This proves the Bolzano's zero theorem. The cut-half technique is useful in tracing a point.
The sequence of intervals $[a_n, b_n]$ satisfying the conditions above is called a sequence of nested intervals.

Theorem 2.3.7 If $[a_n, b_n]$ is a sequence of nested intervals, then there is a unique point $x_0 \in [a_n, b_n]$ for any n.

This means that a sequence of nested intervals will catch a unique point. This is a very useful method to find a particular point.

2.3.4 the least upper bound

To prove Theorem 2.3.5, we need to prepare two theorems.

Definition 2.3.8 A real number x is called a least upper bound of $A \subset \mathbf{R}$ if (1) x is a upper bound of A; (2) it is the least among upper bound, i.e., for any $a < x$, there exists $y \in A$ such that $a < y$. The greatest lower bound can be defined similarly.

The least upper bound is unique if it exists, and denoted by $\sup A$. The greatest lower bound is denoted by $\inf A$.

Theorem 2.3.9 A non-empty set $A \subset \mathbf{R}$ bounded above has the least upper bound.

Proof. The method, the nested intervals, which we used in the proof of Bolzano's zero theorem to find the zero point can be used here also. Take an interval $[a, b]$ where a is not upper bound of A but b is. Consider their middle point m. Then m is either a upper bound or not. If it is, set $a_1 = a$, $b_1 = m$, otherwise $a_1 = m$, $b_1 = b$, i.e., a_1 is not upper bound but b_1 is. Consider their middle point m_1. Then m_1 is either a upper bound or not. If it is, set $a_2 = a_1, b_2 = m$, otherwise $a_2 = m, b_2 = b$. Continue this, we will obtain a sequence of intervals (a_n, b_n) satisfying

1. for any n, $(a_n, +\infty) \cap A \neq \emptyset$, $(b_n, +\infty) \cap A = \emptyset$;

2. $a_n \nearrow$ and $b_n \searrow$;

3. $b_n - a_n = (b - a)/2^n \to 0$.

Since a monotone bounded sequence converges, there exists x such that (a_n) and (b_n) converge to x. It is obvious that x is the least upper bound of A. $\square$

2.3.5 Bolzano-Weierstrass theorem

Theorem 2.3.10 (Bolzano-Weierstrass) A bounded sequence has a convergent subsequence.

Proof. We use the method of nested interval again to trace the limit point and convergent sequence. Fix a sequence $(x_n) \subset [a, b]$. Take the middle point m of $[a, b]$. At least one of $[a, m]$ and $[m, b]$ contains infinitely many terms of (x_n). We denote that interval by $[a_1, b_1]$ and then consider its middle point again. In this way we obtain a sequence of intervals $[a_k, b_k]$, $k \geqslant 1$, satisfying

1. each $[a_k, b_k]$ contains infinitely many terms of (x_n);

2. $a_n \nearrow$ and $b_n \searrow$.

3. $b_n - a_n = (b-a)/2^n \to 0$.

Hence (a_n) and (b_n) have the same limit L. We may pick a point $x_{k_1} \in [a_1, b_1]$, pick a point $x_{k_2} \in [a_2, b_2]$ with $k_2 > k_1$, ..., pick a point $x_{k_n} \in [a_n, b_n]$ with $k_n > k_{n-1}$. It can be done because there are infinitely many points in each $[a_n, b_n]$. Then (x_{k_n}) is a subsequence and

$$a_n \leqslant x_{k_n} \leqslant b_n.$$

Hence $x_{k_n} \to L$. □

2.3.6 the proof of Theorem 2.3.5

To complete the proof of Theorem 2.3.5, we first prove that a continuous function f on $[a, b]$ is bounded by way of contradiction. Suppose that f is unbounded. For any n, there exists $x_n \in [a, b]$ such that $|f(x_n)| > n$. By Bolzano-Weierstrass theorem, (x_n) has a convergent subsequence (x_{k_n}), say, which has limit $c \in [a, b]$. Then by continuity

$$\lim_n f(x_{k_n}) = f(c).$$

This is a contradiction because the left side diverges.
The proof of Theorem 2.3.5: assume that f is continuous on $[a, b]$. Since f is bounded, the set $A = f([a, b])$ is bounded above. It has the least upper bound. Set $M = \sup A$, i.e.,

1. $f(x) \leqslant M$ for any $x \in [a, b]$;

2. for any n, $M - 1/n$ is no longer a upper bound, namely there exists $x_n \in [a, b]$ such that $f(x_n) > M - 1/n$.

We need only to check M can be reached by f. A sequence $(x_n) \subset [a, b]$ is obtained and satisfies

$$M - 1/n < f(x_n) \leqslant M.$$

By Bolzano-Weierstrass theorem, (x_n) has a subsequence (x_{k_n}) converges to some $c \in [a, b]$. Hence

$$M - 1/k_n < f(x_{k_n}) \leqslant M.$$

Then taking limit and by the continuity of f, we have $M = f(c)$.
Actually every bounded function has the least upper bound, which can be seen as the maximum. The problem is whether this value can be reached by the function.
Remark. The rigorous formulation of function limits and continuity was attributed to Cauchy, who was one of the most important mathematicians to make calculus be built on a convincing base.

Exercises

1. If f, g are continuous functions $f(2) = 5$, and $\lim_{x\to 2}[2f(x) - 3g(x)] = 4$, find $g(2)$.

2. For what value of c is the function
$$f(x) = \begin{cases} 1 - x^2, & x < 1, \\ cx, & x \geqslant 1 \end{cases}$$
continuous on $\mathbf{R}$?

3. Let $f(x) = \begin{cases} x, & x \in \mathbf{Q}, \\ 1, & x \in \mathbf{R} \setminus \mathbf{Q}. \end{cases}$ At which point, is f continuous?

4. Let n be an integer and
$$f(x) = \begin{cases} x^n \sin \frac{1}{x}, & x \neq 0, \\ 0, & x = 0. \end{cases}$$
Sketch the graph and prove that f is continuous when $n \geqslant 1$.

5. Show that $f(x) = 3x^5 - x^2 + 10$ has at least one root.

6. Show that $f(x) = x^4 - x - 10$ has at least two roots, in which one is positive and the other is negative.

7. Illustrate that $f(x) = 1/x$ is continuous but unbounded on $(0, 1)$.

8. Illustrate that $f(x) = \begin{cases} x^2, & x \neq 0, \\ 1, & x = 0, \end{cases}$ can reach max but not min on $[-1, 1]$.

9. A sequence $\{x_n\}$ is called a Cauchy sequence if for any $\varepsilon > 0$, there exists an integer $N > 0$ such that
$$|x_n - x_m| < \varepsilon$$
whenever $n, m > N$. Prove that a Cauchy sequence is bounded.

10. Prove that a sequence is a Cauchy sequence if and only if it converges.

Chapter 3

derivatives and application

3.1 how to compute derivatives

3.1.1 derivatives

Here comes the definition of derivative.

Definition 3.1.1 Assume that $y = f(x),\ x \in (a, b)$ is a function. For any fixed point $x \in (a, b)$, if

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

converges to a finite number L, then L is called the derivative of f at x, denoted by $f'(x)$, and we say f has derivative at x or simply f is smooth at x. The procedure to find the derivative of $y = f(x)$ at x is simply said to take derivative of f with respect to x.

The quantity $\frac{f(y)-f(x)}{y-x}$ is the average rate of change of f over $[x, y]$, so derivative is also called the instant rate of change. It is seen that if f has derivative at x then f is continuous at x. Both continuity and derivative mean that a function does not change too dramatically near a point, but having derivative is smoother than being continuous.

The derivative has clear physical and geometric explanation. In physics, if $y = f(t)$ denotes the distance that a particle moves on $[0, t]$, then the derivative $f'(t)$ is the limit of average speed and denotes the instant speed at time t. In geometry, derivative $f'(x)$ is the limit of the slope of secant and denotes the slope of the tangent line of the graph of f at point $(x, f(x))$.

Example 3.1.1 *1. Find the derivative of* $y = x^2$ *at point* x.

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

2. *Consider* $y = |x|,\ x \in \mathbf{R}$. *We need to consider the limit*
$$\lim_{h\to 0} \frac{|x+h| - |x|}{h}.$$
When $x > 0$*, and* $|h|$ *is small,* $x + h$ *is also positive and then*
$$\frac{|x+h| - |x|}{h} = \frac{x+h-x}{h} = 1.$$
When $x < 0$*, and* $|h|$ *is small,* $x + h$ *is also negative and then*
$$\frac{|x+h| - |x|}{h} = \frac{-(x+h)+x}{h} = -1.$$
Hence $f'(x) = 1$ *for* $x > 0$ *and* $f'(x) = -1$ *for* $x < 0$*. When* $x = 0$*,*
$$\frac{|x+h| - |x|}{h} = \frac{|h|}{h}$$
which has right limit 1 and left limit -1*. Hence* f *has no derivative at* $x = 0$*. The function* $y = |x|$ *is continuous but not smooth at 0.*

If $y = f(x)$ has derivative at every point of D, then $y = f'(x)$ is also a function of x on D, called the derivative function of f.

3.1.2 basic formulae for derivatives

Basic formulae for derivatives start from the derivatives of the basic elementary functions.

1. $(x^n)' = nx^{n-1}$, when $n \in \mathbf{N}$;
$$\frac{(x+h)^n - x^n}{h} = \frac{x^n + nx^{n-1}h + \cdots + h^n - x^n}{h} = nx^{n-1} + h(\cdots) \to nx^{n-1}.$$

2. $(a^x)' = a^x \ln a$, $(e^x)' = e^x$;
$$\frac{a^{x+h} - a^x}{h} = a^x \frac{a^h - 1}{h} = a^x \cdot \frac{e^{h \ln a} - 1}{h \ln a} \ln a \to a^x \ln a,$$
where we use the 3rd variant in Example 2.2.3.

3. $(\ln x)' = \frac{1}{x}$;
$$\frac{\ln(x+h) - \ln x}{h} = \frac{\ln(1 + h/x)}{x \cdot h/x} \to \frac{1}{x},$$
where we use the 2rd variant in Example 2.2.3.

4. $(\sin x)' = \cos x$.
$$\begin{aligned}
&\frac{\sin(x+h) - \sin x}{h} \\
=& \frac{\sin x \cos h + \cos x \sin h - \sin x}{h} \\
=& \frac{\sin x(\cos h - 1)}{h} + \cos x \frac{\sin h}{h} \to \cos x,
\end{aligned}$$

where we use the basic limit

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

For other elementary functions, we need to use properties for derivative. First we talk about four operations.

Theorem 3.1.2 (four operations) Assume that functions f, g are smooth at x. Then

1. $(f(x) + g(x))' = f'(x) + g'(x)$, $(cf(x))' = cf'(x)$;
2. $(f(x)g(x))' = f'(x)g(x) + f(x)g'(x)$;
3. $$\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}.$$

Let's prove the division formula. It suffices to calculate the derivative of $\frac{1}{g(x)}$.

$$\begin{aligned}\left(\frac{1}{g(x)}\right)' &= \lim_{h\to 0}\frac{\dfrac{1}{g(x+h)} - \dfrac{1}{g(x)}}{h}\\ &= \lim_{h\to 0}\frac{1}{g(x)g(x+h)}\frac{-(g(x+h)-g(x))}{h} = -\frac{g'(x)}{g(x)^2}.\end{aligned}$$

Example 3.1.2 *Using these formulae, we can easily calculate derivative for many elementary functions.*

$$\begin{aligned}&(x^2 + 2x + 3)' = (x^2)' + (2x)' + 3' = 2x + 2,\\ &(e^x \sin x)' = e^x \sin x + e^x \cos x,\end{aligned}$$

How to calculate the derivative of $y = e^{x \sin x}$? No formula is available. To calculate derivative for all elementary functions, we need chain rule which is used to calculate the derivative of composite function.

Theorem 3.1.3 (chain rule) Assume that f, g are two functions, g is defined on (a, b) and $x \in (a, b)$. If g is smooth at x and f is smooth at $g(x)$, then

$$(f(g(x)))' = f'(g(x))g'(x),$$

where $(f(g(x)))'$ on the left is the derivative of the composite function $f \circ g$ at x and $f'(g(x))$ on the right is the derivative of f at the point $g(x)$.

For a proof, we just need to observe the following:

$$\frac{f(g(x+h)) - f(g(x))}{h} = \frac{f(g(x+h)) - f(g(x))}{g(x+h) - g(x)}\frac{g(x+h) - g(x)}{h}.$$

Set $\widehat{h} := g(x+h) - g(x)$, which goes to 0 as $h \to 0$. Hence

$$\frac{f(g(x+h)) - f(g(x))}{h} = \frac{f(g(x)+\widehat{h}) - f(g(x))}{\widehat{h}} \frac{g(x+h)-g(x)}{h}$$

has limit $f'(g(x)) \cdot g'(x)$.

It is important to distinguish the derivatives on two sides of the formula above. For example, $y = (\sin x)^2$. Taking derivative with respect x gives $2\sin x \cos x$, and taking derivative respect to $\sin x$ means that we treat $\sin x$ as a variable u, and take derivative respect to u, so the answer is $2u = 2\sin x$.

Example 3.1.3

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

Example 3.1.4 *Find the derivative of power function $y = x^a$, $x > 0$. When $a \in \mathbf{N}$, we have $y' = ax^{a-1}$ as in the previous lecture. When $a \in \mathbf{R}$, the previous method does not work. However this function can be viewed as a composite function $y = e^{a\ln x}$. By chain rule, we have*

$$y' = e^{a\ln x} \cdot a(\ln x)' = ax^a \frac{1}{x} = ax^{a-1}.$$

This extends the case where $a \in \mathbf{N}$.

Assume that a function f is denoted by $y = f(x), x \in D$, where x is independent variable. Because the derivative illustrates how y changes as x changes, we may write the derivative as y'_x, the derivative of y with respect x, sometimes x is omitted when no confusion is caused. There are many different ways to write derivative. The notations $f'(x)$, y'_x were invented by I. Newton. It can be denoted by

$$\frac{df(x)}{dx}, \ \frac{dy}{dx},$$

which are Leibniz's symbols. For the derivative at $x = x_0$ we use '$\ldots\big|_{x=x_0}$' to denote, for example $y'_x\big|_{x=x_0}$. The chain rule looks very natural in Leiniz's notation

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

if y is a function of u and u is a function of x. Hence Newton's notation is simple but Leibniz's notation is deep and inspiring.

The chain rule can be executed finite times, for example, taking derivative with respect to x,

$$\left(e^{\sin x^2}\right)' = e^{\sin x^2} \cdot \cos x^2 \cdot 2x.$$

3.1.3 derivatives of inverse functions

Example 3.1.5 1. *$y = \arcsin x$, $|x| \leqslant 1$. The relation is also $x = \sin y$. Then we have an identity: for all $x \in [-1, 1]$,*

$$x = \sin(\arcsin x),$$

where the right side is a compound function. Taking derivative of both sides, by the chain rule,

$$1 = \cos(\arcsin x)(\arcsin x)'.$$

Hence

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

What is $\cos(\arcsin x)$? We know that

$$[\cos(\arcsin x)]^2 + [\sin(\arcsin x)]^2 = 1,$$

and it implies that

$$\cos(\arcsin x) = \sqrt{1 - x^2}.$$

Hence

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

2. *Since $\arcsin x + \arccos x = \pi/2$, we obtain*

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

3. *Look at $y = \arctan x$, $x \in \mathbf{R}$. Then $x = \tan y$, where y is a function of x. Taking derivative with respect to x, we get*

$$1 = (\tan y)' y'_x = \frac{1}{\cos^2 y} y'_x,$$

and then

$$y' = \cos^2 y = \frac{1}{1 + \tan^2 y} = \frac{1}{1 + x^2}.$$

General inverse functions can be done similarly. A strictly monotone and continuous function has inverse function. A function $y = f(x), x \in (a, b)$ may not be strictly monotone on whole interval, but it is usually strictly monotone locally, just like $y = \sin x$, i.e., we may cut (a, b) into finite intervals so that f is strictly monotone on each interval. In this case it has inverse function locally. Assume that $y = f(x)$, $x \in D$, has an inverse function $x = f^{-1}(y)$, $y \in f(D)$, as a function of y.

Conventionally, we are more comfortable to use x to denote the independent variable, so that we would like to exchange two letters x and y and write the inverse function as

$$y = f^{-1}(x),\ x \in f(D).$$

Actually $y = f^{-1}(x)$ is the unique solution of the equation $x = f(y)$.
How to take derivative of f^{-1} with respect to x?
We have

$$x = f(y),$$

where y is a function of x, the inverse function, and the right side is a composite function. Taking derivative of both sides with respect to x, we have by chain rule $1 = f'(y)y'_x$. Then we have the following theorem.

Theorem 3.1.4 The derivative of the inverse function f^{-1} of f is

$$y'_x = \frac{1}{f'(y)},$$

where $y = f^{-1}(x)$.

By Leibniz notation, we have more intuitively

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

This means that the rate of change of y relative to x is the same as the inverse of the rate of change of x relative to y. Note that the inverse function does not always have an explicit expression. For example the function $y = x + \sin x$ is strictly increasing. It is obvious that each y corresponds a unique x, but it is impossible to express x as a function of y in terms of known functions.

Exercises

1.

2. Let n be an integer and

$$f(x) = \begin{cases} x^n \sin \frac{1}{x}, & x \neq 0, \\ 0, & x = 0. \end{cases}$$

Prove that

(a) when $n = 1$, f has no derivative at $x = 0$;

(b) when $n = 2$, f has derivative at $x = 0$, but $f'(x)$ does not converge as $x \to 0$;

(c) when $n \geqslant 3$, f has not only derivative at $x = 0$ but also $f'(x)$ converges as $x \to 0$.

3. Find equations of the tangent line and normal line to the curve $f(x) = x + \sqrt{x}$ at $(1, 2)$.

4. Find derivatives $\frac{dy}{dx}$ and $\frac{dx}{dy}$, the derivatives of inverse functions, when it exists locally.

(a) $y = \sqrt[5]{x} + \sqrt{x^5}$;

(b) $y = \frac{3x-1}{2x+1}$;

(c) $y = \frac{x^2}{1+\sqrt{x}}$;

(d) $y = \sqrt{x + \sqrt{x}}$;

(e) $y = e^{-x^2/2}$;

(f) $y = (1 - x^2)^{10}$.

(g) $y = x + sinx$.

3.1.4 implicit functions and parametrized curves*

We now spend some time to talk about implicit functions. What is an implicit function? The function in the for $y = f(x)$, $x \in D$ explicitly tells the relation, so it is called explicit function. However some relations are given by an equation, for example

$$x^2 + y^2 = 1$$

gives a circle. All points satisfying the equation

$$F(x, y) = 0$$

defines a curve on xy-plane, which gives a relation, not necessarily a function relation, because the function relation requires that each x (as the independent variable) corresponds a unique y. The circle relation does not satisfies this, for example when $x = 1/2$, there are two $y = \pm\sqrt{3}/2$ satisfying the equation. It can be seen that taking any point (x_0, y_0) on the circle, except $x_0 = 1, -1$, the curve in a neighborhood of (x_0, y_0) satisfies this condition. Hence an equation induces a function locally which is called implicit function.
Assume that $F(x, y) = 0$ determines an implicit function $y = f(x)$ locally at (x_0, y_0). We can talk about the derivative of f at a point (x_0, y_0). Note that in addition to indicate x_0, we have to indicate y_0 also, because for different y_0, the function is different. One way is to solve this equation to obtain the explicit expression and then to take derivative. For the circle above it is possible, but it may be impossible in many cases.

Example 3.1.6 *The derivative of implicit functions.*

1. $x^2 + y^2 = 1$. *Taking derivative and using the chain rule, we obtain*

$$2x + 2y \cdot y' = 0, \text{ and } y' = -\frac{x}{y}.$$

Then

$$y'\Big|_{(1/2,\sqrt{3}/2)} = -\frac{1}{\sqrt{3}}, \; y'\Big|_{(1/2,-\sqrt{3}/2)} = \frac{1}{\sqrt{3}}.$$

2. $y = e^{x+y}$. *It is impossible to have the explicit expression of* y *in terms of* x. *Taking derivative and using chain rule*

$$y' = e^{x+y}(1 + y'), \text{ and } y' = \frac{e^{x+y}}{1 - e^{x+y}} = \frac{y}{1 - y}.$$

As long we get a point (x, y) *on the curve, we will have the derivative.*

Before ending the derivative part, it is very important to note that when you take derivative, you should always be clear about which variable the derivative is taken.

Exercises

1. Assume $y = f(x)$ is smooth, $f'(x)$ is continuous in a neighborhood of a and $f'(a) \neq 0$. Prove that $\exists \delta > 0$ such that when $|y - f(a)| < \delta$, $f(x) = y$ has a unique solution. Hence the inverse function $x = f^{-1}(y)$ exists.

3.2 application of derivatives 1

3.2.1 mean-value theorem

We now talk about the application of derivative. It is very useful in analyzing the function properties. A mean-value theorem is given first.

Theorem 3.2.1 (Rolle) Assume that $y = f(x)$, $x \in [a, b]$ is continuous and smooth on (a, b). If $f(b) = f(a)$, there exists $\xi \in (a, b)$ such that $f'(\xi) = 0$.

Proof. If $y = f(x)$ is a constant, the conclusion is true obviously. We assume it is not constant. Then either the maximum or the minimum can be reached in (a, b). We assume that there exists $x_0 \in (a, b)$ such that $f(x_0) \geqslant f(x)$ for any $x \in [a, b]$. Since f is smooth at x_0, the ratio

$$\frac{f(x) - f(x_0)}{x - x_0}$$

has limit as $x \to x_0$. It follows from the fact $f(x) - f(x_0) \leqslant 0$ and the comparison theorem that the right limit and left limit of the ratio is non-positive and non-negative respectively and hence the limit $f'(x_0) = 0$. $\square$

This is the case where the line connecting $(a, f(a))$ and $(b, f(b))$ is horizontal. What if $f(a) \neq f(b)$?

Theorem 3.2.2 (mean-value) Assume that $y = f(x)$, $x \in [a, b]$ is continuous and smooth on (a, b). Then there exists $\xi \in (a, b)$ such that

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

We construct a new function, tilting f to be horizontal,

$$g(x) := f(x) - \frac{f(b) - f(a)}{b - a}(x - a),\ x \in [a, b].$$

It is easy to check $g(a) = g(b)$. Hence by Rolle's theorem, there exists $\xi \in (a, b)$ such that $g'(\xi) = 0$ which means that

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

Example 3.2.1 *A function is defined to be*

$$f(x) = \begin{cases} x \sin \frac{1}{x}, & x \in (0, 1], \\ 0, & x = 0. \end{cases}$$

This function is continuous on $[0, 1]$ *and smooth on* $(0, 1)$ *but not smooth at 0, since*

$$\frac{f(h) - f(0)}{h} = \sin \frac{1}{h}$$

diverges.

Another function $f(x) = \sqrt{x}$, $x \geqslant 0$. *It is continuous, but the derivative at* $x = 0$ *does not exist, since*

$$\frac{f(h) - f(0)}{h} = \frac{\sqrt{h}}{h} = \frac{1}{\sqrt{h}}$$

diverges.

3.2.2 monotonicity

The first application of derivative is to judge the monotonicity of a function.

Theorem 3.2.3 Assume that $y = f(x)$, $x \in (a, b)$ is smooth.

1. If $f'(x) \geqslant 0$ for any $x \in (a, b)$, then f is increasing ($\nearrow$) on (a, b).

2. If $f'(x) > 0$ for any $x \in (a, b)$, then f is strictly increasing on (a, b).

The statements for decreasing are similar.

The inverse of the first statement is also true, but the inverse of the 2nd statement is not true. For example $y = x^3$ is strictly increasing on $\mathbf{R}$ but $f'(0) = 0$.
The proof is simple. For any $b > x_2 > x_1 > a$, by mean-value theorem, there exists $\xi \in (x_1, x_2)$ such that

$$f(x_2) - f(x_1) = f'(\xi)(x_2 - x_1).$$

Then statements 1 and 2 follow directly.

Example 3.2.2 *1.* $y = x^3 - 3x^2 + 1$. $y' = 3x^2 - 6x = 3x(x - 2)$. $y' = 0$ *has two roots: 0,2. On* $(-\infty, 0)$, $f' > 0$, *so* $f \nearrow$, *on* $(0, 2)$, $f' < 0$, *so* $f \searrow$ *and on* $(2, +\infty)$, $f' > 0$ *so* $f \nearrow$.
2. $y = x^{1/3}(2 - x)$. *The function is continuous on* $\mathbf{R}$. *Taking derivative*

$$y' = \frac{1}{3x^{2/3}}(2 - x) - x^{1/3} = \frac{2 - 4x}{3x^{2/3}}.$$

It is smooth except $x = 0$. *We should consider intervals* $(-\infty, 0)$, $(0, 1/2)$ *and* $(1/2, \infty)$. *Note that* $x^{2/3} > 0$ *always. Hence on* $(-\infty, 0)$, $f' > 0$, *so* $f \nearrow$, *on* $(0, 1/2)$, $f' > 0$, *so* $f \nearrow$ *and on* $(1/2, +\infty)$, $f' < 0$ *so* $f \searrow$.

We may use the derivative to prove inequality.

1. Prove that $e^x \geqslant 1 + x$ for any $x \in \mathbf{R}$. Set $f(x) = e^x - (1 + x)$ and take derivative $f'(x) = e^x - 1$. It is seen that $f(x)$ is decreasing when $x < 0$ and increasing when $x > 0$. Hence $f(x) \geqslant f(0) = 0$.

2. Prove that $(1 + x)^a \geqslant 1 + ax$ for $a > 1$ and $x \geqslant 0$. Set $f(x) = (1 + x)^a - (1 + ax)$ and take derivative

   $$f'(x) = a(1 + x)^{a-1} - a = a((1 + x)^{a-1} - 1) > 0.$$

   Hence $f(x) \geqslant f(0) = 0$.

Theorem 3.2.4 Assume that $y = f(x),\ x \in (a, b)$. If $f'(x) = 0$ for any $x \in (a, b)$, then $f(x)$ is constant on (a, b).

In fact fix a point $x_0 \in (a, b)$ and for any other point $x \in (a, b)$, by mean-value theorem, there exists ξ between x_0 and x, such that

$$f(x) - f(x_0) = f'(\xi)(x - x_0) = 0.$$

Hence $f(x) = f(x_0)$.

3.2.3 L'Hôpital's rule

Another application is L'Hôpital's rule, which is an efficient way to compute limit. How to compute the limit

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

when both f and g converge to 0?

Theorem 3.2.5 If $\lim_{x\to a} f(x) = \lim_{x\to a} g(x) = 0$, and f, g are differentiable at a with $g'(a) \neq 0$, then

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

Assuming $f(a) = g(a) = 0$, the formula follows from

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

This formula is more conveniently written as follows when the derivative function is continuous,

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

If the limit on the right side is still $\frac{0}{0}$-type, we could apply the formula again until the condition in Theorem 3.2.5 is satisfied.

Example 3.2.3 *This limit is usually called the type* $\frac{0}{0}$.

$$\lim_{x\to 0}\frac{e^x-1-x}{x^2} = \lim_{x\to 0}\frac{e^x-1}{2x} = \lim_{x\to 0}\frac{e^x}{2} = \frac{1}{2}.$$
$$\lim_{x\to 0}\frac{1-\cos x}{x^2} = \lim_{x\to 0}\frac{\sin x}{2x} = \lim_{x\to 0}\frac{\cos x}{2} = \frac{1}{2}.$$
$$\lim_{x\to 0}\frac{x-\sin x}{x^3} = \lim_{x\to 0}\frac{1-\cos x}{3x^2} = \frac{1}{6}.$$

This formula can also be used to the limit of the type $\frac{\infty}{\infty}$.

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

However the L'Hopital's rule does not always work. For example,

$$\lim_{x\to 0}\frac{x^2\sin(1/x)}{e^x-1}$$

is $\frac{0}{0}$ type and has limit 0, but L'hopital's rule does not work, because it gives

$$\lim_{x\to 0}\frac{x^2\sin(1/x)}{e^x-1} = \lim_{x\to 0}\frac{2x\sin(1/x)-\cos(1/x)}{e^x}$$

which diverges.
Though it is named as L'Hôpital's rule, most people believe that the credit of this theorem should be given to Johann Bernoulli.

Exercises

1. Prove that $y = x - sinx$ is strictly increasing.

2. Find all solutions of
$$\frac{x-1}{x+1} = \ln x.$$

3. Find the limit. Using l'Hopital's rule where appropriate. If there is a more elementary method, using it. If l'Hopital's rule does not work, explain why.

   (a) $\lim_{x\to+\infty} x(\pi - 2\arctan x)$;

   (b) $\lim_{x\to\pi/2} \frac{\cos x}{1-\sin x}$;

   (c) $\lim_{x\to 0} \frac{x-\sin x}{x+\sin x}$;

   (d) $\lim_{x\to 1} \frac{x^a - ax + a - 1}{(x-1)^2}$;

   (e) $\lim_{x\to 0+} x^{x^2}$;

   (f) $\lim_{x\to 0+}(4x+1)^{\cot x}$;

   (g) $\lim_{x\to 0+}(\cos x)^{1/x^2}$.

4. Find the intervals where f is monotone and how many real roots for $f(x) = 0$.

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

   (b) $f(x) = x^4 - 4x^3 - 8x^2 + 2$.

5. prove that $\ln(1+x) \leqslant x$ for $x > -1$. Set $f(x) = x - \ln(1+x)$, $x > 0$. find a suitable constant a so that $f(x) \leqslant ax^2$ holds for any $x > 0$.

6. For $b > a > 1$, which one is bigger: a^b and b^a? prove.

3.3 application of derivatives 2

3.3.1 max/min of a continuous function on a closed interval

The maximum and minimum,which we defined before, are called global extremum, because they are the biggest and smallest respectively on whole domain. The point where the function reaches global extremum is called the global extremal point. A continuous function on a closed interval reaches the global extremum and in this section, we shall discuss how to find the global extremal points. To this end, we need to discuss local extremum first.

Definition 3.3.1 Assume that $y = f(x), x \in (a, b)$. We say f reaches local maximum at $x_0 \in (a, b)$ if there exists $\delta > 0$ such that

$$f(x_0) \geqslant f(x)$$

for any $x \in (x_0 - \delta, x_0 + \delta)$.

Intuitively f reaches maximum at x_0 in a neighborhood of x_0 or locally. The local minimum may be defined similarly. The local maximum and local minimum are called local extremum and the point where the function reaches local extremum is called a local extremal point.

The following lemma was proved in the previous lecture when we proved Rolle's theorem.

Lemma 3.3.2 (Fermat) If $y = f(x),\ x \in (a, b)$ reaches a local extremum at $c \in (a, b)$ and f is smooth at c, then $f'(c) = 0$.

Definition 3.3.3 Assume that $y = f(x),\ x \in (a, b)$. A point $c \in (a, b)$ is called critical if either f is smooth at c and $f'(c) = 0$ or f is not smooth at c.

Theorem 3.3.4 Assume that $y = f(x),\ x \in (a, b)$. If f reaches a local extremum at $c \in (a, b)$ then c is critical.

Note a critical point may not be a local extremal point. For example, $y = x^3,\ x \in \mathbf{R}$ for which $x = 0$ is a critical point but this function is strictly increasing on whole line. How to tell a function reaches a local extremum at a critical point? It is obvious that if the function is increasing on one side and decreasing on the other side, then the point is a local extremal point, and if the function is increasing on both sides or decreasing on both sides, then it is not.

Example 3.3.1 *1.* $y = x^3 - 3x^2 + 1$. $y' = 3x^2 - 6x = 3x(x - 2)$. $y' = 0$ *has two roots: 0,2, which are critical points. At* $x = 0$*, the function is increasing on the left and decreasing on the right. Hence it reaches local maximum at* 0*. The same reason tells that* f *reaches local minimum at* 2*.*

2. $y = x^{1/3}(2 - x)$*. The function is continuous on* $\mathbf{R}$*. This function also has two critical points* $0, 1/2$*, in which* 0 *is a non-smooth point and* $1/2$ *is a smooth point,* $f'(1/2) = 0$*. At* 0*, the function* f *is increasing on both sides. Hence* 0 *is not a local extremal point of* f*. But* $1/2$ *is a local maximal point of* f*.*

3. The function $y = |x|$ *has a non-smooth point* $x = 0$*, which is also a local minimal point.*

We know that a continuous function $y = f(x)$ on closed interval $[a, b]$ will reach its global maximum and minimum. How to find the points where f reaches the global extremum?

Theorem 3.3.5 The global extremum of a continuous function on a closed interval must be attained at either critical points or boundaries a, b.

Here gives a routine.Assume that $y = f(x)$, $x \in [a, b]$.

