School of Mathematics and Statistics

MATH 1021

Calculus of One Variable

THE UNIVERSITY OF SYDNEY

Revised February 2021

Table of contents

Acknowledgements

The material in these notes has been developed over many years by the following members of the School of Mathematics and Statistics:

Eduardo Altmann	**Mary Myerscough**
Sandra Britton	**Nigel O'Brian**
Chris Durrant	**Sharon Stephen**
Dave Galloway	**Fernando Viera**
Jenny Henderson	**Haotian Wu**
Andrew Mathas	

Introduction

Calculus is one of the major achievements of the 17th century. It plays a key role in almost every instance where mathematics is applied in the sciences, engineering, or in economics. Without calculus, we would not have cars, computers, televisions or mobile phones; Einstein would never have penned his theory of relativity; we would not know of the existence of DNA; we would never have landed on the moon. The list goes on.

Whether you end up continuing in mathematics or majoring in another field it will be important for you to learn and understand the meaning of calculus. The reason for this is quite simple. In high school you can progress simply by memorizing formulas; in university there will be times when you need to develop formulas for yourself, and this is where a proper understanding of calculus will be a definite asset.

These notes are intended to supplement the lectures of MATH1021. Your lecturers will almost certainly use different examples and they will also explain some of the material in the course slightly differently from these notes. In some places these notes go into more detail than your lectures; at other times your lecturer will go into more detail.

Reading mathematics is not like reading a novel; we have to think and struggle with every sentence. We are professional mathematicians and we are not ashamed to say that in our research there have been times when we have spent more than a day trying to understand a single line of mathematics! You will be pleased to know that in this course you should not have to spend this long on a single sentence; however, there will be times when you do have to think quite hard to understand what is going on. If you do get stuck then go and ask your lecturer or tutor to explain it to you!

In addition to thinking when you read mathematics you should also work through the calculations yourself using pen and paper.

At some places in the notes and in the Appendices we have included material which is more "advanced" than we expect you to know or understand. You are free to either read these sections or skip over them, as you wish.

Tutorial problems and Exercise sheets

There are plenty of problems with full solutions for you to practice.

a) **Worked examples** with full solutions have been included in these lecture notes throughout all chapters.
b) **Exercises** are available at the end of the chapters in these notes and answers to Selected Exercises can be found in Appendix G.

c) **Exercise sheets** containing problems to be solved before the tutorial session are available on the MATH1021 web page. Full solutions will be available online at the end of the corresponding week.
A detailed list of mathematical objectives (knowledge, understanding and skills) for a given chapter is provided in the weekly Exercise Sheets.
d) **Board tutorial sheets** will be handed out during tutorials with problems to be solved during the tutorial class. Full solutions will be available online at the end of the corresponding week.
e) **Solutions** will be provided to assignments 1 and 2.
f) **Questions and solutions** to selected past exam papers will be made available near the end of semester.

Why study mathematics
The study of mathematics enhances your ability to think logically and analytically, move from the particular to the general, work quantitatively and improve problem-solving skills. By reading and working carefully through the material in these notes you will develop the following additional generic skills:

- Generalise simple and familiar ideas to more complex settings.
- Use geometric/visual techniques to help understand new concepts.
- Apply simple techniques in unfamiliar situations.
- Estimate values by using suitable approximation techniques.
- Recognise that bounds on the error are an important part of any good approximation.

A note about definitions
A mathematical definition is a precise description of some mathematical concept. Historically, many concepts in mathematics have been used extensively before a precise definition of the concept has been formulated.

While precision in definitions is certainly important, learning a definition off by heart, without an understanding of the concept, is unlikely to be helpful. It is important to spend some time thinking about a definition in order to gain this understanding.

CHAPTER 1

Real and Complex Numbers

Mathematics includes not only the study of logic, structure and geometry, but also ideas about numbers. Real numbers in particular, are fundamental to calculus and many other branches of mathematics. In this chapter we review the concepts of sets and extend previous work on numbers, particularly the real numbers, before introducing the set of complex numbers.

1.1 Sets

Set notation is a convenient and precise way to write about collections of numbers. We start by talking about general sets.

> **Definition**
> A "set" is a collection of objects which are called "members" or "elements" of the set.

Example 1.1a A set can be written as a list, for example, $A = \{a, b, c, d\}$, where

- A is the name of the set,
- a, b, c, d are the elements of the set enclosed in braces and separated by commas.
- If the list of elements is large, three dots may be used to mean **'and so on'**. For example, the set of natural numbers may be denoted by $\mathbb{N} = \{0, 1, 2, 3, \dots\}$.

$\Diamond$

1.2 Number Sets

Our understanding of numbers, what they are and how they work, develops from simple counting through fractions and negative numbers to an appreciation of irrational numbers and real numbers. Mathematically, different types of numbers belong to different sets.

The set of "natural numbers" $\{0,1,2,3,4,\ldots\}$, is denoted by the symbol $\mathbb{N}$. It is *closed* under the operations of addition and multiplication. That is, adding two natural numbers gives another natural number, as does multiplying them together.

The set of "integers" $\{\ldots,-4,-3,-2,-1,0,1,2,3,4,\ldots\}$, denoted by $\mathbb{Z}$, is the set of whole numbers, including both positive whole numbers, negative whole numbers and zero. The set of integers is closed under the operations of addition, subtraction and multiplication.

The set of "rational numbers", denoted by $\mathbb{Q}$, is the set of all numbers of the form n/m where n and m are integers and $m \neq 0$. Some examples are $\frac{1}{2}, -\frac{1}{4}, \frac{4}{3}$. Rational numbers include decimals which either terminate or repeat. Note that the integers are a subset of the rational numbers, since they are of the form n/m where $m = 1$. The set of rational numbers is closed under the operations of addition, subtraction, multiplication and division, provided that division by zero is excluded.

The set of "real numbers", denoted by $\mathbb{R}$, includes all rational numbers and all irrational numbers. Irrational numbers cannot be expressed as n/m, where m and n are integers, although some may be interpreted geometrically. For example, $\sqrt{2}$ is the length of a diagonal of a unit square. The irrational number π is the ratio of the circumference of a circle to the circle's diameter.

The set of "complex numbers", denoted by $\mathbb{C}$, contains all the other number sets mentioned above and all the imaginary numbers to be introduced in Section 1.4.

In fact we can summarise these numbers sets diagrammatically as shown in Figure 1.1.

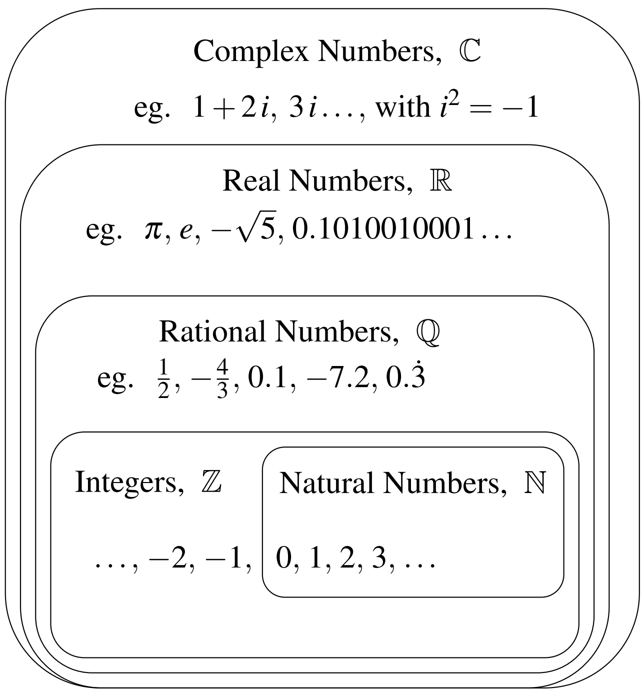


Figure 1.1: Number Sets

Set notation

- **Element of a set –** The symbol $\in$ means "is an element of". For example, $-3 \in \mathbb{Z}$ is read as "-3 is an element of the set of integers"; $y \in B$ is read as "y is an element of the set B" or "y is a member of the set B".

- **Subset of a set –** The symbol $\subseteq$ should be read as "is a subset of". For example, $\mathbb{N} \subseteq \mathbb{Z}$ is read as "the set of natural numbers is a subset of the set of integers" or "the set of natural numbers is contained in the set of integers".

- **Strictly a subset –** Sometimes you may see the symbol $\subset$ which means that the smaller set is strictly a subset of the larger; the two cannot be equal. For example, it is most precise to write $\mathbb{N} \subset \mathbb{Z}$ as the two sets are not the same.

- **Contains a set –** The reversed symbol $\supseteq$ means "contains". For example, $\mathbb{R} \supseteq \mathbb{Q}$ reads "the set of real numbers contains the set of rational numbers". (There is also a symbol $\supset$ which means "contains, but is not equal to".) If $A \subseteq B$ then $B \supseteq A$.

- **Not an element of a set –** A forward slash through any of these symbols above means "not". For example, $-1 \notin \mathbb{N}$ is read as "-1 is not an element of the set of natural numbers".

- **Not a subset –** Another example, $\mathbb{R} \not\subseteq \mathbb{Z}$, is read as "the set of real numbers is not a subset of the set of integers."

There are other symbols which describe sets formed from other sets:

- **Union of sets –** The expression $A \cup B$ denotes the union of set A with set B. The "union" of two sets is the set of elements which are members of either one or both of the sets. If an element occurs in both sets, it is only listed once in the union. For example

$$\{1,2,3,4,\} \cup \{3,4,5,6\} = \{1,2,3,4,5,6\}$$

- **Intersection of sets –** The intersection of sets A and B is written $A \cap B$. The "intersection" of two sets is the set of elements which are members of *both* of the sets. For example

$$\{1,2,3,4,\} \cap \{3,4,5,6\} = \{3,4\}$$

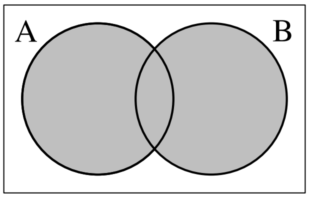
- **Subtraction of sets –** The symbol $\setminus$ which is read "minus" or "without", is used to indicate the set of elements which are in one set but not in another. That is, $A \setminus B$ is the set of all elements which are in A but not in B. So for example,

$$\{1,2,3,4,\} \setminus \{3,4,5,6\} = \{1,2\}.$$

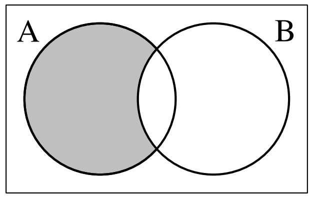
Venn Diagrams

A set can be represented in a simple, graphical way by a "Venn diagram". Each set is drawn as a circle, a square or some other closed shape. Shapes representing sets may overlap one another if sets have elements in common. Sometimes, the elements of the sets are written on the Venn diagram but often they are not. Different parts of a Venn diagram can be shaded to illustrate different parts of the set.