1. Find the points where f is not smooth.

2. Take derivative and find the solution of $f'(x) = 0$.

3. Put all critical points and a, b in a set A.

4. Compute $f(c)$ for all $c \in A$. The biggest is the maximum and the smallest is the minimum.

Example 3.3.2 *1.* $y = x^3 - 3x^2 + 1$, $x \in [-1, 2]$. *The critical point is* 0*, and boundaries are* $-1, 2$*. Compute* $f(-1) = -3$*,* $f(0) = 1$*,* $f(2) = -3$*. Hence the maximum value is* 1*, reached at* 0 *and minimum value is* -3*, reach at boundary* $-1, 2$*.*

2. $y = x^{1/3}(2 - x)$, $x \in [-1, 1]$*. The points we need to pay attention are* $-1, 0, 1/2, 1$*. Compute* $f(-1) = -3$*,* $f(0) = 0$*,* $f(1/2) = 3/(2 \cdot 2^{1/3}) > 1$*,* $f(1) = 1$*. Hence the maximum is attained at* $1/2$ *and the minimum is reached at* -1*.*

3. $y = (x - 1)|x^2 - 1|$, $x \in [-1.1, 2]$*. The roots of* $x^2 - 1 = 0$ *is non-smooth points. When* $x^2 - 1 > 0$*, i.e.,* $x \in (-\infty, -1) \cup (1, \infty)$*,* $y = (x - 1)(x^2 - 1)$*, and*

$$y' = (x^2 - 1) + 2x(x - 1) = (x - 1)(3x + 1) > 0.$$

When $x^2 - 1 < 0$*, i.e.,* $x \in (-1, 1)$*,* $y = -(x - 1)(x^2 - 1))$*, and*

$$y' = -(x - 1)(3x + 1) \begin{cases} < 0, & x < -1/3, \\ > 0, & x > -1/3. \end{cases}$$

The line is divided into 4 intervals

$$\begin{array}{cccc} (-\infty, -1), & (-1, -1/3), & (-1/3, 1), & (1, +\infty), \\ \nearrow & \searrow & \nearrow & \nearrow. \end{array}$$

critical points $\{-1, -1/3, 1\}$*, boundary* $\{-1.1, 2\}$*. Comparing the values on these points, the maximum 3 at* $x = 2$ *and the minimum* $-32/27$*.*

3.3.2 higher order derivatives and applications

We have seen that the derivative of a function can be used to find where the function is monotone and where the function reaches its extremum. If the function $y = f(x),\ x \in (a, b)$ is smooth, then its derivative is still a function $y = f'(x),\ x \in (a, b)$. If f' is smooth on (a, b), we may take derivative again to obtain the derivative of derivative, or the 2nd order derivative, denoted by $f''(x)$ and y''. Similarly if the function is nice enough, we may take derivative of n times, called n-th order derivative and denoted by $f^{(n)}(x)$ and $y^{(n)}$.

Example 3.3.3 *1.* $y = x^n$, $y'' = n(n-1)x^{n-2}$.

2. $y = e^{ax}$, $y^{(n)} = e^{ax}a^n$.

3. $y = e^{x^2}$, $y' = e^{x^2} \cdot 2x$, $y'' = e^{x^2}(4x^2 + 2)$.

4. $y = \sin x$, $y' = \cos x$, $y'' = -\sin x$,

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

The distance y a particle moves is a function of time x: $y = f(x)$. The derivative y' will the velocity and the 2nd derivative y'' will be the rate of change of velocity, i.e., acceleration. The 2nd derivative describes the convexity of a function.

Definition 3.3.6 The function $y = f(x),\ x \in (a, b)$ is called convex if for any $x_1, x_2 \in (a, b)$,

$$f(\frac{x_1 + x_2}{2}) \leqslant \frac{f(x_1) + f(x_2)}{2}.$$

The function f is called concave if $y = -f(x)$ is convex. The point x_0 is called an inflection point of f if f exhibits different convexities on two sides of x_0.

It can be proved by induction that when f is convex on (a, b), for any $n \in \mathbf{N}$ and $x_1, x_2, \cdots x_n \in (a, b)$ it holds that

$$f(\frac{1}{n}\sum_{j=1}^{n} x_j) \leqslant \frac{1}{n}\sum_{j=1}^{n} f(x_j).$$

The next result is more useful.

Theorem 3.3.7 Assume f is continuous on (a, b). Then f is convex if and only if for any $x_1, x_2 \in (a, b)$, the secant connecting $(x_1, f(x_2))$ and $(x_2, f(x_2))$ is above f on (x_1, x_2).

The statement that the secant connecting $(x_1, f(x_2))$ and $(x_2, f(x_2))$ is above f on (x_1, x_2), abbreviated as 'the secant is above f on (x_1, x_2)', means that for any $t \in (0, 1)$,

$$f(tx_1 + (1-t)x_2) \leqslant tf(x_1) + (1-t)f(x_2).$$

The definition of convexity requires only $t = 1/2$. The proof is explained roughly. Suppose that there exists $c \in (x_1, x_2)$, such that $(c, f(c))$ is above the secant. Then there are c_1, c_2

satisfying $x_1 \leqslant c_1 < c < c_2 \leqslant x_2$ such that the secant is below f on (c_1, c_2) and that contradicts to the definition.

It is hard to tell if a function is convex by the definition or theorem above by trying a few examples like $y = x^2$ and $y = e^x$. The second derivative provides a more efficient criterion. The following lemma is a very intuitive fact.

Lemma 3.3.8 Assume that f is smooth on (a, b). If for two points $x_1 < x_2$ in (a, b), the secant is above f on (x_1, x_2), then

$$f'(x_1) \leqslant \frac{f(x_2) - f(x_1)}{x_2 - x_1} \leqslant f'(x_2).$$

Proof. Since the secant is above f, for any $x \in (x_1, x_2)$ it holds that

$$\frac{f(x) - f(x_1)}{x - x_1} \leqslant \frac{f(x_2) - f(x_1)}{x_2 - x_1}.$$

The limit of the left side is $f'(x_1)$ and then the conclusion follows from the comparison theorem. □

This lemma implies that f is convex if and only if f is above the tangent line of f at any $x_0 \in (a, b)$. This lemma implies also that the derivative f' is increasing.

Theorem 3.3.9 Assume that $y = f(x)$, $x \in (a, b)$ is smooth. If f' is increasing on (a, b), then f will be convex.

We will give an illustration. Assume that $y = f(x)$, $x \in (a, b)$, and f' increases on (a, b). We claim that for any $a < x_1 < x_2 < b$, the secant is above f on (x_1, x_2). Without loss of generality, we may assume that the secant is horizontal, $f(x_1) = f(x_2)$. Then there exists $c \in (x_1, x_2)$ such that $f'(c) = 0$. Hence

$$f'(x) \leqslant 0,\ x < c \text{ and } f'(x) \geqslant 0,\ x > c.$$

It means that f decreases on (x_1, c) and increases on (c, x_2). This implies that the curve is below the secant between (x_1, x_2).

The following conclusion is obvious.

Theorem 3.3.10 Assume that f is 2-nd order differentiable on (a, b). If $f'' \geqslant 0$ on (a, b), then f is convex. The inverse is also true. If x_0 is an inflection point, then $f''(x_0) = 0$.

Example 3.3.4 *1.* $y = e^x$*, is convex on* $\mathbf{R}$*.*

Then for any $x_1, \cdots, x_n$*, we have*

$$e^{\frac{1}{n}(x_1+\cdots+x_n)} \leqslant \frac{1}{n}(e^{x_1} + \cdots + e^{x_n}).$$

Set $a_j := e^{x_j} > 0$ *and the mean-value inequality holds*

$$\sqrt[n]{a_1 \cdots a_n} \leqslant \frac{1}{n}(a_1 + \cdots + a_n).$$

2. $y = x^n$. $y'' = n(n-1)x^{n-2}$. *When* n *is odd,* $y = x^n$ *is convex for* $x > 0$, *concave for* $x < 0$ *and* 0 *is an inflection point. When* n *is even,* $y = x^n$ *is convex.* 0 *is not an inflection point.*

3. $y = \sin x$, $x \in (0, 2\pi)$. $y'' = -\sin x$. *Then* $\sin x$ *is concave on* $(0, \pi)$, *convex on* $(\pi, 2\pi)$ *and* π *is an inflection point.*

3.3.3 differentiation

Fix $x_0 \in (a, b)$, $\Delta x = x - x_0$ is called an increment of x and $\Delta y = f(x) - f(x_0)$ is called the corresponding increment of y. The derivative $f'(x_0)$ is the limit of the ratio of increments

$$\frac{\Delta y}{\Delta x}$$

if exists. Therefore Leibniz intuitively turned Δ into 'd' and wrote the derivative as

$$\frac{dy}{dx} := \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}.$$

When the limit exists, called f 'differentiable at x_0'. In Leibniz's language, taking derivative is called 'differentiating', this action is called 'differentiation' and dy is called the differential of y.
It is seen that

$$\Delta y \sim dy,$$

i.e., the increment of y can be approximated by the differential of y. Since

$$\frac{dy}{dx} = f'(x_0),$$

the differential $dy = f'(x_0)dx$, which is easier to compute than the increment

$$\Delta y = f(x_0 + \Delta x) - f(x_0).$$

This gives us a method to compute the value of f near x_0 approximately

$$f(x_0 + \Delta x) \approx f(x_0) + f'(x_0)\Delta x.$$

Approximate computation was important before computer was developed, but it is no longer an important aspect of Calculus.

Exercises

1. Find the global maximum/minimum of f on the given interval.

   (a) $f(x) = 2x^3 - 3x^2 - 12x + 1$, $[0, 3]$.

   (b) $f(x) = x^4 - 4x^2 + 2$, $[-3, 2]$.

   (c) $f(x) = x\sqrt{4 - x^2}$, $[-1, 2]$.

2. Find the n-th order derivatives for $y = \sqrt{1 + x}$ at $x = 0$.

3. Analyse the monotonicity and convexity of $y = \frac{x}{1+x^2}$, and sketch the graph.

4. Assume f is smooth on (a, b). Prove that f is convex if and only if for any $x_0 \in (a, b)$, f is above the tangent line of f at x_0.

Chapter 4

integral and application

4.1 indefinite integrals 1

4.1.1 anti-derivatives

The anti-derivative is the inverse operation of taking derivative.

Definition 4.1.1 Assume that $y = f(x)$ and $y = g(x)$ are defined on (a, b). If $(g(x))' = f(x)$ for all $x \in (a, b)$, then we say g is an anti-derivative, or primitive function or infinite integral of f, denoted by $\int f(x)dx$. The function f is called the integrand. Computing an anti-derivative or indefinite integral is called to integrate f. The action is called integration.

Hence

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

The anti-derivative is not unique but they differ by a constant.

Theorem 4.1.2 The different anti-derivatives of f differ by a constant.

If g_1 and g_2 are two anti-derivatives of f, then for any $x \in (a, b)$,

$$(g_2(x) - g_1(x))' = g_2'(x) - g_1'(x) = f(x) - f(x) = 0$$

and it follows from Theorem 3.2.4 that $g_2 - g_1$ is a constant. Hence the function $y = f(x) + c$ is always an anti-derivative of f', i.e.,

$$\int f'(x)dx = f(x) + c.$$

For example, it is not surprising that both $\sin^2 x$ and $-\cos^2 x$ are anti-derivatives of $\sin 2x$, because $\sin^2 x = -\cos^2 x + 1$. Hence both

$$\int \sin 2x dx = \sin^2 x + c, \text{ and } \int \sin 2x dx = -\cos^2 x + c,$$

are correct. When you integrate a function, it suffices to obtain one correct anti-derivative. The constant $+c$ is added just for the completeness of answer. Leibniz's notation plays a role here. Since $\dfrac{df(x)}{dx} = f'(x)$, $df(x) = f'(x)dx$ or better to write $f'(x)dx = df(x)$. For example, we should be able quickly to write the derivative in the reversal direction,

$$2xdx = d(x^2),\ e^x dx = de^x,\ \frac{1}{x}dx = d\ln x,\ \sin x dx = -d\cos x.$$

Then

$$\int df(x) = \int f'(x)dx = f(x) + c.$$

It seems that $\int$ and d cancel each other and they are inverse each other: informally

$$\int^{-1} = d,\ \int = d^{-1}.$$

It is much more difficult to compute anti-derivative than to compute derivative. In many cases, it is even impossible to find explicit expression for anti-derivatives. For example, the anti-derivative

$$\int e^{-x^2}dx$$

is no longer an elementary function. It is like saying that we can not solve an equation, for example $-x = e^x$. It does not mean that the equation has no solution. The equation has solution but there is no way to know what precisely it is.

4.1.2 basic integral formulae

From derivative formulae it is immediate to have the following basic formulae.

1. $\int 0dx = c$.
2. If $a \neq -1$,
$$\int x^a dx = \frac{x^{a+1}}{1+a} + c.$$
3. Since $(\ln x)' = \frac{1}{x}$ for $x > 0$ and $(\ln(-x))' = \frac{1}{-x} \cdot (-1) = \frac{1}{x}$ for $x < 0$, we have
$$\int x^{-1}dx = \ln|x| + c.$$
4. If $a > 0$, $a \neq 1$, then
$$\int a^x dx = \frac{a^x}{\ln a} + c.$$
5. $\int \sin x dx = -\cos x + c$.
6. $\int \cos x dx = \sin x + c$.

7.
$$\int \frac{1}{\cos^2 x} dx = \tan x + c.$$

8.
$$\int \frac{1}{\sin^2 x} dx = -\cot x + c.$$

9. Since $(\arcsin x)' = \frac{1}{\sqrt{1-x^2}}$,
$$\int \frac{1}{\sqrt{1-x^2}} dx = \arcsin x + c.$$

10. Since $(\arctan x)' = \frac{1}{1+x^2}$,
$$\int \frac{1}{1+x^2} dx = \arctan x + c.$$

Some functions are absent from this list, for example, $\ln x$, $\tan x$ and $\arcsin x$ etc, because we do not know immediately what their primitive functions are. Students may add more to this list afterwards as long as you think they are basic enough to remember.
The reasons that taking derivative is no longer a problem are four operations and chain rule for derivatives, which are no longer available for anti-derivative.

4.1.3 linearity of integration

Two operations are still available: addition and scalar multiplication.

$$\int af(x)dx = a\int f(x)dx.$$
$$\int (f(x)+g(x))dx = \int f(x)dx + \int g(x)dx.$$

To prove these formulae, it suffices to verify that both sides have the same derivative. But this is obvious by the definition of anti-derivatives.

Example 4.1.1

$$\int (2+x+3x^2)dx = \int 2dx + \int xdx + 3\int x^2 dx = 2x + \frac{x^2}{2} + x^3 + c.$$
$$\int (2\sin x + 3x^{-1})dx = 2\int \sin x dx + 3\int x^{-1}dx = -2\cos x + 3\ln x + c.$$

But these formulae are not powerful enough to integrate functions like xe^x, $x\sin x$, $\ln x$. We need more techniques.

4.1.4 integration by parts

Fortunately, though we can not have a complete product rule, the product rule of derivative can be partly helpful.
The product rule of derivative:

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

In Leibniz's notation:

$$d(f(x)g(x)) = g(x)f'(x)dx + f(x)g'(x)dx = g(x)df(x) + f(x)dg(x).$$

Then integrate both sides we have the following theorem.

Theorem 4.1.3 (integration by parts)

$$f(x)g(x) = \int g(x)df(x) + \int f(x)dg(x).$$

This formula is called integration by parts. Students may be puzzled about how to use this formula. Actually the formula does not give us a complete solution, but it tells us that if one of two integrals in the right side can be integrated, then the other can also be integrated. This would help us to integrate many functions.

Example 4.1.2 *1. Starting with* $\ln x$*,*

$$\begin{aligned}\int \ln x dx &= x\ln x - \int x d\ln x \\ &= x\ln x - \int x\cdot x^{-1}dx = x\ln x - \int dx = x\ln x - x + c.\end{aligned}$$

2. $\int xe^x dx = \int xde^x = xe^x - \int e^x dx = xe^x - e^x + c.$

3. Using integration by parts

$$\int e^x \sin x dx = \int \sin x de^x = e^x \sin x - \int e^x d\sin x = e^x \sin x - \int e^x \cos x dx.$$

Using it again,

$$\int e^x \cos x dx = \int \cos x de^x = e^x \cos x - \int e^x d\cos x = e^x \cos x + \int e^x \sin x dx.$$

Come back to the same integral, but with different sign, and then

$$\int e^x \sin x dx = \frac{1}{2}e^x(\sin x - \cos x) + c.$$

Students may take derivative of the right side to verify the answer.

We said that this is not a complete solution, because it does not always work. For example, you may not be able to integrate

$$\int \tan x dx \text{ and } \int \arctan x dx,$$

by using integration by parts only.

Exercises

1. Find the anti-derivative of the function.
   (a) $f(x) = x - 3$;
   (b) $f(x) = 6\sqrt{x} - \sqrt[6]{x}$;
   (c) $f(u) = \cos u - 5 \sin u$;
2. Find f.
   (a) $f''(x) = 6x + 12x^2$;
   (b) $f'''(t) = 60t^2$;
   (c) $f''(x) = 24x^2 + 2x + 10,\ f(1) = 5,\ f'(1) = -3$;
   (d) $f''(x) = 2 - 12x,\ f(0) = 9,\ f(2) = 15$.
3. Verify by definition
$$\int \cos^3 x dx = \sin x - \frac{1}{3}\sin^3 x + C.$$
4. Find the indefinite integral.
   (a) $\int \frac{3x-4}{x^2+1} dx$;
   (b) $\int \sqrt{1+x^2} x^5 dx$;
   (c) $\int x^2 \cos 3x dx$;
   (d) $\int x^{-2} \ln x dx$;
   (e) $\int x^4 (\ln x)^2 dx$.

4.2 indefinite integrals 2

4.2.1 change of variable

Another incomplete solution comes from the chain rule of derivative. The chain rule is

$$(f(g(x)))' = f'(g(x))g'(x),$$

but in Leibniz's notation

$$df(g(x)) = f'(g(x))g'(x)dx = f'(g(x))dg(x) = f'(y)dy$$

where a substitution $y = g(x)$ is used. This means that the anti-derivative of the function which can be written into

$$f'(g(x))g'(x)$$

is $f(g(x))$. For example $(\sin x^2)' = 2x\cos x^2$, it implies that

$$\int 2x\cos(x^2)dx = \int \cos(x^2)d(x^2) = \sin(x^2) + c,$$

where in our mind, we replace x^2 with y and use the basic formula. Precisely

$$\int f'(g(x))g'(x)dx = \int f'(g(x))dg(x) = f(g(x)) + c.$$

This formula gives us an idea to integrate, called change of variable, or substitution rule.

Theorem 4.2.1 (change of variable) Set $y = g(x)$. we have identity

$$\int f(g(x))g'(x)dx = \int f(y)dy.$$

We may use the simple linear substitution $y = ax + b$ to simplify the integrand. Note that to simplify the integrand is the important step to integrate because this makes us know what is the essential problem we need to attack.

Example 4.2.1

$$\int (2x+4)^3 dx = \frac{1}{2}\int (2x+4)^3 d(2x+4) = \frac{1}{2}\int y^3 dy,\ y = 2x+4;$$

$$\int \ln(3x-1)dx = \frac{1}{3}\int \ln y dy,\ y = 3x-1,\ dy = 3dx,$$

$$\begin{aligned}\int \frac{1}{x^2+x+1}dx &= \int \frac{dx}{(x+1/2)^2+3/4} = \frac{1}{3/4}\int \frac{dx}{(\frac{x+1/2}{\sqrt{3/4}})^2+1} \\ &= \frac{2}{\sqrt{3}}\int \frac{1}{y^2+1}dy,\ y = \frac{x+1/2}{\sqrt{3/4}},\ dy = \frac{dx}{\sqrt{3/4}}.\end{aligned}$$

$$\int \frac{1}{(2x+3)^n}dx = \frac{1}{2}\int \frac{1}{u^n}du,\ u = 2x+3.$$

Similarly this formula does not give us a complete solution and it only tells us that if we can integrate one side, then we have the answer for the other side.

Example 4.2.2 *Let's see if this formula works in some cases.*

1.

$$\begin{aligned}\int \tan x dx &= \int \frac{\sin x}{\cos x}dx = -\int \frac{d\cos x}{\cos x} \\ &= -\int \frac{dy}{y} = -\ln|y| + c = -\ln|\cos x| + c.\end{aligned}$$

2. *Using both formulae,*

$$\begin{aligned}\int \arctan x dx &= x\arctan x - \int x d\arctan x\\ &= x\arctan x - \int \frac{x}{1+x^2}dx\\ &= x\arctan x - \frac{1}{2}\int \frac{d(1+x^2)}{1+x^2}\\ &= x\arctan x - \frac{1}{2}\ln(1+x^2)+c.\end{aligned}$$

3. *Using both formulae,*

$$\begin{aligned}\int \arcsin x dx &= x\arcsin x - \int x d\arcsin x\\ &= x\arcsin x - \int \frac{xdx}{\sqrt{1-x^2}}\\ &= x\arcsin x + \frac{1}{2}\int \frac{dy}{\sqrt{y}}, \quad y = 1-x^2,\ dy = -2xdx\\ &= x\arcsin x + \frac{1}{2}(-\frac{1}{2})y^{-1/2+1}+c\\ &= x\arcsin x - \frac{1}{4}(1-x^2)^{1/2}+c.\end{aligned}$$

When $\int$ *disappears, you should add* $+c$ *there.*

In this example, we use change of variable formula in theorem above to integrate the right side and then obtain the answer for the left side, i.e., the right helps the left. However there are also examples later where the left helps the right.

4.2.2 rational functions

There is no problem to integrate a polynomial. A polynomial here means a polynomial with real coefficients. The ratio of two polynomials is called a rational function, for example $\frac{x^3+1}{x-1}$ and $\frac{x^2+1}{x^3+x-2}$. We would say every rational function can be integrated. When the degree of numerator is less than the degree of denominator, it is called a proper rational function. Through polynomial division, a rational function can be written into the sum of a polynomial and a proper rational function.

Example 4.2.3 *We will see how to do division first.*

1. $\dfrac{x^3+1}{x-1} = \dfrac{x^3-x^2+x^2-x+x-1+2}{x-1} = x^2+x+1+\dfrac{2}{x-1}.$

2. $\dfrac{x^3+1}{x^2+1} = \dfrac{x^3+x-x+1}{x^2+1} = x+\dfrac{-x+1}{x^2+1}.$

Hence we may only consider proper rational functions.
Students may have felt that polynomials are like integers, rational functions are like rational numbers and proper rational functions are like proper fraction. Yes they are similar. As every integer is a product of some prime numbers with multiples, for example $72 = 3^2 \times 2^3$, every polynomial is a product of some prime polynomials with multiples, for example

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

A prime polynomial, which is defined to be a polynomial with no non-trivial factors, is either a polynomial of either degree 1, $x - a$, or of degree 2, $x^2 + bx + c$, but no real roots. When the degree of the denominator polynomial is 1 or 2, we can do polynomial division to make the degree of numerator polynomial is less. A fraction can be represented by

$$\frac{5}{72} = \frac{5}{8} - \frac{5}{9} = \frac{1}{2} + \frac{1}{2^3} - \frac{1}{3} - \frac{2}{3^2}.$$

Similarly any proper rational function is a linear combination of rational functions like

$$\frac{1}{(x-a)^n}, \quad \frac{1}{(x^2+bx+c)^n}, \quad \frac{x}{(x^2+bx+c)^n},$$

where $x^2 + bx + c = 0$ has no real roots.

Example 4.2.4 *1. Consider*

$$\frac{1}{x^3-1} = \frac{1}{(x-1)(x^2+x+1)} = \frac{a}{x-1} + \frac{bx+c}{x^2+x+1},$$

where a, b, c *are to be determined. With common denominator, we have*

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

Comparing the coefficients, we obtain the equations

$$\begin{cases} a + b = 0, \\ a + c - b = 0, \\ a - c = 1. \end{cases}$$

The solution $a = 1/3, b = -1/3, c = -2/3$. *Hence*

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

which can be integrated according to the previous example.

2. Consider

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

which can be integrated easily.

3. *Some are difficult.*

$$\frac{1}{(x^4-1)^2} = \frac{1}{(x-1)^2(x+1)^2(x^2+1)^2}$$
$$= \frac{a}{x-1} + \frac{b}{(x-1)^2} + \frac{c}{x+1} + \frac{d}{(x+1)^2} + \frac{e_1x+e_0}{x^2+1} + \frac{f_1x+f_0}{(x^2+1)^2}.$$

We leave it to students to find these coefficients.

The discussions above are summarized as follows. The problem of integrating a rational function is reduced to integrating the following three types of rational functions by substitution in Example 4.2.1.3,

$$x^{-n}, \quad \frac{1}{(x^2+1)^n}, \quad \frac{x}{(x^2+1)^n}.$$

The first type is easy and the third type can be integrated by substitution rule

$$\int \frac{x}{(x^2+1)^n}dx = \frac{1}{2}\int \frac{d(x^2+1)}{(x^2+1)^n} = \frac{1}{2}\int \frac{dy}{y^n}, \; y = x^2+1,$$

which is easy too.

The last problem is how to integrate

$$\int \frac{1}{(x^2+1)^n}dx.$$

We need powerful trigonometric substitutions introduced in next lecture.

Exercises

1. Find the indefinite integral
   (a) $\int xe^{x^2}dx$;
   (b) $\int \frac{x^2}{\sqrt{5+x^3}}dx$;
   (c) $\int \frac{e^x dx}{\sqrt{e^x+1}}$;
   (d) $\int \frac{e^x}{e^{2x}+1}dx$.
2. Determine a, b
$$\frac{x+5}{x^2+x-2} = \frac{a}{x-1} + \frac{b}{x+2}.$$
   and then integrate it.
3. Determine the polynomial $S(x)$ and $R(x)$
$$\frac{x^3-x}{x^2+1} = S(x) + \frac{R(x)}{x^2+1},$$
   where the degree of $R(x)$ is less than 2, and then integrate it.

4. Evaluate
$$\int \frac{x^2+2x-1}{2x^3-3x^2-2x}dx.$$

5. Evaluate $\int \frac{x^n}{x^4+1}dx$ for $n = 0, 1, 2, 3$. Hint: $x^4 + 1 = (x^2 + 1)^2 - 2x^2$.

4.3 indefinite integrals 3

4.3.1 trigonometric substitution

Recall the substitution rule: by a substitution $y = g(x)$, we have

$$\int f(g(x))dg(x) = \int f(y)dy.$$

As we said before, sometimes the left helps the right and sometimes the other way around. The trigonometric substitutions

$$y = \sin x,\ y = \frac{1}{\sin x},\ y = \tan x,$$

from $x \to y$ and $y \to x$, are very useful in integrating some particular types of functions, which contain $\sqrt{1-x^2}$, $\sqrt{x^2-1}$, x^2+1. It should be remembered that

$$d(\sin x) = \cos x dx,\ d(\sin x)^{-1} = -\frac{\cos x}{\sin^2 x}dx,\ d(\tan x) = \frac{1}{\cos^2 x}dx.$$

We have seen some examples before. There is no general rule here but we may obtain experiences from examples.

Example 4.3.1 *We may try the substitution* $x = \sin t$ *when we see* $1 - x^2$.

1.

$$\begin{aligned}\int \sqrt{1-x^2}dx &= \int \sqrt{1-\sin^2 t}d\sin t,\ x = \sin t,\\ &= \int \cos^2 tdt\\ &= \frac{1}{2}\int(1+\cos 2t)dt = \frac{1}{2}\left(\int dt + \frac{1}{2}\int \cos 2td(2t)\right)\\ &= \frac{1}{2}t + \frac{1}{4}\sin 2t + c\\ &= \frac{1}{2}\arcsin x + \frac{1}{2}x\sqrt{1-x^2} + c,\end{aligned}$$

where $\sin 2t = 2\sin t\cos t = 2x\sqrt{1-x^2}$.

2. If we do not use the basic formula, how do we integrate $\frac{1}{\sqrt{1-x^2}}$*?*

$$\int \frac{1}{\sqrt{1-x^2}}dx = \int \frac{d\sin t}{\cos t} = \int dt = t + c = \arcsin x + c.$$

As two examples show, no matter what substitutions are used, the final answer should be a function of x if we integrate a function of x in the beginning. However for saving time, we shall stop writing whenever we know how to integrate the function

Example 4.3.2 *1. To integrate the second type integral left in rational functions, we use substitution* $x = \tan y$*,*

$$\begin{aligned}
\int \frac{1}{(x^2+1)^n} dx &= \int \frac{d\tan y}{(\tan^2 y + 1)^n} \\
&= \int \frac{1}{(\cos y)^{-2n}} \cdot \frac{1}{\cos^2 y} dy \\
&= \int \cos^{2n-2} y dy \\
&= \int \cos^2 y dy, \ \textit{when } n = 2.
\end{aligned}$$

$$\begin{aligned}
\int \sqrt{1+x^2} dx &= \int (\cos y)^{-1} d\tan y \\
&= \int \frac{1}{\cos^3 y} dt = \int \frac{d\sin y}{(1-\sin^2 y)^2} \\
&= \int \frac{du}{(1-u^2)^2}, \ u = \sin y, \textit{a rational function}; \\
&= \int \frac{1}{4}\left(\frac{1}{1-u} + \frac{1}{(1-u)^2} + \frac{1}{1+u} + \frac{1}{(1+u)^2}\right) du \\
&= \frac{1}{2}(x\sqrt{1+x^2} + \ln(x + \sqrt{x^2+1}) + c.
\end{aligned}$$

$$\begin{aligned}
\int \frac{1}{\sqrt{1+x^2}} dx &= \int \cos y d\tan y \\
&= \int \frac{1}{\cos y} dy, \ \textit{similar to } \int \frac{1}{\sin y} dy.
\end{aligned}$$

2. Using substitution $x = \frac{1}{\cos y}$*, we have*

$$x^2 - 1 = \tan^2 y, \ dx = \frac{\sin y}{\cos^2 y} dy.$$

$$\begin{aligned}
\int \frac{1}{\sqrt{x^2-1}} dx &= \int \frac{1}{\tan y}\frac{\sin y}{\cos^2 y} dy = \int \frac{1}{\cos y} dy; \\
\int \sqrt{x^2-1} dx &= \int \tan y \frac{\sin y}{\cos^2 y} dy = \int \frac{dy}{\cos^3 y} - \int \frac{1}{\cos y} dy; \\
\int \frac{dy}{\cos^3 y} &= \int \frac{d\sin y}{\cos^4 y}, \ \sin y = u, \\
&= \int \frac{du}{(1-u^2)^2}; \ \textit{a rational function},
\end{aligned}$$

4.3.2 integrating trigonometric functions

We have seen that it is important to integrate trigonometric functions and their powers. To do that, we usually need to use semi-angle, double-angle formulae and also trigonometric substitutions.

Example 4.3.3 *We know how to integrate* $\sin x$*,* $\sin^2 x$*. It is similar to integrate* $\cos x$ *and* $\cos^2 x$*. How to integrate* $\sin^n x$ *and* $\cos^n x$*?*

$$\int \sin^3 x dx = -\int (1-\cos^2 x) d\cos x = -\int (1-y^2) dy, \ y = \cos x,$$

$$\int \sin^4 x dx = \int \left(\frac{1-\cos 2x}{2}\right)^2 dx = \frac{1}{4}\int (1 - 2\cos 2x + \cos^2 2x) dx$$
$$= \frac{1}{4}(x - \sin 2x + \frac{1}{2}\int (1+\cos 4x) dx).$$

From these examples, we may learn how to integrate $\sin^{2n} x$ *and* $\sin^{2n+1} x$*.*

Example 4.3.4 *We know how to integrate* $(\cos x)^{-1}$*,* $(\cos x)^{-2}$ *and* $(\cos x)^{-3}$*. It is similar to integrate* $(\sin x)^{-1}$*,* $(\sin x)^{-2}$ *and* $(\sin x)^{-3}$*. We may learn how to integrate* $(\sin x)^{-n}$*.*

$$\int \frac{1}{\sin^4 x} dx = -\int \frac{1}{\sin^2 x} d\cot x = -\int (1+\cot^2 x) d\cot x = -(\cos x + \frac{1}{3}\cot^3 x) + c$$
$$\int \frac{1}{\sin^5 x} dx = -\int \frac{d\cos x}{(1-\cos^2 x)^3} = -\int \frac{dy}{(1-y^2)^3}, \ y = \cos x,$$

and the rational function is decomposed into

$$\frac{1}{(1-y^2)^3} = \frac{1}{8}\left(\frac{1}{(1+y)^3} + \frac{1}{(1-y)^3} + 3\frac{1}{(1+y)^2(1-y)} + 3\frac{1}{(1+y)(1-y)^2}\right)$$
$$= \frac{1}{8}\left(\frac{1}{(1+y)^3} + \frac{1}{(1-y)^3} + 3\frac{\frac{1}{2}y+1}{(1+y)^2} + 3\frac{\frac{-1}{2}y+1}{(1-y)^2}\right).$$

Then the integration can be completed.

Example 4.3.5 *How to integrate* $\tan^n x$*? We use* $y = \tan x$ *and the identity* $1+y^2 = \frac{1}{\cos^2 x}$*.*

$$\int (\tan x)^n dx = \int \tan^n x \cos^2 x d\tan x = \int \frac{y^n dy}{1+y^2};$$
$$= \int (y - \frac{y}{1+y^2}) dy = \frac{y^2}{2} - \frac{1}{2}\ln(1+y^2) + c, \ n = 3.$$

It is seen that sometimes we need to use rational functions to integrate trigonometric function and sometimes vice versa.