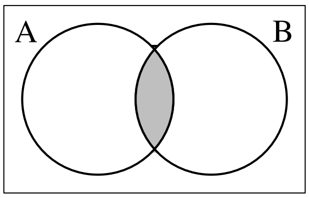
Venn diagrams are a useful way to represent relations between sets. Note that $A\backslash B$ is not the same as $B\backslash A$.



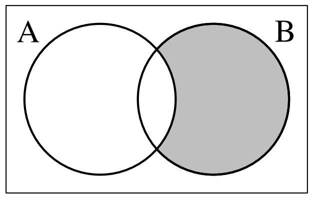
$A \cup B$



$A\backslash B$



$A \cap B$



$B\backslash A$

Conditions on Sets

If we want to specify a set whose elements fulfil a certain condition then we do this in the way illustrated in the following examples.

- If we want to express that “A is the set of all rational numbers x such that x is positive”, we write
$$A = \{x \in \mathbb{Q} \mid x > 0\}.$$
The vertical slash should be read as **such that**.

- Let W be the set of words in English. Then
$$B = \{x \text{ in } W \,|\, x \text{ begins with the letter “P”}\}$$
reads “B is the set of all elements x of the set of English words such that x begins with P” or “B is the set of all English words that begin with P.”

- If we want to say that "C is the set of all integers x such that $x/2$ is an integer" or "C is the set of all even integers", we write

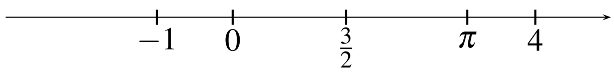
$$C = \{x \in \mathbb{Z} \mid \frac{x}{2} \in \mathbb{Z}\}.$$

- If we want to say that "D is the set of all real numbers which are greater than -1 and less or equal to 1", we write

$$D = \{x \in \mathbb{R} \mid -1 < x \leq 1\}.$$

1.3 The real number line – Intervals

- **The real number line –** Every real number can be located on the "real number line". For example:



It is straightforward to sketch sets that are written using interval notation on the real number line.

Note that an open dot is used if the end point of the interval is not included in the set. If the endpoint is part of the set, then a closed dot is used.

- **Interval notation –** Sets of real numbers which lie between two end points can be represented using "interval notation". For example

$$D = \{x \in \mathbb{R} \mid -1 < x \leq 1\} = (-1, 1]$$

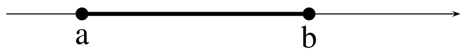
A **curved bracket** is used to show that an endpoint (such as -1 in this example) *is not* included in the set and a **square bracket** is used when the endpoint *is* part of the set.

- **Open interval –** An interval where neither endpoint is part of the set is called an "open interval".

$$(a, b) = \{x \in \mathbb{R} \mid a < x < b\}$$

The interval $(a,b) = \{x \in \mathbb{R} \mid a < x < b\}$ and the point (a,b) in the Cartesian plane are written in exactly the same way. They are not, however, the same thing. It is usually clear from the context whether (a,b) represents a point or an interval.

- **Closed interval –** If both endpoints are part of the interval it is called a "closed interval".



$[a, b] = \{x \in \mathbb{R} \mid a \leq x \leq b\}$

It is also possible that one endpoint will be in the set and the other will not be. For example,

$(a, b] = \{x \in \mathbb{R} \mid a < x \leq b\}$

$[a, b) = \{x \in \mathbb{R} \mid a \leq x < b\}$

- **Semi-infinite intervals –** There is special notation for sets of the number line that extend infinitely in one direction or the other.

$$(a,\infty) = \{x \in \mathbb{R} \mid x > a\}; \quad (-\infty,a) = \{x \in \mathbb{R} \mid x < a\}$$

$$[a,\infty) = \{x \in \mathbb{R} \mid x \geq a\}; \quad (-\infty,a] = \{x \in \mathbb{R} \mid x \leq a\}$$

Note that ∞ is *not* a number, rather, it represents infinity.

Both ∞ and $-\infty$ always take a round bracket.

Examples 1.3a

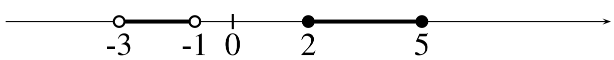
i) $A = [7,29] = \{x \in \mathbb{R} \mid 7 \leq x \leq 29\}$

$[7, 29]$

ii) $S = (2,\infty) = \{x \in \mathbb{R} \mid x > 2\}$

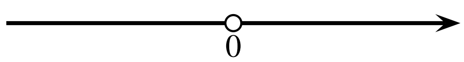
$(2, \infty)$

iii) $V = (-3,-1) \cup [2,5] = \{x \in \mathbb{R} \mid -3 < x < -1 \;\text{ or }\; 2 \leq x \leq 5\}$



$(-3, -1) \cup [2, 5]$

iv) $T = (-\infty,0) \cup (0,\infty) = \{x \in \mathbb{R} \mid x \neq 0\} = \mathbb{R} \setminus \{0\}$. As you can see there may be a number of ways of writing down a set.



$\mathbb{R} \setminus \{0\}$

$\Diamond$

Modulus or absolute value

The "modulus" or "absolute value" $|x|$ of a real number x gives the distance on the real number line from x to zero. The modulus of x is defined in this way:

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

For example $|5| = 5$ and $|-10| = 10$.

The distance between two numbers on the number line can also be expressed using modulus. The distance between x and y is given by $|x-y| = |y-x|$. For example, the distance between 3 and -4 is $|3-(-4)| = |3+4| = 7$ which is what we intuitively expect to be the distance from -4 to 3. Alternatively we could have written $|-4-3| = |-7| = 7$.

1.4 Complex numbers - Cartesian form

Suppose that you are asked to solve the equation

$$x^2 + 1 = 0.$$

Your first response might be to say that there will be two solutions as it is a quadratic equation. Very quickly you might write down the line

$$x^2 = -1.$$

At that point you might conclude, correctly, that there are no real solutions to the equation, because in the real number system, we cannot take square roots of negative numbers. But what if we agree that there exists a number x such that $x = \sqrt{-1}$?

Such a number does indeed exist, although it is not a real number. It is known as an "imaginary number". We denote it by i (although some branches of engineering use j instead) and we'll assume that the usual rules for algebraic manipulation apply.

Imaginary unit
The number denoted by i that satisfies the condition $\quad i^2 = -1 \quad$ is called the *imaginary unit*. It follows that

$$i = \sqrt{-1}.$$

The equation $x^2 + 1 = 0$ now has two imaginary solutions, namely i and $-i$. To check that $x = \pm i$ are solutions, substitute into the equation

$$x^2 + 1 = (\pm i)^2 + 1 = -1 + 1 = 0.$$

What about the equation $x^2+9=0$? In this case

$$x^2+9=0 \quad \Longrightarrow \quad x=\pm\sqrt{-9}=\pm\sqrt{-1\times 9}=\pm\sqrt{-1}\sqrt{9}=\pm 3\,i.$$

It is easy to show by substitution into $x^2+9=0$ that $x=\pm 3\,i$ are both solutions.

Properties of i – The imaginary unit satisfies the following useful relations:

$$i^2=(-i)^2=-1$$

$$i^3=(i^2.\,i)=(-1.\,i)=-i$$

$$i^4=(i^2.\,i^2)=(-1).(-1)=1$$

$$i^8=(i^4.\,i^4)=(1.\,1)=1, \quad \text{and so on.}$$

We are now in a position to introduce a new number set:

> **Imaginary numbers**
> Any non–zero real multiple of i is called a **purely imaginary number** or just **imaginary number**. The square of an imaginary number is a negative real number.

For example

$$3i,\ -20i,\ -i/5 \quad \text{and} \quad \pi i$$

are all imaginary numbers, and their squares

$$(3i)^2=-9, \quad (-20i)^2=-400, \quad (-i/5)^2=-1/25, \quad (\pi\,i)^2=-\pi^2,$$

are all negative real numbers.

Complex numbers

Suppose now that you are given this equation to solve:

$$x^2-4x+5=0.$$

Completing the square and rearranging gives $(x-2)^2=-1$; that is, $x-2=\pm i$ or $x=2\pm i$. These solutions can also be obtained by applying the familiar quadratic formula:

$$x=\frac{4\pm\sqrt{16-20}}{2}=\frac{4\pm\sqrt{-4}}{2}=\frac{4\pm\sqrt{-1}\,\sqrt{4}}{2}=\frac{4\pm 2\,i}{2}=2\pm i.$$

These solutions are not purely imaginary, although they do involve an imaginary number. The solutions $2+i$ and $2-i$ are called **complex numbers**.

Cartesian form of a complex number
A complex number expressed in the form $a+ib$ is said to be in Cartesian form.

- **Real numbers** are a special case of complex numbers when $b=0$.
- **Imaginary numbers** are a special case of complex numbers when $a=0$.

The complex number $2+i$ in the above example is written in Cartesian form with $a=2$ and $b=1$.

Real and Imaginary parts

Given a complex number in Cartesian form $z=a+ib$:

- The real number a is called the **real part** of z and we write $\mathrm{Re}(z)=a$.
- The real number b is called the **imaginary part** of z and we write $\mathrm{Im}(z)=b$.

Examples 1.4a

i) If $z=2+i$ then $\mathrm{Re}(z)=2$ and $\mathrm{Im}(z)=1$.

ii) If $w=3+8i$ then $\mathrm{Re}(w)=3$ and $\mathrm{Im}(w)=8$.

iii) If $z=\frac{1}{2}-5i$ then $\mathrm{Re}(z)=\frac{1}{2}$ and $\mathrm{Im}(z)=-5$.

iv) The purely imaginary number $z=-7i$ can be written as $z=0-7i$. Therefore $\mathrm{Re}(z)=0$ and $\mathrm{Im}(z)=-7$.

v) For the real number $z=4=4+0i$ we have $\mathrm{Re}(4)=4$ and $\mathrm{Im}(4)=0$. $\Diamond$

Quadratic equations

If we allow complex numbers as solutions to quadratic equations with real coefficients then every such quadratic equation will always have two solutions, and they will be either both real or both complex.

We can see this in general if we look at the quadratic formula. The solution to the quadratic equation $ax^2+bx+c=0$, where a, b and c are reals, is given by

$$x = \frac{-b \pm \sqrt{b^2-4ac}}{2a}.$$

Whether $ax^2+bx+c=0$ has (purely) real or complex roots depends on the sign expression b^2-4ac which is known as the discriminant of the quadratic.

$$x \text{ is } \begin{cases} \text{real} & \text{if } b^2-4ac \geq 0 \\ \text{complex} & \text{if } b^2-4ac < 0. \end{cases}$$

Example 1.4b The solutions of $x^2+6x+25=0$ must be complex since $b^2-4ac=-64<0$. Using the quadratic formula, the solutions are found to be $-3+4i$ and $-3-4i$. These complex numbers are related; they are *complex conjugates* of each other. This will be examined further in the next section. $\Diamond$

1.5 Arithmetic in Cartesian form

Complex numbers can be added or multiplied together, subtracted from one another or divided by one another.

Consider two complex numbers $z=a+bi$ and $w=c+di$. Here the real part of z is a and the imaginary part of z is b; the real part of w is c and the imaginary part of w is d.

Addition

$$\begin{aligned} z+w &= (a+bi)+(c+di) \\ &= (a+c)+(b+d)i \end{aligned}$$

Rule: Add real parts to real parts and imaginary parts to imaginary parts.

Example 1.5a

$$\begin{aligned} (3-4i)+(1+2i) &= 3+1+(-4+2)i \\ &= 4-2i \end{aligned}$$

$\Diamond$

Subtraction

$$\begin{aligned} z-w &= (a+bi)-(c+di) \\ &= (a-c)+(b-d)i \end{aligned}$$

Rule: Subtract real parts from real parts and imaginary parts from imaginary parts.

Example 1.5b

$$(3-4i)-(1+2i) = 3-1+(-4-2)i = 2-6i$$

$\Diamond$

Multiplication

$$\begin{aligned} zw &= (a+bi)(c+di) \\ &= ac+adi+bci+(bd)i^2 \\ &= (ac-bd)+(ad+cb)i \end{aligned}$$

Rule: Expand the brackets in the normal way, remembering that i^2 can be simplified to -1, and collect terms into real and imaginary parts.

Example 1.5c

$$(3-4i)(1+2i) = 3-4i+6i-8i^2 = 3+2i+8 = 11+2i$$

$\Diamond$

Complex conjugate and division

To divide one complex number by another we have to introduce the complex conjugate of a complex number.

> **Complex conjugate**
> The complex conjugate of the number $z = a+ib$ is the complex number defined by $\overline{z} = a-ib$.

The geometric interpretation of the complex conjugate $\overline{z}$ is the reflection of z about the real axis. The following properties of the complex conjugate can be easily proved from the definition,

Properties of conjugates

$$\overline{z+w} = \overline{z} + \overline{w} \qquad \overline{zw} = \overline{z}\,\overline{w} \qquad \overline{z^n} = \overline{z}^n$$

If $z = a+ib$ then $z\overline{z} = (a+ib)(a-ib) = a^2+b^2$.

Examples 1.5d

i) $\overline{3+5i} = 3-5i \qquad \overline{2-7i} = 2+7i$

ii) Verify the first property of conjugates in the box above when $z = 1+2i$ and $w = 3+i$

$$\overline{z+w} = \overline{(1+2i)+(3+i)} = \overline{4+3i} = 4-3i$$

$$\overline{z} + \overline{w} = \overline{(1+2i)} + \overline{(3+i)} = (1-2i)+(3-i) = 4-3i$$

iii) If z is a real number then $\overline{z} = z$. For example, if $z = \sqrt{2}$ then $\overline{z} = \sqrt{2}$.

iv) If z is a purely imaginary number then $\overline{z} = -z$. For example $\overline{3i} = -3i$. ◊

Division

If $w \neq 0$ then to find $\frac{z}{w}$ we multiply both top and bottom by the complex conjugate of w.

$$\begin{aligned}
\frac{z}{w} &= \frac{z}{w}\frac{\overline{w}}{\overline{w}} \\
&= \frac{(a+bi)}{(c+di)}\frac{(c-di)}{(c-di)} \\
&= \frac{ac-adi+cbi-(bd)i^2}{c^2-cdi+cdi-d^2i^2} \\
&= \frac{(ac+bd)+(cb-ad)i}{c^2+d^2} \\
&= \frac{ac+bd}{c^2+d^2} + \frac{cb-ad}{c^2+d^2}\,i
\end{aligned}$$

This process is similar to rationalising the denominator of a quotient of surds. Multiplying by the complex conjugate of the divisor produces a real number in the denominator and allows the number to be written in the Cartesian form $a+bi$.

Example 1.5e

$$\begin{aligned}\frac{5-10i}{1+2i} &= \frac{(5-10i)(1-2i)}{(1+2i)(1-2i)} \\ &= \frac{5-20i+20i^2}{1-2i+2i-4i^2} \\ &= \frac{-15-20i}{5} \\ &= -3-4i\end{aligned}$$

$\Diamond$

Equality of complex numbers

Two complex numbers are equal to each other if and only if both their real and imaginary parts are equal. In other words, if $z = a+bi$ and $w = c+di$, then $z = w$ if and only if $a = c$ and $b = d$.

1.6 The set of complex numbers

We have seen that a real number is a particular type of complex number, one with zero imaginary part. The complex numbers include real numbers and form a set which contains the set of real numbers and hence all of the other number sets we have mentioned.

Set of complex numbers
The "set of complex numbers" $\mathbb{C}$ is the set of all numbers of the form $a+bi$ where a and b are real numbers and $i^2 = -1$.
Using set notation we can write:

$$\mathbb{C} = \{a+bi \mid a,b \in \mathbb{R},\ i^2 = -1\}.$$

As we have shown in Figure 1.1, we also have

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}.$$

The set of complex numbers, like the set of real numbers, is closed under addition, subtraction, multiplication and division. This means that the sum of two complex numbers is another complex number, and so on.

The set of complex numbers is not ordered

Complex numbers lack an important property of the real numbers.

The set of real numbers is "ordered"; that is, if we have any two real numbers x and y we can say that either

$$x > y \quad \text{or} \quad x < y \quad \text{or} \quad x = y.$$

Because of this property we are able to represent real numbers on the real number line.

The set of complex numbers is not ordered. Consider the two complex numbers $2-3i$ and $-1+5i$. It does not make sense to write

$$2-3i \;>\; -1+5i \quad \text{or} \quad -3i \;<\; -1+5i.$$

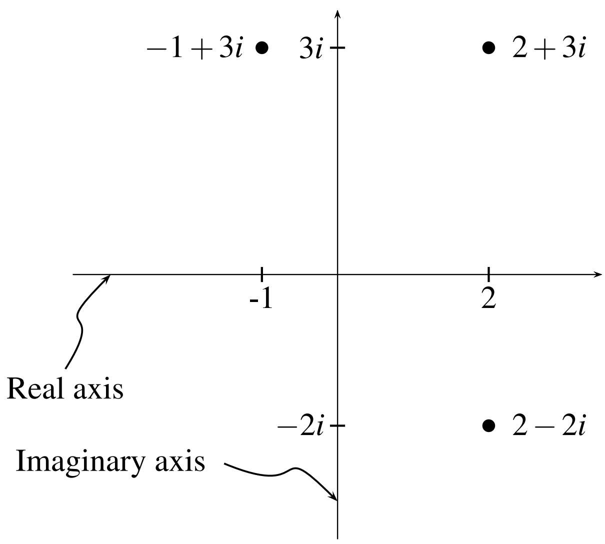
However, we may write $\text{Re}(2-3i) \;>\; \text{Re}(-1+5i)$ and $\text{Im}\,(2-3i) \;<\; \text{Im}\,(-1+5i)$ because the real part and the imaginary part of a complex numbers are both real numbers.

Because the set of complex numbers is not ordered, complex numbers cannot be represented as points on a line. Instead, complex numbers are represented as points on a plane.

The complex plane

The "complex plane" or "Argand diagram" allows complex numbers to be represented graphically. The horizontal axis in the complex plane is called the "real axis". All real numbers lie on the horizontal axis in the complex plane; positive numbers to the right of the origin, negative numbers to its left.

The vertical axis is known as the "imaginary axis". All purely imaginary numbers lie on the vertical axis. Each point in the complex plane corresponds to a single complex number. For example:



The modulus of a complex number

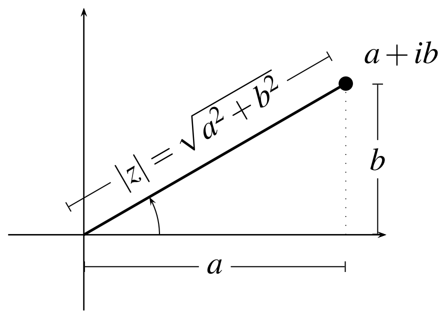
In Section 1.3 we defined the "modulus" or "absolute value" $|x|$ of a real number x as the distance on the real number line from x to zero, given by

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

In the case of complex numbers the modulus is also defined to give the distance from the origin:

Modulus
For a complex number $z = a + bi$ the modulus is defined by $|z| = \sqrt{a^2 + b^2}$.

By Pythagoras' Theorem, this formula gives the distance from z to the origin of the complex plane.



We can express $|z|$ in terms of z and its complex conjugate as follows,

$$z\,\bar{z} = (a+bi)(a-bi) = a^2 + b^2 = |z|^2 \quad \Longrightarrow \quad |z| = \sqrt{z\,\bar{z}}.$$

Properties of the modulus

For all complex numbers $z = a + ib$ and $w = c + id$, we have

a) $|zw| = |z|\,|w|$,

b) $|z/w| = |z|/|w|$,

c) $|z+w| \leq |z| + |w|$. This is called the **triangle inequality**,

d) $|z-w| \geq |z| - |w|$.