4.3.3 strategy for integration

Integration is much more difficult than taking derivative. All techniques to attack a integration problem are above: 1. basic formulae, 2. integration by parts, 3. substitution rule, but there is no general rule to learn it depends essentially on our own wisdom. The observation and experience are always important, though they do not guarantee that an anti-derivative can be found.

Then the following strategy may help you to integrate.

1. basic formulae and basic techniques.

2. to simplify the integrand as simple as possible.

3. do more exercises and get more experiences. remember the types of functions you have integrated as many as possible.

4. if fail, try different approach.

We shall integrate $x^n\sqrt{x^2+1}$ to show you that the approaches are very different for different n though those functions look similar.

Example 4.3.6 *We know how to integrate* $\sqrt{x^2+1}$.

$$\begin{aligned}
\int x\sqrt{x^2+1}dx &= \int \frac{1}{2}(1+x^2)^{1/2}d(1+x^2) = \frac{1}{2}\frac{(1+x^2)^{3/2}}{3/2}+c,\\
\int x^2\sqrt{x^2+1}dx &= \int (x^2+1)^{3/2}dx - \int \sqrt{x^2+1}dx,\\
\int (x^2+1)^{3/2}dx &= \int \frac{1}{\cos^5 y}dy,\ \ x=\tan y,\\
&= \int \frac{1}{(1-\sin^2 y)^3}d\sin y\\
&= \int \frac{1}{(1-u^2)^3}du,\ \ u=\sin y.\\
\int x^3\sqrt{x^2+1}dx &= \int x(x^2+1)^{3/2}dx - \int x\sqrt{x^2+1}dx.
\end{aligned}$$

Exercises

1. Find the integral
   (a) $\displaystyle\int \frac{dx}{\cos^5 x}$;
   (b) $\displaystyle\int \frac{dx}{\cos^6 x}$;
   (c) $\displaystyle\int \frac{dx}{\tan^5 x}$;
   (d) $\displaystyle\int \sin 8x\cos 5x dx$

2. Find the integral.
   (a) $\displaystyle\int \frac{\sqrt{x^2-9}}{x^3}dx$;

(b) $\int x\sqrt{1-x^4}dx$;

(c) $\int \frac{\sqrt{1+x^2}}{x}dx$;

4.4 Riemann integrals

4.4.1 the area under curve and Riemann sums

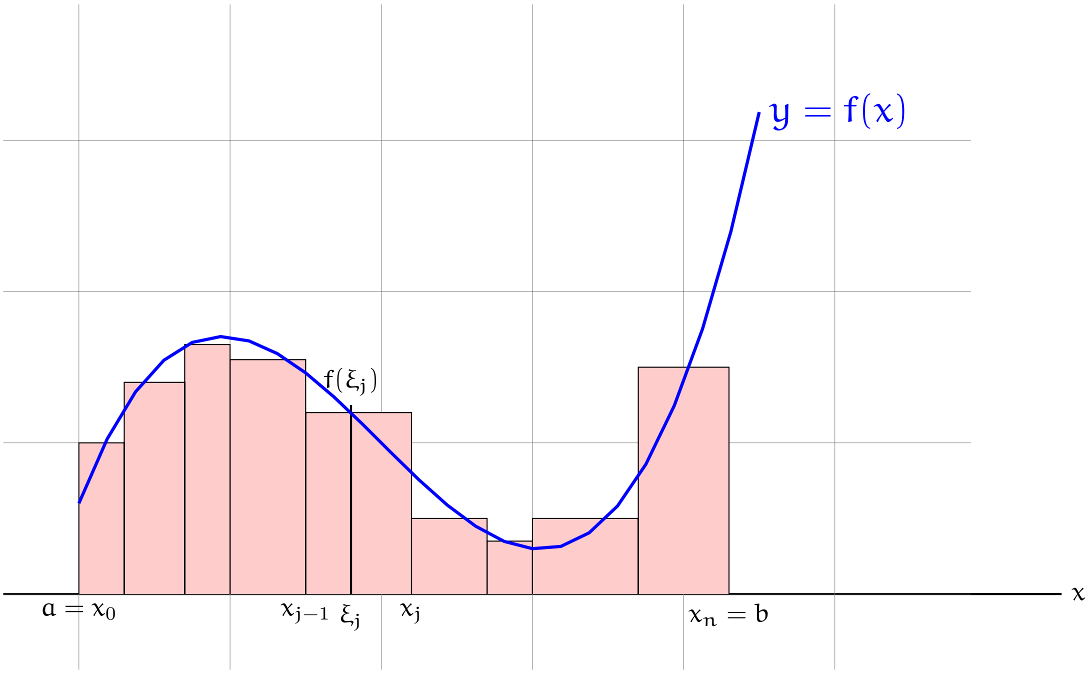
Assume that $y = f(x)$ is a function on $[a, b]$. Its graph is a curve over x-axis. Then we may talk about the 'area' of the region surrounded by this curve and lines $y = 0$, $x = a$ and $x = b$. This area is different from the usual area. The area we talk about now is signed, i.e., it is positive when the region is above x-axis and negative when it is below x-axis. It is natural to approximate the area by rectangle slides. The first step is to cut $[a, b]$ into n small pieces

$$a = x_0 < x_1 < x_2 \cdots < x_{n-1} < x_n = b,$$

called a partition $P = (x_j)$. When the function is good enough, the region under the curve on each piece $[x_{j-1}, x_j]$ is very much like a rectangle, and has the area about $(x_j - x_{j-1})f(\xi_j)$ approximately where ξ_j is a point picked from $[x_{j-1}, x_j]$ arbitrarily. Then the second step is to add those small rectangles together

$$\sum_{j=1}^{n} f(\xi_j)(x_j - x_{j-1})$$

to approximate the area under curve.



This sum is called a **Riemann sum** of f on P, with the choice of points (ξ_j) over P. When the partition gets smaller, it is believed that the sum approximate the area under curve better. This is a rough idea how to approximate the area under curve. But we need a precise definition, which was formulated by Bernhard Riemann(1826-1866), one of the greatest mathematicians.

How to describe the smallness of partition? We denote the length of a partition $P = (x_j)$

$$|P| = \max_j (x_j - x_{j-1}).$$

Definition 4.4.1 (Riemann) Given a function $y = f(x)$ on $[a, b]$. We say that the area under curve exists or f is integrable, if there exists $A \in \mathbf{R}$ such that for any $\varepsilon > 0$, there exists $\delta > 0$, such that for any partition $P = (x_j)$ with $|P| < \delta$ and any choice (ξ_j), it holds that

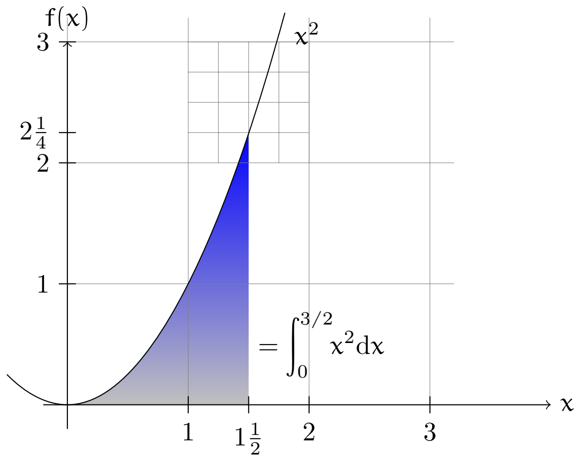
$$\left| \sum_{j=1}^{n} f(\xi_j)(x_j - x_{j-1}) - A \right| < \varepsilon.$$

In this case, A is called the area under curve or more precisely the (definite) integral, or Riemann integral, of f with lower limit a and upper limit b, or simply, integrate f from a to b, and denoted by

$$\int_a^b f(x)dx, \text{ or simply } \int_a^b f,$$

where the integral symbol $\int$ is a deformed S. There are different ways to express integral limits

$$\int_a^b = \int_{[a,b]} = \int_{a \leqslant x \leqslant b}.$$



We simply say that the Riemann sum has limit when the partition goes to zero independent of choice of points on the partition. It will be seen later why definite integrals and indefinite integrals use the same name and same notation, while they seem to have nothing to do with each other. Intuitively when taking limit, $\sum_{j=1}^n$ becomes $\int_a^b$, $f(\xi_j)$ becomes $f(x)$ and $(x_j - x_{j-1}) = \Delta x_j$ becomes dx:

$$\begin{array}{cccc} & \displaystyle\sum_{j=1}^{n} & f(\xi_j) & \Delta x_j \\ & \downarrow & \downarrow & \downarrow \\ \longrightarrow & \displaystyle\int_a^b & f(x) & dx. \end{array}$$

In this notation, only a, b and the rule f are essential and the letter x, called integral variable or dummy variable, is not essential, and can be replaced by any other letter, i.e.,

$$\int_a^b f(x)dx = \int_a^b f(t)dt = \int_a^b f(y)dy.$$

In this definition, we do not see any relation between indefinite integral and definite integral though they use the same notation.

According to the definition, the area is actually signed, namely when $f(x) < 0$ on some interval $[c, d]$, the Riemann sum is obviously negative and hence the area is negative.

Are all functions integrable? No. We present an example which is not integrable. Reading the definition, we see that if f is integrable, the difference between the Riemann sums with different choices of (ξ_j) can be arbitrarily small, since the difference between the limit S and any Riemann sum can be arbitrarily small. This forces f to vary not too much locally. Let

$$f(x) = \begin{cases} 1, & x \in \mathbf{Q}, \\ 0, & x \in \mathbf{R} \setminus \mathbf{Q}. \end{cases}$$

Consider its Riemann sum

$$\sum f(\xi_j)(x_j - x_{j-1})$$

on an arbitrary partition (x_j) of $[0,1]$. On any interval $[x_{j-1}, x_j]$, there are rational numbers and irrational numbers. When we choose $\xi_j \in \mathbf{Q}$, the Riemann sum is 1, and when we choose $\xi_j \in \mathbf{R} \setminus \mathbf{Q}$, the Riemann sum is 0. Hence it is impossible that the Riemann sum has a limit. What functions are integrable? Continuous function because it varies small enough locally. Hence any elementary function on $[a, b]$, which is entirely in its natural domain, is integrable. The proof will be given later.

Theorem 4.4.2 A continuous function on $[a, b]$ is integrable.

A function f on $[a, b]$ is called piece-wise continuous if there exists a partition

$$a < x_1 < x_2 < \cdots < x_n = b$$

such that f is continuous on each interval (x_{j-1}, x_j) and has left and right limits at each point x_j. Of course a piece-wise continuous function on $[a, b]$ is integrable.

Assume that f is integrable on $[a, b]$. We may integrate f from b to a, $\int_b^a f(x)dx$, i.e., the limit of Riemann sum

$$\sum_j f(\xi_j)(x_{j-1} - x_j).$$

Obviously

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

The following properties are intuitive and easy to verify by using the definition.

1. for any three points a, b, c, it holds that

$$\int_a^c f(x)dx = \int_a^b f(x)dx + \int_b^c f(x)dx,$$

   as long as these integrals are well defined.

2. If f is non-negative on $[a, b]$, then the Riemann sum is non-negative and

$$\int_a^b f(x)dx \geqslant 0.$$

3. $\left|\int_a^b f(x)dx\right| \leqslant \int_a^b |f(x)|dx.$

4. Assume that f_1, f_2 are integrable and c_1, c_2 are real numbers.

$$\int_a^b (c_1 f_1(x) + c_2 f_2(x))dx = c_1 \int_a^b f_1(x)dx + c_2 \int_a^b f_2(x)dx.$$

5. If f is even on $[-a, a]$, then

$$\int_{-a}^{a} f(x)dx = 2\int_{0}^{a} f(x)dx.$$

6. if f is odd on $[-a, a]$, then

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

4.4.2 the limit of Riemann sum

We define the integral of f to be the limit of Riemann sum, but it is not so easy to compute for most functions.

Example 4.4.1 *1.* $\int_0^1 x dx$ *is the area under line* $y = x$ *on* $[0, 1]$*, which is a right triangle. The area is 1/2 with simple formula we learned in middle high school. Let's verify by Riemann sum*

$$\sum_{j=1}^{n} \frac{j}{n}\frac{1}{n}$$

where $[0, 1]$ *is equally divided into* n *pieces and* $\xi_j = j/n$*. Why can we do this way? Because when* f *is integrable, no matter how we take the partition and how we take the points, limit should be the same. Then the limit of Riemann sum is*

$$\frac{1}{n^2}\sum_{j=1}^{n} j = \frac{1}{n^2}\frac{n(n+1)}{2} \to \frac{1}{2}.$$

2. $\int x^2 dx$ *is the area under parabola* $y = x^2$ *on* $[0, 1]$*, which we do not know the answer. Let's compute it by Riemann sum*

$$\sum_{j=1}^{n} \left(\frac{j}{n}\right)^2 \frac{1}{n} = \frac{1}{n^3}\sum_{j=1}^{n} j^2.$$

Since

$$\sum_{j=1}^{n} j^2 = \frac{n(n+1)(2n+1)}{6},$$

the limit of Riemann sum is

$$\frac{1}{n^3}\frac{n(n+1)(2n+1)}{6} \to \frac{1}{3}.$$

3. $\int_0^1 a^x dx$*. Riemann sum is a sum of geometric sequence and the limit is*

$$\sum_{j=1}^{n} a^{j/n}\frac{1}{n} = \frac{1}{n}a^{1/n}\frac{a^{n/n}-1}{a^{1/n}-1} = a^{1/n}(a-1)\frac{1/n}{e^{(1/n)\ln a}-1} \to \frac{a-1}{\ln a}.$$

It seems to work, but if you look at it carefully, you will find that to calculate $\int_0^1 x^k dx$ by Riemann sum, you need a formula for $\sum_{j=1}^n j^k$ which is difficult to obtain. In general, to calculate $\int_0^1 f(x)dx$, we need a formula for $\sum_{j=1} f(j/n)$, which is not available in most cases. Hence it is almost impossible to integrate f by the limit of Riemann sum.

4.4.3 Newton-Leibniz formula

Here comes the most important theorem in calculus, called the fundamental theorem of calculus, the greatest contribution to mathematics made by Newton-Leibniz. This theorem connects the definite integral to indefinite integral.

Theorem 4.4.3 Assume that $y = f(x)$ is continuous on $[a, b]$ and $g'(x) = f(x)$ or $\int f(x)dx = g(x)$ on $[a, b]$. Then

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

Proof. The proof is easy. Take a partition $(x_j : 1 \leqslant j \leqslant n)$ and by the mean-value theorem

$$\begin{aligned} g(b) - g(a) &= \sum_{j=1}^n g(x_j) - g(x_{j-1}) = \sum_{j=1}^n g'(\xi_j)(x_j - x_{j-1}) \\ &= \sum_{j=1}^n f(\xi_j)(x_j - x_{j-1}). \end{aligned}$$

The right side, the Riemann sum of f on P, converges to $\displaystyle\int_a^b f(x)dx$ since f is continuous. □

Moreover we may apply this formula to the integral of f on $[a, x]$ for $x \in [a, b]$

$$\int_a^x f(u)du = g(x) - g(a),$$

where we use u as the dummy variable. The integral is a function of x, more precisely a primitive function of f. Hence the relation between derivative and integral is illustrated as

$$\left(\int_a^x f(u)du\right)' = \frac{d\int_a^x f(u)du}{dx} = f(x).$$

We usually write $g(b)-g(a)$ into $g\Big|_a^b$. In this notation, the Newton-Leibniz formula is written into

$$\int_a^b f(x)dx = \int f(x)dx\Big|_a^b.$$

Now the problem of definite integral becomes the problem of indefinite integral.

Example 4.4.2 *1.* $\displaystyle\int_0^1 xdx = \int xdx\Big|_0^1 = \frac{1}{2}x^2\Big|_0^1 = \frac{1}{2}.$

2. $\displaystyle\int_0^1 x^n dx = \int x^n dx\Big|_0^1 = \frac{1}{n+1}x^{n+1}\Big|_0^1 = \frac{1}{n+1}.$

3. $\int_0^1 a^x dx = \int a^x dx\Big|_0^1 = \frac{a^x}{\ln a}\Big|_0^1 = \frac{a-1}{\ln a}.$

4. $\int_1^e \ln x dx = \int \ln x dx\Big|_1^e = (x\ln x - x)\Big|_0^1 = 1.$

5. $\int_{-1}^1 \sqrt{1-x^2}dx$ *is the area of half unit disk. By Example 4.3.1,*

$$\begin{aligned}\int_{-1}^1 \sqrt{1-x^2}dx &= \int_{-1}^1 \sqrt{1-x^2}dx \\ &= \frac{1}{2}\arcsin x + \frac{1}{2}x\sqrt{1-x^2}\Big|_{-1}^1 = \frac{\pi}{2}.\end{aligned}$$

6. *If we integrate a function with absolute value, be careful.*

$$\begin{aligned}\int_0^2 |x^2-1|dx &= \int_0^1 (1-x^2)dx + \int_1^2 (x^2-1)dx \\ &= (x - x^3/3)\Big|_0^1 + (x^3/3 - x)\Big|_1^2 = 2.\end{aligned}$$

Newton-Leibniz theorem is surely the most efficient way to calculate definite integrals. However it does not always work because we may not be able to find primitive functions explicitly for most functions. There are other clever approaches to find the definite integral of a function without an explicit primitive function, for example

$$\int_{\infty}^{\infty} e^{-x^2/2}dx = \sqrt{2\pi},$$

though we are not able to find the anti-derivative of $e^{-x^2/2}$. We will see this later.

4.4.4 integration by parts and substitution rule

To do the definite integral, we can find the primitive function first and then use the Newton-Leibniz formula. However we may also carry the integral limits when we use integration by parts and substitution rule. Of course these two methods are essentially the same. You do not have to learn this.

1.
$$\int_a^b f(x)dg(x) = f(x)g(x)\Big|_a^b - \int_a^b g(x)df(x).$$

2. If g is smooth on $[a,b]$ and $g(a)=u,\ g(b)=v$, then

$$\int_a^b f(g(x))dg(x) = \int_u^v f(y)dy.$$

Example 4.4.3 *1. In this example, we use integration by parts*

$$\begin{aligned}\int_1^e x\ln x dx &= \frac{1}{2}\int_1^e \ln x dx^2\\ &= \frac{1}{2}(x^2\ln x\Big|_1^e - \int_1^e x^2\frac{1}{x}dx)\\ &= \frac{1}{2}(e^2 - \frac{1}{2}(e^2-1)) = \frac{1}{4}(e^2+1).\end{aligned}$$

2. In this example $y = \sqrt{x+1}$, $y^2 = x+1$, $2ydy = dx$,

$$\begin{aligned}\int_0^1 \frac{x}{\sqrt{1+x}} &= \int_1^{\sqrt{2}} \frac{y^2-1}{y} 2ydy = 2\int_1^{\sqrt{2}} (y^2-1)dy\\ &= 2(y^3/3 - y)\Big|_1^{\sqrt{2}} = \frac{2}{3}(2-\sqrt{2}).\end{aligned}$$

Exercises

1. Evaluate the integral

   (a) $\displaystyle\int_1^9 \frac{x-1}{\sqrt{x}}dx$;

   (b) $\displaystyle\int_0^{\pi/4} \frac{1}{\cos^2 x}dx$.

2. Is the following work correct? Why?

$$\int_{-1}^2 x^{-4}dx = \frac{x^{-3}}{-3}\Big|_{-1}^2 = -\frac{3}{8}.$$

3. Prove that for any two integers n, m,

$$\int_0^{2\pi} \sin nx \cos mx dx = 0.$$

4. Prove that for any two integers $n \neq m$,

$$\int_0^{2\pi} \sin nx \sin mx dx = 0.$$

5. Evaluate the integral $\displaystyle\int_0^{\pi/2} \sin x \sin 2x dx$.

6. Set for $n \geqslant 1$

$$a_n = \int_0^{\pi} \sin^n x dx$$

   and prove that $\lim a_n = 0$.

4.5 application of integrals

4.5.1 area surrounded by curves

Definite integral can be used to compute the area of the region surrounded by curves. Assume that $f(x) \geqslant g(x)$ for $x \in [a, b]$. Then

$$\int_a^b (f(x) - g(x))dx$$

is the area of the region surrounded by curve $y = f(x)$, $y = g(x)$, $x = a$ and $x = b$, or written into

$$g(x) \leqslant y \leqslant f(x), \quad a \leqslant x \leqslant b,$$

which is called a rectangle with curved y-sides.

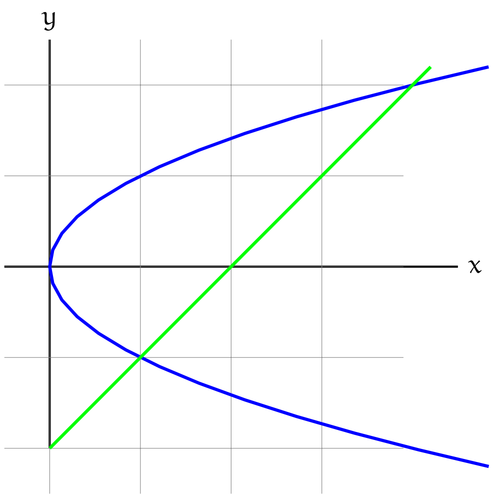
Example 4.5.1 *1.* $\sqrt{x} \geqslant x^2$ *on* $[0, 1]$. *The area of the region is*

$$\int_0^1 (\sqrt{x} - x^2)dx = \frac{1}{1 + 1/2} - \frac{1}{3} = \frac{1}{3}.$$

2. $\sin x \geqslant \cos x$ *on* $[\pi/4, 3\pi/4]$. *The area of the region is*

$$\int_{\pi/4}^{3\pi/4} (\sin x - \cos x)dx = (-\cos x - \sin x)\Big|_{\pi/4}^{3\pi/4} = 2\sqrt{2}.$$

3. We shall always choose a simpler way to compute the integral if possible. The region surrounded by curves $x = y^2$ *and* $x = y + 2$.



Look at the graph and we need to separate it into two parts $x \in [0, 1]$, *where the lower curve is* $y = -\sqrt{x}$, *and* $x \in [1, 4]$, *where the lower curve is* $y = x - 2$. *The area of the* *region is*

$$\int_0^1 (\sqrt{x} - (-\sqrt{x}))dx + \int_1^4 (\sqrt{x} - (x-2))dx = 2\frac{x^{3/2}}{3/2}\Big|_0^1 + (\frac{x^{3/2}}{3/2} - \frac{1}{2}x^2 + 2x)\Big|_1^4.$$

The region is easier to write when we treat x *as a function of* y*. In this case, the area of the region is*

$$\int_{-1}^2 (y + 2 - y^2)dy = (y^2/2 + 2y - y^3/3)\Big|_{-1}^2 = \frac{27}{6} = \frac{9}{2}.$$

If you are not comfortable, you may exchange x, y*, i.e., flip the graph about* $y = x$*.*

4.5.2 idea of Riemann sum: cut and add

The area of the region surrounded by $y = f(x)$, $y = g(x)$, $x = a$ and $x = b$ is

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

More generally given a region A, we may cut it by vertical lines into many thin pieces and each piece can be approximated by the area of rectangle as in the definition of Riemann sum. Assume that the region is between $x = a$ and $x = b$. The length of intersection of the vertical line at point x and the region A is denoted by $L(x)$ and then the area is the following integral

$$\int_a^b L(x)dx.$$

An ancient Chinese mathematician, Zugeng(456-536), made the following observation, which is called Cavalieri's theorem, which is also called Zugeng's principle.

Theorem 4.5.1 (Cavalieri) Given two regions A and B on plane. If for every straight line ℓ with a fixed slope, the length of $\ell \cap A$ is the same as the length of $\ell \cap B$, then A and B have the same area.

The idea of Riemann sum is very powerful, because the same reason will give us a formula to compute the volume of a solid.

Let us put a solid S and x-axis on space. Assume that the solid is between the plane at a and the plane at $x = b$. A plane at x_0 means a plane perpendicular to x-axis at x_0. The intersection of the plane at x and the solid S is called the section of the solid S at x and its area is denoted by $A(x)$. The volume of S is the limit of Riemann sum

$$\sum_{j=1}^n A(x_j)(x_j - x_{j-1}),$$

where (x_j) is a partition of $[a, b]$ and hence we have Cavalieri's theorem for solids.

Theorem 4.5.2 The volume of S is

$$\int_a^b A(x)dx.$$

Roughly speaking, the volume is equal to integrating the areas of sections.

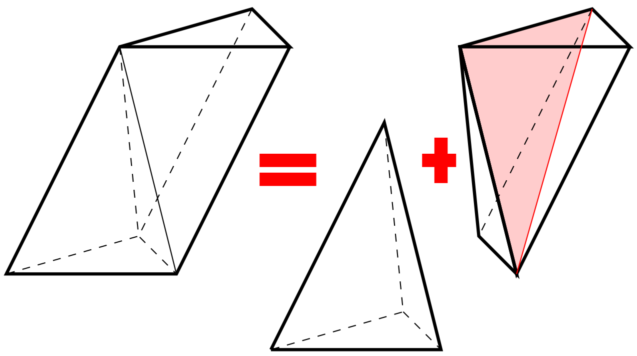
Example 4.5.2 1. *Let us compute the volume of a cone with base area A and height h. We put this cone on space with the base perpendicular to x-axis and the vertex being the origin. Then the cone is between $[0, h]$ What is the area $A(x)$ of the section of the plane at x on the cone? It is not hard to verify that*

$$A(x) = \frac{x^2}{h^2}A.$$

The volume of the cone is

$$\int_0^h A(x)dx = \int_0^h \frac{x^2}{h^2}Adx = \frac{1}{3}Ah.$$

Of course this can be shown by an elementary approach. By the technique of Riemann sum we need only to show that the formula holds for a cone with triangle base. This can be seen in the following graph, in which three equal volume cones make up a cylinder.



2. *Let us compute the volume of unit ball. Put the ball on space with the center being the origin of xyz-coordinate system. We only consider the positive part S of the ball ($x > 0$). The plane at $x < 1$ cuts the ball to obtain a disk with radius $\sqrt{1-x^2}$ and the area is $A(x) = \pi(1-x^2)$. Then the volume the half ball is*

$$\int_0^1 A(x)dx = \int_0^1 \pi(1-x^2)dx = \pi(1-1/3) = 2\pi/3.$$

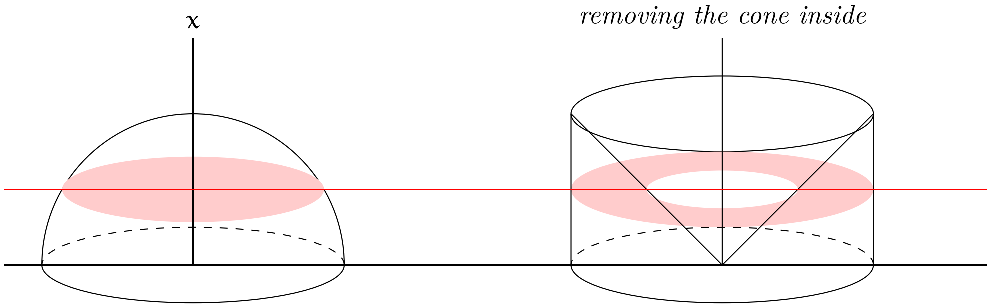
Here we shall explain how to use Cavalieri's theorem or Zugeng's principle to find the volume of ball. Place a cylinder with radius 1 and the central line being x-axis. Keep the cylinder between $x = 0$ and $x = 1$. Removing a cone C, with the center at the origin *and the base being the cylinder base, from the cylinder obtains a solid, denoted by* S_1*. Then the section of* S_1 *at* $x \leqslant 1$ *is a ring with area*

$$A_1(x) = \pi - \pi x^2 = \pi(1 - x^2) = A(x),$$

which is exactly the area of the ball at x*. Then Zugeng made the observation, which is exactly Cavalieri's theorem, that the volume of* S *is equal to the difference of volumes of the cylinder and the cone*

$$\pi - \frac{1}{3}\pi = \frac{2}{3}\pi,$$

where the volume of cone is a third of the volume of cylinder.

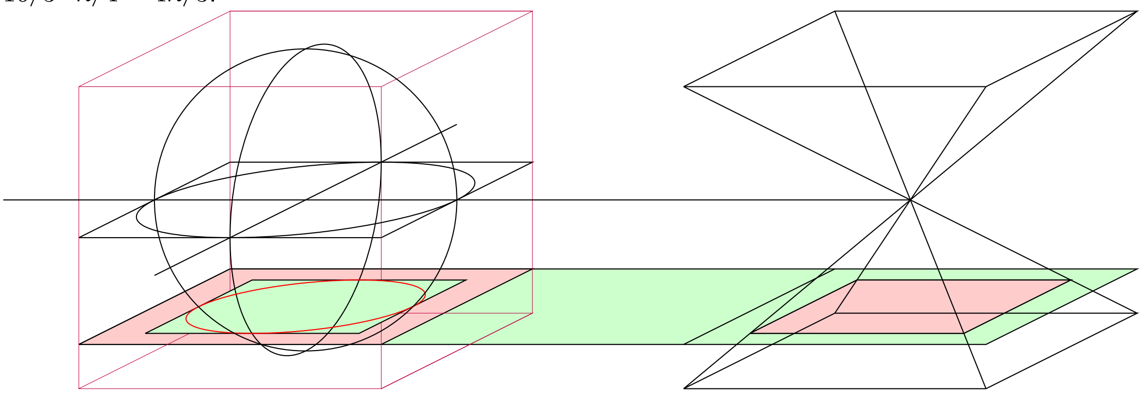


the area of red disk=the area of red ring

Remark 4.5.3 The formula for volume of ball was given by the great Greek mathematician Archimedes, about 700 years earlier than Zugeng, who obtained the formula independently. In ancient China, there were some scattered ideas on limit and calculus. Zugeng's approach to calculate the volume of ball is an example and well worth a remark. The book, which was written by Zugeng and his father Zuchongzhi about their main achievements, was lost in history. The existing materials about Zugeng's idea comes from other people's words. Zugeng's original sentence is . It is arguable what this simple ancient chinese sentence means literally. Combining what Zugeng did, the modern Chinese experts translate it into roughly 'equal area at each level implies equal volume', named Zugeng's principle.

Historically the story about volume of ball is more interesting. To be fair, Liuhui, a mathematician 200 years earlier than Zugeng, should be mentioned. His method to compute π, named Liuhui's technique of circle partition, made him one of few greatest ancient Chinese mathematicians. He pointed out that the formula of volume of ball in the ancient famous book entitled (The nine chapters on the mathematical art, the greatest book in Chinese mathematical history, written about 100 years before Liuhui and the author was unknown) was wrong and he proposed a road to compute. Let the unit ball B be inside the cube C with side 2 placed on the plane. A solid, called S, is obtained by the intersection of two cylinders with radius 1 containing the ball and directed in two horizontal and perpendicular sides of the cube. The solid S is named , 'dual square umbrellas' in English roughly. Each horizontal plane cuts S and B to obtain a square and a disk, and the ratio of square area to disk area is always $4:\pi$. Liuhui then asserted that the ratio of volume of S to volume of B was also $4:\pi$. From this conclusion, we would say that Liuhui actually had the idea of Zugeng's principle, though he did not achieve his goal: find the volume of S and then the volume of B.

Zugeng solved this problem. Assume now that two cones based on the lower and upper faces of the cube are placed with the center of ball being their common vertex, as shown in the graph. Zugeng observed surprisingly that the same plane above cuts both C outside S and the cone, and two sections (red regions in the graph) obtained have the same areas. Zugeng's principle implies that the difference of the volume of C subtracting the volume of S is the same as the volume of two cones. It follows that the volume of S is $8-8/3=16/3$. Hence the volume of B is $16/3\cdot\pi/4=4\pi/3$.



volume of cube=8 volume of two cones=8/3

Though Liuhui and Zugeng's idea to compute the volume of ball is really great and smart, it is only an isolated idea, which was inspired in solving some isolated examples. In the ancient China, there was no systematic mathematics, but there were some clever ideas appearing from time to time. Of course, we should not ask too much, since in long history of human-beings, mathematics appeared only in ancient Greek.

4.5.3 volume of rotating body

Given a curve $y=f(x)$, we may let it rotate around x-axis or y-axis to form a rotating body. The idea of Riemann sum can be used to express the volume of any rotating body as integral. Assume that $y=f(x)$, $x\in[a,b]$ rotates around x-axis. Then the area of section of the rotating body at x is $\pi f(x)^2$ and hence the volume of rotating body is

$$\int_a^b \pi f(x)^2 dx.$$

Example 4.5.3 *1. The volume of a bowl. The parabola* $y = \sqrt{x}$, $x \in [0,1]$ *rotates around* x*-axis and the volume of the rotating body is*

$$\int_0^1 \pi(\sqrt{x})^2 dx = \frac{\pi}{2}.$$

2. The volume of the rotating body obtained by $y = x^2$, $x \in [0,1]$*, is*

$$\int_0^1 \pi(x^2)^2 dx = \frac{\pi}{5}.$$

3. The volume of a torus obtained by rotating $(x-2)^2 + y^2 = 1$ *around* y*-axis. For any* $y \in [-1,1]$*, the area of the section of the rotating body at* y *is*

$$\pi[(2+\sqrt{1-y^2})^2 - (2-\sqrt{1-y^2})^2] = 8\pi\sqrt{1-y^2}.$$

Hence the volume is, by Example 4.3.1

$$8\pi\int_{-1}^1 \sqrt{1-y^2}dy = 8\pi\frac{1}{2}(\arcsin y + y\sqrt{1-y^2})\Big|_{-1}^1 = 4\pi^2.$$

4.5.4 length of a curve

Is it possible to find the length of a curve given by a smooth function $y = f(x)$ on $[a,b]$? What is the length of a curve? Fix a partition (x_j). We think that the length of curve $y = f(x)$ on $[x_{j-1}, x_j]$ is approximately

$$\sqrt{(\Delta x_j)^2 + (\Delta y_j)^2},$$

where $\Delta x_j = x_j - x_{j-1}$ and $\Delta y_j = f(x_j) - f(x_{j-1})$. Hence the length of the curve on $[a,b]$ is approximately

$$\sum_{j=1}^n \sqrt{(\Delta x_j)^2 + (\Delta y_j)^2} = \sum_{j=1}^n \sqrt{1 + \left(\frac{\Delta y_j}{\Delta x_j}\right)^2}\Delta x_j$$

which converges to

$$\int_a^b \sqrt{1+f'(x)^2}dx,$$

when $y = f'(x)$ is continuous on $[a,b]$.

Example 4.5.4 *Let us compute the length of the curve* $y = x^2$ *on* $[0,1]$*. By the formula above, it is*

$$\int_0^1 \sqrt{1+(2x)^2}dx.$$

By Example 4.3.2, it is

$$\begin{aligned}\frac{1}{2}\int_0^1 \sqrt{1+(2x)^2}d2x &= \frac{1}{2}\int_0^2 \sqrt{1+y^2}dy \\ &= \frac{1}{4}(y\sqrt{1+y^2}+\ln(y+\sqrt{1+y^2}))\Big|_0^2 \\ &= \frac{1}{4}(2\sqrt{5}+\ln(2+\sqrt{5})).\end{aligned}$$

Exercises

1. Sketch the region enclosed by the given curves. Decide to integrate with respect to x or y.
   (a) the region enclosed by $y = 5x - x^2$ and $y = x$;
   (b) $y = \sqrt{x+3}$, $y = (x+3)/2$;
   (c) $y = \cos x$, $y = \sin 2x$, $x = 0$, $x = \pi/2$;
   (d) $y = 1/x^2$, $y = x$, $y = x/8$.
2. Evaluate $\int_0^{\pi/2} |\sin x - \cos 2x| dx$.
3. Find the volume of the solid obtained by rotating the region bounded by the given curves about the specified line.
   (a) $y = 1 + \frac{1}{\cos x}$, $y = 3$, about $y = 1$;
   (b) $y = |x-1| - 1$, $y = 0$, about $x = 0$; about $y = -2$;
   (c) $x = y^2$, $x = 1$, about $x = 1$.

4.6 improper integrals

4.6.1 improper integrals

The Riemann integral above is defined on a finite closed interval and is called normal integral or proper integral. Other types of integrals are abnormal integrals, or improper integrals. An integral is called improper if either the interval is infinite or the function goes to infinity when x approaches to some point in the interval so that the integral can not be defined by the normal Riemann sum.

The first type of improper integral is the integral of a function on a half line $(-\infty, a]$ or $[a, +\infty)$.

Definition 4.6.1 Assume that $y = f(x)$ is a function on $[a, \infty)$ and for every $u > a$, f is integrable on $[a, u]$. The notation

$$\int_a^{+\infty} f(x)dx$$

denotes the improper integral. If

$$\lim_{u \to +\infty} \int_a^u f(x)dx$$

exists, then we say that f is integrable on $[a, \infty)$, or the improper integral $\int_a^\infty f(x)dx$ exists or converges, otherwise we say the improper integral does not exist or diverges.

Actually what we are concerned for an improper integral

$$\int_a^\infty f(x)dx$$

is whether it converges, which is a property of f near ∞ and does not depend on the left end-point a. The convergence of the improper integral

$$\int_{-\infty}^b f(x)dx$$

can be defined similarly. It is easy to see that if $\int_a^\infty f_1(x)dx$ and $\int_a^\infty f_2(x)dx$ converge and c_1, c_2 are constant, then

$$\int_a^\infty (c_1 f(x) + c_2 f_2(x))dx = c_1 \int_a^\infty f_1(x)dx + c_2 \int_a^\infty f_2(x)dx.$$

Example 4.6.1 *1. Consider* $y = e^{-ax}$ *for* $x \geqslant 0$. *Assume* $a > 0$.

$$\int_0^\infty e^{-ax}dx = \lim_{y \to +\infty} \int_0^y e^{-ax}dx = \lim_{y \to \infty} \left.\frac{-e^{-ax}}{a}\right|_0^y = \frac{1}{a}.$$

For saving notation, we simply write $\int_0^\infty e^{-ax}dx = \left.\frac{-e^{-ax}}{a}\right|_0^\infty$ *where* ∞ *on the right refers to taking limit automatically.*

2. Consider $y = x^a$ *on* $[0, \infty)$. *When* $a \geqslant 0$,

$$\int_0^\infty x^a dx = \left.\frac{y^{a+1}}{a+1}\right|_0^\infty = +\infty.$$

The improper integral does not exist. Because in this case, the function is increasing, the area under curve on $[0, +\infty)$ *is obviously infinite. Now let's consider* $a < 0$. *Because the domain of* $y = x^a$ *is* $(0, \infty)$, *we consider the improper integral on* $[1, +\infty)$. *When* $-1 < a < 0$,

$$\int_1^{+\infty} x^a dx = \left.\frac{x^{a+1} - 1}{a+1}\right|_1^{+\infty} = +\infty.$$

When $a = -1$,

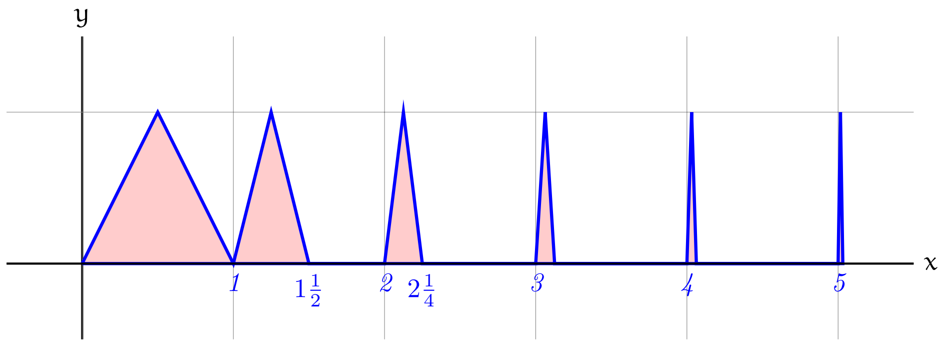
$$\int_1^{+\infty} x^{-1} dx = \ln x \Big|_1^{+\infty} = +\infty.$$

In either case, the improper integral diverges. When $a < -1$,

$$\int_1^{\infty} x^a dx = \frac{x^{a+1} - 1}{a+1} \Big|_1^{+\infty} = \frac{-1}{a+1}.$$

The improper integral exists.

3. *It seems that the area under* f *on* $[a, +\infty)$ *being finite will force* f *to go to zero at infinity. But it is not true. Define piece-wisely a function* $f(x)$ *as in the following graph*



The n*-th triangle is formed by vertices* $(n, 0)$, $(n + \frac{1}{2^{n+1}}, 1)$, *and* $(n + \frac{1}{2^n}, 0)$. *This function is continuous and has no limit when* $x \to +\infty$, *but* $\int_0^\infty f(x)dx$ *is the sum of areas of all triangles, which equals*

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

i.e., the improper integral converges.

In most cases, it is impossible to find the value of integral as the examples above. There are some methods to judge if an improper integral converges or not. If $f(x) \geqslant 0$ for any $x \geqslant a$, then $\int_a^u f(x)dx$ is increasing as a function of u. Hence the improper integral $\int_a^{+\infty} f(x)dx$ converges if and only if $\int_a^u f(x)dx$, as a function of u, is bounded, and in this case we write

$$\int_a^\infty f(x)dx < \infty.$$

Be careful this makes sense only when f is non-negative.

Theorem 4.6.2 (comparison) If $0 \leqslant g(x) \leqslant f(x)$ for large x, then $\displaystyle\int_a^\infty f(x)dx < \infty$ implies that $\displaystyle\int_a^\infty g(x)dx < \infty$.

Example 4.6.2 *1. When $x > 1$, $e^{-x^2} < e^{-x}$, and it follows that*

$$\int_0^\infty e^{-x^2} dx < \infty.$$

The improper integral on $(-\infty, a]$ is defined similarly. Hence

$$\int_{-\infty}^\infty e^{-x^2} dx = 2\int_0^\infty e^{-x^2} dx < \infty.$$

2. Consider $y = \frac{\ln x}{x^2}$, $x > 2$. We may prove that

$$\lim_{x\to+\infty} \frac{\ln x}{\sqrt{x}} = 0$$

by L'Hospital's rule. When x is large enough, $\ln x < \sqrt{x}$. Actually we may prove that for any $a > 0$, when x is large enough, $\ln x < x^a$. Hence when x is large enough,

$$\frac{\ln x}{x^2} \leqslant \frac{\ln x}{\sqrt{x}} \frac{1}{x^{3/2}} \leqslant x^{-3/2}$$

and then

$$\int_2^\infty \frac{\ln x}{x^2} dx < \infty.$$

Actually for any $a > 0$,

$$\int_2^\infty \frac{\ln x}{x^{1+a}} dx < \infty.$$

There are other types of improper integrals. Assume that $y = f(x)$ is continuous on $[a, b)$ and $\lim_{x\to b} f(x) = +\infty$. Then the integral

$$\int_a^b f(x) dx$$

is called an improper integral. It is said to converge if

$$\lim_{y\to b-} \int_a^y f(x) dx$$

exists. The improper integral of function $y = f(x)$ which is continuous on $(a, b]$ can be defined similarly.

Example 4.6.3 *1. We still consider $y = x^{-a}$, with $a > 0$, on $(0, 1]$.*

$$\int_y^1 x^{-a} dx = \begin{cases} -\ln y, & a = 1, \\ \dfrac{1 - y^{-a+1}}{-a+1}, & a \neq 1. \end{cases}$$

Consider the limit when $y \to 0$ and it is seen that $\int_0^1 x^{-a} dx$ converges when $a < 1$ and diverges when $a \geqslant 1$. Similarly we also write

$$\int_0^1 x^{-a} dx = \left.\frac{1 - x^{-a+1}}{-a+1}\right|_0^1$$

and the right side should refer to taking limit.

2. *Consider Γ-function*

$$\Gamma(a) := \int_0^{+\infty} x^{a-1} e^{-x} dx.$$

When $a \geqslant 1$, the integrand is continuous on $[0, \infty)$. When $a < 1$, the integrand is not continuous at 0 and hence this integral has two types of improper integral: 0 and $+\infty$,

$$\int_0^{+\infty} x^{a-1} e^{-rx} dx = \int_0^1 + \int_1^{+\infty} x^{a-1} e^{-x} dx.$$

It follows from the example above that when $a > 0$, the improper integral converges.

4.6.2 absolute/conditional convergence

Definition 4.6.3 If

$$\int_a^\infty |f(x)| dx < \infty,$$

we say the improper integral

$$\int_a^\infty f(x) dx$$

converges absolutely. If $\int_a^\infty f(x) dx$ converges but $\int_a^\infty |f(x)| dx$ diverges, we say the improper integral $\int_a^\infty f(x) dx$ converges conditionally.

If $\int_a^\infty f(x) dx$ converges absolutely, it must converge. In fact, denote by $f^+(x)$ and $f^-(x)$ the positive part and negative part of f, respectively. When $f(x) \geqslant 0$, define $f^+(x) = f(x)$ and $f^-(x) = 0$, and conversely when $f(x) < 0$, define $f^+(x) = 0$ and $f^-(x) = -f(x)$. Then

$$f(x) = f^+(x) - f^-(x), \text{ and } |f(x)| = f^+(x) + f^-(x).$$

If the improper integral converges absolutely, then

$$\int_a^\infty [f^+(x) + f^-(x)] dx < \infty,$$

and it follows that both $\int_a^\infty f^+(x) dx$ and $\int_a^\infty f^-(x)$ converge. Hence by the property stated above

$$\int_a^{+\infty} f(x) dx = \int_a^{+\infty} f^+(x) dx - \int_a^{+\infty} f^-(x) dx.$$

Example 4.6.4 *Consider*

$$y = \frac{\sin x}{x}, \ x > 0,$$

which has right limit 1 *at* 0. $\int_0^\infty \frac{|\sin x|}{x}dx = +\infty$.

$$\begin{aligned}\int_\pi^{n\pi} \frac{|\sin x|}{x}dx &= \sum_{j=1}^{n-1}\int_{j\pi}^{(j+1)\pi} \frac{|\sin x|}{x}dx \\ &\geqslant \sum_{j=1}^{n-1}\frac{1}{(j+1)\pi}\int_{j\pi}^{(j+1)\pi}|\sin x|dx \\ &= \sum_{j=1}^{n-1}\frac{1}{(j+1)\pi}2 \geqslant 2\sum_{j=1}^{n-1}\int_{(j+1)\pi}^{(j+2)\pi}\frac{1}{x}dx = 2\int_{2\pi}^{(n+1)\pi}\frac{1}{x}dx \to +\infty.\end{aligned}$$

Example 4.6.5 $\int_0^\infty \frac{\sin x}{x}dx$ *converges. Let*

$$a_n = \int_0^{2n\pi}\frac{\sin x}{x}dx.$$

We shall verify (a_n) *is increasing and bounded.*

1. $\int_{2n\pi}^{2(n+1)\pi}\frac{\sin x}{x}dx > 0$ *and* $a_n \nearrow$.
2. (a_n) *is bounded, since*

$$a_n = \int_0^\pi + \int_\pi^{3\pi} + \cdots + \int_{(2n-3)\pi}^{(2n-1)\pi} + \int_{(2n-1)\pi}^{2n\pi}\frac{\sin x}{x}dx \leqslant \int_0^\pi \frac{\sin x}{x}dx,$$

where only the first integral is positive.

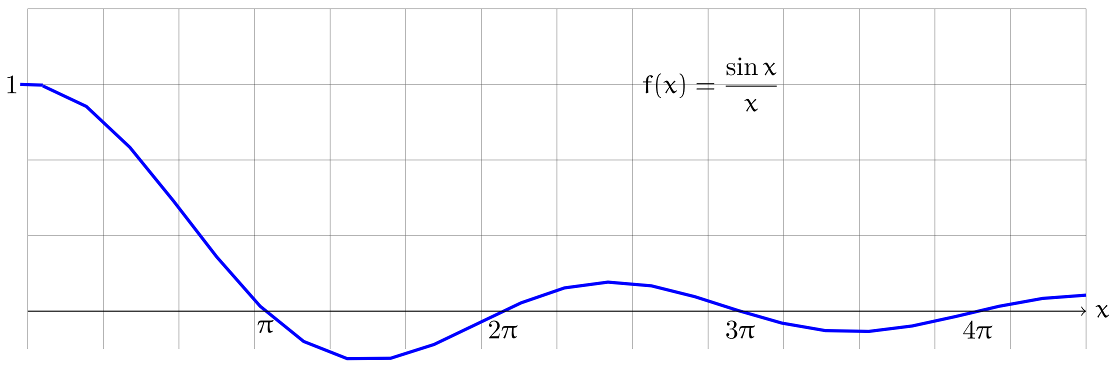
Hence (a_n) *converges. When* $u \in (2n\pi, (2n+2)\pi)$, *since*

$$\int_0^u \frac{\sin x}{x}dx - a_n = \int_{2n\pi}^u \frac{\sin x}{x}dx,$$

we have

$$\left|\int_0^u \frac{\sin x}{x}dx - a_n\right| \leqslant \int_{2n\pi}^{(2n+2)\pi}\frac{|\sin x|}{x}dx \leqslant \int_{2n\pi}^{(2n+2)\pi}\frac{1}{x}dx \leqslant \frac{1}{n} \to 0,$$

as $u \to +\infty$. *It follows that the improper integral exists.*



Intuitively the area under $y = \frac{\sin x}{x}$ on $[0, +\infty)$ is positive and negative alternatively. Adding all areas in their absolute values gives infinity but adding all signed areas will give a finite number.

Exercises

1. Find the length of the curve $y = \ln x$ for $x \in [1, 2]$.

2. Determine whether the integral is convergent. Evaluate those that are convergent.

   (a) $\displaystyle\int_0^\infty \sin x dx$; (b) $\displaystyle\int_0^\infty e^{-x} \sin x dx$; (c) $\displaystyle\int_{-\infty}^0 xe^{2x} dx$;

   (d) $\displaystyle\int_{-2}^2 \frac{1}{x^3} dx$; (e) $\displaystyle\int_{-1}^8 x^{-1/3} dx$; (f) $\displaystyle\int_0^2 z^2 \ln z dz$;

   (g) $\displaystyle\int_0^1 \frac{1}{\sqrt{x(1-x)}} dx$.

3. For what values of p, does $\displaystyle\int_0^1 x^p \ln x dx$ converge? If does, find the integral.

4. express the function f in Example 4.6.1-3 explicitly.

5. prove that if g is continuous and decreasing to 0 on $[a, \infty)$, then the improper integral $\int_a^\infty g(x) \sin x dx$ converges.

6. Prove that the following limit exists

$$\lim_{y \to 0+} \int_y^1 \sin \frac{1}{x} dx.$$

4.7 analysis 3: Riemann integrability

4.7.1 uniform continuity

The limit of Riemann sum is not a usual sequence or function limit. To prove that any continuous function on $[a, b]$ is integrable, we need to find A first and then prove that the Riemann sum is close to A when $|P|$ is small. Hence we shall compare the Riemann sum on the same partition and also on different partition. On the same partition (x_j) of $[a, b]$, choose different points $\xi_j, \eta_j \in [x_{j-1}, x_j]$ and the difference of their Riemann sums is

$$\sum_{j=1}^n [f(\xi_j) - f(\eta_j)](x_j - x_{j-1}).$$

Roughly, for each j, when the interval $[x_{j-1}, x_j]$ is small enough, we would expect the difference $f(\xi_j) - f(\eta_j)$ is small enough by the continuity. However if we want to write this argument down, we will find that this smallness has to be true for all j simultaneously when $|P|$ is small. This is called uniform continuity.

Definition 4.7.1 A function $y = f(x)$ on D is called uniformly continuous, if for any $\varepsilon > 0$, there exists $\delta > 0$ such that

$$|f(x) - f(y)| < \varepsilon,$$

for any $x, y \in D$ with $|x - y| < \delta$, **uniformly**.

The usual continuity of f is defined for a point x_0. Thus for each $\varepsilon > 0$, the $\delta > 0$ we found depends also on the point x_0. It may not be able to find a common δ. For example $y = x^2$, for given $\varepsilon > 0$, it is impossible to find $\delta > 0$ satisfying for any $a \in \mathbf{R}$, when $|x - a| < \delta$, $|x^2 - a^2| < \varepsilon$, because $y = x^2$ increases faster and faster, so that the bigger is $|a|$, the smaller δ has to be. The key of the definition above is that the δ depends only on ε, and suits every point of D. It is called a uniform property.

Theorem 4.7.2 A continuous function f on a closed interval $[a, b]$ is uniformly continuous.

Proof. Prove by contradiction. Suppose that f is not uniformly continuous. What does it mean? There must be an $\varepsilon_0 > 0$, for which we can not find a δ as in definition, namely no matter how small $\delta > 0$ is, there are two points $x, y \in [a, b]$ with $|x - y| < \delta$, but

$$|f(x) - f(y)| > \varepsilon_0.$$

Then for each $n \in \mathbf{N}$, δ is chosen to be $1/n$, there are two point $x_n, y_n \in [a, b]$ satisfying for any n,

$$|x_n - y_n| < 1/n, \text{ and } |f(x_n) - f(y_n)| > \varepsilon_0.$$

It follows from Bolzano-Weiertrass theorem that (x_n) has a convergent subsequence $x_{k_n} \to x_0$. Since $|x_{k_n} - y_{k_n}| < 1/k_n$, $y_{k_n} \to x_0$. Applying this to $|f(x_{k_n} - f(y_{k_n}| > \varepsilon_0$, we have by the continuity of f

$$|f(x_0) - f(x_0)| \geqslant \varepsilon_0,$$

which is obviously a contradiction. □

4.7.2 integrability of continuous functions

Proof. Assume that f is continuous on $[0, 1]$ without loss of generality. For a partition $P = (x_j)$ with $|P| = \max_j(x_j - x_{j-1})$ and any choice $\xi_j \in [x_{j-1}, x_j]$, denote the Riemann sum by $R(P)$, namely

$$R(P) = \sum f(\xi_j)(x_j - x_{j-1})$$

which depends also on (ξ_j). According to the definition, we need to find A satisfying that for any $\varepsilon > 0$, $\exists\, \delta > 0$ such that

$$|R(P) - A| < \varepsilon,$$

whenever $|P| < \delta$ and any choice of (ξ_j).

The proof will be completed in a few steps. Students should notice where the uniform continuity is used.

1. The Riemann sum on a partition is between the upper and lower sums. In a Riemann sum on a partition P, if $f(\xi_j)$ is the maximum (resp., minimum) of f on $[x_{j-1}, x_j]$, this Riemann sum is called the upper sum (resp., lower sum) on P, denoted by $\overline{R}(P)$ (resp. $\underline{R}(P)$). Though the Riemann sum depends on the choice of point (ξ_j), it always hold that

$$\underline{R}(P) \leqslant R(P) \leqslant \overline{R}(P).$$

2. Adding more points to a partition makes the upper sum smaller and the lower sum bigger. It is obvious that when $[c, d] \subset [a, b]$,

$$\begin{aligned} &\max\{f(x) : x \in [c, d]\} \leqslant \max\{f(x) : x \in [a, b]\}, \\ &\min\{f(x) : x \in [c, d]\} \geqslant \min\{f(x) : x \in [a, b]\}. \end{aligned}$$

Assume that P is a partition. Adding more points gives us a denser partition Q, $Q \supset P$. Then the lower sum will be bigger and the upper sum will be smaller, i.e.,

$$\underline{R}(P) \leqslant \underline{R}(Q) \leqslant \overline{R}(Q) \leqslant \underline{R}(P).$$

3. To find A, take a special partition sequence: 'cut-half' each time

$$\begin{aligned} P_1 :& 0 < \frac{1}{2} < 1; \\ P_2 :& 0 < \frac{1}{4} < \frac{2}{4} < \frac{3}{4} < 1; \\ & \cdots\cdots; \\ P_n :& 0 < \frac{1}{2^n} < \frac{2}{2^n} < \cdots < \frac{2^n - 1}{2^n} < 1; \\ & \cdots\cdots. \end{aligned}$$

This is a sequence of partitions getting denser and denser. The sequence $(\underline{R}(P_n))$ is increasing and bounded and hence has a limit A. Clearly

$$\underline{R}(P_n) \leqslant A \leqslant \overline{R}(P_n).$$

4. The uniform continuity of f will make the difference of the upper sum and lower sum as small as possible. For $\varepsilon > 0$, we may choose $\delta > 0$ as in the definition of uniform continuity. Fix any partition $P = (x_j)$ with $|P| = \max_j(x_j - x_{j-1}) < \delta$ and any choice $\xi_j \in [x_{j-1}, x_j]$. Then the difference between the maximum M_j and minimum m_j on any $[x_{j-1}, x_j]$ is no bigger than ε, and it implies that
$$\overline{R}(P) - \underline{R}(P) \leqslant \sum_{j=1}^{n} (M_j - m_j)\Delta x_j \leqslant \varepsilon.$$
Applying to P_n, take $n \in \mathbf{N}$ so that $|P_n| = \frac{1}{2^n} < \delta$. Then by the conclusion above
$$\underline{R}(P_n) \leqslant A \leqslant \overline{R}(P_n),\ \ \overline{R}(P_n) - \underline{R}(P_n) < \varepsilon.$$

5. Then let $Q_n = P_n \cup P$ which is a partition denser than P and P_n. Then
$$\begin{cases} \underline{R}(P) \leqslant \underline{R}(Q_n) \leqslant \overline{R}(Q_n) \leqslant \overline{R}(P), \\ \underline{R}(P_n) \leqslant \underline{R}(Q_n) \leqslant \overline{R}(Q_n) \leqslant \overline{R}(P_n). \end{cases} \tag{4.1}$$
Finally the interval $[\underline{R}(P), \overline{R}(P)]$, which contains $R(P)$, and $[\underline{R}(P_n), \overline{R}(P_n)]$, which contains A, have length less than ε and have common points by the property (4.1), so that $|R(P) - A| < 2\varepsilon$.

The proof is completed. □

Exercises

1. prove that if f is integrable on $[a, b]$, then it is bounded.

2. Is the following function uniformly continuous on the given interval?

   (a) $y = \sin x,\ x \in \mathbf{R}$; (b) $y = x^2,\ x \in \mathbf{R}$; (c) $y = x^{-1},\ x \in (0, 1]$;

   (d) $y = x \sin x,\ x \in \mathbf{R}$; (e) $y = \dfrac{\sin x}{x}, x \in (0, \pi)$.

3. Assume that $y = f(x)$ is uniformly continuous on (a, b) and $a \in \mathbf{R}$.

   (a) prove that f is bounded on (a, b).

   (b) prove that $f(x)$ converges when $x \downarrow a$.

4. Assume f is a bounded function on $[a, b]$ which is continuous on $(a, b]$. prove that f is integrable on $[a, b]$. A function on $[a, b]$ with only finite discontinuous points is called piecewise continuous. It follows that a bounded and piecewise continuous function is integrable.

Chapter 5

series and Taylor expansions

5.1 series

5.1.1 series

Let's come back to sequences, but now we will consider their sum.

Definition 5.1.1 Assume that (a_n) is a sequence. The form $\sum\limits_{n\geqslant 1} a_n$, $\sum\limits_{n=1}^{\infty} a_n$, or simply $\sum a_n$, is called a series. The sum

$$s_n = \sum_{j=1}^{n} a_n = a_1 + a_2 + \cdots + a_n$$

is also a sequence and called the partial sum sequence of (a_n). If (s_n) converges (to S), then we say that the series converges (to S), i.e., if $S = \lim_n s_n$, then we write

$$\sum_{n\geqslant 1} a_n = \lim_{n\to\infty} \sum_{j=1}^{n} a_j.$$

If (s_n) diverges, we say the series diverges.

The series above starts from $n = 1$, but this is not essential, it could start from any integer, 0, 100, or even -10. Sometimes we simply write the series as $\sum a_n$ if we do not care where it starts. Whether a series converges does not depend on where the series starts, but the value of the sum does. For example,

$$\sum_{n\geqslant 1}^{\infty} 2^{-n} = 1, \text{ and } \sum_{n\geqslant 0} 2^{-n} = 2.$$

Theorem 5.1.2 If $\sum a_n$ converges, then $\lim_n a_n = 0$.

The proof is simple. Since $a_n = s_n - s_{n-1}$, when (s_n) converges

$$\lim_n a_n = \lim_n s_n = \lim_n s_{n-1} = 0.$$

Hence we have a test for divergence.

Theorem 5.1.3 If a_n does not converge to 0, then $\sum a_n$ diverges.

However that a_n converges to 0 does not mean that $\sum a_n$ converges. For example

$$\sum \frac{1}{n} = +\infty$$

which will be proven later.

Example 5.1.1 *1.* $\sum_{n \geqslant 1} 1 = +\infty$, $\sum_{n \geqslant 1} n = \infty$.

2. By an elementary formula $\displaystyle\sum_{j=0}^{n} x^j = \frac{1 - x^{n+1}}{1 - x}$, *we have when* $|x| < 1$,

$$\sum_{n \geqslant 0} x^n = \lim_n \frac{1 - x^{n+1}}{1 - x} = \frac{1}{1 - x}$$

converges and when $|x| \geqslant 1$, *it diverges.*

The properties of series are similar to integrals.

Theorem 5.1.4 1. If $\sum_n a_n$ converges, and c is constant, then $\sum c a_n = c \sum a_n$ converges.

2. If $\sum a_n$ and $\sum b_n$ converge, then $\sum (a_n + b_n)$ converge and $\sum (a_n + b_n) = \sum a_n + \sum b_n$.

5.1.2 series with positive terms

When every $a_n \geqslant 0$, we say the series $\sum_n a_n$ is a series with positive terms. The partial sum of a series with positive terms is increasing, so that such a series converges if and only if the partial sum is bounded. In this case, the convergence and divergence are written into

$$\sum a_n < \infty \text{ and } \sum a_n = +\infty$$

respectively. Note that this makes sense only for series with positive terms.

Example 5.1.2 *The series* $\sum_n \frac{1}{n^a}$ *is important. It diverges obviously when* $a \leqslant 0$. *Assume now that* $a > 0$. *For any* $j \in \mathbf{N}$, *it is clear that*

$$\frac{1}{(j+1)^a} \leqslant \int_j^{j+1} \frac{1}{x^a} dx \leqslant \frac{1}{j^a}.$$

With this comparison, the improper integral $\int_1^{\infty} \frac{1}{x^a} dx$ *converges if and only if* $\sum_n \frac{1}{n^a}$ *converges. Hence*

$$\sum_n \frac{1}{n^a} \begin{cases} = +\infty, & a \leqslant 1, \\ < \infty, & a > 1. \end{cases}$$

Hence the harmonic series diverges $\sum_{n=1}^{\infty} \frac{1}{n} = +\infty$. It is quite strange for many students why cumulating these small numbers will increase to infinity. The following elementary proof, due to Jakob Bernoulli, might be easier to explain this.

$$\begin{aligned} \sum_{n=1}^{8} \frac{1}{n} &= 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) \\ &> 1 + \frac{1}{2} + \left(\frac{1}{4} + \frac{1}{4}\right) + \left(\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}\right) \\ &= 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} = 1 + \frac{3}{2}. \end{aligned}$$

Next

$$\sum_{n=9}^{16} > \frac{8}{16} = \frac{1}{2}, \quad \sum_{n=17}^{32} \frac{1}{n} > \frac{1}{2}$$

and generally

$$\sum_{n=2^{k-1}+1}^{2^k} \frac{1}{n} > \frac{2^{k-1}}{2^k} = \frac{1}{2}.$$

Then it is seen that

$$\sum_{n=1}^{2^k} \frac{1}{n} > 1 + \frac{k}{2}.$$

The example above shows that the series and improper integrals are related as indicated in the theorem below.

Theorem 5.1.5 (integral test) When $y = f(x)$ is continuous, positive and decreasing on $[1, \infty)$,

$$\int_1^{\infty} f(x) dx < \infty \text{ if and only if } \sum_{n \geqslant 1} f(n) < \infty.$$

The comparison test is a powerful tool for series with positive terms.

Theorem 5.1.6 (comparison test) Assume that $a_n \geqslant 0, \ b_n \geqslant 0$.

1. If $a_n \leqslant b_n$ for n large enough, then that $\sum b_n$ converges implies that $\sum a_n$ converges, or that $\sum a_n$ diverges implies that $\sum b_n$ diverges.

2. If $\lim_n a_n/b_n = c > 0$, then $\sum b_n$ converges if and only if $\sum a_n$ converges.

To use comparison, we need to have some suitable series to compare, either compare ($\leqslant$) to a convergent series, or compare ($\geqslant$) to a divergent series.

Example 5.1.3 1. *Consider* $\sum_n \frac{\ln n}{n^a}$. *It converges if and only if* $a > 1$. *When* $a \leqslant 1$, $\frac{\ln n}{n^a} \geqslant \frac{1}{n^a}$. *When* $a > 1$, *we write* $a = 1 + b$ *with* $b > 0$ *and*

$$\frac{\ln n}{n^a} = \frac{1}{n^{1+b/2}} \cdot \frac{\ln n}{n^{b/2}} < \frac{1}{n^{1+b/2}},$$

because when n *large,* $\ln n < n^{b/2}$.