To show that (a) is true, we calculate $|zw|$ and $|z|\,|w|$ separately and show they are equal.

$$\begin{aligned}
|zw| &= |(a+ib)(c+id)| \\
&= |(ac-bd)+i(ad+bc)| \\
&= \sqrt{(ac-bd)^2+(ad+bc)^2} \\
&= \sqrt{a^2c^2+b^2d^2-2abcd+a^2d^2+b^2c^2+2abcd} \\
&= \sqrt{a^2c^2+b^2d^2+a^2d^2+b^2c^2},
\end{aligned}$$

while

$$\begin{aligned}
|z|\,|w| &= \sqrt{a^2+b^2}\sqrt{c^2+d^2} \\
&= \sqrt{a^2c^2+b^2d^2+b^2c^2+a^2d^2} \\
&= |zw|.
\end{aligned}$$

We will show (c) algebraically (a geometric argument can also be used, and this is left as an exercise). Note that since all quantities are non-negative, proving $|z+w| \le |z|+|w|$ is equivalent to proving $|z+w|^2 \le (|z|+|w|)^2$. Now

$$|z+w|^2 = |(a+c)+i(b+d)|^2 = (a+c)^2+(b+d)^2 = a^2+b^2+c^2+d^2+2(ac+bd).$$

Similarly,

$$(|z|+|w|)^2 = |z|^2+|w|^2+2|z|\,|w| = a^2+b^2+c^2+d^2+2|z|\,|w|.$$

So we must prove that $ac+bd \le |z|\,|w|$. Now

$$\begin{aligned}
|z|\,|w| &= \sqrt{a^2+b^2}\sqrt{c^2+d^2} \\
&= \sqrt{(a^2+b^2)(c^2+d^2)} \\
&= \sqrt{a^2c^2+b^2d^2+a^2d^2+b^2c^2} \\
&= \sqrt{(ac+bd)^2-2abcd+a^2d^2+b^2c^2} \\
&= \sqrt{(ac+bd)^2+(ad-bc)^2} \\
&\ge \sqrt{(ac+bd)^2} \\
&= |ac+bd|
\end{aligned}$$

Now every real number k is less than or equal to its own absolute value $|k|$. Hence

$$ac+bd \le |ac+bd| \le |z|\,|w|,$$

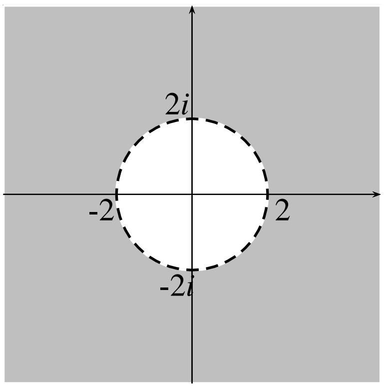
and the proof is complete. The proofs of (b) and (d) are left as an exercise.

Subsets of the complex plane

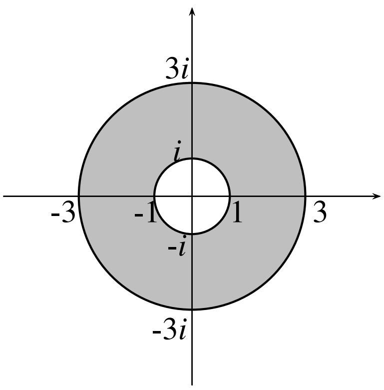
The modulus can be used to specify subsets of the set of complex numbers which can be graphed in the complex plane.

Examples 1.6a

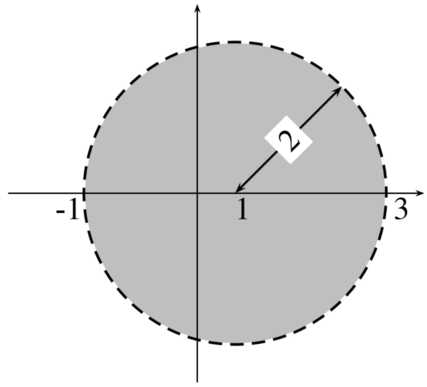
i) $\{z \in \mathbb{C} \mid |z| > 2\}$ is the set of complex numbers z such that z is more than 2 units distant from the origin. The set is represented by the shaded area below, which extends indefinitely.



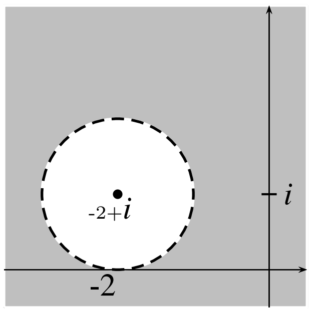
ii) $\{z \in \mathbb{C} \mid 1 \leq |z| \leq 3\}$ is the set of complex numbers which are between one and three units distant from the origin.



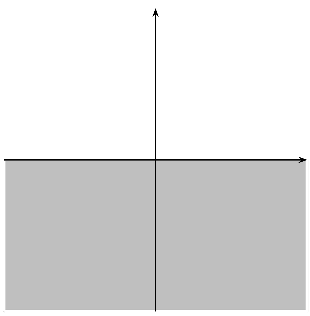
iii) $\{z \in \mathbb{C} \mid |z-1| < 2\}$. As with real numbers, $|z-1|$ is exactly the distance from z to 1. Hence, this is the set of all complex numbers whose distance from 1 is less than 2. Geometrically, these are all points in the complex plane that are inside the circle, centre 1, radius 2.



iv) $\{z \in \mathbb{C} \mid |z+2-i| > 1\}$. Here $|z+2-i| = |z-(-2+i)|$ is the distance from the complex number z to $-2+i$. So this set is the set of all complex numbers whose distance from $-2+i$ is greater than one unit. In other words, this is the set of points in the complex plane with are strictly outside the circle of radius 1 and centre $-2+i$.



v) Here is a different type of subset of the complex numbers: $\{z \in \mathbb{C} \mid \operatorname{Im} z \leq 0\}$ is the set of all complex number whose imaginary part is less or equal to zero.



◊

Summary of Chapter 1

- **Set theory** is the natural language used to describe and manipulate numbers.
- **Complex numbers** in Cartesian form $a + ib$ were introduced to find solutions of quadratic equations with no real roots.
- **Arithmetic operations** of addition, subtraction, multiplication and division were introduced in Cartesian form.
- **The complex conjugate** of a complex number $z = a + ib$ was defined as $\bar{z} = a - ib$.
- **The modulus** of $z = a + ib$ was defined as $|z| = \sqrt{a^2 + b^2}$.
- **The complex plane**, also known as the **Argand diagram**, gives us a geometric representation of a complex number $z = a + bi$ as a point (a, b) in the Cartesian plane.

Exercises

1.1 In each of the following exercises, perform the indicated operations and give the final answer in the form $x + yi$.

a) $(5-2i)+(2+3i)$
b) $(2-i)-(6-3i)$
c) $(2+3i)(-2-3i)$
d) $-i(5+i)$
e) $1/i$
f) $(a+ib)(a-ib)$
g) $6i/(6-5i)$
h) $(a+ib)/(a-ib)$
i) $1/(3+2i)$
j) $i^2, i^3, i^4, \ldots, i^{10}$
k) $(1+i)/(1-i)$
l) $[i/(1-i)]+[(1-i)/i]$
m) $(1/i)-3i/(1-i)$
n) $i^{123}-4i^9-4i$

1.2 If $z = 5 + 12i$ and $w = 3 + 4i$, express

$$w+z, \quad z-w, \quad zw \quad \text{and} \quad z/w$$

in the form $a + bi$. Use these results to verify that

a) $|zw| = |z||w|$

b) $|z/w| = |z|/|w|$

c) $|z+w| \leq |z| + |w|$

d) $|z-w| \geq |z| - |w|$

1.3 If $z = x + yi$, express each of the following explicitly in terms of x and y.

a) $\operatorname{Re}(z/\overline{z})$

b) $|(z/\overline{z})|$

c) $\operatorname{Im} z^3$

d) $\operatorname{Re} z^4$

e) $|z^6|$

f) $|(z+1)/(z-1)|$

g) $\operatorname{Re}(1/z^2)$

1.4 Simplify the following expressions.

a) $\operatorname{Im} \dfrac{1}{1+i}$

b) $\operatorname{Re} \dfrac{(1-i)^2}{1+2i}$

c) $|\cos\theta + i\sin\theta|$, where θ is any angle

d) $\left|\dfrac{1+3i}{3+i}\right|$

e) $\left|\dfrac{(1+i)^6}{i^3(1+4i)^2}\right|$

1.5 Solve the following equations using the quadratic formula.

a) $y^2 + 2y + 5 = 0$

b) $z^2 + 3z + 8 = 0$

c) $t^2 + t - 1 = 0$

d) $7a^2 + 8a + 4 = 0$

1.6 If $z = 3 - 2i$, plot z, $-z$, $\overline{z}$ and $-\overline{z}$ as points in the complex plane.

1.7 Show that for any complex number z, $|\overline{z}| = |z|$.

1.8 If $\overline{z} = z$, what can you say about z?

1.9 Prove properties (ii) and (iv) of the modulus, given at the end of the chapter.
[Hint for (iv): write $|z| = |(z+w) - w|$ and use property (iii).]

1.10 Give a geometric justification of the **triangle inequality**:

$$|z_1 + z_2| \leq |z_1| + |z_2|,$$

where z_1 and z_2 are any two complex numbers.

1.11 In each of the following cases, find the set of all points in the complex plane satisfying the given condition (describe the set, sketch it, and give its cartesian equation, if appropriate).

a) $\operatorname{Im} z \geq 0$

b) $0 < \operatorname{Im}(z+1) \leq 2\pi$

c) $-1 \leq \operatorname{Re} z < 1$

d) $\operatorname{Re}(iz) = 3$

e) $\operatorname{Re}(z+2) = -1$

f) $|z-5| = 6$

g) $|z+2i| \geq 1$

h) $|z+i| = |z-i|$

i) $|z+3| + |z+1| = 4$

j) $|z+3| - |z+1| = \pm 1$

1.12 If z is a variable complex number, mark clearly on an Argand diagram (i.e., on the complex plane) the regions described by

a) $\operatorname{Re} z \geq -2$ and $0 \leq \operatorname{Im} z \leq 3$

b) $\operatorname{Re} z \geq -2$ or $0 \leq \operatorname{Im} z \leq 3$

c) $2 < |z| < 3$ and $\operatorname{Re} z < 2.5$

d) $|z-2+i| > 1$ and $\operatorname{Re} z > 2$

e) $1 < |z-2+i| < 3$ and $\operatorname{Im} z \geq 0$.