2. *Let* $b > 0$, $a > 1$. *Consider* $a_n = \frac{n^b}{a^n}$. *We may write*

$$a_n = \frac{n^b}{\sqrt{a}^n} \frac{1}{\sqrt{a}^n} < \frac{1}{\sqrt{a}^n}$$

when n *is large enough, since* $\lim_n n^b/\sqrt{a}^n = 0$. *This implies that* $\sum a_n$ *converges.*

3. *Take any* $x > 0$ *and consider*

$$\sum_{n \geqslant 0} \frac{x^n}{n!}.$$

Because $\lim_n \frac{x^n}{n!} = 0$ *and*

$$\frac{x^n}{n!} = \frac{1}{n(n-1)} \cdot \frac{x^2 \cdot x^{n-2}}{(n-2)!} < \frac{1}{n(n-1)},$$

for n *large enough,* $\sum_{n \geqslant 0} \frac{x^n}{n!}$ *converges. We now use the limit*

$$\lim_n \left(1 + \frac{x}{n}\right)^n = e^x$$

to find the value of series. By Newton's binomial expansion,

$$\begin{aligned} \left(1 + \frac{x}{n}\right)^n &= \sum_{j=0}^{n} \frac{x^j}{j!} \frac{(n-1)\cdots(n-j+1)}{n^{j-1}} \\ &\leqslant \sum_{j=0}^{n} \frac{x^j}{j!} < \sum_{n \geqslant 0} \frac{x^n}{n!}. \end{aligned}$$

For any fixed integer k, *when* $n > k$,

$$\begin{aligned} \left(1 + \frac{x}{n}\right)^n &= \sum_{j=0}^{n} \frac{x^j}{j!} \frac{(n-1)\cdots(n-j+1)}{n^{j-1}} \\ &\geqslant \sum_{j=0}^{k} \frac{x^j}{j!} \frac{(n-1)\cdots(n-j+1)}{n^{j-1}}. \end{aligned}$$

Let $n \to \infty$ *and it follows that for any* $k \in \mathbf{N}$,

$$e^x \geqslant \sum_{j=0}^{k} \frac{x^j}{j!}.$$

Hence

$$e^x = \sum_{n \geqslant 0} \frac{x^n}{n!}.$$

5.1.3 absolute/conditional convergence

Similar to the improper integral, there are also absolute convergence and conditional convergence in series.

Definition 5.1.7 Given a series $\sum a_n$.

1. If $\sum |a_n| < \infty$, we say $\sum a_n$ converges absolutely.
2. If $\sum |a_n| = \infty$ but $\sum a_n$ converges, we say $\sum a_n$ converges conditionally.

For any n, if $a_n \geqslant 0$, define $a_n^+ = a_n$, $a_n^- = 0$, otherwise define $a_n^+ = 0$, $a_n^- = -a_n$. The sequence (a_n^+) and (a_n^-) are two non-negative sequences called the positive part and negative part of (a_n) respectively. Clearly

$$a_n = a_n^+ - a_n^-, \ |a_n| = a_n^+ + a_n^-.$$

It is obvious that when $\sum |a_n| < \infty$, both $\sum a_n^+$ and $\sum a_n^-$ converges so that $\sum a_n$ converges too.

Theorem 5.1.8 If $\sum a_n$ converges conditionally, then both positive part $\sum a_n^+$ and negative part $\sum a_n^-$ diverge.

The reason is that one of them converges implies both converge and then $\sum |a_n| < \infty$, since

$$|a_n| = 2a_n^+ - a_n, \ |a_n| = -2a_n^- + a_n.$$

We knew that the improper integral

$$\int_0^\infty \frac{\sin x}{x} dx$$

converges conditionally. Actually

$$\sum_{n \geqslant 0} \int_{n\pi}^{(n+1)\pi} \frac{\sin x}{x} dx$$

is a series, which converges conditionally.

Exercises

1. Prove that if $\sum a_n$ and $\sum b_n$ converge, then $\sum (a_n + b_n)$ converges. Give a counterexample to show that the inverse is not true.

2. Express the number as a series and then a fraction.

   (a) $0.\overline{2} = 0.22222\cdots$;

   (b) $3.01\overline{214} = 3.01214214214\cdots$;

   (c) $0.\overline{a_1 \cdots a_n}$, a cyclic expression.

3. Determine whether the series is divergent or convergent.

   (a) $\sum_{n \geqslant 1} \sqrt[n]{2}$;

   (b) $\sum_{n \geqslant 1} \ln \frac{n^2+1}{n^2+2}$;

   (c) $\sum_{n \geqslant 1} (e^{1/n} - e^{1/(n+1)})$.

4. If the partial sum s_n of $\sum_{n \geqslant 1} a_n$ is
$$s_n = \frac{n-1}{n+1},$$
find a_n and $\sum_{n \geqslant 1} a_n$.

5. Let
$$s_n = \sum_{j=1}^{n} \frac{1}{j}.$$
Prove that $e^{s_n} > 1 + n$ which implies that $\lim s_n = +\infty$.

5.2 regroup and rearrangement

5.2.1 alternating series

There are not so many effective tests as for series with positive terms to tell if a general series converges. For example, the comparison test does not work in general. The idea which proves that the improper integral above converges can be used to prove the following theorem for alternating series, a very special kind.

Theorem 5.2.1 (alternating series test) If (a_n) is monotone and converges to 0, then the alternating series
$$\sum_n (-1)^n a_n$$
converges.

Proof. Assume that a_n decreases to 0. The partial sum $s_n := \sum_{j=0}^{n} (-1)^j a_j$. The subsequence (s_{2n-1}) can be written into
$$s_{2n-1} = (a_0 - a_1) + (a_2 - a_3) + \cdots + (a_{2n-2} - a_{2n-1})$$
$$s_{2n-1} = a_0 - (a_1 - a_2) - \cdots - (a_{2n-3} - a_{2n-2}) - a_{2n-1} \leqslant a_0.$$

It is seen that it is increasing and bounded. Hence (s_{2n-1}) converges. Since for any n,

$$s_{2n} - s_{2n-1} = a_{2n},$$

(s_{2n}) converges and

$$\lim_n s_{2n} = \lim_n s_{2n-1}.$$

Then $\lim s_n$ exists. □

Example 5.2.1 *The alternating harmonic series*

$$\sum_{n \geqslant 1} \frac{(-1)^{n-1}}{n}$$

converges conditionally.

5.2.2 regroup of series

Given a series $\sum_{n\geqslant 1} a_n$. We may take a subsequence (k_n) of $\mathbf{N}$ and regroup the series

$$(a_1 + \cdots a_{k_1}) + (a_{k_1+1} + \cdots + a_{k_2}) + \cdots = \sum_n b_n$$

where the sum of the n-th group

$$b_n = a_{k_{n-1}+1} + \cdots + a_{k_n}, \ n \geqslant 1,$$

and the number of terms $k_n - k_{n-1}$ is the length of this group. The series $\sum b_n$ is called a regroup of the series $\sum a_n$. Grouping a series may change the convergence. For example, the series $\sum(-1)^n$ diverges but its regroup

$$(1 - 1) + (1 - 1) + \cdots$$

converges. However the partial sum of a regroup series is a subsequence of the original series and hence we have the following theorem.

Theorem 5.2.2 If $\sum a_n$ converges, then any regroup $\sum_n b_n$ converges.

The example $\sum_n (-1)^n$ shows that the inverse is not true. But it is true by adding a condition.

Theorem 5.2.3 If $\sum_n b_n$ converges and one additional condition is satisfied

$$\lim_n \sum_{j=k_{n-1}+1}^{k_n} |a_j| = 0,$$

then $\sum_n a_n$ converges.

In fact, we need to verify that the partial sum (s_n) converges. We know that a sequence s_{k_n} converges. For any n, there exists m such that

$$k_{m-1} < n \leqslant k_m,$$

and

$$|s_n - s_{k_{m-1}}| \leqslant \sum_{j=k_{n-1}+1}^{k_n} |a_j|.$$

When $n \to \infty$, $m \to \infty$ and hence

$$\lim_n s_n = \lim_m s_{k_{m-1}}.$$

5.2.3 rearrangement of series

For addition the commutative law holds $a + b = b + a$. In general for a finite sum

$$x_1 + x_2 + \cdots + x_n$$

the commutative law tells us that the sum does not depend on the order of summation A series is an infinite sum, in which case, we may not be able to change the order of summation. Changing the order of summation is called a rearrangement.

Take the alternating harmonic series, and denote its sum as a

$$a = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots + \frac{(-1)^{n-1}}{n} + \cdots, \tag{5.1}$$

and regroup it

$$a = \sum_{n \geqslant 1} \left(\frac{1}{2n-1} - \frac{1}{2n} \right). \tag{5.2}$$

Clearly $a > 0$. Multiplying the series (5.1) by 1/2,

$$\frac{a}{2} = \sum_{n \geqslant 1} \frac{(-1)^{n-1}}{2n}. \tag{5.3}$$

Add (5.2) and (5.3) together we have

$$\frac{3a}{2} = \sum_{n \geqslant 1} \left(\frac{1}{2n-1} - \frac{1}{2n} + \frac{(-1)^{n-1}}{2n} \right).$$

This gives

$$\frac{3a}{2} = 1 + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} + \frac{1}{7} - \frac{1}{4} + \frac{1}{9} + \frac{1}{11} - \frac{1}{6} + \cdots.$$

The series on the left side is a rearrangement of the series (5.1): every term in the alternating harmonic series appears in this series only once. This rearragement gives a different sum.

Exercises

1. Assume that $\sum_n (a_n - b_n)$ converges. Prove that $\sum_n a_n$ converges if and only if $\sum_n b_n$ converges.

2. Does $\displaystyle\sum_{n\geqslant 1} \frac{(-1)^{n^2}}{n}$ converge?

3. Consider the series
$$\sum_{n\geqslant 1} \frac{(-1)^{[\sqrt{n}]}}{n}.$$
We shall prove it converges in the following steps.

   (a) Write down the first 20 terms of the series.

   (b) Regroup the series according to the integer part of $\sqrt{n}$ into an alternating series $\sum_{n\geqslant 1} (-1)^n b_n$ where $b_n > 0$, and write down b_n explicitly.

   (c) prove (b_n) decreases and converges to zero. Hint:
$$\frac{1}{n+1} < \int_n^{n+1} \frac{1}{x} dx < \frac{1}{n}.$$

   (d) Prove that Theorem 5.2.3 and then $\displaystyle\sum_{n\geqslant 1} \frac{(-1)^{[\sqrt{n}]}}{n}$ converges.

5.3 analysis 4: rearrangement of series and upper limit

5.3.1 Riemann rearrangement theorem

What is a rearrangement? Roughly, it is to sum the same sequence with a different order. Precisely a rearrangement is a rearrangement of $\mathbf{N}$, or a one-to-one correspondence $\sigma : \mathbf{N} \to \mathbf{N}$. This induces a rearrangement of a sequence

$$a_1, a_2, a_3, \cdots, a_n, \cdots \longmapsto a_{\sigma(1)}, a_{\sigma(2)}, a_{\sigma(3)}, \cdots, a_{\sigma(n)}, \cdots.$$

Any a_n appears in the right side once and only once. Then the series $\sum a_{\sigma(n)}$ is called a re-arrangement of the series $\sum a_n$. The following is the Riemann's rearrangement theorem.

Theorem 5.3.1 (Riemann) Given a series $\sum a_n$.

1. If $\sum a_n$ converges absolutely, then any re-arrangement converges to the same sum
$$\sum a_{\sigma(n)} = \sum a_n.$$

2. If $\sum a_n$ converges conditionally, then for any $S \in \mathbf{R}$, there exists a re-arrangement σ such that $\sum a_{\sigma(n)}$ converges to S. This is true when $|S| = \infty$.

The proof of 1: Let $s_n = \sum_{j=1}^n a_n$, and $s'_n = \sum_{j=1}^n a_{\sigma(j)}$. We need to prove that

$$\lim_n s'_n = S$$

where $S = \lim s_n$. Since $\sum a_n$ converges absolutely, i.e., $\sum_{j=1}^\infty |a_n| = L$, for any $\varepsilon > 0$, $\exists N$ such that

$$\sum_{j>N} |a_j| = L - \sum_{j=1}^N |a_j| < \varepsilon.$$

Then when $n \geqslant N$, $|s_n - S| < \varepsilon$. We may find N' such that

$$\{1, 2 \cdots, N\} \subset \{\sigma(1), \sigma(2), \cdots, \sigma(N')\}.$$

Actually $N' = \max\{j : 1 \leqslant \sigma(j) \leqslant N\}$. When $n > N'$, set $A_n = \{\sigma(j) : 1 \leqslant j \leqslant n\}$ and

$$B_n = A_n \setminus \{1, 2 \cdots, N\} \subset \{j \in \mathbf{N} : j > N\}.$$

Then

$$|s'_n - s_N| = \left| \sum_{j \in A_n} a_j - \sum_{j=1}^N a_j \right| = \left| \sum_{j \in B_n} a_j \right| \leqslant \sum_{j \in B_n} |a_j| \leqslant \sum_{j>N} |a_j|$$

and

$$|s'_n - S| \leqslant |s'_n - s_N| + |s_N - S| \leqslant \sum_{j>N} |a_j| + |s_N - S| < 2\varepsilon.$$

The proof of 2: We shall give a rough idea. Students who are interested should learn to write the proof down carefully. Assume that $\sum_n a_n$ converges conditionally. We shall separate (a_n) into two parts: the non-negative pool

$$A^+ = \{a_{k_1}, a_{k_2}, \cdots, \ a_{k_n}, \cdots\},$$

and the negative pool

$$A^- = \{a_{j_1}, a_{j_2}, \cdots, \ a_{j_n}, \cdots\}.$$

We know that $\sum_n a_{k_n} = +\infty$ and $\sum_n a_{j_n} = -\infty$, but a_n goes to 0.
Assume that $S > 0$. We start to take out numbers one by one in order from the pool A^+ and add together until the sum is bigger than S. Then we take out numbers one by one in order from the pool A^- and add to decrease the sum until the sum is smaller than S. Going on this way recursively and we will empty both pools and obtain a rearrangement. The last number we take from the pool each time changes the sum from $<$ to $>$, or from $>$ to $<$, and is called a crossing number. The rearrangement series has the sum S, because the difference between the partial sum of the rearrangement and S is no bigger than the crossing number (this statement should be justified carefully), which goes to 0.

5.3.2 the upper limit

In this section, we shall introduce the notion of upper limits, which will be used to give a criterion that a series converges. Given a sequence (x_n), we may define for any n

$$y_n = \sup\{x_k : k \geqslant n\} = \sup\{x_n, x_{n+1}, x_{n+2}, \cdots\},$$

which is simply called the tail superimum sequence. If (x_n) is bounded above, then (y_n) is a decreasing sequence, otherwise $y_n = +\infty$ for any n.

Definition 5.3.2 If the tail superimum sequence (y_n) converges, we say that its limit is the upper limit of (x_n) and denoted by $\overline{\lim}\, x_n$.

When (x_n) is not bounded above, we would say that the upper limit of (x_n) is $+\infty$, and when x_n goes to $-\infty$, y_n goes to $-\infty$ and we would say the upper limit of (x_n) is $-\infty$.Though the limit of a sequence may not exist, the upper limit always exists, but may not be finite. The following theorem is intuitive and useful.

Theorem 5.3.3 Assume that $\overline{\lim}\, x_n = L$. Then for any $\varepsilon > 0$, there are only finite many x_n bigger than $L + \varepsilon$, or $x_n < L + \varepsilon$ for large n, but there are infinitely many x_n bigger than $L - \varepsilon$.

Proof. Since $\lim y_n = L$, $\exists N$, such that for $n > N$,

$$L - \varepsilon < y_n = \sup\{x_n, x_{n+1}, x_n + 2, \cdots\} < L + \varepsilon.$$

The right $<$ tells us that $x_n < L + \varepsilon$ when $n > N$. The left $<$ tells us that for any $n > N$, there exists $k \geqslant n$ such that $x_k > L - \varepsilon$. Therefore there are infinitely many x_n bigger than $L - \varepsilon$. □

Assume $\overline{\lim}\, x_n = L$. From this theorem it follows that there is a subsequence of (x_n) which converges to L. Actually any point which is the limit of some subsequence of (x_n) is called a limit point of (x_n) and the upper limit is the biggest limit point, the least upper bound of limit points, of (x_n). Similarly we may define the lower limit of (x_n) to be the limit of the tail infimum sequence

$$\underline{\lim}\, x_n = \lim_n \inf\{x_k : k \geqslant n\},$$

which is the least limit point of (x_n).

Example 5.3.1 1. *The sequence* $x_n = 1/n$ *has one limit point* $\{0\}$ *and the upper limit is the same as limit.*

2. $x_n = n(-1)^n$. *The upper limit is* $+\infty$ *and lower limit is* $-\infty$.

3. $x_n = n^{(-1)^n}$. *The upper limit is* $+\infty$ *and lower limit is 0.*

4. *For the sequence* $x_n = (-1)^n(n+1)/n$*, the set of limit points is* $\{-1, 1\}$ *and the upper limit is 1.*

5. *The sequence* $x_n = n$ *has no limit point. Both the upper limit and lower limit are* $+\infty$*.*

6. *We said that every real number may be approximated by a sequence of rational numbers. The set of rational numbers can be listed as a sequence* (x_n)*. How? This means that the set of limit points of this sequence is* $\mathbf{R}$*.*

5.3.3 root test and ratio test

For a general series $\sum a_n$, we also have test for absolute convergence. The following theorem is called the root test.

Theorem 5.3.4 (root test) Assume that the root limit $\overline{\lim}\ \sqrt[n]{|a_n|} = r$ (could be $+\infty$). When $r < 1$, $\sum a_n$ converges absolutely. When $r > 1$, $\sum a_n$ diverges.

Proof. If $r < 1$, then, by the theorem above, when n is large enough,

$$\sqrt[n]{|a_n|} < \frac{1+r}{2} = s < 1.$$

Hence $|a_n| < s^n$, for large n, and it follows that $\sum a_n$ converges absolutely. If $r > 1$, then, by the theorem above, we take $\varepsilon = r - 1$, there are infinitely many n such that

$$\sqrt[n]{|a_n|} > r - \varepsilon = 1.$$

Hence for those n, $|a_n| > 1$ and then by Theorem 5.1.2, $\sum a_n$ diverges, because a_n does not converges to 0. □

When $r = 1$, it fails to give a definite answer. Here is an example. When $a_n = n^{-a}$ with $a > 0$, $\sqrt[n]{a_n}$ converges to 1. However $\sum a_n$ may converge and diverge. That's why we say it tells nothing when $r = 1$.

The following theorem gives ratio test. The ratio limit is easier to get sometimes.

Theorem 5.3.5 (ratio test) If $a_n \neq 0$ for any n and the ratio limit $\lim \left|\frac{a_{n+1}}{a_n}\right|$ exists, then

$$\lim \sqrt[n]{|a_n|} = \lim \left|\frac{a_{n+1}}{a_n}\right|.$$

Proof. Assume that the ratio limit is r. For any $r > \varepsilon > 0$, $\exists N$ such that when $n \geqslant N$,

$$\left|\frac{|a_{n+1}|}{|a_n|} - r\right| < \varepsilon, \text{ or } r - \varepsilon < \frac{|a_{n+1}}{|a_n|} < r + \varepsilon.$$

Now for $n > N$,

$$|a_n| = |a_N|\frac{|a_{N+1}|}{|a_N|}\frac{|a_{N+2}|}{|a_{N+1}|}\cdots\frac{|a_n|}{|a_{n-1}|},$$

and

$$|a_N|(r-\varepsilon)^{n-N} < |a_n| < |a_N|(r+\varepsilon)^{n-N}.$$

Then

$$\sqrt[n]{|a_N|}(r-\varepsilon)^{1-N/n} < \sqrt[n]{|a_n|} < \sqrt[n]{|a_N|}(r+\varepsilon)^{1-N/n}.$$

Let $n \to +\infty$ now. Since a_N is a constant, $\sqrt[n]{a_N} \to 1$ and we obtain

$$r - \varepsilon \leqslant \lim_n \sqrt[n]{|a_n|} \leqslant r + \varepsilon.$$

Since ε is arbitrarily small, $\lim \sqrt[n]{|a_n|} = r$. □

It is seen from this theorem whenever the ratio limit exists, the root limit also exists and hence the ratio test is not as powerful as the root test. However the ratio test is more convenient sometimes.

Exercises

1. Find the upper limit $\overline{\lim}_n (-1)^{[\sqrt{n}]} \dfrac{n}{n+1}$.
2. Find an example where

$$\overline{\lim}\,(a_n + b_n) \neq \overline{\lim}\, a_n + \overline{\lim}\, b_n,$$

   i.e., the upper limit is not additive.

3. Determine whether the series diverges or converges.
   (a) $\displaystyle\sum_n \frac{3^n + 2^n + 1}{4^n}$;
   (b) $\displaystyle\sum_n \frac{10^n}{2^{n^2}}$;
   (c) $\displaystyle\sum_n 2^{-n(-1)^n}$;
   (d) $\displaystyle\sum_n \frac{n!}{2^{n^2}}$.

5.4 power series

Example 5.4.1 *1.* $a_n = n^a x^n$. *Then*

$$\sqrt[n]{|a_n|} = (\sqrt[n]{n})^a |x| \to |x|.$$

When $|x| < 1$, $\sum a_n$ *converges. When* $|x| > 1$, *it diverges.*

2. $\sum_n \left(\frac{2n+10}{3n+20}\right)^n$. *It is better to use the root test,*

$$\left[\left(\frac{2n+10}{3n+20}\right)^n\right]^{1/n} = \frac{2n+10}{3n+20} \to \frac{2}{3} < 1$$

and hence the series converges. In this case, it is obvious that the root test is more convenient.

3. $a_n = \frac{n!}{n^n}$. *Then*

$$\sqrt[n]{a_n} = \frac{\sqrt[n]{n!}}{n}.$$

It is difficult to get this limit. In this case the ratio test is more convenient

$$\frac{a_{n+1}}{a_n} = \frac{(n+1)!}{(n+1)^{n+1}} \frac{n^n}{n!} = \left(\frac{n}{n+1}\right)^n \to \frac{1}{e}.$$

Hence the series $\sum a_n$ *converges and we have limit*

$$\lim \frac{\sqrt[n]{n!}}{n} = e^{-1}.$$

The strategy to tell whether a series $\sum_n a_n$ converges:

1. $a_n \to 0$? if not, diverges.

2. whether or not $\sum |a_n|$ converges? Comparison to known series, root test, ratio test. It is hard to say which is better.

3. If $\sum |a_n|$ converges, stop.

4. If $\sum |a_n| = \infty$, whether or not $\sum a_n$ converges? If it is an alternating series, we have a test.

5. None of above, we need to solve it by other methods.

5.4.1 power series

A linear combination of $1, x, x^2, \cdots, x^n$,

$$a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

is called a polynomial in x. Some function can be written into an infinite sum of power functions

$$\frac{1}{1-x} = 1 + x + x^2 + \cdots + x^n + \cdots = \sum_{n \geqslant 0} x^n,$$

$$e^x = 1 + x + \frac{x^2}{2} + \cdots + \frac{x^n}{n!} + \cdots .$$

They can be approximated by polynomials. This is particularly useful in engineering computation. For example, to approximate the Euler's number e, it is much easier to use the series

$$e = \sum_{n\geqslant 0} \frac{1}{n!} = \lim_n \sum_{j=0}^{n} \frac{1}{j!}$$

than to use limit

$$e = \lim_n \left(1 + \frac{1}{n}\right)^n.$$

Definition 5.4.1 A series

$$\sum_{n\geqslant 0} a_n (x-a)^n$$

is called a power series centered at x or about a. If $a = 0$, it is called a power series. The number a_n is the coefficient of $(x-a)^n$ for $n \geqslant 0$.

Usually we need only to study power series and the conclusions are similar for power series centered at a. If x is viewed as a variable, then the power series is a function of x

$$f(x) = \sum_{n\geqslant 0} a_n (x-a)^n$$

when the series converges. In this case the function f is called the sum function of the series. However we are more interested in the other way around, namely expanding a given function into a power series.

5.4.2 radius of convergence

We prove a key theorem.

Theorem 5.4.2 If $x_0 \neq 0$ and $\sum a_n x_0^n$ converges, then for any $|x| < |x_0|$, $\sum a_n x^n$ converges absolutely.

Proof. Since $\sum a_n x_0^n$ converges, $\lim a_n x_0^n = 0$. Hence

$$|a_n x^n| = |a_n x_0^n| \left(\frac{|x|}{|x_0|}\right)^n \leqslant \left(\frac{|x|}{|x_0|}\right)^n,$$

when n is large enough, and by the comparison theorem $\sum a_n x^n$ converges absolutely when $|x|/|x_0| < 1$. $\square$

Then the following theorem is obvious.

Theorem 5.4.3 There are three cases for a power series $\sum_{n\geqslant 0} a_n (x-a)^n$.

1. The series converges only at $x = a$.

2. The series converges for all $x \in \mathbf{R}$.

3. There exists a positive number R such that the series converges when $|x-a|<R$ and diverges when $|x-a|>R$.

Definition 5.4.4 The number $0 \leqslant R \leqslant +\infty$ above is called the radius of convergence of the power series. The set

$$\{x \in \mathbf{R} : \sum a_n (x-a)^n \text{ converges}\}$$

is an interval, called the interval of convergence or the field of convergence.

Actually

$$R = \sup\{|x-a| : \sum_{n \geqslant 0} a_n (x-a)^n \text{ converges}\}.$$

This tells us only the existence of R. How to get the value of R? By the root test, the series converges when

$$\overline{\lim}\ \sqrt[n]{|a_n(x-a)^n|} = |x-a| \cdot \overline{\lim}\ \sqrt[n]{|a_n|} < 1,$$

and diverges when the upper limit > 1. Hence we have

$$R = \frac{1}{\overline{\lim}\ \sqrt[n]{|a_n|}}$$

which is equal to $\lim a_n/a_{n+1}$, if it exists, The radius R is 0 or ∞ when the upper limit is ∞ or 0. When $0 < R < \infty$, the convergence of the power series at two points $\{a-R, a+R\}$, where the root test fails, are to be determined by other approaches.

Example 5.4.2 *After obtaining the radius of convergence, we need to see the convergence when $x = -R$ and $x = R$ and obtain the interval of convergence.*

1. $\sum x^n$. $R = 1$ *and the interval of convergence is* $(-1, 1)$.
2. $\sum \frac{x^n}{n}$. $R = 1$ *and the interval of convergence is* $[-1, 1)$.
3. $\sum \frac{x^n}{n^2}$. $R = 1$ *and the interval of convergence is* $[-1, 1]$.
4. $\sum \frac{x^n}{n!}$.
$$R = \lim \frac{(n+1)!}{n!} = \lim(n+1) = +\infty.$$
5. $\sum n! x^n$. $R = 0$ *and the series converges only at* $x = 0$.

Note when $|x-a| < R$, the power series converges absolutely but when $|x-a| = R$, it may converge conditionally, as shown in the example above.

5.4.3 power series expansion

A power series $\sum_{n\geqslant 0} a_n(x-a)^n$ can be seen as a function with the domain being the interval of convergence.

Definition 5.4.5 Assume that $y = f(x)$ defined on a neighborhood of a. If

$$f(x) = \sum_{n\geqslant 0} a_n(x-a)^n \tag{5.4}$$

for any x in a neighborhood of a, then we say that f has a power series expansion or can be expanded into a power series centered at a.

Be careful! In the definition above, (5.4) should hold in an open interval containing a. Actually a neighborhood of a or near a means that an interval $(a-\delta, a+\delta)$ for some $\delta > 0$. If f has a power series expansion centered at a, then it is necessary that the power series must have a positive radius of convergence. It is known that

$$\sum_{n\geqslant 0} x^n = \frac{1}{1-x}, \ |x| < 1, \quad \text{and} \ \sum_{n\geqslant 0} \frac{x^n}{n!} = e^x, \ x \in \mathbf{R}.$$

Hence $y = \frac{1}{1-x}$ and $y = e^x$ have a power series expansion centered at 0. We may use these two expansions to get other expansions.

Example 5.4.3 1. *Assume that* $a \neq 0$.

$$\frac{1}{a-x} = \frac{1}{a}\frac{1}{1-x/a} = \frac{1}{a}\sum_{n\geqslant 0} x^n/a^n.$$

The radius of convergence is $R = |a|$.

2. *When* $a = 0$, $\frac{1}{x}$ *is not continuous at* $x = 0$. *Can it be expanded into a power series centered at* 0*? The answer is no and we shall prove it later. But it can be expanded into a power series about* $c \neq 0$,

$$\frac{1}{x} = \frac{1}{c+(x-c)} = \frac{1}{c}\frac{1}{1-(-(x-c)/c)} = \frac{1}{c}\sum_{n\geqslant 0}\frac{(-1)^n(x-c)^n}{c^n}.$$

3. *Consider the following expansion*

$$\frac{1}{1-x^2} = \sum_{n\geqslant 0} x^{2n} = \sum_{n\geqslant 0} a_n x^n,$$

where $a_{2n} = 1$ *and* $a_{2n+1} = 0$. *Hence* $\overline{\lim}\ \sqrt[n]{|a_n|} = 1$ *but* $\lim_n \sqrt[n]{|a_n|}$ *does not exist.*

How to expand the functions such as $y = \sin x$, $y = \ln x$, $y = \sqrt{x}$ and other elementary functions? We need more techniques.

Exercises

1. Determine whether the series converges. If it does, does it converge absolutely or conditionally?

   (a) $\sum_{n\geqslant 1} \frac{\sin n}{n^2}$;

   (b) $\sum_{n\geqslant 1} \frac{\sin(n\pi/2)}{n}$;

   (c) $\sum_{n\geqslant 1} (1-1/n)^{n^2}$.

2. Prove that for any $b \in \mathbf{R}$, the power series $\sum_n a_n x^n$ and $\sum_n a_n n^b x^n$ have the same radius of convergence. Give an example where they have different intervals of convergence.

3. Find the radius of convergence of the series
$$\sum_n \left(1 + \frac{1}{2!} + \cdots + \frac{1}{n!}\right) x^n.$$

4. Expand $y = \frac{1}{x}$ into a power series about $x = 9$.

5. Analyze the product of two power series $\sum_n a_n x^n$ and $\sum_n b_n x^n$ to obtain a method to expand the function $y = \dfrac{e^x}{1-x}$ about $x = 0$.

5.5 Taylor expansions

5.5.1 term-by-term differentiation and integration

The following theorem looks intuitive but hard to prove. We will skip the proof now.

Theorem 5.5.1 Assume that
$$f(x) = \sum_{n\geqslant 0} a_n (x-a)^n$$
for $|x-a| < R$ where $R > 0$ is the radius of convergence.

1. term-by-term differentiation: for $|x-a| < R$,
$$f'(x) = \sum_{n\geqslant 0} n a_n (x-a)^{n-1}.$$

2. term-by-term integration: for $|x-a| < R$,
$$\int_a^x f(y)\,dy = \sum_{n\geqslant 0} \frac{a_n}{n+1} (x-a)^{n+1}.$$

The radii of convergence of both power series above are R.

Because $(a_n(x-a)^n)' = na_n(x-a)^{n-1}$ and

$$\int_a^x a_n(y-a)^n dy = \frac{a_n}{n+1}(x-a)^{n+1},$$

the formulae above can be written more intuitively into

1. term-by-term differentiation

$$\left(\sum_{n\geqslant 0} a_n(x-a)^n\right)' = \sum_{n\geqslant 0}(a_n(x-a)^n)'.$$

2. term-by-term integration

$$\int_a^x \sum_{n\geqslant 0} a_n(y-a)^n dy = \sum_{n\geqslant 0}\int_a^x a_n(y-a)^n dy.$$

From these two formulae, we will obtain more power series expansions.

Example 5.5.1 *We start from the well-known power series expansion*

$$\frac{1}{1-x} = \sum_{n\geqslant 0} x^n, \ |x| < 1.$$

1. term-by-term differentiation: for $|x| < 1$,

$$\frac{1}{(1-x)^2} = \sum_{n\geqslant 0} nx^{n-1} = \sum_{n\geqslant 0}(n+1)x^n.$$
$$\frac{1}{(1-x)^3} = \sum_{n\geqslant 0}(n+1)nx^{n-1} = \sum_{n\geqslant 0}(n+2)(n+1)x^n.$$

2. term-by-term integration

$$-\ln(1-x) = \int_0^x \frac{1}{1-y}dy = \sum_{n\geqslant 0}\int_0^x y^n dy = \sum_{n\geqslant 0}\frac{x^{n+1}}{n+1}.$$

In the examples above, we know that each expansion holds for $|x| < 1$ but we do not know whether the equality holds when $|x| = 1$ even if the power series converges there. In the 2nd example above, the right side converges at $x = -1$. Hence it is natural to conjecture a very beautiful result for the exact value of the alternating harmonic series

$$\ln 2 = -\sum_{n\geqslant 1}\frac{(-1)^n}{n} = \sum_{n\geqslant 1}\frac{(-1)^{n-1}}{n},$$

which we shall prove later.

5.5.2 Taylor expansions

Given a function f, there are two questions: when and where can f be expanded into a power series at a? The following theorem gives a necessary condition and the uniqueness.

Theorem 5.5.2 If f has a power series expansion centered at a

$$f(x) = \sum_{n \geqslant 0} a_n (x-a)^n,$$

then f is infinitely differentiable in a neighborhood of a and

$$a_n = \frac{f^{(n)}(a)}{n!}, \ n \geqslant 0.$$

The coefficients in the power series are determined by f and this implies that the power series is unique.