CHAPTER 2

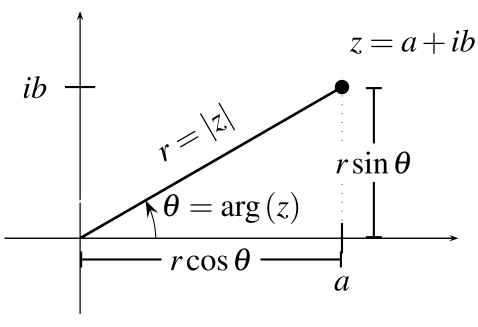
Polar Forms of Complex Numbers

In the last chapter we introduced the set of complex numbers to solve quadratic equations and showed how they can be represented as points on the complex plane using the Cartesian form $a+ib$. The polar forms introduced in this chapter simplify arithmetic calculations of multiplication and division as well as the calculation of integer powers and roots of complex numbers.

2.1 Standard Polar form

To position the complex number $z = a+bi$ in the complex plane we used the real part a and the imaginary part b as Cartesian coordinates on the plane. It is also possible to plot the same number using polar coordinates, that is,

- the distance $r = |z| = \sqrt{a^2+b^2}$ of the point from the origin, called the modulus and
- the angle θ of the line from the point to the origin, measured anti-clockwise from the positive real axis. This angle is called the argument of the complex number and denoted $\arg(z)$ as shown in the figure below:



Elementary trigonometry shows that

$$\tan\,\theta = \frac{b}{a} \quad \text{and therefore} \quad \arg(z) = \theta = \tan^{-1}\left(\frac{b}{a}\right),$$

where $\tan^{-1}$ is the inverse function of tan. It also shows that

$$a = r\cos\theta \quad \text{and} \quad b = r\sin\theta,$$

and therefore,

$$\begin{aligned} z &= a + ib \\ &= r\cos\theta + ir\sin\theta \\ &= r(\cos\theta + i\sin\theta). \end{aligned}$$

Standard polar form
The "standard polar form" of a complex number z is given by

$$z = r(\cos\theta + i\sin\theta)$$

where r is the modulus and θ its argument.

Recall that the form $z = a + ib$ introduced in the last chapter is called the "Cartesian form", in which z is specified by its real part a and its imaginary part b.

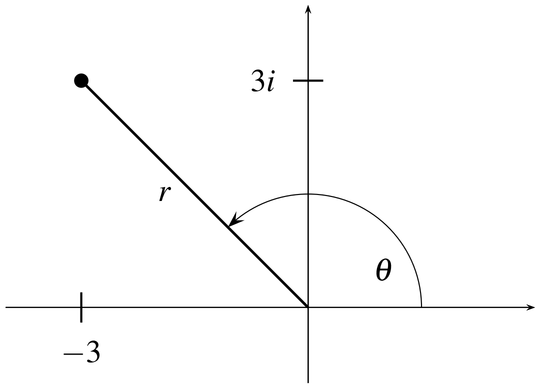
Calculating the argument θ

Always plot the complex number to find θ.
Because $\tan\theta$ has the same values in the first and third quadrants and in the second and fourth quadrants **it is essential that you plot the complex number on the complex plane when you are finding its argument**. Otherwise you may get the wrong value of θ.

For example, $z_1 = 2 + 2i$ is in the first quadrant and $z_2 = -2 - 2i$ in the third quadrant. However, $\tan\theta_1 = \dfrac{2}{2} = 1$ and $\tan\theta_2 = \dfrac{-2}{-2} = 1$ have the same value. The only way to determine the correct quadrant is to plot z_1 and z_2 in the complex plane.

Examples 2.1a

i) Write $-3 + 3i$ in polar form.
Here $r = |-3 + 3i| = \sqrt{(-3)^2 + 3^2} = \sqrt{18} = 3\sqrt{2}$.
Also, we find that $\tan\theta = \dfrac{3}{-3} = -1$. Using the calculator in radian mode, it will tell you that $\tan^{-1}(-1) \approx -0.7854$ which is $-\dfrac{\pi}{4}$.
This is not the correct angle. Plotting $-3 + 3i$ on the complex plane gives the following picture:

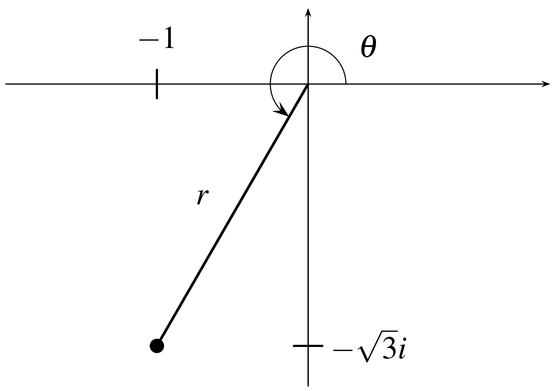


which shows that $\theta = \arg(-3+3i)$ is an angle in the second quadrant. By inspection we can see that $\arg(-3+3i) = 3\pi/4$. The diagram easily distinguishes between right and wrong answers. So the standard polar form is

$$-3+3i = 3\sqrt{2}\,(\cos 3\pi/4 + i\sin 3\pi/4)\,.$$

ii) Write $-1-\sqrt{3}i$ in polar form.

The modulus is given by $r = \sqrt{(-1)^2 + (\sqrt{3})^2} = \sqrt{1+3} = 2$. We plot $-1-\sqrt{3}i$ in the complex plane.

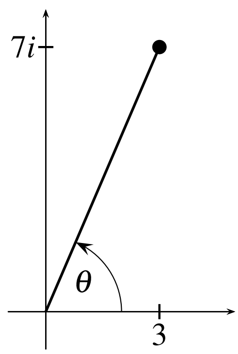


We see that $\arg(-1-\sqrt{3}i)$ lies in the third quadrant. Since $\tan\theta = \sqrt{3}$, the value of θ is $4\pi/3$. (We could also write $\theta = -2\pi/3$ equally correctly.) Therefore the standard polar form is

$$-1-\sqrt{3}i = 2(\cos 4\pi/3 + i\sin 4\pi/3).$$

iii) Find the modulus and argument of $3+7i$.

The modulus is $r = \sqrt{3^2+7^2} = \sqrt{58}$. In the complex plane $3+7i$ lies in the first quadrant.

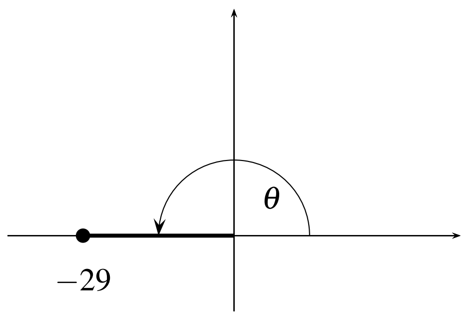


We find that $\tan\theta = \frac{7}{3}$ and so $\theta = \tan^{-1}\frac{7}{3} \approx 1.17$, therefore the polar form is

$$3+7i = \sqrt{58}\left(\cos(\tan^{-1}\frac{7}{3}) + i\sin(\tan^{-1}\frac{7}{3})\right) \approx \sqrt{58}\Big(\cos 1.17 + i\sin 1.17\Big).$$

iv) Write -29 in polar form.

Although -29 is a real number it can still be written in polar form. Clearly $|-29| = 29$ and from the complex plane we see $\arg(-29) = \pi$.

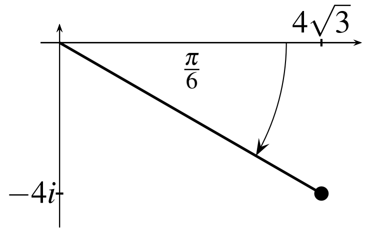


Hence $-29 = 29\Big(\cos\pi + i\sin\pi\Big)$ in polar form.

v) Convert $8\Big(\cos(-\pi/6) + i\sin(-\pi/6)\Big)$ to Cartesian form.

It is usually much simpler to convert a complex number from polar form to Cartesian form than to convert a complex number from Cartesian to polar form. All that needs to be done is to evaluate the cosine and sine and simplify the resulting expression. So

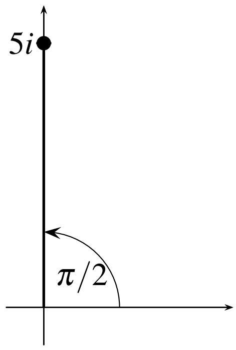
$$8\Big(\cos(-\pi/6) + i\sin(-\pi/6)\Big) = 8\left(\frac{\sqrt{3}}{2} - \frac{1}{2}i\right) = 4\sqrt{3} - 4i.$$



vi) Convert $5\Big(\cos(\pi/2)+i\sin(\pi/2)\Big)$ into Cartesian form.

Since $\cos(\pi/2)=0$ and $\sin(\pi/2)=1$, we obtain

$$5\Big(\cos(\pi/2)+i\sin(\pi/2)\Big)=5\big(0+i\big)=5i$$



$\Diamond$

Principal Argument Arg(z)

If we add 2π to the argument $\arg(z)$ of a complex number z, we come back to the same point on the complex plane. In fact, adding or subtracting any integer multiple of 2π gives the same complex number again. Therefore a complex number has an infinite number of arguments which differ by integer multiples of 2π.

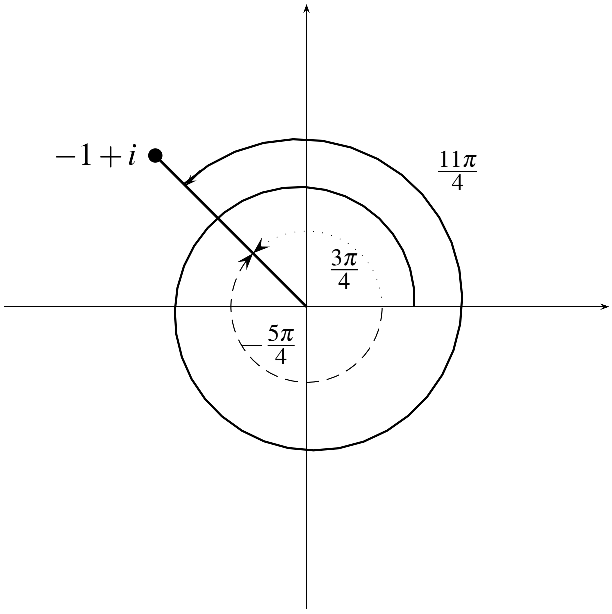
The most general form of the argument may be expressed in the form

$$\arg(z)=\theta+2k\pi \quad \text{where} \quad k\in\mathbb{Z},$$

where θ is any argument of z. This fact will become important when we take roots of complex numbers later in this chapter.

Example 2.1b Find an argument θ of the complex number $z=-1+i$ and then write the most general form of the argument.

First plot z in the complex plane as shown in the figure below.



We can choose $\theta = 3\pi/4,\ 11\pi/4,\ -5\pi/4,$ and so on. The most general argument may be expressed in any of the forms

$$\arg(z) = 3\pi/4 + 2k\pi \quad \text{where} \quad k \in \mathbb{Z}, \quad \text{or}$$

$$\arg(z) = 11\pi/4 + 2k\pi \quad \text{where} \quad k \in \mathbb{Z}, \quad \text{or}$$

$$\arg(z) = -5\pi/4 + 2k\pi \quad \text{where} \quad k \in \mathbb{Z}.$$

$\Diamond$

To eliminate the problem of having an infinite number of arguments we make the following definition.

Principal argument
The "principal argument" of z, denoted Arg (z), is the unique argument that satisfies the condition

$$-\pi < \mathrm{Arg}\,(z) \leq \pi.$$

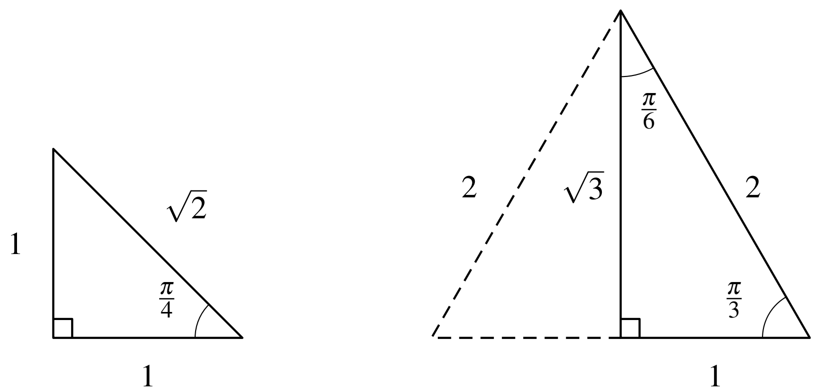
Referring to Example 2.1b we see that only the argument $\theta = 3\pi/4$ satisfies the condition $-\pi < \theta \leq \pi$ and therefore the principal argument Arg $(z) = 3\pi/4$.

A note on special angles

In the examples above you will see that most of the polar angles that we used were angles with exact sines or cosines, sometimes known as special angles. For example, any angle which is a multiple of $\pi/2$ has either sine or cosine equal to zero. So $\cos(\pi/2) = 0$, $\sin(\pi/2) = 1$ and $\cos\pi = -1$, $\sin\pi = 0$.

The angles $\pi/6$ and $\pi/3$, which correspond to 30 and 60 degrees respectively (and any angles that are multiples of these) have special values for sine and cosine, as does $\pi/4$ (45 degrees) and its multiples.

You will have learnt about these special cases at high school. As they are used extensively in this chapter, it is important that you revise them as soon as possible if you have forgotten about them. You may find it helpful to look at the right-angle triangles with angle $\pi/4$ or $\pi/3$ and $\pi/6$.



It is easy to find sines and cosines from these triangles using the basic definitions

$$\sin\theta = \frac{\text{opposite}}{\text{hypotenuse}} \qquad \cos\theta = \frac{\text{adjacent}}{\text{hypotenuse}} \qquad \tan\theta = \frac{\text{opposite}}{\text{adjacent}}.$$

In summary,,

- $\cos(\pi/2) = 0 \quad \sin(\pi/2) = 1,$
- $\cos(\pi) = -1 \quad \sin(\pi) = 0,$
- $\cos(\pi/4) = 1/\sqrt{2}, \quad \sin(\pi/4) = 1/\sqrt{2}$
- $\cos(\pi/3) = 1/2, \qquad \sin(\pi/3) = \sqrt{3}/2$
- $\cos(\pi/6) = \sqrt{3}/2, \quad \sin(\pi/6) = 1/2.$

2.2 Polar exponential form - Euler's formula

There is a very useful expression known as Euler's formula that we will prove in Chapter 8 on Taylor series. It is named after Leonhard Euler, a Swiss mathematician, physicist, astronomer, geographer, logician and engineer (1707 – 1783):

Euler's formula

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

Using this formula we can write a complex number in polar exponential form as follows:

$$z = r(\cos\theta + i\sin\theta) = re^{i\theta}.$$

Polar exponential form
The polar exponential form of a complex number $z = a + ib$ is

$$z = re^{i\theta},$$

where $r = \sqrt{a^2+b^2}$ is the modulus and $\theta = \arg(z)$ is an argument.

Examples 2.2a The Cartesian, standard polar and polar exponential forms of the complex numbers in Examples 2.1a are, respectively,

i) $-3+3i \;=\; 3\sqrt{2}\big(\cos 3\pi/4 + i\sin 3\pi/4\big) \;=\; 3\sqrt{2}\,e^{i(3\pi/4)}$

ii) $-1-\sqrt{3}i \;=\; 2(\cos 4\pi/3 + i\sin 4\pi/3) \;=\; 2\,e^{i(4\pi/3)}$

iii) $3+7i = \sqrt{58}\Big(\cos(\tan^{-1}\tfrac{7}{3}) + i\sin(\tan^{-1}\tfrac{7}{3})\Big) = \sqrt{58}\,e^{i(\tan^{-1}\frac{7}{3})}$

iv) $-29 = 29(\cos\pi + i\sin\pi) = 29\,e^{i\pi}$

v) $5\Big(\cos(\pi/2) + i\sin(\pi/2)\Big) = 5\,e^{i(\pi/2)}$ $\Diamond$

Of course, once we have the polar exponential form, the standard polar and the Cartesian forms can be written down immediatelly using Euler's formula.

2.3 Arithmetic in polar form

The polar forms are particularly useful when multiplying or dividing complex numbers, or raising a complex number to a power. However, using polar forms for addition and subtraction is more complicated and gives no extra insight into the problem. Therefore, we will not consider addition or subtraction in polar form.

Multiplication and division

Let $z = re^{i\theta}$ and $w = se^{i\phi}$ be any two non-zero complex numbers. Then

Multiplication

$$zw = re^{i\theta} \times se^{i\phi} = (rs)e^{i(\theta+\phi)}$$

Multiply the moduli and add the arguments

Division

$$\frac{z}{w} = \frac{re^{i\theta}}{se^{i\phi}} = \left(\frac{r}{s}\right)e^{i(\theta-\phi)}$$

Divide the moduli and subtract the arguments

Example 2.3a Let $z = 6e^{i(\pi/3)}$ and $w = 2e^{i(\pi/6)}$. Calculate zw and z/w.

$$zw = (6e^{i(\pi/3)})\,(2e^{i(\pi/6)}) = (6\times 2)\,e^{i(\pi/3+\pi/6)} = 12\,e^{i(\pi/2)}$$

$$z/w = 6e^{i(\pi/3)}\,/\,2e^{i(\pi/6)} = (6/2)\,e^{i(\pi/3-\pi/6)} = 3\,e^{i(\pi/6)}$$

◊

In standard polar notation the above rules become

Multiplication

$$zw = r\Big(\cos(\theta)+i\,\sin(\theta)\Big)s\Big(\cos(\phi)+i\,\sin(\phi)\Big) = (rs)\Big(\cos(\theta+\phi)+i\,\sin(\theta+\phi)\Big)$$

Division

$$\frac{z}{w} = \frac{cos(\theta)+i\,\sin(\theta)}{\cos(\phi)+i\,\sin(\phi)} = \left(\frac{r}{s}\right)\Big(\cos(\theta-\phi)+i\,\sin(\theta-\phi)\Big).$$

Notice that the polar exponential form is more intuitive and concise than the standard polar form.

Raising to an integer power

For every positive integer n, we have

$$z^n = (re^{i\theta})^n = (re^{i\theta}) \times (re^{i\theta}) \times \ldots\ldots \times (re^{i\theta}) = r^n e^{in\theta}.$$

In fact, this holds for all integers n, whether positive, negative or zero.

> **Integer power**
> $$z^n = (re^{i\theta})^n = r^n e^{in\theta}.$$
> To raise a complex number to any integer, raise the modulus to the integer and multiply the argument by the integer.

In the special case when a complex number of modulus 1 is raised to an integer power, we have "De Moivre's theorem", named after the French mathematician Abraham De Moivre (1667–1754).

> **De Moivre's theorem**
> If a complex number has modulus 1 then $\left(e^{i\theta}\right)^n = e^{in\theta}$. Using Euler's formula on both sides of the equation gives
> $$(\cos\theta + i\sin\theta)^n = \cos(n\theta) + i\,\sin(n\theta), \quad \text{for any} \quad n \in \mathbb{Z}.$$
> This last expression is called De Moivre's theorem.

Example 2.3b Use the polar exponential form to find z^8 when $z = 1 + \sqrt{3}i$. Write the final answer in Cartesian form.

We find that $|z| = 2$ and $\arg(z) = \pi/3$, so $z = 2\,e^{\,i\,\pi/3}$. Hence

$$\begin{aligned}
z^8 = \left(2\,e^{i\,\pi/3}\right)^8 &= 2^8\,e^{\,i\,8\pi/3} \\
&= 256\,e^{\,i\,8\pi/3} \\
&= 256\,e^{\,i\,2\pi/3} \quad \text{since} \quad \left(2\pi/3 = 8\pi/3 - 2\pi\right) \\
&= 256\,\left(\cos 2\pi/3 + i\sin 2\pi/3\right) \\
&= -\,128 + 128\sqrt{3}\,i.
\end{aligned}$$

◊

Notice that to write the final answer in Cartesian form, it is easier to transform the complex exponential first to the standard polar form.

Example 2.3c Use the polar exponential form to find z^2w^3 and z^2/w^3 when $z = 2\,e^{\,i\,\pi/4}$ and $w = 3\,e^{\,i\,3\pi/2}$.

We first calculate z^2 and w^3, to obtain

$$z^2 = 4\,e^{\,i\,\pi/2} \quad \text{and} \quad w^3 = 27\,e^{\,i\,9\pi/2}.$$

Therefore

$$z^2w^3 = 4\,e^{\,i\,\pi/2} \times 27\,e^{\,i\,9\pi/2} = 108\,e^{\,i\,5\pi} = 108\Big(\cos 5\pi + i\,\sin 5\pi\Big) = -108.$$

Similarly,

$$\frac{z^2}{w^3} = \frac{4\,e^{\,i\,\pi/2}}{27\,e^{\,i\,9\pi/2}} = \frac{4}{27}\,e^{\,-i\,4\pi} = \frac{4}{27}.$$

$\Diamond$

Example 2.3d Simplify $e^{\,i\,15\pi/7}$.