Proof. It is obvious $f(a) = a_0$ by setting $x = a$ above. Because the series has positive radius of convergence, we can take derivative term-by-term

$$f'(x) = a_1 + 2a_2(x-a) + 3a_3(x-a)^2 + \cdots,$$

and hence $f'(a) = a_1$. Do it again and we obtain $f''(a) = 2a_2$. Finally we have the formula above. □

Definition 5.5.3 When $y = f(x)$ is infinitely differentiable at a, the power series

$$\sum_{n \geqslant 0} \frac{f^{(n)}(a)}{n!}(x-a)^n$$

is well-defined and called the Taylor series or expansion of f about a. It is also known as MacLaurin series or expansion when $a = 0$.

The theorem above tells us that if f can be expanded into a power series, the series must be its Taylor series. Unfortunately the condition that f is infinitely differentiable does not guarantee that f can be expanded.

Example 5.5.2 *(Cauchy's example published in 1822) assume that*

$$f(x) = \begin{cases} e^{-x^{-2}}, & x \neq 0, \\ 0, & x = 0. \end{cases}$$

Then f is infinitely differentiable on $\mathbf{R}$ and, in particular, for any $n \geqslant 0$, $f^{(n)}(0) = 0$. In fact, when $x \neq 0$,

$$f'(x) = e^{-x^{-2}} \frac{2}{x^3}, \ f''(x) = e^{-x^{-2}} \left(\frac{4}{x^6} - \frac{6}{x^4} \right), \ f'''(x) = \cdots.$$

It can be verified by induction that

$$f^{(n)}(x) = e^{-x^{-2}} g_n(1/x),$$

where g_n *is a polynomial, and*

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

Hence the Taylor series of f *is 0, namely* $a_n = 0$ *for all* n*, and* f *does not have a power series expansion by the definition.*

We now want to know that for what x, f and its Taylor expansion coincide

$$f(x) = \sum_{n \geqslant 0} \frac{f^{(n)}(a)}{n!}(x-a)^n.$$

Note that two sides are two functions, i.e., f, defined on its domain, and its Taylor expansion which is defined on the interval of convergence. For example, we know $\ln(1-x)$ has power series expansion

$$\ln(1-x) = -\sum_{n \geqslant 0} \frac{x^{n+1}}{n+1} == -\sum_{n \geqslant 1} \frac{x^n}{n},$$

for $|x| < 1$, but we do not know if it holds at $x = -1$.

Let

$$R_n(x) = f(x) - \sum_{j=0}^{n} \frac{f^{(j)}(a)}{j!}(x-a)^j$$

which is called the remainder of Taylor series. It is seen that for any x,

$$f(x) = \sum_{n \geqslant 0} \frac{f^{(n)}(a)}{n!}(x-a)^n$$

if and only if the remainder $R_n(x) \longrightarrow 0$ as $n \to +\infty$. Fortunately the remainder can be expressed in terms of f.

5.5.3 the expression of the remainder

The following theorem is an integral expression of the remainder, which is called the integral remainder.

Theorem 5.5.4 Assume that $y = f(x)$ is differentiable in any order (infinitely differentiable) on a neighborhood U of a. For any $n \geqslant 0$ and $x \in U$,

$$R_n(x) = \frac{1}{n!} \int_a^x f^{(n+1)}(y)(x-y)^n dy.$$

Proof. For simplicity, assume $a = 0$. Then for $x > 0$, by the integration by parts formula,

$$\begin{aligned} f(x) - f(a) &= \int_a^x f'(y)dy = -\int_a^x f'(y)d(x-y) \\ &= -f'(y)(x-y)\Big|_a^x + \int_a^x (x-y)f''(y)dy \\ &= f'(a)(x-a) + \int_a^x f''(y)(x-y)dy, \ n = 1; \\ &= f'(a)(x-a) - \frac{1}{2}\int_a^x f''(y)d(x-y)^2 \\ &= f'(a)(x-a) + \frac{1}{2}\left(-f''(y)(x-y)^2\Big|_a^x + \int (x-y)^2 f'''(y)dy\right) \\ &= f'(a)(x-a) + \frac{1}{2}f''(a)(x-a)^2 + \frac{1}{2}\int_a^x f'''(y)(x-y)^2 dy, \ n = 2. \end{aligned}$$

Similarly we can prove the formula for any n. □

For other expressions of the remainder $R_n(x)$, we need to prepare a theorem, which is called the integral mean-value theorem.

Theorem 5.5.5 If $y = f(x)$ is continuous on $[a, b]$, then there exists $\xi \in (a, b)$ such that

$$f(\xi) = \frac{1}{b-a}\int_a^b f(x)dx.$$

More generally if g is a positive continuous function on $[a, b]$, then there exists $\xi \in (a, b)$ such that

$$f(\xi)\int_a^b g(x)dx = \int_a^b f(x)g(x)dx.$$

Proof. The first result is a special case of the second, when $g = 1$. However the second result is an easy consequence of the intermediate value theorem of continuous functions because

$$m \leqslant \frac{\int_a^b f(x)g(x)dx}{\int_a^b g(x)dx} \leqslant M$$

where m, M are the minimum and maximum of f on $[a, b]$. □

Carefully looking, the second formula holds when g is either positive identically or negative identically between a, b for even $a > b$.

Applying the integral mean-value theorem directly gives

$$R_n(x) = \frac{(x-a)(x-\xi)^n f^{(n+1)}(\xi)}{n!},$$

where ξ is between a and x. This expression of the remainder is called the Cauchy remainder. By the generalized integral mean-value theorem, there exists ξ between a and x such that

$$\begin{aligned} R_n(x) &= \frac{1}{n!}\int_a^x f^{(n+1)}(y)(x-y)^n dy \\ &= \frac{1}{n!}f^{(n+1)}(\xi)\int_a^x (x-y)^n dy \\ &= \frac{1}{(n+1)!}f^{(n+1)}(\xi)(x-a)^{n+1}. \end{aligned}$$

This expression of the remainder is called the Lagrange remainder, which is more commonly used than others. Notice that ξ in either expression depends on x and also on n.
Recall that the Taylor expansion holds

$$f(x) = \sum_{n\geqslant 0} \frac{f^{(n)}(a)}{n!}(x-a)^n$$

whenever $\lim_n R_n(x) = 0$.

Exercises

1. Find the power series expansion for $y = \arctan x$ about $x = 0$ by its derivative
$$(\arctan x)' = \frac{1}{1+x^2}.$$

2. Define
$$y = \begin{cases} \frac{\ln(1+x)}{x}, & x > -1, x \neq 0, \\ 1, x = 0. \end{cases}$$
Prove that this function is infinitely differentiable at $x = 0$ and find $f^{(n)}(0)$ for any $n \geqslant 1$.

3. Find the power series of $y = x^{-2}$ about $x = 1$.

4. Find Taylor remainder R_2 about the given point.
   (a) $y = \dfrac{e^x}{1-x}$, $x = 0$.
   (b) $y = e^x \sin x$, $x = 0$.

5. Assume that
$$f(x) = \begin{cases} e^{-x^{-2}}, & x \neq 0, \\ 0, & x = 0. \end{cases}$$
   (a) Prove that f is infinitely differentiable at $x = 0$ and $f^{(n)}(0) = 0$.
   (b) Explain why f does not have a power series expansion by the definition.

5.6 function sequences

Example 5.6.1 *We consider some functions as examples.*

1. $f(x) = \dfrac{1}{1-x}$. $f^{(n)}(x) = \dfrac{n!}{(1-x)^{n+1}}$. *Its Taylor series is* $\sum_{n\geqslant 0} x^n$ *and its Lagrange remainder is*

$$R_n(x) = \frac{1}{(n+1)!}(n+1)!\frac{x^{n+1}}{(1-\xi)^{n+2}} = \left(\frac{x}{1-\xi}\right)^{n+1}(1-\xi)^{-1}.$$

Actually it is easy to prove that $R_n(x)$ *converges to 0 when* $x \in (-1, 1/2)$ *but not so easy for* $x \in [1/2, 1)$. *On the other hand, its Cauchy remainder is*

$$R_n(x) = \frac{1}{n!}\frac{(n+1)!}{(1-\xi)^{n+2}}x(x-\xi)^n = (n+1)x(1-\xi)^{-2}\left(\frac{x-\xi}{1-\xi}\right)^n,$$

where ξ *is between* 0 *and* x. *It is not hard to prove that* $R_n(x)$ *converges to 0 when* $x \in (-1, 1)$, *because in this case*

$$\frac{|x-\xi|}{|1-\xi|} \leqslant |x| < 1.$$

From this example, we know that the different remainder expressions have different applications.

A remark is necessary for why we need to require

$$\frac{|x-\xi|}{|1-\xi|} \leqslant |x|$$

here. When $x \in (-1, 1)$, *and* ξ *is between* $0, x$, *it is obvious that* $\dfrac{|x-\xi|}{1-\xi} < 1$, *but this is not enough to have*

$$\left(\frac{|x-\xi|}{1-\xi}\right)^n \longrightarrow 0$$

because ξ *depends not only on* x *but also on* n. *For example* $1-1/n < 1$ *but* $(1-1/n)^n$ *does not converge to zero.*

2. $f(x) = e^x$. $f^{(n)} = e^x$ *and its Taylor series is* $\displaystyle\sum_{n\geqslant 0}\frac{x^n}{n!}$. *When does* $e^x = \displaystyle\sum_{n\geqslant 0}\frac{x^n}{n!}$? *Its Lagrange remainder is*

$$R_n(x) = \frac{1}{(n+1)!}e^{\xi}x^n,$$

where ξ *is between* $0, x$. *Since* $|\xi| < |x|$, *we have*

$$|R_n(x)| \leqslant \frac{1}{(n+1)!}e^{|x|}|x|^n$$

which converges to 0 for any $x \in \mathbf{R}$.

3. *$f(x) = \sin x$. $f^{(2n)} = (-1)^n \sin x$, $f^{(2n+1)} = (-1)^n \cos x$. Hence for any n and x, $|f^{(n)}(x)| \leqslant M$ and we have Taylor expansion*

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots + \frac{(-1)^n x^{2n+1}}{(2n+1)!} + \cdots.$$

Similarly

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

5.6.1 exchange of limit and integral

We shall now prove the term-by-term integration and differentiation formula and discuss whether the Taylor expansion holds at the end points of its interval of convergence.

The partial sum $S_n(x)$ of a power series is a function sequence, i.e., for any n, $S_n(x)$ is a function on D. We shall discuss when the following exchange between $\lim_n$ and $\int_a^b$ is true

$$\int_a^b \lim_n S_n(x)dx = \lim_n \int_a^b S_n(x)dx,$$

and the exchange between the limit and differentiation.

Assume that $\{f_n(x)\}$ is a function sequence on D. If for any $x \in D$, $f_n(x)$ converges and the limit is denoted by $f(x)$, which is a function on D, we say the function sequence $f_n(x)$ converges to $f(x)$ pointwise on D, denoted by $f_n \longrightarrow f$ pointwise on D.

Example 5.6.2 1. *Assume that $f_n(x) = x^n$, $D = [0, 1]$, Then f_n converges pointwise on D to the function*

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

Clearly f_n is continuous, but f is not continuous at $x = 1$.

2. *Define*

$$f_n(x) = \begin{cases} (n+1)x^n, & x < 1, \\ 0, & x = 1. \end{cases}$$

Then f_n converges to 0 on $[0, 1]$ and

$$\int_0^1 (n+1)x^n dx = 1.$$

Hence

$$\lim \int_0^1 f_n(x)dx \neq \int_0^1 f(x)dx.$$

This example shows that the pointwise limit of a sequence of continuous functions is not continuous and the pointwise limit does not commute with the integral. Hence we need a stronger convergence.

5.6.2 uniform convergence

Assume that $f_n \longrightarrow f$ pointwise on D, i.e., for any $x \in D$ and $\varepsilon > 0$, $\exists N$ such that when $n \geqslant N$,

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

The integer N we found depends on ε but also on x. We need a convergence where N does not depend on $x \in D$.

Definition 5.6.1 Assume that $\{f_n\}$ is a function sequence and f a function on D. We say that f_n converges to f uniformly on D if for any $\varepsilon > 0$, there exists N such that when $n \geqslant N$, it holds that

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

for all $x \in D$ **uniformly**.

Exercises

1. Using the formula
$$\arcsin x = \int_0^x \frac{dy}{\sqrt{1-y^2}}$$
to derive Taylor expansion for $y = \arcsin x$.

2. Assume that
$$f_n(x) = \sqrt{x^2 + 1/n}, \ x \in \mathbf{R}.$$
Find the pointwise limit f of f_n and determine whether it converges uniformly. Observe if f_n and f are differentiable.

3. Find the Taylor series of the following functions and prove that f equals its Taylor series on $(-1, 1)$.

   (a) $f(x) = (1-x)^{-a}$, where $a > 0$;

   (b) $f(x) = (1-x^2)^{-1/2}$;

   (c) $f(x) = \arcsin x$.

4. Let $f_n(x) = (\sin(1/x))^n$, $x > 0$. Ask where the pointwise limit of f_n exists.

5.7 analysis 5: uniform convergence

Theorem 5.7.1 If $\{f_n\}$ is a sequence of continuous functions on $[a, b]$ and converges to f uniformly on $[a, b]$, then f is continuous on $[a, b]$ and

$$\lim_n \int_a^b f_n(x)dx = \int_a^b f(x)dx.$$

Proof. By the uniform convergence, for any $\varepsilon > 0$, $\exists N$ such that

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

holds for $n \geqslant N$ and all $x \in [a, b]$. Since f_N is continuous on $[a, b]$, it is uniformly continuous, i.e., $\exists \delta > 0$ such that

$$f_N(x) - f_N(y)| < \varepsilon$$

for all $x, y \in [a, b]$ with $|x - y| < \delta$. Now for such x, y,

$$|f(x) - f(y)| \leqslant |f(x) - f_N(x)| + |f_N(x) - f_N(y)| + |f_N(y) - f(y)| < 3\varepsilon.$$

Hence f is continuous on $[a, b]$.

Now when $n \geqslant N$,

$$\left| \int_a^b f_n(x)dx - \int_a^b f(x)dx \right| \leqslant \int_a^b |f_n(x) - f(x)|dx < \varepsilon(b - a).$$

That completes the proof. □

Theorem 5.7.2 Assume that f_n is a function sequence smooth on (a, b). If (f_n) converges pointwise to f on (a, b) and its derivative (f_n') is a sequence of continuous function which converges uniformly to g on (a, b), then f is smooth on (a, b) and

$$f'(x) = g(x), \; x \in (a, b).$$

Proof. Take any point $a < u < y < b$, by the theorem above, we have

$$f_n(y) - f_n(u) = \int_u^y f_n'(x)dx \longrightarrow \int_u^y g(x)dx.$$

It implies that

$$f(y) - f(u) = \int_u^y g(x)dx,$$

and then $f'(x) = g(x)$ for any $x \in (a, b)$. □

The conclusion can be written into

$$(\lim_n f_n(x))' = \lim_n f_n'(x), \; x \in (a, b).$$

Hence we say that the limit and derivative commute under this condition.

Consider two function sequences in the previous example. They are not uniformly convergent because their limits are not continuous. But they both converge uniformly on $[0, a]$ when $a < 1$, because

$$0 \leqslant (n+1)x^n \leqslant (n+1)a^n,$$

and the right side does not depend on x. It is obvious that if

$$|f_n(x) - f(x)| \leqslant a_n$$

for all $x \in D$ and $\lim_n a_n = 0$, then f_n converges to f uniformly on D.

5.7.1 uniform convergence of power series

A power series $\sum_{n\geqslant 0} a_n x^n$ is a particular type of function sequence

$$f_n(x) = \sum_{j=0}^{n} a_j x^j, \ n \geqslant 0.$$

When f_n converges uniformly, we say the power series converges uniformly. Since

$$\sum_{j=0}^{n} x^j = \frac{1-x^{n+1}}{1-x},$$

the series $\sum x^n$ converges pointwise on $(-1, 1)$ and not uniformly converges, but uniformly converges on $[-a, a]$ when $a < 1$.

Theorem 5.7.3 Assume that its radius of convergence is $R > 0$. Then for any $r \in (0, R)$, the series converges uniformly on $[-r, r]$.

Proof. We may assume that

$$f(x) = \sum_{n\geqslant 0} a_n x^n, \ x \in (-R, R).$$

Then $\sum a_n r^n$ converges absolutely and for $|x| \leqslant r$,

$$|f(x) - f_n(x)| = \left|\sum_{j>n} a_j x^j\right| \leqslant \sum_{j>n} |a_n||r|^n \longrightarrow 0.$$

Hence f_n converges uniformly to f on $[-r, r]$. □

Since term-by-term differentiation $\sum n a_n x^{n-1}$ and integration $\sum \frac{a_n}{n+1} x^{n+1}$ have the same radius of convergence R, we can apply the theorems above to prove Theorem 5.5.1 directly.

5.7.2 Abel's theorem and Leibniz series

Let's prove the value of the alternating harmonic series

$$\ln 2 = \sum_{n\geqslant 1} \frac{(-1)^{n-1}}{n}.$$

It suffices to show that the power series

$$\ln(1-x) = -\sum_{n\geqslant 0} \frac{x^{n+1}}{n+1} == -\sum_{n\geqslant 1} \frac{x^n}{n},$$

converges uniformly on $[-1,0]$. This is a consequence of Abel's theorem.

Theorem 5.7.4 (Abel) If $\sum_{n\geqslant 0} a_n$ converges, then

$$\sum_{n\geqslant 0} a_n x^n$$

converges uniformly on $[0,1]$.

Write $f_n(x) = \sum_{j=0}^{n} a_j x^j$. By the assumption, $f_n(x)$ converges pointwise on $[0,1]$ to a function, denoted by $f(x)$. Set the remainder $A_n = \sum_{k=n}^{\infty} a_k$, $n \geqslant 1$. Since $\sum a_n$ converges, there exists N such that when $n > N$,

$$|A_n| < \frac{\varepsilon}{2}.$$

Write $a_j = A_j - A_{j+1}$. Since $x \in [0,1]$, $1 \geqslant x^j \geqslant x^{j+1}$ and we have for $n > N$

$$\begin{aligned}
|f(x) - f_n(x)| &= \left|\sum_{j=n}^{\infty} a_j x^j\right| = \left|\sum_{j\geqslant n} (A_j - A_{j+1}) x^j\right| \\
&= \left|(A_n - A_{n+1})x^n + (A_{n+1} - A_{n+2})x^{n+1} + (A_{n+2} - A_{n+3})x^{n+2} + \cdots\right| \\
&= \left|A_n x^n - A_{n+1}(x^n - x^{n+1}) - A_{n+2}(x^{n+1} - x^{n+2}) - \cdots\right| \\
&\leqslant |A_n| x^n + |A_{n+1}|(x^n - x^{n+1}) + |A_{n+2}|(x^{n+1} - x^{n+2}) + \cdots \\
&\leqslant \frac{\varepsilon}{2} x^n + \frac{\varepsilon}{2}[(x^n - x^{n+1}) + (x^{n+1} - x^{n+2}) + \cdots] \\
&< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.
\end{aligned}$$

This proves the uniform convergence. Hence $f(x) = \sum a_n x^n$ is continuous on $[0,1]$, and

$$\sum a_n = \lim_{x\to 1-} \sum a_n x^n.$$

Example 5.7.1

$$\begin{aligned}
\arctan x &= \int_0^x \frac{1}{1+y^2} dy \\
&= \int_0^x (1 - y^2 + y^4 - y^6 + \cdots) dy \\
&= \sum_{n=0}^{\infty} (-1)^{n+1} \frac{x^{2n+1}}{2n+1}.
\end{aligned}$$

The series above converges at $x = 1$ *and then by Abel's theorem*

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots$$

The formula above is called Leibniz formula and the series on the right is called Leibniz series.

Exercises

1. Let
$$f_n(x) = \begin{cases} 0, & x \in (1/n, 1], \\ n[1 - |2nx - 1|], & x \in [0, 1/n]. \end{cases}$$
   (a) sketch the graph of f_n.
   (b) Is $f_n(x)$ continuous on $[0, 1]$?
   (c) Prove that $f_n(x) \to 0$ for any $x \in [0, 1]$.
   (d) Evaluate
   $$\int_0^1 f_n(x)dx.$$
   (e) whether does it hold that
   $$\lim_n \int_0^1 f_n(x)dx = \int \lim_n f_n(x)dx?$$
   (f) Does f_n converge uniformly? Why?

2. Prove that if
$$|f_n(x) - f(x)| \leqslant a_n$$
for all $x \in D$ and $\lim_n a_n = 0$, then f_n converges to f uniformly on D.

Chapter 6

Multi-variate calculus

We now start Calculus of multi-variate functions or functions with several variables. This is natural because in real world, a variable usually depends on many other variables. In mathematics, the basic analysis tools will be similar: limit, continuity, differentiation and integration. We shall go over the whole theory briefly in the remaining lectures, and not stick to the mathematical rigor which is harder to follow in multi-dimensional case. In this course, we shall only focus on partial derivatives and multiple integrals with their applications.

6.1 functions and graphs

6.1.1 Euclidean distance on $\mathbf{R}^n$

Before talking about multivariate functions, we introduce multi-dimensional space. We know that there is a natural coordinate system on the plane. Any point on the plane is denoted by its coordinate (x, y) and conversely any ordered pair (x, y) determines a point on the plane uniquely. This gives the notion of product space. The plane is usually written into

$$\{(x, y) : x, y \in \mathbf{R}\} = \mathbf{R} \times \mathbf{R} = \mathbf{R}^2,$$

which is the product space of $\mathbf{R}$ and $\mathbf{R}$. When $A \subset \mathbf{R}$, $B \subset \mathbf{R}$, the product

$$A \times B = \{(x, y) : x \in A, y \in B\}$$

is a subset of $\mathbf{R}^2$, which is conventionally called a rectangle with sides A and B. Similarly

$$\mathbf{R}^n = \{(x_1, x_2, \cdots, x_n) : x_j \in \mathbf{R}, 1 \leqslant j \leqslant n\},$$

which is called n-dimensional space.

An interval on $\mathbf{R}$ means a subset of the points between two points a, b with the boundaries included or excluded. It is easy to define and understand. But a region D on $\mathbf{R}^2$ is not so easy to define. Roughly a region is a subset of $\mathbf{R}^2$ surrounded by a few closed continuous curves, with the curves excluded. These curves are called the boundary of the region. The set with the region and its boundary together is called a closed region. A region is called bounded if it is contained in a disk. Be careful, these sentences are not really a definition because a closed continuous curve has not precisely defined.
We shall introduce some concepts briefly.

- distance: there is a natural distance on $\mathbf{R}^n$ called Euclidean distance which is the straight line distance

$$d(P, Q) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2},$$

where $P = (x_1, y_1)$, $Q = (x_2, y_2)$ are two points in $\mathbf{R}^2$. When $n = 1$, $d(x, y) = |y - x|$. The key property of distance is the triangular inequality, for any three points $P, Q, R \in \mathbf{R}^2$,

$$d(P, Q) \leqslant d(P, R) + d(R, Q).$$

This means that the distance is the shortest distance.

- neighborhood: given a point $P_0 = (x_0, y_0) \in \mathbf{R}^2$, the ball centered at P with radius $r > 0$ is the set

$$\{P = (x, y) \in \mathbf{R}^2 : d(P, P_0) < r\}$$

which is also called an r-neighborhood of P_0. Hence a neighborhood of P is a disk centered at P.

- limit: a sequence of points $(P_n) \subset \mathbf{R}^2$ is said to converge to P, denoted by $P_n \longrightarrow P$ or $\lim_n P_n = P$, if for any $\varepsilon > 0$, there exists N, such that $d(P_n, P) < \varepsilon$ for any $n > N$. A sequence of points $(P_n) \subset \mathbf{R}^2$ is said to converge if there exists $P \in \mathbf{R}^2$ such that $\lim P_n = P$.

Theorem 6.1.1 If (P_n) converges, then the limit is unique.

Proof. Suppose (P_n) has two different limit $P \neq Q$. Then for any $\varepsilon > 0$, there exists N, when $n > N$, $d(P_n, P) < \varepsilon$ and $d(P_n, Q) < \varepsilon$. Then by the triangular inequality

$$0 < d(P, Q) \leqslant d(P_n, P) + d(P_n, Q) < 2\varepsilon.$$

This is a contradiction because the right side could be arbitrarily small. □

Assume that $A \subset \mathbf{R}^2$.

- A point P is an interior point of A, if some neighborhood of P is contained in A.
- A point $P \in \mathbf{R}^2$ is a limit point of A, if any neighborhood of P contains infinitely many points of A. The set of all limit points of A is denoted by A'.
- If any point of A is an interior point, then we say A is open. The set of all interior points of A is denoted by A°.
- If A contains all limit point of A, then we say A is closed.
- The set $A \cup A'$ is closed and is called the closure of A, denoted by $\overline{A}$.

Actually P is a limit point of A as long as any neighborhood of it contains at least one point in A other than P.

Theorem 6.1.2 Assume that $A \subset \mathbf{R}^2$. Then A is open if and only if $A^c = \mathbf{R}^2 \setminus A$ is closed.

Given $A \subset \mathbf{R}^2$. All points on $\mathbf{R}^2$ are classified into 3 kinds:

1. A°: the interior of A;
2. $(A^c)^\circ = (\overline{A})^c$: the interior of A^c, which is called the exterior of A;
3. $\partial A = \overline{A} \setminus A^\circ$, which is called the boundary of A.

Example 6.1.1 1. *Let* $B = \{(x, y) : x^2 + y^2 < 1\}$. *Then* B *is open but not closed because any boundary point of* B *is limit point not in* B. *Moreover* $B^\circ = B$, $\overline{B} = \{(x, y) : x^2+y^2 \leqslant 1\}$ *and the boundary of* B, ∂B, *is the unit circle.*

2. *The set* $A = B \cup \{(0, 1)\}$ *is not open nor closed.*

3. *If* A *is the set of all points with rational components, then* $A^\circ = \emptyset$, $\overline{A} = \mathbf{R}^2$ *and the boundary* $\partial A = \mathbf{R}^2$.

In this lecture notes we refer a region to an open set and a closed region to the closure of a region in the sequel. In n-dimensional space, Bolzano-Weierstrass theorem is also true.

Theorem 6.1.3 A bounded sequence of points (P_n) in $\mathbf{R}^2$ has a convergent subsequence.

It is easy to prove the theorems in this section.

6.1.2 functions and graphs

The function $y = f(x)$, $x \in D \subset \mathbf{R}$, is called a function of one variable. In the real world, there are many functions with more variables. For example, $s = vt$, the distance depends simultaneously on both the velocity v and the time t, and $r = \sqrt{x^2+y^2+z^2}$, the distance of a point in space (x, y, z) from the origin.

We shall start a function with two independent variables. The functions with more variables are similar and called multivariate functions.

Definition 6.1.4 A function f with the domain $D \subset \mathbf{R}^2$ is a mapping

$$f : D \longrightarrow \mathbf{R},$$

which is called a function with two variables. We denote such a function by $z = f(x, y)$, $(x, y) \in D$, where x, y are independent variables and z is dependent variable.

Similarly the letters x, y, z in $z = f(x, y)$ are chosen to indicate how f maps and they can be replaced by any other letters.

Actually an elementary function of two variables is also obtained by a finite times of four operations and compositions of elementary functions of x and of y. Every elementary function has its natural domain.

Example 6.1.2 *1.* $f(x, y) = x + ye^x$.

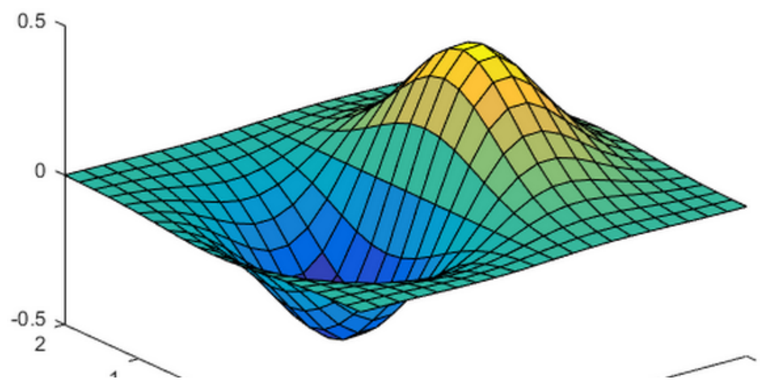
2. $f(x, y) = e^{xy^2} \ln(x + y)$

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

4. $f(x, y) = \sin(x + y)$

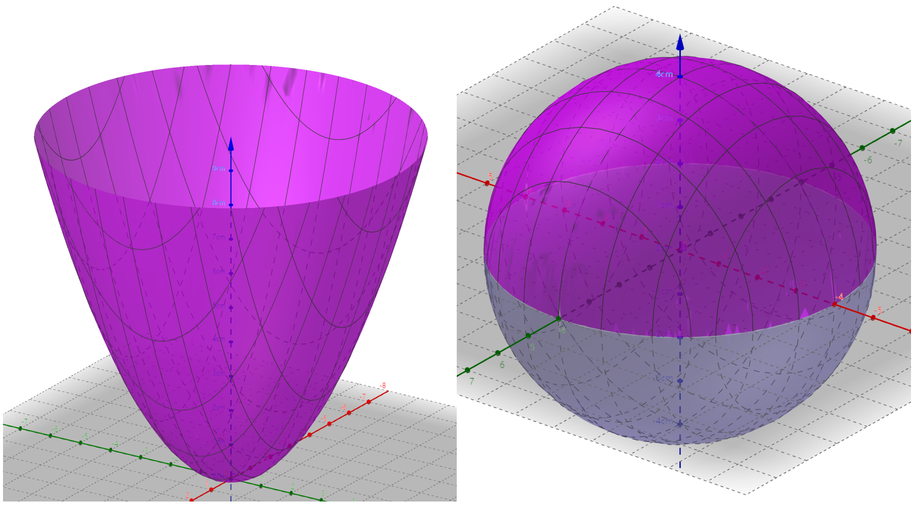
5. $f(x, y) = \dfrac{xy}{x^2 + y^2}$

The graph of a function $y = f(x)$, $x \in D \subset \mathbf{R}$ with one variable is the set of points $(x, f(x))$, $x \in D$, on xy-coordinate system. The graph of a function $z = f(x, y)$, $(x, y) \in D$ with two variables is the set of points $(x, y, f(x, y))$, $(x, y) \in D$, on xyz-coordinate system, a 3-dimensional space and it is usually a surface.

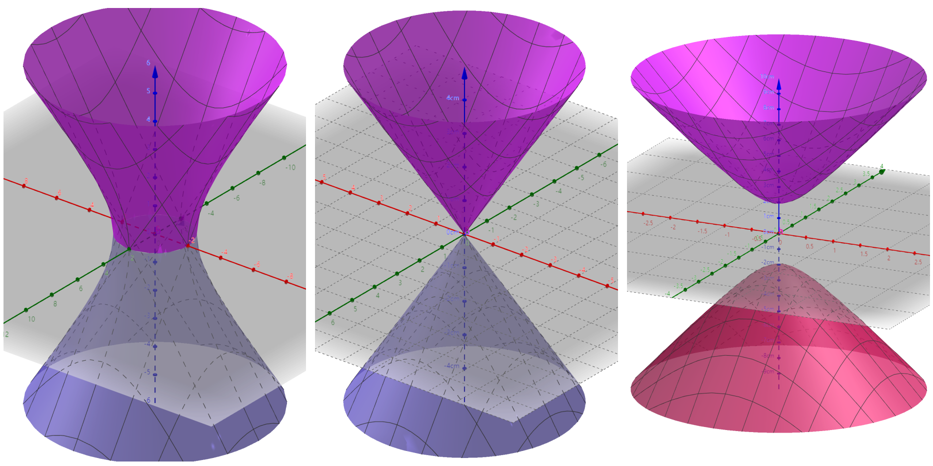


When $x = x_0$ is fixed, it is a function of y, $z = f(x_0, y)$. When $y = y_0$ is fixed, it is a function of x. The curves on the surface in the graphs above are such functions, which are obtained by the intersection of the surface and the plane $x = x_0$ and $y = y_0$. It is not easy but still possible to draw the graph of a surface on 3-dim space by hand but there are any softwares now which can be used to do this. However it is impossible to draw the graph of a function with more variables. The following functions are commonly used and their graphs would better be remembered.

Example 6.1.3 1. *A bowl* $z = x^2 + y^2$.



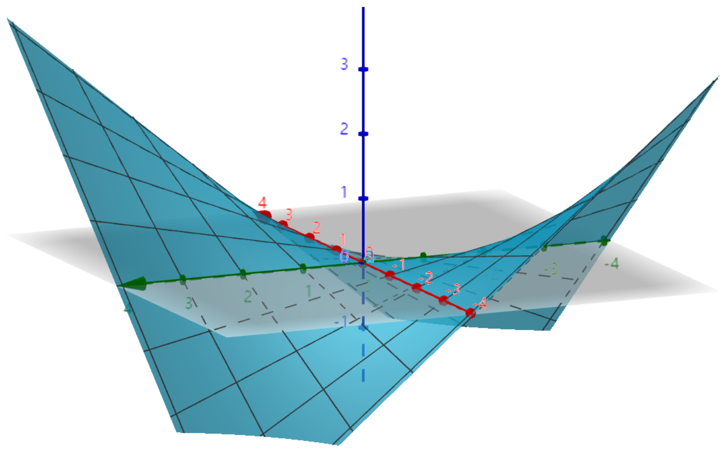
2. *A ball* $z = \sqrt{r^2 - x^2 - y^2}$, *the upper half.*
3. *A hyperboloid of one sheet* $z = \sqrt{x^2 + y^2 + a}$, *the upper half,* $a < 0$.



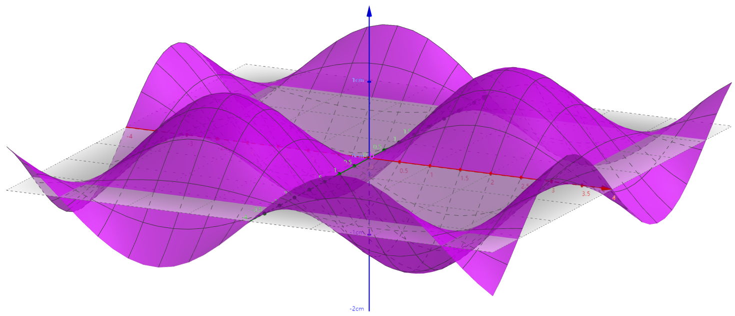
4. *A cone* $z = \sqrt{x^2 + y^2 + a}$, *the upper half,* $a = 0$.

5. *A hyperboloid of two sheets* $z = \sqrt{x^2 + y^2 + a}$, *the upper half,* $a > 0$.

6. *A saddle surface* $z = xy$



7. $z = \sin x \sin y$.



Exercises

1. Prove that a set A is open if and only if A^c is closed.

2. Let $A \subset \mathbf{R}^2$. Prove that the set of limit points A' is closed.

3. Let $A \subset B \subset \mathbf{R}^d$. Prove that if B is closed, then $\overline{A} \subset B$.

4. Prove that $A = \{(x, y) : x^2 + y^2 \leqslant 1\}$ is closed.

5. Prove that $(A^c)^\circ = (\overline{A})^c$.

6. Prove that a bounded sequence of points (P_n) in $\mathbf{R}^2$ has a convergent subsequence.

6.2 limit and derivatives

6.2.1 function limit

A δ-neighborhood of (x_0, y_0) is the disk centered at (x_0, y_0) and with radius δ

$$\{(x, y) : (x - x_0)^2 + (y - y_0)^2 < \delta^2\}.$$

A neighborhood is a δ-neighborhood for some $\delta > 0$. We can define the function limit similarly.

Definition 6.2.1 Assume that a function $z = f(x, y)$ on D and (a, b) is a limit point of D, not may not be in D. We say that f converges to L as $(x, y) \to (a, b)$, denoted by

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

if for any $\varepsilon > 0$, $\exists\delta > 0$, such that

$$|f(x, y) - L| < \varepsilon$$

whenever

$$0 < \sqrt{(x - a)^2 + (y - b)^2} < \delta \text{ and } (x, y) \in D.$$

Intuitively no matter how small $\varepsilon > 0$ is, we can find a small $\delta > 0$, such that

$$f(x, y) \in (L - \varepsilon, L + \varepsilon)$$

when $(x, y) \in D$ is in δ-neighborhood of (a, b) without the center.

Example 6.2.1 $z = xy$.

$$\lim_{(x,y)\to(a,b)} xy = ab.$$

Because the limit is a local property, we may consider a neighborhood of (a, b), *say* $(x-a)^2 + (y - b)^2 < 1$. *Then* $|x| < 1 + |a|$ *and* $|y| < 1 + |b|$. *Now we need to find* $\delta > 0$ *so that* $|xy - ab| < \varepsilon$, *when* $0 < (x - a)^2 + (y - b)^2 < \delta^2$. *By the inequality*

$$\begin{aligned}|xy - ab| &\leqslant |xy - xb| + |xb - ab| \leqslant |x||y - b| + |b||x - a| \\ &< (1 + |a|)\delta + |b|\delta = (1 + |a| + |b|)\delta,\end{aligned}$$

we know that it is enough to have $(1 + |a| + |b|)\delta < \varepsilon$.

As long as we have defined function limit, we will have the generic properties of limit: the uniqueness of limit, the comparison theorem and sandwich theorem, four operations of limit. Assume that $y = g(x)$ is a function and $\lim_{x\to a} g(x) = b$. When x goes to a, $(x, g(x))$ goes to (a, b). In this case we say that (x, y) goes to (a, b) along with the curve $y = g(x)$. The function $f(x, g(x))$ is a composite function and it is a function of x. We have the following theorem which states the relation between limit of two variable functions and one variable functions.

Theorem 6.2.2 If

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

then

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

Intuitively the function $f(x, y)$ has limit L whenever (x, y) goes to (a, b) along with any curve $y = g(x)$,

Example 6.2.2 $f(x, y) = \dfrac{xy}{x^2 + y^2}$. *The function is not defined at* $(0, 0)$. *Let's consider its limit when* $(x, y) \to (0, 0)$. *Take a function* $y = kx$. *Clearly the value of* f *on this curve is*

$$f(x, kx) = \frac{xkx}{x^2 + (kx)^2} = \frac{k}{1 + k^2}.$$

Of course the limit is also $k/(1 + k^2)$. *It is not constant. Hence by theorem above,* f *diverges when* (x, y) *goes to* $(0, 0)$.

The definition of continuity is similar to one-variable case.

Definition 6.2.3 A function $z = f(x, y)$, defined on D, is said to be continuous at $(a, b) \in D$ if

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

A function $z = f(x, y)$ is continuous on D if it is continuous at every point of D.

Similarly any elementary function has its natural domain and it is continuous at any point in its natural domain.

6.2.2 partial derivatives

Taking derivative of a two-variable function with respect to one variable is called partial derivative. Precisely, assume that $z = f(x, y)$ is a function on a region D. Fix y, z is a function of x, which is a function of one variable. Then we may take derivative, which is the partial derivative with respect to x, denoted by

$$\frac{\partial f}{\partial x}, \ \frac{\partial z}{\partial x}, \quad f'_x, \ z'_x.$$

The partial derivative with respect to y is defined similarly. Hence partial derivatives are not different from derivatives.

Example 6.2.3 *1.* $z = xy$. $z'_x = y$, $z'_y = x$.

2. $z = x^2 + y^2$, $z'_x = 2x$, $z'_y = 2y$.

3. $z = \frac{xy}{x^2+y^2}$, *domain* $x^2 + y^2 \neq 0$.

$$\frac{\partial z}{\partial x} = \frac{y}{x^2+y^2} - xy\frac{2x}{(x^2+y^2)^2} = \frac{y(x^2+y^2) - 2x^2y}{(x^2+y^2)^2} = \frac{y(y^2-x^2)}{(x^2+y^2)^2}.$$

The partial derivative of $z = f(x, y)$ is also a function of x and y and we may take partial derivative again, called the 2nd order partial derivatives. Taking partial derivative of the partial derivative $\frac{\partial z}{\partial x}$, we have

$$\frac{\partial}{\partial x}\frac{\partial z}{\partial x}, \text{ and } \frac{\partial}{\partial y}\frac{\partial z}{\partial x}$$

and they are denoted by

$$\frac{\partial^2 z}{\partial x^2} = z''_{xx} = f''_{xx}, \text{ and } \frac{\partial^2 z}{\partial y \partial x} = z''_{xy} = f''_{xy}.$$

Similarly we would have other two partial derivatives

$$\frac{\partial^2 z}{\partial x \partial y} = z''_{yx} = f''_{yx} \text{ and } \frac{\partial^2 z}{\partial y^2} = z''_{yy} = f''_{yy}.$$

Totally we have four 2nd-order partial derivatives, and then eight 3rd-order partial derivatives.

Example 6.2.4 1. $z = x^3y^2$. $z'_x = 3x^2y^2$, $z'_y = 2x^3y$.

$$\frac{\partial^2 z}{\partial x^2} = 6xy^2, \frac{\partial^2 z}{\partial y \partial x} = 6x^2y, \frac{\partial^2 z}{\partial x \partial y} = 6x^2y, \frac{\partial^2 z}{\partial y^2} = 2x^3.$$

2. $z = \frac{xy}{x^2+y^2}$, *domain* $x^2 + y^2 \neq 0$.

$$\begin{aligned}
\frac{\partial z}{\partial x} &= \frac{y}{x^2+y^2} - xy\frac{2x}{(x^2+y^2)^2} = \frac{y(x^2+y^2) - 2x^2y}{(x^2+y^2)^2} = \frac{y(y^2-x^2)}{(x^2+y^2)^2}.\\
\frac{\partial z}{\partial y} &= \frac{x(x^2-y^2)}{(x^2+y^2)^2}.\\
\frac{\partial^2 z}{\partial x^2} &= \frac{-2xy(x^2+y^2)^2 - y(y^2-x^2)2(x^2+y^2)2x}{(x^2+y^2)^4} = -2xy\frac{3y^2-x^2}{(x^2+y^2)^3}.\\
\frac{\partial^2 z}{\partial y \partial x} &= \frac{(3y^2-x^2)(x^2+y^2)^2 - y(y^2-x^2)2(x^2+y^2)2y}{(x^2+y^2)^4}\\
&= \frac{(3y^2-x^2)(x^2+y^2) - 4y^2(y^2-x^2)}{(x^2+y^2)^3}\\
&= \frac{6y^2x^2 - x^4 - y^4}{(x^2+y^2)^3}.\\
\frac{\partial^2 z}{\partial x \partial y} &= \frac{6y^2x^2 - x^4 - y^4}{(x^2+y^2)^3}.\\
\frac{\partial z^2}{\partial y^2} &= -2xy\frac{3x^2-y^2}{(x^2+y^2)^3}.
\end{aligned}$$

Theorem 6.2.4 (Clairaut) For $z = f(x, y)$, we have

$$\frac{\partial^2 z}{\partial x \partial y} = \frac{\partial^2 z}{\partial y \partial x}.$$

6.2.3 chain rule

When taking partial derivatives, the four operations still work, but the chain rule looks different.

Assume that $z = f(x, y)$ and $x = x(t)$, $y = y(t)$. Then by composition, z is a function of t. Let us see how to take derivative of z against t. This is just a particular example of composition of multi-variate functions but all other cases are essentially the same,

$$\begin{aligned} z'_t = \frac{dz}{dt} &= \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt} \\ &= f'_x \cdot x' + f'_y \cdot y', \end{aligned}$$

where $x' = x'(t)$, $y' = y'(t)$.

Roughly

$$\begin{aligned} &\frac{f(x(t+h), y(t+h)) - f(x(t), y(t))}{h} \\ =& \frac{f(x(t+h), y(t+h)) - f(x(t), y(t+h))}{h} + \frac{f(x(t), y(t+h)) - f(x(t), y(t))}{h}, \end{aligned}$$

and it is seen that the first is to take derivative of f against t with y fixed nd the second is to take derivative of f against t with x fixed. This is exactly the formula above by using the chain rule for functions with one-variable.

Example 6.2.5 1. *$z = e^x \sin y$, $x = s + t^2$, $y = s^2 + t$. The formula can surely be used to compute the following derivatives. For example to take derivative of z against s we may treat x, y as functions of s and use chain rule above*

$$\frac{\partial z}{\partial s} = e^x \sin y \cdot 1 + e^x \cos y \cdot 2s = e^x(\sin y + 2s \cos y).$$

$$\frac{\partial z}{\partial t} = e^x \sin y \cdot 2t + e^x \cos y \cdot 1 = e^x(2t \sin y + \cos y).$$

The students may verify this by writing z into a function of s, t

$$z = e^{s+t^2} \sin(s^2 + t)$$

and then taking partial derivative directly.

2. *The function $z = f(x(t), y(t))$ is the function $z = f(x, y)$ over the curve $x = x(t), y = y(t)$. Then the following derivative*

$$\begin{aligned} z'(t) = \frac{dz}{dt} &= \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} \\ &= f'_x \cdot x' + f'_y \cdot y', \end{aligned}$$

is the derivative of the function f *along with the curve* $x = x(t), y = y(t)$*. When the curve is a straight line* $x = at, y = bt$*, the derivative above is called the directional derivative* $\frac{dz}{dt} = af'_x + bf'_y$*. The second-order derivative is more complicated and will not be required. It is written below because it will be used later in a proof. Applying the product rule and chain rule, we have*

$$\begin{aligned} z''(t) &= (f'_x \cdot x')'_t + (f'_y \cdot y')'_t \\ &= (f''_{xx} \cdot x' + f''_{xy} \cdot y') \cdot x' + f'_x \cdot x'' + (f''_{xy} \cdot x' + f''_{yy} \cdot y') \cdot y' + f'_y \cdot y'' \\ &= f''_{xx} \cdot (x')^2 + 2f''_{xy} \cdot x'y' + f''_{yy} \cdot (y')^2 + f'_x \cdot x'' + f'_y \cdot y'', \end{aligned}$$

where all derivatives x', x''*,* y, y'' *are taken against* t*.*

6.2.4 differentiation*

Though differentiation is an important notion, it will not be used in this course. Hence the students may skip this part. Assume that f is defined on an open set D and $(x_0, y_0) \in D$.

Definition 6.2.5 It is said that f is differentiable at (x_0, y_0) if there are $A, B \in \mathbf{R}$ such that

$$\frac{f(x,y) - f(x_0,y_0) - A(x - x_0) - B(y - y_0)}{r}$$

converges to 0 as $r = \sqrt{(x - x_0)^2 + (y - y_0)^2}$ goes to zero.

It is important to know the relationship among three key notions: continuity, partial derivative, and differentiability. If f is differentiable at (x_0, y_0), then f is continuous at (x_0, y_0) and its partial derivatives at (x_0, y_0) must exist, because we may set $x - x_0 = h \to 0$ and $y = y_0$ to prove that

$$\left.\frac{\partial f}{\partial x}\right|(x_0, y_0) = A,$$

for example. However that the partial derivatives of f at (x_0, y_0) exist can not guarantee that f is differentiable. The reason is that the differentiability of f characterizes the property of f around (x_0, y_0) while the partial derivatives of f characterizes only the property of f along with two lines $x = x_0$ and $y = y_0$ around (x_0, y_0).

Example 6.2.6 *Let*

$$f(x,y) = \begin{cases} x, & y = 0, \\ y, & x = 0, \\ 1, & \textit{otherwise}. \end{cases}$$

Clearly f *has the partial derivatives at* $(0,0)$*, but* f *is discontinuous, hence not differentiable at* $(0,0)$*.*

Exercises

1. Find $\lim_{(x,y)\to(0,0)} \frac{4xy^2}{x^2+y^2}$ if exists.
2. Show that $\lim_{(x,y)\to(0,0)} \frac{4xy^2}{x^2+y^4}$ does not exist.
3. Let $f(x,y) = x\cos y + ye^x$. Find $f''_{xx}, f''_{xy}, f''_{yy}, f]]_{yx}$.
4. Assume that the equation $F(x,y) = 0$ determines an implicit function $y = f(x)$. Using the chain rule to prove the implicit function derivative formula
$$y'_x = -\frac{F'_x}{F'_y}.$$

6.3 extremum of multi-variate functions

6.3.1 local min/max

Assume that $z = f(x,y)$ is a function on D and (a,b) is an interior point of D, i.e., D contains a neighborhood of (a,b).

Definition 6.3.1 We say that f reaches a local maximum (resp., minimum) at (a,b) if there exists $\delta > 0$ such that when $(x-a)^2 + (y-b)^2 < \delta^2$,

$$f(a,b) \geqslant \text{(resp., } \leqslant) f(x,y).$$

Such a point (a,b) is called a local extremum (maximum or minimum) point of f.

Obviously if f reaches the local extremum at (a,b), then $x \mapsto f(x,b)$ reaches a local extremum at $x = a$ and $y \mapsto f(a,y)$ reaches a local extremum at $y = b$. Similar to one-variable case, we have the following theorem, which follows from Fermat's theorem.

Theorem 6.3.2 If f reaches a local maximum or minimum at (a,b), where f has both partial derivatives, then

$$f'_x(a,b) = f'_y(a,b) = 0.$$

A point (a,b) where $f'_x(a,b) = f'_y(a,b) = 0$ is called a critical point of f. A local extremum point of f must be a critical point, but a critical point may not be a local extremum point.

Example 6.3.1 *Refer to the graphs in Example 6.1.3.*

1. *The surface $z = x^2 + y^2$ is a bowl. $(0,0)$ is a critical point and a local minimum point.*
2. *The surface $z = xy$ is a saddle surface. $(0,0)$ is a critical point but not a local extremum point.*

How to tell if a critical point of f is a local extremum point? In one-variable case, the monotonicity of the derivative near critical point a tells the answer, which is actually related to the sign of the 2nd order derivative. If $f''(a) > 0$, then $f''(x) > 0$ near a, and $f'(x)$ increases near a. It implies that f reaches a local minimum at a. The same reason tells that if $f''(a) < 0$, f reaches a local maximum at a. But when $f''(a) = 0$, it tells no conclusion. However in two-variable case, we have another useful indicator: the Hessian

$$H(x, y) = \frac{\partial^2 z}{\partial x^2}\frac{\partial^2 z}{\partial y^2} - \left(\frac{\partial^2 z}{\partial x \partial y}\right)^2 = f''_{xx}f''_{yy} - (f''_{xy})^2.$$

For example,

1. $z = x^2 + y^2$ reaches local min at $(0, 0)$ and
$$H(0, 0) = 2 \cdot 2 - 0 = 4,\ z''_{xx}(0, 0) = 2.$$
2. $z = -x^2 - y^2$ reaches local max at $(0, 0)$ and
$$H(0, 0) = 2 \cdot 2 - 0 = 4,\ z''_{xx}(0, 0) = -2.$$
3. $z = xy$ has a saddle at $(0, 0)$, and
$$H(0, 0) = 0 - 1 = -1.$$

Theorem 6.3.3 Assume that f has 2nd-order partial derivatives and $P = (a, b)$ is a critical point. If

$$H(a,b)\begin{cases} > 0 \begin{cases} \left.\dfrac{\partial^2 z}{\partial x^2}\right|_{(a,b)} > 0, & \text{P is a local minimum point,} \\ \left.\dfrac{\partial^2 z}{\partial x^2}\right|_{(a,b)} < 0, & \text{P is a local maximum point,} \end{cases} \\ < 0, \ \text{P is NOT a local extremum point, say a saddle,} \\ = 0, \ \text{nothing to tell.} \end{cases}$$

We give a rough explanation. Assume that $a = b = 0$ without loss of generality. Let (x_0, y_0) be in the δ-neighborhood of $(0, 0)$ and $y = kx$ the line connecting (x_0, y_0) and $(0, 0)$. Its parametrization equation is $x = t, y = kt$ where $kx_0 = y_0$. Then we consider the function $z = f(t, kt)$ and compare $f(x_0, y_0) = f(x_0, kx_0)$ and $f(0, 0)$. Write $h(t) = f(t, kt)$. We shall see if h reaches local extremum at $t = 0$. By the formula in Example 6.3.1-2 of the previous section, we have

$$\begin{aligned} h'(t) &= f'_x + kf'_y, \\ h''(x) &= f''_{xx} + 2kf''_{xy} + k^2 f''_{yy}, \end{aligned}$$

since $x' = 1$, $y' = k$ and $x'' = y'' = 0$. Hence $h'(0) = 0$ and

$$h''(0) = f''_{xx}(0,0) + 2kf''_{xy}(0,0) + k^2 f''_{yy}(0,0).$$

This is a quadratic function of k with the discriminant being the negative Hessian $-H(0,0)$. If $H(0,0) > 0$, the quadratic function will not change its sign, either $h''(0) > 0$ for any $k \in \mathbf{R}$ when $f''_{xx}(0,0) > 0$, or $h''(0) < 0$ for any $k \in \mathbf{R}$ when $f''_{x,x}(0,0) < 0$. In the former case, h reaches local minimum at 0 and $f(x_0, y_0) > f(0,0)$, i.e., f reaches local minimum at $(0,0)$. In the latter case, f reaches local maximum at $(0,0)$.
If $H(0,0) < 0$, then $h''(0)$ will be positive for some k and negative for some other k. This implies that the function $z = f(x, kx)$ reaches local minimum at $x = 0$ for some k and local maximum at $x = 0$ for other k. Hence $(0,0)$ is a saddle point of f.

Example 6.3.2 *1. Consider $z = \sin x \sin y$. Due to periodic property, it is enough to restrict to the square $0 \leqslant x, y < 2\pi$. Refer to the graph in Example 6.1.3-7. $z'_x = \cos x \sin y$ and $z'_y = \sin x \cos y$. Then the set of critical points is*

$$\begin{aligned}\{(x,y) : z'_x = z'_y = 0\} &= \{(x,y) : \sin x = \sin y = 0\} \cup \{(x,y) : \cos x = \cos y = 0\} \\ &= \{(j_1\pi, j_2\pi) : j_1 = 0,1; j_2 = 0,1\} \cup \{(k_1\pi + \frac{\pi}{2}, k_2\pi + \frac{\pi}{2}) : k_1 = 0,1; k_2 = 0,1\}.\end{aligned}$$

Totally 8 points. Since $z''_{xx} = -\sin x \sin y$, $z''_{yy} = -\sin x \sin y$, and $z''_{xy} = \cos x \cos y$, we have

$$H(x,y) = \sin^2 x \sin^2 y - \cos^2 x \cos^2 y$$

and then

$$\begin{aligned}H(j_1\pi, j_2\pi) &= -1, \\ H(k_1\pi + \frac{\pi}{2}, k_2\pi + \frac{\pi}{2}) &= 1.\end{aligned}$$

Hence $(j_1\pi, j_2\pi)$ is not a local extremum point and $(k_1\pi + \pi/2, k_2\pi + \pi/2)$ is a local extremum point. When $k_1 + k_2$ is even, it is a local maximum point. When $k_1 + k_2$ is odd, it is a local minimum point. All phenomena can be seen from the graph.

2. $z = x^4 + y^4 - 4xy$. Then $z'_x = 4x^3 - 4y = 0$ and $z'_y = 4y^3 - 4x = 0$ give 3 solutions: $x = y = 0$, $x = y = 1$ and $x = y = -1$. The 2nd order partial derivatives

$$z''_{xx} = 12x^2,\ z''_{xy} = -4,\ z''_{yy} = 12y^2.$$

Then the Hessian $H(x,y) = 144x^2y^2 - 16$. It is seen that $H(0,0) = -16$, $H(1,1) = H(-1,-1) = 128$. Hence the function reaches local minimum at $x = y = 1$ and $x = y = -1$. The point $(0,0)$ is not a local extremum.

3. *The functions* $z = x^2y^2$ *and* $z = x^3y^2$. *Clearly* $(0,0)$ *is a critical point and the Hessian* $H(0,0)$ *is 0 for both. But* $(0,0)$ *is a local minimum point for the 1st function and is NOT a local extremum point for the 2nd one.*

6.3.2 global max/min

A continuous function on a closed interval $[a,b]$ reaches the global maximum and minimum. We have a similar result on $\mathbf{R}^n$. The proof is similar too.

Theorem 6.3.4 A continuous function f on a bounded closed region of $\mathbf{R}^n$ reaches the global maximum and minimum.

How to find the global max/min for a continuous function on a closed region? The method is similar to one-dimensional case. Assume that f is a continuous function on a bounded closed region $D \subset \mathbf{R}^2$. If f reaches a global extremum at point (x_0, y_0) of D, then either $(x_0, y_0) \in D^\circ$, in this case when f has partial derivatives at (x_0, y_0),

$$f'_x(x_0, y_0) = f'_y(x_0, y_0) = 0$$

or $(x_0, y_0) \in \partial D$. The interior points with partial derivative zero or where f has no partial derivatives are called critical points. Hence the global max/min will be reached among critical points and boundary points. The boundary of a closed region is usually (not always) a continuous curve. Finding the global max/min is a problem for a continuous function on a closed interval.

Therefore the strategy to find global extremum is to find critical points and extremum points on boundary where only the function may reach max/min and then to compare the function values on these points. In particular, it is not necessary to distinguish whether a critical point is a local extremum point. The following two examples will explain step by step how to find the global max/min.

Example 6.3.3 *Find max/min for* $f(x,y) = x^2 - y^2 - xy$ *on the triangle* D *determined by* $(0,0)$, $(0,2)$ *and* $(3,0)$.

1. *find critical points inside: to find critical points,* $f'_x = 2x - y = 0$ *and* $f'_y = -2y - x = 0$. *The solution is* $(0,0)$ *which is on the boundary. There is no critical point inside.*

2. *find the max/min on boundary: Consider* f *on the boundary. We shall consider three sides respectively. Their boundaries are points* $(0,0)$, $(0,2)$, $(3,0)$ *at which the function values are* $0, -4, 9$ *respectively. Then we should find the critical points of* f *on each side of the triangle.*

- *on the line connecting* $(0,0)$ *and* $(0,2)$, *i.e.,* $f(0,y) = -y^2$, $0 \leqslant y \leqslant 2$. *The critical point is* $y = 0$ *which is* $(0,0)$ *on the boundary.*
- *on the line connecting* $(0,0)$ *and* $(3,0)$, *i.e.,* $f(x,0) = x^2$, $0 \leqslant x \leqslant 3$. *No critical points when* $x \in (0,3)$.
- *On the line connecting* $(0,2)$ *and* $(3,0)$, $y = 2 - 2x/3$, $x \in [0,3]$ *and the function is*

$$f(x, 2 - 2x/3) = x^2 - (2 - 2x/3)^2 - x(2 - 2x/3), \; x \in [0,3].$$

Its derivative is

$$2x + 8(1 - x/3)/3 - 2 + 4x/3 = \frac{22}{9}x + \frac{2}{3}$$

and there is no critical point in $[0,3]$.

3. *comparing the values: we need only to compare the function values at three points* $(0,0), (0,2), (3,0)$ *and see that its maximum is 6 reached at* $(3,0)$ *and its minimum is* -4 *reached at* $(0,2)$.

Example 6.3.4 *Find max/min for* $f(x,y) = 2x^3 + y^4$ *on* $x^2 + y^2 \leqslant 1$.

1. *find critical points inside the region: the only critical point is* $(0,0)$ *and the function value is 0.*
2. *find max/min on boundary: The boundary of the domain is a unit circle:* $y^2 = 1 - x^2$ *for* $x \in [-1,1]$, *which should be separated into two pieces. But here the values of* f *on both pieces are the same*

$$g(x) = 2x^3 + (1 - x^2)^2 = 2x^3 + 1 - 2x^2 + x^4, \; x \in [-1,1].$$

Now

$$g'(x) = 6x^2 - 4x + 4x^3 = 2x(2x - 1)(x + 2)$$

and the critical points in $[-1,1]$ *are* 0 *and* $1/2$. *The max and min of* g *are 2, -2, reached on* $x = 1$ *and* -1.

3. *Compare the value at the critical point inside and the max/min on the boundary: it is seen that the global max/min of* f *on the region are* $2/-2$.

Exercises

1. Find all critical points and determine what they are.

(a) $f(x, y) = x^3 + y^3 + 3xy$;

(b) $f(x, y) = x^3 - y^3 - 2xy$;

(c) $f(x, y) = x^3 + y^3 + 3x^2 - 3y^2$;

(d) $f(x, y) = 4xy - x^4 - y^4$;

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

2. Find the global extremum of $f(x, y) = 4x - 8xy + 2y$

   (a) on the triangle bounded by $x = 0, y = 0$ and $x + y = 1$;

   (b) on the region bounded by the curves $y + x = 2$ and $y = x^2$.

3. Find two numbers $a \leqslant b$ such that

$$\int_a^b (6 - x - x^2) dx$$

has its largest value.

6.4 multiple integrals

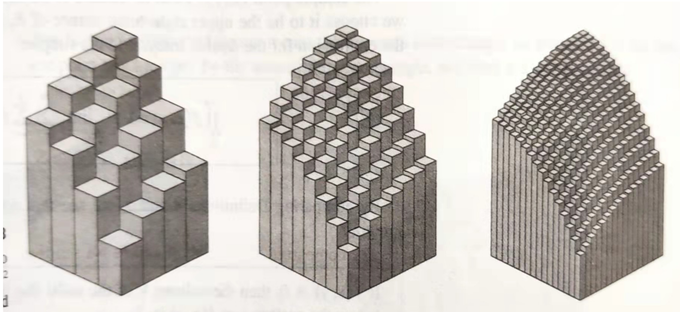
6.4.1 volume of solid under surface

The graph of a function $z = f(x, y)$ on a region D is a surface over D. It is natural to form a solid by all vertical segments $[0, f(x, y)]$ for any $(x, y) \in D$. The volume of this solid is called the volume under the surface on D.

Assume that the region D is bounded, say $D \subset [a, b] \times [c, d]$. Cut $[a, b]$ by a partition $(x_j : 1 \leqslant j \leqslant n)$ and $[c, d]$ by a partition $(y_k : 1 \leqslant k \leqslant m)$. If the small rectangle $[x_{j-1}, x_j] \times [y_{k-1}, y_k]$, named $r(j, k)$, is entirely in D, we say the rectangle $r(j, k)$ is good, and bad otherwise. Denote the length of sides $\Delta x_j = x_j - x_{j-1}$ and $\Delta y_k = y_k - y_{k-1}$. If the rectangle $r(j, k)$ is good, we choose any point (ξ_j, η_k) from it, i.e., $\xi_j \in [x_{j-1}, x_j]$ and $\eta_k \in [y_{k-1}, y_k]$. Intuitively the Riemann sum

$$\sum_{j,k:r(j,k)\text{good}} f(\xi_j, \eta_k)\Delta x_j \Delta y_k,$$

where $f(\xi_j, \eta_k)\Delta x_j \Delta y_k$ is the volume of cylinder with base $r(j, k)$ and height $f(\xi_j, \eta_k)$, approximates the volume under the surface $z = f(x, y)$ on D.



Similar to one-variable function, we have the following definition.

Definition 6.4.1 (Riemann) Assume that $y = f(x, y)$ is a function defined on a region D. We say that the volume under surface f exists or f is integrable, if there exists $V \in \mathbf{R}$ such that for any $\varepsilon > 0$, there exists $\delta > 0$, such that for any partition $(x_j : 1 \leqslant j \leqslant n)$ on x-axis and $(y_k : 1 \leqslant k \leqslant m)$ with

$$\max_j (x_j - x_{j-1}) < \delta, \ \max_k (y_k - y_{k-1}) < \delta$$

and any choice $\xi_j \in [x_{j-1}, x_j]$, $\eta_k \in [y_{k-1}, y_k]$ it holds that

$$\left| \sum_{j,k:r(j,k)\text{good}} f(\xi_j, \eta_k) \Delta x_j \Delta y k - V \right| < \varepsilon.$$

In this case, V is called the volume under surface or more precisely the (definite) integral, or Riemann integral, of f, called the integrand, over D, called the integral region, denoted by

$$\int\int_D f(x, y) dxdy, \text{ or simply } \int\int_D f dxdy, \ \int\int_D f.$$

The integral above is simply called the double integral or multiple integral (even for the functions of more variables). Intuitively when the partition goes to 0, the union of good rectangles goes to D and $\Delta A_{j,k}$ goes to $dxdy$. That's why the Riemann sum goes to the integral $\int\int_D f(x, y) dxdy$. Obviously when D has no interior points, there is no good rectangles in Riemann sum and the integral on D must be zero. The following properties are easy to verify by using the definition.

1. If f is non-negative on D, then the Riemann sum is non-negative and

$$\int\int_D f(x, y) dxdy \geqslant 0.$$

2. $\left| \int\int_D f(x, y) dxdy \right| \leqslant \int\int_D |f(x, y)| dxdy.$

3. Assume that f, g are integrable and a, b are real numbers.

$$\int\int_D (af(x,y) + bg(x,y))dxdy = a\int\int_D f(x,y)dxdy + b\int\int_D g(x,y)dxdy.$$

4. If $D = D_1 \cup D_2$ where D_1, D_2 have no common interior points, then

$$\int\int_D f(x,y)dxdy = \int\int_{D_1} f(x,y)dxdy + \int\int_{D_2} f(x,y)dxdy.$$

Though the definition of integral is clear, it is not realistic to compute the integral with Riemann sum. How do we compute the double integral? We shall briefly introduce two methods: the iterated integral formula and change of variables.
It is the very base for a multiple integral to understand the integral region properly. In one-variable case, it is not a problem because the region is an interval, but in multi-variable case, it really is. There are many way to express the region D. One way is to say it is bounded by some curves, for example D is bounded by $y = x$ and $y = x^2$. Another way is to express by some inequalities, for example D: $0 \leqslant y \leqslant x \leqslant 1$. Sometimes it is difficult to understand and sketch the region correctly.