We have

$$\begin{aligned} e^{\,i\,15\pi/7} &= e^{\,i\,\left(\pi/7+2\pi\right)} \\ &= e^{\,i\,\pi/7} \times e^{\,i\,2\pi} \\ &= e^{\,i\,\pi/7}\,\big(\cos 2\pi + i\sin 2\pi\big) \\ &= e^{\,i\,\pi/7}\,(1 + i\,0) \\ &= e^{\,i\,\pi/7}. \end{aligned}$$

$\Diamond$

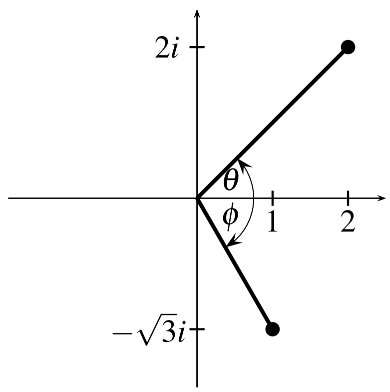
The last example demonstrates once again that arguments are only determined up to integer multiples of 2π; that is, $e^{\,i\,\theta}$ is the same as $e^{\,i\,\phi}$ when θ and ϕ differ by an integer multiple of 2π. It is useful to remember that $e^{2\pi i} = 1$ and that for every integer n, $e^{2n\pi i} = 1$.

Equality of complex numbers
If $r\,e^{\,i\,\theta} = s\,e^{\,i\,\phi}$, then $r = s$ and $\theta = \phi + 2k\pi$, for some integer k.

Examples 2.3e

i) Find $(2+2i)(1-\sqrt{3}i)$ in polar exponential and standard polar forms.

First, let us put both numbers into polar form. This simplifies the multiplication and we will also need these numbers in polar form for the next example. It is essential to draw a diagram:



Here $|2+2i| = \sqrt{4+4} = \sqrt{8} = 2\sqrt{2}$ and $|1-\sqrt{3}i| = \sqrt{1+3} = 2$. From the diagram, $\theta = \arg(2+2i)$ is in the first quadrant and $\phi = \arg(1-\sqrt{3}i)$ is in the fourth quadrant.

Since $\tan\theta = 1$, $\theta = \pi/4$ and since $\tan\phi = \sqrt{3}$, $\phi = -\pi/3$. So we have

$$\begin{aligned}
(2+2i)(1-\sqrt{3}i) &= 2\sqrt{2}\, e^{i\pi/4}\, 2e^{-i\pi/3} \\
&= e^{i\left(\pi/4+(-\pi/3)\right)} \\
&= 4\sqrt{2}\, e^{-i(\pi/12)} \\
&= 4\sqrt{2}\big(\cos(-\pi/12) + i\sin(-\pi/12)\big).
\end{aligned}$$

ii) Find $(2+2i)/(1-\sqrt{3}i)$.

The numbers $(2+2i)$ and $(1-\sqrt{3}i)$ are already in polar form from the previous example.

$$\begin{aligned}
\frac{(2+2i)}{(1-\sqrt{3}i)} &= \frac{2\sqrt{2}e^{i\pi/4}}{2e^{-i\pi/3}} \\
&= \sqrt{2}\, e^{i\left(\pi/4-(-\pi/3)\right)} \\
&= \sqrt{2}\, e^{i(7\pi/12)}.
\end{aligned}$$

iii) Find $\big((2+2i)/(1-\sqrt{3}i)\big)^6$.

The quotient has already been calculated in polar form in the previous example.

$$\begin{aligned}
\left(\frac{(2+2i)}{(1-\sqrt{3}i)}\right)^6 &= \left(\sqrt{2}\, e^{i(7\pi/12)}\right)^6 \\
&= \left(2^{\frac{1}{2}}\right)^6 e^{i(7\pi/2)} \\
&= 2^3\, e^{i(7\pi/2)}.
\end{aligned}$$

◊

2.4 Roots of complex numbers

Recall the process of finding the n^{th} root of a positive real number: we say that x is a n^{th} root of a if $x^n = a$ and we write $x = a^{1/n}$. For example, $3 = 9^{1/2}$ because $9 = 3^2$.

By analogy with roots of real numbers, the n^{th} root of a complex number w is a complex number z such that $z^n = w$ and we write $z = w^{1/n}$.

Every non–zero complex number (which includes every real number) has two complex square roots, three complex cube roots, four complex fourth roots and so on. In general:

Roots of complex numbers
Every non–zero complex number has exactly n distinct complex nth roots.

Therefore if we seek to find all cube roots of a complex number, for example, we know that there will be three of them. Knowing how many roots to look for is useful in deciding how many different values of k to use in finding the roots explicitly.

IMPORTANT – When finding roots of complex numbers, express it in polar form $z = e^{i\phi}$ first, where ϕ is any value of the argument and then add an integer multiple of 2π,

$$z = e^{\phi + 2k\pi}, \quad k \in \mathbb{Z}.$$

Then calculate

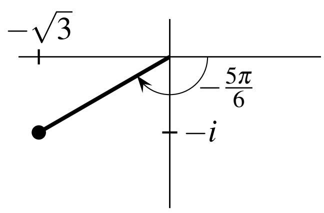
$$z^{1/n} = e^{(\phi+2k\pi)/n}.$$

The n complex roots are obtained by taking n consecutive values of k.

Example 2.4a Find all fifth roots of $-\sqrt{3} - i$, that is, all z such that $z^5 = -\sqrt{3} - i$.

First, we put $-\sqrt{3} - i$ into polar form. The modulus of $-\sqrt{3} - i$ is $|-\sqrt{3} - i| = 2$.

Plotting $-\sqrt{3} - i$ on the complex plane we see that the principal argument of $-\sqrt{3} - i$ is $-\dfrac{5\pi}{6}$.



Therefore

$$-\sqrt{3}-i = 2\,e^{-i\,\left(5\pi/6\right)} = 2\,e^{-i\,\left(5\pi/6+2k\pi\right)}.$$

Taking the fifth root, gives

$$z = 2^{1/5}\,e^{-i\,\left(5\pi/6+2k\pi\right)/5} = 2^{1/5}\,e^{-i\,\left(\pi/6+2k\pi/5\right)}.$$

For $k = 0,1,2,3,4$. we obtain the five different values of z:

$$z = 2^{1/5}\,e^{-i\left(\pi/6\right)},$$

$$z = 2^{1/5}\,e^{-i\left(17\pi/30\right)},$$

$$z = 2^{1/5}\,e^{-i\left(29\pi/30\right)},$$

$$z = 2^{1/5}\,e^{-i\left(41\pi/30\right)},$$

$$z = 2^{1/5}\,e^{-i\left(53\pi/30\right)}.$$

These are the five fifth roots of $-\sqrt{3}-i$.

$\lozenge$

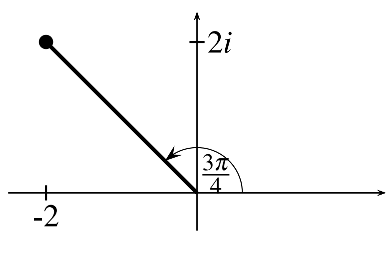
Example 2.4b Find the cubic roots of $w = -2+2i$.

Write the number in polar form: The modulus is

$$|w| = |-2+2i| = \sqrt{8}.$$

The diagram shows that the principal argument is $\theta = 3\pi/4$. However, remember that when finding roots the most general form of the argument must be used by adding integer multiples of 2π, that is,

$$\theta = 3\pi/4 + 2k\pi, \quad k \in \mathbb{Z}.$$



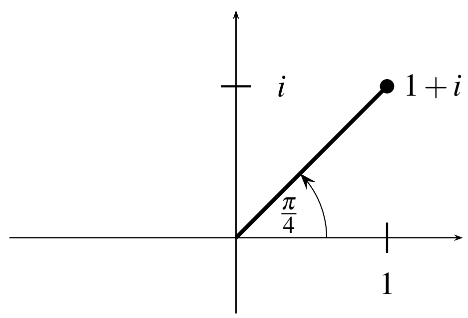
Therefore,

$$\begin{aligned} z = \left[\sqrt{8}\,e^{i(3\pi/4+2k\pi)}\right]^{1/3} &= \left(\sqrt{8}\right)^{1/3}\,e^{i\left(3\pi/4+2k\pi\right)/3} \\ &= \sqrt{2}\,e^{i\left(\pi/4+2k\pi/3\right)} \\ &= \sqrt{2}\,e^{i\phi} \quad \text{where} \quad \phi = \pi/4 + 2k\pi/3. \end{aligned}$$

Since k can be any integer it appears at first sight that there are infinitely many complex numbers z whose cube is $-2+2i$. However, close inspection shows that after three consecutive values of k we come back to the same point in the complex plane. To see this, let $k=0,1,2$. When $k=0$, we obtain

$$\begin{aligned} z &= \sqrt{2}\, e^{i\,\pi/4} \\ &= \sqrt{2}(\cos\pi/4 + i\sin\pi/4) \\ &= \sqrt{2}\Big(\frac{1}{\sqrt{2}} + i\frac{1}{\sqrt{2}}\Big) \\ &= 1+i. \end{aligned}$$

Plotting this answer on the complex plane we get:



When $k=1$, we obtain

$$\begin{aligned} z &= \sqrt{2}\, e^{i\,\left(\pi/4+2\pi/3\right)} \\ &= \sqrt{2}\, e^{i\,\left(11\pi/12\right)} \\ &= \sqrt{2}\left(\cos\frac{11\pi}{12} + i\sin\frac{11\pi}{12}\right) \end{aligned}$$

When $k=2$, we obtain

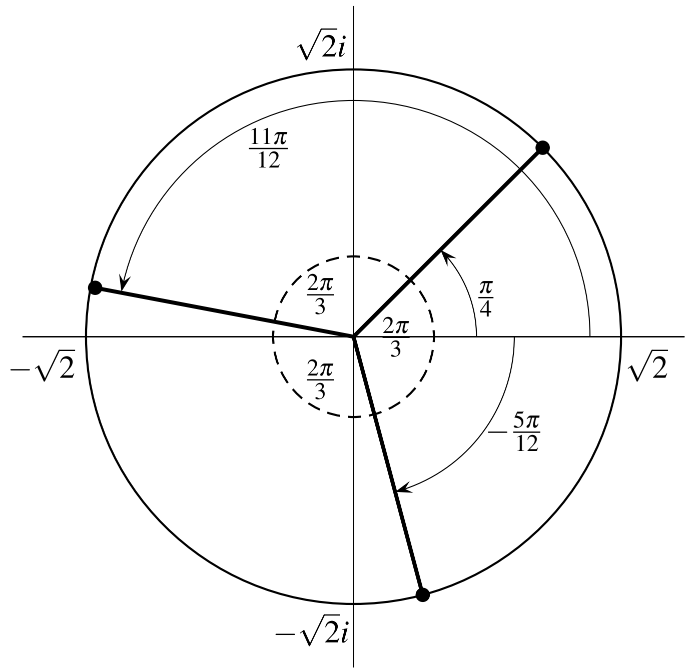
$$\begin{aligned} z &= \sqrt{2}\, e^{i\,\left(\pi/4+4\pi/3\right)} \\ &= \sqrt{2}\, e^{i\,\left(19\pi/12\right)} \\ &= \sqrt{2}\, e^{-i\,\left(5\pi/12\right)} \quad \text{(after subtracting } 2\pi\text{)}, \\ &= \sqrt{2}\left(\cos\left(\frac{-5\pi}{12}\right) + i\sin\left(\frac{-5\pi}{12}\right)\right) \end{aligned}$$

If we choose other values of k it turns out that we simply replicate one of the three values of z that we've already calculated. For example, if $k=-1$, then

$$\begin{aligned} z &= \sqrt{2}\, e^{\,i\,\left(\pi/4+(-2\pi/3)\right)} \\ &= \sqrt{2}\, e^{\,-i\,\left(5\pi/12\right)}, \end{aligned}$$

which is one of the values already found.