6.4.2 iterated integrals

One approach is to turn a double integral into an iterated integral, i.e., to integrate $f(x,y)$ against x and then integrate against y or other way around.
Let's consider a very special case where the region

$$D = [a,b] \times [c,d] = \{(x,y) : x \in [a,b], y \in [c,d]\},$$

i.e., D is a rectangle obtained by the product of interval $[a,b]$ and $[c,d]$. In this case all small rectangles $[x_{j-1}, x_j] \times [y_k, y_{k-1}]$ are good and the Riemann sum is

$$\sum_{j,k} f(\xi_j, \eta_k)\Delta_{j,k} = \sum_j \left(\sum_k f(\xi_j, \eta_k)(y_k - y_{k-1})\right)(x_j - x_{j-1}).$$

The sum inside converges to

$$\int_c^d f(\xi_j, y)dy$$

and the Riemann sum converges to

$$\int_a^b \left(\int_c^d f(x,y)dy\right)dx.$$

We have a formula

$$\int\int_{[a,b]\times[c,d]} f(x,y)dxdy = \int_a^b \left(\int_c^d f(x,y)dy\right)dx.$$

The right side is called an iterated integral, to integrate f against y first then against x next. When we integrate f against y, x is treated as a constant, and after integrating, $\int_c^d f(x,y)dy$ is a function of x, which will be integrated against x. Similarly we may integrate against x first then y,

$$\int\int_{[a,b]\times[c,d]} f(x,y)dxdy = \int_c^d \left(\int_a^b f(x,y)dx\right) dy.$$

This formula makes us apply Newton-Leibniz formula.

Example 6.4.1

$$\begin{aligned}\int\int_{[0,1]\times[0,1]} xe^{xy}dxdy &== \int_0^1 \left(\int_0^1 xe^{xy}dy\right) dx\\ &= \int_0^1 e^{xy}\Big|_{y=0}^{y=1} dx = \int_0^1 (e^x - 1)dx\\ &= (e^x - x)\Big|_0^1 = e - 1 - (1-0) = e - 2.\end{aligned}$$

Moreover if $f(x,y) = f_1(x)f_2(y)$, then we have

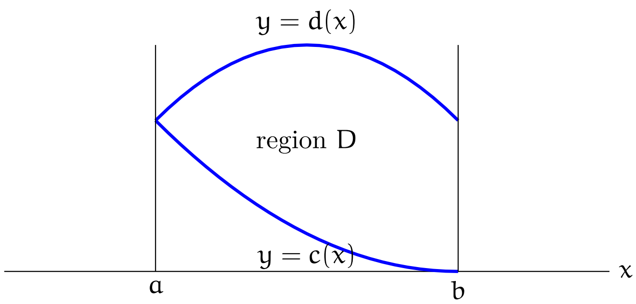
$$\begin{aligned}\int\int_{[a,b]\times[c,d]} f_1(x)f_2(y)dxdy &= \int_a^b \left(\int_c^d f_1(x)f_2(y)dy\right) dx\\ &= \int_a^b f_1(x)dx \cdot \int_c^d f_2(y)dy.\end{aligned}$$

In this special case, the double integral is the product of two one-variable integrals.

In most cases, the region D is not a rectangle, but can be written into

$$D: \ c(x) \leqslant y \leqslant d(x), \ a \leqslant x \leqslant b,$$

where $y = c(x)$ and $y = d(x)$ are two function curves. Such a region is called a rectangle with curved y-sides.



D: bounded by $y = c(x)$,$y = d(x)$, $x = a$ and $x = b$

Of course D may also be written into

$$D: \ a(y) \leqslant x \leqslant b(y), \ c \leqslant y \leqslant d.$$

Similarly this region is called a rectangle with curved x-sides. In either case, we say that D is a rectangle with curved sides. More generally D may be cut into finite pieces of rectangle with curved sides.

Now assume that D is as in the first case,

$$D: \ c(x) \leqslant y \leqslant d(x), \ a \leqslant x \leqslant b.$$

Look at the Riemann sum, choose $\xi_j = x_j$ and $\eta_k = y_k$ and then we have

$$\sum_{j,k:\text{good}} f(x_j, y_k)\Delta A_{j,k} = \sum_j \left(\sum_{c(x_j)\leqslant y_k \leqslant d(x_j)} f(x_j, y_k)\Delta y_k \right) \Delta x_j.$$

The Riemann sum is composed by two sums: the sum inside and sum outside. The sum inside is a Riemann sum which converges to

$$\int_{c(x)}^{d(x)} f(x, y)dy$$

which is a function of x, and the sum out side is also a Riemann sum which converges to the integral

$$\int_a^b \left(\int_{c(x)}^{d(x)} f(x, y)dy \right) dx.$$

This is only an rough idea of the following theorem and the rigorous proof is difficult.

Theorem 6.4.2 If $D = \{(x, y) : c(x) \leqslant y \leqslant d(x), \ a \leqslant x \leqslant b$, then integrate y first from curve $y = c(x)$ to curve $y = d(x)$ and then x from a to b,

$$\int\int_D f(x, y)dxdy = \int_a^b \left(\int_{c(x)}^{d(x)} f(x, y)dy \right) dx.$$

The iterated integral can also be written into

$$\int_a^b dx \int_{c(x)}^{d(x)} f(x, y)dy.$$

If $D = \{(x, y) : a(y) \leqslant x \leqslant b(y), \ c \leqslant y \leqslant d$, then integrate x first and then y

$$\int\int_D f(x, y)dxdy = \int_c^d \left(\int_{a(y)}^{b(y)} f(x, y)dx \right) dy.$$

The Iterated integral formula is to integrate in different order. In most cases, the order makes no difference, but in some cases, the order does make difference.

What is the integral inside the iterated integral, i.e.,

$$A(x) = \int_{c(x)}^{d(x)} f(x, y)dy?$$

This is the area of the section of the solid under surface cut by the plane at x perpendicular to x-axis. From this point-view, this theorem is exactly the same as Theorem 4.5.2, which says that the volume of a solid is obtained by integrating the areas of parallel sections

$$V = \int_a^b A(x)dx.$$

Example 6.4.2 1. *The volume of a solid under saddle surface* $z = xy$ *on a region* D *surrounded by lines* $x = 0, y = 0$ *and* $x + y = 1$. *The region can be written into* $0 \leqslant y \leqslant 1 - x$, $0 \leqslant x \leqslant 1$. *Then we use the iterated integral formula*

$$\begin{aligned}
\int\int_D xy dx dy &= \int_0^1 x dx \int_0^{1-x} y dy = \int_0^1 x dx \frac{y^2}{2}\Big|_0^{1-x} \\
&= \int_0^1 x\frac{(1-x)^2}{2} dx = \frac{1}{2}\int_0^1 (x - 2x^2 + x^3) dx \\
&= \frac{1}{2}(x^2/2 - 2x^3/3 + x^4/4)\Big|_0^1 \\
&= \frac{1}{2}(1/2 - 2/3 + 1/4) = \frac{1}{24}.
\end{aligned}$$

2. *How to compute the volume of the unit ball by integration? The equation of unit sphere is* $x^2 + y^2 + z^2 = 1$ *and the upper part is a function* $z = \sqrt{1 - x^2 - y^2}$. *Hence the one-eighth of it is*

$$\int\int_D \sqrt{1 - x^2 - y^2} dx dy$$

where D *is a quarter of unit disk*

$$0 \leqslant y \leqslant \sqrt{1 - x^2}, 0 \leqslant x \leqslant 1.$$

By the iterated integral formula

$$\begin{aligned}
\int\int_D \sqrt{1 - x^2 - y^2} dx dy &= \int_0^1 dx \int_0^{\sqrt{1-x^2}} \sqrt{1 - x^2 - y^2} dy \\
&= \int_0^1 \sqrt{1 - x^2} dx \int_0^{\sqrt{1-x^2}} \sqrt{1 - \left(\frac{y}{\sqrt{1 - x^2}}\right)^2} dy, \ \textit{set}\ u = \frac{y}{\sqrt{1 - x^2}}, \\
&= \int_0^1 (1 - x^2) dx \int_0^1 \sqrt{1 - u^2} du \\
&= \frac{2}{3} \cdot \frac{1}{2}(\arcsin u + u\sqrt{1 - u^2})\Big|_0^1 = \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{\pi}{2} = \frac{\pi}{6}.
\end{aligned}$$

The volume of the unit ball is $8 \cdot \pi/6 = 4\pi/3$.

3. *To integrate* $\int\int_D xy dx dy$ *where* D *is the region surrounded by two circles*

$$(x - 1)^2 + y^2 = 1 \ \textit{and}\ x^2 + (y - 1)^2 = 1,$$

we need to make clear how to express the region D*. We draw two circles on plane and see that* D *is formed by two curves: the lower half of* $x^2 + (y-1)^2 = 1$ *and the upper half of* $(x-1)^2 + y^2 = 1$*, i.e.,*

$$-\sqrt{1-x^2}+1 \leqslant y \leqslant \sqrt{1-(x-1)^2}$$

on $0 \leqslant x \leqslant 1$*. Hence*

$$\begin{aligned}
\iint_D xy\,dxdy &= \int_0^1 xdx \int_{-\sqrt{1-x^2}+1}^{\sqrt{1-(x-1)^2}} y\,dy = \int_0^1 xdx \frac{y^2}{2}\Big|_{-\sqrt{1-x^2}+1}^{\sqrt{1-(x-1)^2}} \\
&= \frac{1}{2}\int_0^1 xdx\left[(1-(x-1)^2) - (-\sqrt{1-x^2}+1)^2\right] \\
&= \frac{1}{2}\int_0^1 x(2x-2+2\sqrt{1-x^2})dx \\
&= \int_0^1 (x^2 - x + x\sqrt{1-x^2})dx \\
&= \frac{x^3}{3} - \frac{x^2}{2} - \frac{(1-x^2)^{3/2}}{3}\Big|_0^1 = \frac{1}{6}.
\end{aligned}$$

4. *Usually it is not important which order to take, integrating* x *first or* y *first, but sometimes it is. To integrate*

$$\iint_D e^{-x^2}dxdy$$

where D*:* $0 \leqslant y \leqslant x \leqslant a$*, we can integrate* y *first*

$$\int_0^a dx \int_0^x e^{-x^2}dy = \int_0^a xe^{-x^2}dx = -\frac{1}{2}e^{-x^2}\Big|_0^a = 1/2(1-e^{-a^2}).$$

But integrating x *first can not solve the problem because we can not find the primitive function of* e^{-x^2}*.*

Exercises

1. Sketch the region and evaluate the integral.

   (a) $\iint_D x\sin y\,dxdy$, where D: $0 \leqslant y \leqslant x \leqslant \pi$;

   (b) $\iint_D 3y^3e^{xy}dxdy$, where D is bounded by the curves $y^2 = x$, $x = 0$ and $y = 1$;

   (c) $\iint_D (y-2x^2)dxdy$, where D is region bounded by the curve $|x| + |y| \leqslant 1$;

   (d) $\iint_D (y-2x^2)dxdy$, where D is region bounded by the curves $x+y = 2$ and $y = x^2$.

(e) $\int\int_{0\leqslant x\leqslant y\leqslant 1-x} \frac{1}{\sqrt{xy}} dxdy.$

2. Sketch the region and reverse the order of iterated integral

$$\int_0^2 dy \int_0^{4-y^2} y dx.$$

3. Sketch the region and evaluate the following integrals. Reverse the order if necessary.

(a) $\int_0^1 x dx \int_0^{1-x} \sqrt{y} dy.$

(b) $\int_0^{\pi} dy \int_0^{y} \frac{\sin x}{x} dx.$

6.5 change of variables

6.5.1 change of variables

The iterated integral formula is very powerful to double integrals, because we then can apply the Newton-Leibniz formula. We now introduce another important integral technique, change of variable in multiple integrals, which may make double integral easier or make it possible when the iterated formula fails.

In the case of one-variable integral, a change of variable $x = g(u)$ gives

$$\int_0^1 f(x)dx = \int_a^b f(g(u))g'(u)du,$$

where g is differentiable on $[a, b]$, strictly increasing, and $0 = g(a)$ and $1 = g(b)$. What does this mean? It can be seen from Riemann sum that given a partition (x_j) on $[0, 1]$,

$$\sum_j f(x_{j-1})(x_j - x_{j-1}) = \sum_j f(g(u_{j-1}))(g(u_j) - g(u_{j-1})),$$

where $x_j = g(u_j)$ for any j. When the partition is small,

$$g(u_j) - u(u_{j-1}) \sim g'(u_{j-1})(u_j - u_{j-1}) \sim g'(u)du.$$

Therefore a change of variable formula is essentially a formula changing the length element

$$dx = g'(u)du.$$

In two-variable integral, the integral element $dxdy$ is called an area element. A change of variables is a bijective mapping or a transform

$$x = x(u, v), \;\; y = y(u, v),$$

which maps a region G on uv-plane to a region D on xy-plane. It is similar that the function $f(x, y)$ of x, y becomes the function $f(x(u, v), y(u, v))$ of u, v. The major problem is the relationship between the area element $dxdy$ and the area element $dudv$, which is given below

$$dxdy = \left\| \begin{matrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{matrix} \right\| dudv,$$

where the $|\cdot|$ inside is a determinant, called Jacobi determinant of the given transform, and the $|\cdot|$ outside is absolute value. However we would simply write only one $|\cdot|$ to save notations.

Theorem 6.5.1 With assumptions and notations above, we have the change of variable formula

$$\iint_D f(x, y)dxdy = \iint_G f(x(u, v), y(u, v)) \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} dudv.$$

The formula can be used in both ways, depending on which side is easier to compute.

The rigorous proof is difficult and will be skipped. However it can be seen by an example. Consider linear mapping

$$\begin{cases} x = au + bv, \\ y = cu + dv, \end{cases}$$

where a, b, c, d are constant. This mapping maps any point in uv-plane to a point in xy-plane. Since

$$(0, 0) \mapsto (0, 0), \ (1, 0) \mapsto (a, c), \ (0, 1) \mapsto (b, d), \ (1, 1) \mapsto (a + b, c + d),$$

the unit square $[0, 1] \times [0, 1]$ on uv-plane is mapped to a parallelogram on xy-plane generated by vectors $(a, c), (b, d)$. The area of the parallelogram is exactly the Jacobi determinant

$$\begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = |ad - bc|.$$

In fact the area of the parallelogram is

$$S = \sqrt{a^2 + c^2}\sqrt{b^2 + d^2} \sin\theta$$

where θ is the angle between vectors (a, c) and (b, d). It is known that the inner product of the vectors (a, c) and (b, d) is

$$ab + cd = \sqrt{a^2 + c^2}\sqrt{b^2 + d^2} \cos\theta.$$

Hence $S^2 + (ab + cd)^2 = (a^2 + c^2)(b^2 + d^2)$ and it follows that

$$S^2 = (ad - cb)^2.$$

Similarly a general transform

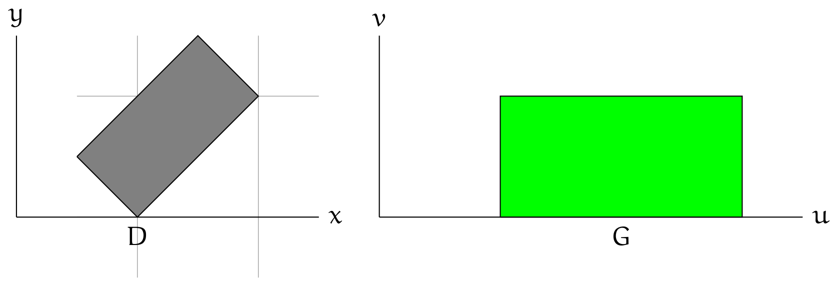
$$\begin{cases} x = x(u, v), \\ y = y(u, v), \end{cases}$$

is locally linear

$$\begin{cases} dx = \dfrac{\partial x}{\partial u} du + \dfrac{\partial x}{\partial v} dv, \\ dy = \dfrac{\partial y}{\partial u} du + \dfrac{\partial y}{\partial v} dv, \end{cases}$$

and it sends a small rectangle with sides du and dv to a region close to a parallelogram with area given by the Jacobi determinant.

Example 6.5.1 *Find* $\int\int_D xy dx dy$ *where* $D: 1 \leqslant x+y \leqslant 3, 0 \leqslant x-y \leqslant 1$. *to calculate this integral, we may use the iterated integral formula. Since* D *is a tilted rectangle, we have to cut it into three parts first and it is very complicated. Using a substitution* $u = x+y, v = x-y$, *or* $x = (u+v)/2, y = (u-v)/2$, *the region* D *becomes* $G: 1 \leqslant u \leqslant 3, 0 \leqslant v \leqslant 1$ *and* $dx dy = 1/2 du dv$.



Hence we have

$$\int\int_D xy dx dy = \int\int_G \frac{1}{8}(u^2 - v^2) du dv = \frac{1}{8}\left(\int_1^3 u^2 du \int_0^1 dv - \int_1^3 du \int_0^1 v^2 dv\right)$$
$$= \frac{1}{8}(\frac{26}{3} - 2 \cdot \frac{1}{3}) = 1.$$

Students may ask if there is any trick to choose a substitution. It is hard to say what substitution can solve a problem or if there is any, but we may choose a substitution according to the concrete problem so that it simplifies the integral region or the integrand or both if we are lucky enough.

6.5.2 polar coordinates

Another popular coordinate system is the polar coordinate system. A vector (x, y) can be represented by its length r and the angle θ, from the horizontal line to the vector, called argument,

$$x = r\cos\theta, y = r\sin\theta.$$

This is a one-to-one correspondence between $\mathbf{R} \setminus \{0\}$ and $0 < r < \infty, 0 \leqslant \theta < 2\pi$, which is called the polar coordinate transform. Since

$$\frac{\partial x}{\partial r} = \cos\theta, \ \frac{\partial x}{\partial \theta} = -r\sin\theta,$$
$$\frac{\partial y}{\partial r} = \sin\theta, \ \frac{\partial y}{\partial \theta} = r\cos\theta,$$

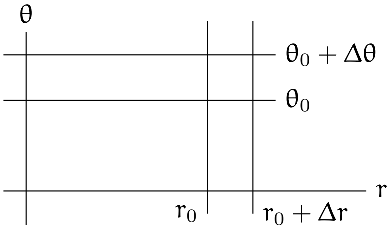
we have the relation between area elements

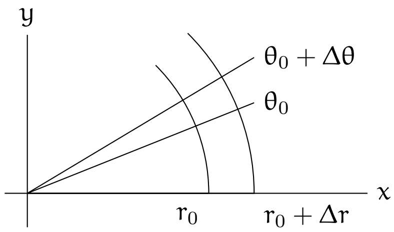
$$dxdy = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} drd\theta = rdrd\theta.$$

It is known that the area of the image of $r_0 < r < r_0 + \Delta r$ and $\theta_0 < \theta < \theta_0 + \Delta\theta$ under the mapping above is

$$(r_0 + \Delta r)^2 \Delta\theta/2 - r_0^2 \Delta\theta/2 = r_0 \Delta r \Delta\theta,$$

which is exactly what the Jacobi determinant tells us.





How to use this formula to solve problems?

Example 6.5.2 *1. Take the integral we did before*

$$I = \iint_D \sqrt{1 - x^2 - y^2} dxdy$$

where $D : 0 \leqslant y \leqslant \sqrt{1 - x^2}, 0 \leqslant x \leqslant 1$. *The polar coordinate transform above is a bijection from* $G : 0 \leqslant r \leqslant 1, 0 \leqslant \theta \leqslant \pi/2$ *to* D *and then*

$$\begin{aligned} I &= \iint_G \sqrt{1 - r^2} rdrd\theta \\ &= \int_0^1 r\sqrt{1 - r^2} dr \int_0^{\pi/2} d\theta = -\frac{1}{2} \frac{(1 - r^2)^{3/2}}{3/2} \Big|_0^1 \cdot \frac{\pi}{2} = \frac{\pi}{6}. \end{aligned}$$

2. The volume of the solid bounded by xy*-plane* $z = 0$ *and the bowl upside down* $z = 1 - x^2 - y^2$ *is*

$$V = \iint_{x^2 + y^2 \leqslant 1} (1 - x^2 - y^2) dxdy = \int_0^{2\pi} \int_0^1 (1 - r^2) rdrd\theta = \pi/2.$$

3. *The volume of the solid under surface* $z = x^2 + y^2$ *on the region* $D : (x-1)^2 + y^2 \leqslant 1$ *is*

$$V = \int\int_{(x-1)^2+y^2\leqslant 1} (x^2 + y^2) dxdy.$$

A substitution is chosen to make either the region or the integrand look better. The substitution

$$\begin{cases} x - 1 = r\cos\theta, \\ y = r\sin\theta, \end{cases}$$

makes the region look better, which takes $G : 0 \leqslant r \leqslant 1, 0 \leqslant \theta < 2\pi$ *to* D. *Then* $dxdy = rdrd\theta$*, and*

$$\begin{aligned} V &= \int\int_G ((1 + r\cos\theta)^2 + r^2 \sin\theta) rdrd\theta \\ &= \int\int_G (r^2 + 2r\cos\theta + 1) rdrd\theta = 2\pi(\frac{r^4}{4} + \frac{r^2}{2})\Big|_0^1 + 0 = \frac{3\pi}{2}. \end{aligned}$$

The following substitution $x = r\cos\theta, y = r\sin\theta$ *makes the integrand look better. In this case,* D *becomes* $G : 0 \leqslant r \leqslant 2\cos\theta,\ \theta \in [-\pi/2, \pi/2]$ *and we use the iterated formula*

$$\int_D (x^2 + y^2) dxdy = \int_G r^2 rdrd\theta = \int_{-\pi/2}^{\pi/2} d\theta \int_0^{2\cos\theta} r^3 dr = 8\int_0^{\pi/2} \cos^4\theta d\theta = \frac{3\pi}{2}.$$

It is obvious that the polar coordinate transform may be useful when $x^2 + y^2$ appear in the integral region and/or integrand.

One more remark: under change of variable, the interval becomes interval in one-variable case, but the rectangle is no longer a rectangle in more variable case. In the Riemann sum of multiple integrals, we use rectangles to cut the region D. However to make the theory of multiple integration rigorous, for example to make change of variables work, we have to consider cutting the region D into small pieces of any shape. This is much more complicated.

6.5.3 Gaussian density

To end this course, we shall consider a convergent improper integral

$$I = \int_{-\infty}^{+\infty} e^{-x^2/2} dx = \lim_{R\to+\infty} \int_{|x|\leqslant R} e^{-x^2/2} dx.$$

This is the most important density function in probability theory, called Gaussian (or normal) density function, and can not be computed by Newton-Leibniz formula because the primitive function of the integrand is not an elementary function.

What follows is a marvelous idea to compute this integral. Instead of computing I, we compute I^2 which can be viewed as an integral on plane. Let $R > 0$. By the iterated formula,

$$\int\int_{|x|\leqslant R, |y|\leqslant R} e^{-(x^2+y^2)/2} dxdy = \int_{-R}^{R} e^{-x^2/2} dx \cdot \int_{-R}^{R} e^{-y^2/2} dy = \left(\int_{-R}^{R} e^{-x^2/2} dx\right)^2,$$

Hence we have

$$I^2 = \lim_{R\to+\infty} \iint_{|x|\leqslant R, |y|\leqslant R} e^{-(x^2+y^2)/2} dx dy.$$

However instead of computing the integral on the square, we compute the integral on the disk $x^2 + y^2 \leqslant R^2$, by the polar coordinate transform $x = r\cos\theta, y = r\sin\theta$. Then we have

$$\begin{aligned}\iint_{x^2+y^2\leqslant R^2} e^{-(x^2+y^2)/2} dx dy &= \iint_{r\leqslant R, \theta\in[0,2\pi)} e^{-r^2/2} r dr d\theta \\ &= \int_0^{2\pi} d\theta \int_0^R e^{-r^2/2} r dr = 2\pi(-e^{-r^2/2})\Big|_0^R = 2\pi(1 - e^{-R^2/2}),\end{aligned}$$

which has limit 2π. Though $\iint_{|x|\leqslant R, |y|\leqslant R} f \neq \iint_{x^2+y^2\leqslant R^2} f$, where $f = e^{-(x^2+y^2)/2}$, their limits as $R \to \infty$ are the same, since

$$2\pi(1 - e^{-R^2/2}) = \iint_{x^2+y^2\leqslant R^2} f \leqslant \iint_{|x|\leqslant R, |y|\leqslant R} f \leqslant \iint_{x^2+y^2\leqslant 2R^2} f = 2\pi(1 - e^{-2R^2/2}).$$

Hence $I = \sqrt{2\pi}$.

Remark. To end this course, the author would like to make one more remark. We learn this course for application in economics. However mathematics was born not only for application, but also for its own value and beauty. This result

$$\int_{-\infty}^{+\infty} e^{-x^2/2} dx = \sqrt{2\pi}$$

is extremely beautiful. The left and right side contain the Euler's number e and the circumference ratio π, respectively and these two numbers, the most important numbers in mathematics, seem to have nothing to do with each other according to their definitions. The beauty of mathematics is one of beauties in life, which is a little hard to appreciate. The more beauty you feel, the better you understand mathematics.

Exercises

1. find

   (a) $\iint_D xye^{x+y} dx dy$ where D is given by $x + y \leqslant 0$ and $0 \leqslant y - 2x \leqslant 3$.

   (b) $\iint_D \frac{1}{\sqrt{xy}} e^{-(x+y)} dx dy$ where D is given by $0 \leqslant x \leqslant y \leqslant 1 - x$.

2. Using polar coordinates transform to evaluate the following integrals.

   (a) $\int_{-1}^0 dx \int_{-\sqrt{1-x^2}}^0 \frac{2}{1 + \sqrt{x^2 + y^2}} dy;$

(b) $\int_0^2 dx \int_0^{\sqrt{1-(x-1)^2}} \frac{x+y}{x^2+y^2} dy;$

(c) $\int_0^1 dx \int_0^{\sqrt{1-x^2}} e^{-(x^2+y^2)} dy.$

3. compute $\int_{-\infty}^{+\infty} x^n e^{-x^2/2} dx$ for $n \in \mathbf{N}$.

Bibliography

[1] J. Stewart, CALCULUS, 6th edition.

[2] W. Dunham, THE CALCULUS GALLERY: MASTERPIECES FROM NEWTON TO LEBESGUE, Princeton University Press, 2005

appendix: midterms and exams

2022

midterm 1

1. write 2/7 into decimal expression.

2. Does $\sqrt[n]{n}$ converge? If yes, find the limit.

3. Does $\lim\limits_{x \to 0} x \sin(1/x)$ exist? if yes, find the limit.

4. Does $\lim\limits_{x \to +\infty} x \sin(1/x)$ exist? if yes, find the limit.

5. prove that $x^3 - 3x + 1 = 0$ has at least one solution.

6. Let $f(x) = \ln(\sqrt{1+x^2} - x)$. Find $f'(x)$.

7. prove that for $n \in \mathbb{N}$, $\ln(1 + \frac{1}{n}) < \frac{1}{n}$.

8. Find the monotonicity of $f(x) = x^3 - 3x$ and max/min of f on $[-1, 2]$.

9. Let
$$a_n = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2}.$$
prove that a_n converges.

10. find the least real number c so that for all $x \geqslant 0$, it holds that
$$x - \ln(1+x) \leqslant cx^2.$$

midterm 2

1. (20) Let $f(x) = x^3 - 6x^2 + 9x - 2$.

   (a) Find the intervals where f is monotone, f is convex or concave.

(b) Find the points where f reaches local extremum and find the global extremum of f on $[0, 3]$.

2. (40) Find the following integrals.

 (a) $\int (\tan x)^2 dx$.

 (b) $\int (\tan x)^3 dx$.

 (c) $\int (\sin x)^3 dx$.

 (d) $\int (\sin x)^4 dx$.

3. (30)Find the integrals

 (a) $\int_0^{+\infty} x^2 e^{-x} dx$.

 (b) $\int_{-2}^{2} |x - x^3| dx$.

 (c) $\int_0^{\pi} (\sin x)^2 (\cos x)^9 dx$.

4. (10) Prove that

$$\lim_n \int_0^1 (4x - 4x^2)^n dx = 0.$$

Final

1. find limit $\lim_n \frac{1}{n} \sum_{k=1}^{n} \ln \frac{n}{k}$.

2. Assume that $\frac{1}{\sqrt{1-x}} = \sum_{n=0}^{\infty} a_n x^n$ for $|x| < 1$. Find a_0, a_1, a_2, a_3, a_4.

3. Set $f(x) = arcsinx$. Find $f^{(5)}(0)$, $f^{(6)}(0)$, $f^{(7)}(0)$, $f^{(8)}(0)$, $f^{(9)}(0)$.

4. find $\int (\cos x)^{-1} dx$, $\int (\cos x)^{-3} dx$, $\int (\cos x)^{-4} dx$.

5. The closed region D is bounded by the curves $x + y = 2$ and $y = x^2$. Sketch the region and find the global max/min of $f(x, y) = 4x - 8xy + 2y$ on D.

6. sketch the region $D : x > 0, y > 0, x + y \leqslant 1, y \geqslant x$ and find

$$\int\int_D \frac{1}{\sqrt{xy}} dxdy, \int\int_D \frac{e^{-(x+y)}}{\sqrt{xy}} dxdy.$$

 Hint: use change of variable according to the description of D if necessary.

7. prove that (1) if $\sum_n a_n$ converges absolutely then $\sum_n a_n$ converges; (2) if $\sum_n a_n$ converges, then the radius of convergence of $\sum_n a_n x^n$ is at least 1.

2021

midterm 1

1. $A = \{1, 2, 3, 4, 5\}$, $B = \{x \in \mathbf{R} : x^2 - 8x + 12 = 0\}$, the set of real roots of $x^2 - 8x + 12 = 0$. Write down B, $A \cup B$, $A \cap B$ and $A \setminus B$.

2. (1) Find the limit $a_n = \left(\dfrac{n^2 + 2n + 100}{n^2}\right)^n$, (2) Prove that

$$b_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} - \ln(n+1)$$

   is increasing.

3. State the definition of $\lim_{x\to a} f(x) = L$ and prove carefully that $\lim_{x\to a} x^2 = a^2$ according to definition.

4. (1) $y = e^{\sqrt{1-x^2}}$, find y'_x. (2) $x^2 + e^{x-y} = y$, find $y_x\Big|_{(1,1)}$.

5. $y = \sqrt{|x|}(x^2 + x - 2)$, $x \in [-2, 1]$. (1) find critical points. (2) find intervals where the function is monotone. (3) find local extremum. (4) find global extremum.

midterm 2

1. $\displaystyle\int x \sin x dx$,

2. $\displaystyle\int \sin 2x \sin x dx$

3. $\displaystyle\int_0^{\pi} \sin 3x \sin 2x \sin x dx$

4. compute the length of curve $y = e^x$ for $0 \leqslant x \leqslant 1$.

5. Prove that $\displaystyle\int_0^{+\infty} x^{10} \sin x e^{-x} dx$ converges.

final

1. (5)Prove that $\sqrt{2}$ is not rational.

2. (10)Prove that for any $a < b$, there are infinitely many rational numbers and irrational numbers in (a, b).

3. (5)Prove that $\sqrt[n]{n}$ converges to 1.

4. (5)Find the derivative of $y = \sqrt{1 + x^2}$ at $x == 0$.

5. (5)Find the max/min of $y = x^3 - 3x + 1$ on $[-1, 2]$.

6. (5)Find the integral
$$\int_0^{\pi/4} \frac{1}{\cos y} dy.$$

7. (5)Find the improper integral $\int_0^1 \ln x dx$.

8. (30)Assume that $f_n(x) = x^n$ on $[0,]1$.

   (a) Write down its pointwise limit on $[0, 1]$;

   (b) prove that it does not converge uniformly on $[0, 1]$;

   (c) prove that it converges uniformly on $[0, r]$ when $0 < r < 1$.

9. (10)Find the interval of convergence of the power series
$$\sum_{n=0}^{n} \frac{1}{n}(x-1)^n.$$

10. (10)Find the max/min of $z = x^2 - y^2$ on the region D surrounded by $y = x + 1$ and $y = x^2 - 1$.

11. (10)Find the integral of $z = x^2 - y^2$ on the region D surrounded by $y = x + 1$ and $y = x^2 - 1$.

2020

1. (10)Find limits
$$(i)\ \lim_n (1 + \frac{1}{3n^2})^{4n} \text{ and } (ii)\ \lim_n \sqrt[2n]{n}.$$

2. (10)

   (a) Let $y = x^3 - 3x^2$. Find its maximum and minimum on $[-1, 2]$.

   (b) $f(x) = \dfrac{1 - \sqrt{1 - 4pqx^2}}{2x}$, where $1 > p > q > 0$ and $p + q = 1$. find $f'(1)$, derivative of f at 1.

3. (10) Find
$$(i)\ \int \frac{1}{9 - x^2} dx, \text{ and } (ii) \int_0^{\pi} x \cos x dx.$$

4. (16) converges or diverges?
$$(i)\ \sum_{n=1}^{\infty} \left(\frac{n^2}{2n^2 + 2}\right)^n \text{ and } (ii)\ \sum_n \frac{\ln n}{n}$$

5. (20) Given a series $\sum_n a_n$

   (a) (10)state definition of absolute convergence and conditional convergence.

   (b) (10)prove that if it converges, then $\lim a_n = 0$.

6. (14)

   (a) expand $f(x) = (3 + x)^{-1}$ into power series (center 0) and find its domain of convergence.

   (b) expand the function $f(x) = \ln x$ into a power series with center 3 and find its domain of convergence.