If all three distinct solutions are plotted on the complex plane we see that all lie on the circle of radius $\sqrt{2}$ centred on the origin, and each is separated from the others by an angle of $2\pi/3$.



$\diamondsuit$

Complex roots of real numbers

How many complex roots does a real number have? Let us look at the fourth roots of 16. You already know that $2^4 = (-2)^4 = 16$. Hence 2 and -2 are fourth roots of 16. Are there other fourth roots?

Write the number 16 in polar form: The modulus is 16 and the principal argument is 0. The most general form of the argument is $\theta = 0 + 2k\pi, \quad k \in \mathbb{Z}$, therefore

$$16 = 16\, e^{i\left(0+2k\pi\right)} \qquad \Longrightarrow \qquad z = 16^{1/4}\, e^{i\left(0/4+2k\pi/4\right)} = 2\, e^{k\pi/2}$$

where k is any integer. We shall now choose various values of k to find explicit values of z.

When $k = 0$ we obtain

$$z = 2\, e^{i\,0} = 2(\cos 0 + i \sin 0) = 2.$$

When $k = 1$ we obtain

$$z = 2\, e^{i\pi/2} = 2\left(\cos\frac{\pi}{2} + i \sin\frac{\pi}{2}\right) = 2i.$$

When $k = 2$ we obtain

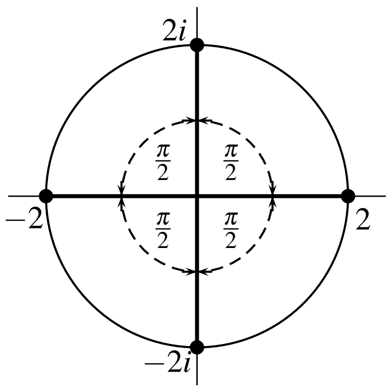
$$z = 2\, e^{i\pi} = 2(\cos\pi + i \sin\pi) = -2,$$

and when $k = 3$ we obtain

$$z = 2e^{i3\pi/2} = \left(\cos\frac{3\pi}{2} + i\sin\frac{3\pi}{2}\right) = -2i.$$

All other values of k give one of the four answers already found, namely $\pm 2, \pm 2i$. For example, if $k = 7$, we obtain $z = 2\big(\cos 7\pi/2 + i\sin 7\pi/2\big) = -2i$.

Therefore, 16 has two *real* fourth roots $\big(\pm 2\big)$ but it has four *complex* fourth roots $\big(\pm 2, \pm 2i\big)$. When these roots are plotted on the complex plane, they all lie on the circle of radius 2 centred at 0, spaced $\pi/2$ apart.



2.5 Roots of polynomial equations

In Chapter 1 we looked at the solutions of quadratic equations to motivate the introduction of complex numbers. Quadratic equations are a special case of more general equations called polynomial equations.

Polynomial expressions
A polynomial in z is an expression of the form

$$a_n z^n + a_{n-1} z^{n-1} + a_{n-2} z^{n-2} + \cdots + a_1 z + a_0$$

where z is the "variable" and the numbers $a_n, a_{n-1}, \ldots, a_0$ are the coefficients.

Polynomial equations
If we set the expression equal to zero, we obtain a polynomial equation

$$a_n z^n + a_{n-1} z^{n-1} + a_{n-2} z^{n-2} + \cdots + a_1 z + a_0 = 0.$$

If $a_n \neq 0$ then the polynomial is said to have "degree" n. The term $a_n z^n$ is known as the "leading term".

Roots of polynomial equations
The "roots of a polynomial equation" are the numbers z which satisfy

$$a_n z^n + a_{n-1} z^{n-1} + a_{n-2} z^{n-2} + \cdots + a_1 z + a_0 = 0.$$

A polynomial equation of degree n has at most n complex roots. All, some or none of these roots may be real.

For example, the problem "solve $z^4 - 16 = 0$ over the real numbers", or equivalently,

$$\text{find the real roots of } z^4 - 16 = 0$$

has the answer $z = 2$ or -2. If the problem is changed slightly to read "solve $z^4 - 16 = 0$ over the complex numbers", or

$$\text{find the complex roots of } z^4 - 16 = 0,$$

then the correct answer is $z = 2,\ 2i, -2$ or $-2i$. Clearly this polynomial equation has **two real roots** but has **four complex roots**.

Complex roots of real polynomials

Polynomial equations of degree n have at most n distinct complex roots, as some roots might be repeated. For example, the polynomial

$$z^2 - 2z + 1 = (z-1)^2$$

has a **double root** at $z = 1$, also called a root of **multiplicity** 2.

If n is a positive integer then the equation $z^n = 0$ has a root of multiplicity n at $z = 0$. It is true that every polynomial equation of degree n has *exactly* n complex roots, *counted with multiplicity*.

We have already seen in Chapter 1 that when a quadratic equation with real coefficients has non–real complex roots, then these roots come in complex conjugate pairs. So for example, the equation $z^2 + 4z + 5 = 0$ has roots $z = -2 + i$ and $z = -2 - i$.

In fact:

> **Complex roots of real polynomials**
> If the coefficients in a polynomial equation are *all real* then all of the non–real complex roots occur in complex conjugate pairs.

For example, the polynomial equation $z^4 - 16 = 0$ that was discussed earlier in this chapter has two real roots (2 and -2) and two imaginary roots ($2i$ and $-2i$), and these imaginary roots are complex conjugates of each other. The coefficients of this polynomial, 1 and -16 are both real and so we expect complex roots will occur in complex conjugate pairs.

By contrast, the polynomial $z^3 - i = 0$, does not have all real coefficients; the coefficients are 1, which is real, and $-i$, which is not real. The roots of $z^3 - i = 0$ are

$$\frac{\sqrt{3}}{2} + \frac{1}{2}i, \quad \frac{-\sqrt{3}}{2} + \frac{1}{2}i, \quad -i.$$

Although they are all non–real complex numbers, they do not occur in complex conjugate pairs.

> **Complex conjugate roots**
> If one complex root of a polynomial equation with real coefficients is known then its complex conjugate can immediately be written down to give another root.

Example 2.5a Find all roots of $z^4 - 18z^2 + 192z - 175 = 0$, given that $3 - 4i$ is a root.

Since the coefficients of the polynomial are real, if $3 - 4i$ is a root, then its complex conjugate $3 + 4i$ is also a root. We construct the quadratic expression with these two roots,

$$\big[z - (3 - 4i)\big]\big[z - (3 + 4i)\big] = z^2 - 6z + 25.$$

Next, we use polynomial long division to divide this quadratic into the original polynomial as follows,

$$\begin{array}{rrrrrr}
 & & & z^2 & +6z & -7 \\ \cline{2-6}
z^2 - 6z + 25 \; \big) & z^4 & & -18z^2 & +192z & -175 \\
 & z^4 & -6z^3 & +25z^2 & & \\ \cline{2-6}
 & & 6z^3 & -43z^2 & +192z & \\
 & & 6z^3 & -36z^2 & +150z & \\ \cline{3-6}
 & & & -7z^2 & +42z & -175 \\
 & & & -7z^2 & +42z & -175 \\ \cline{4-6}
 & & & & & 0
\end{array}$$

Therefore $\dfrac{z^4-18z^2+192z-175}{z^2-6z+25} = z^2+6z-7.$ The quotient is another quadratic which

can be easily factorized $z^2+6z-7=(z-1)(z+7)$ and therefore

$$\begin{aligned} z^4-18z^2+192z-175 &= (z^2+6z-7)(z^2-6z+25) \\ &= (z-1)(z+7)\big[z-(3-4i)\big]\big[z-(3+4i)\big]. \end{aligned}$$

So the four roots of the original degree 4 polynomial are: $1,\quad -7,\quad 3-4i,\quad 3+4i.$

$\diamondsuit$

Technical aside It is not difficult to prove rigorously that if a polynomial equation with real coefficients has complex roots then these roots occur in complex conjugate pairs.

First we need to show that $\overline{z+w}=\overline{z}+\overline{w}$ and that $\overline{zw}=\overline{z}\,\overline{w}$. Try doing this by writing $z=a+ib$ and $w=c+id$ and calculating $\overline{z}+\overline{w}$ and $\overline{z}\,\overline{w}$.

Then, let us consider a polynomial

$$p(z)=a_nz^n+a_{n-1}z^{n-1}+a_{n-2}z^{n-2}+\cdots+a_1z+a_0$$

where a_n, a_{n-1}, a_{n-2}, $\ldots$, a_0 are all real.

Suppose there is some complex number v which is a root of the polynomial, so that $p(v)=0$. If we take complex conjugates of both sides of the equation we have $\overline{p(v)}=\overline{0}=0$ and hence

$$\begin{aligned} 0 &= \overline{p(v)} \\ &= \overline{a_nv^n+a_{n-1}v^{n-1}+a_{n-2}v^{n-2}+\cdots+a_1v+a_0} \\ &= \overline{a_nv^n}+\overline{a_{n-1}v^{n-1}}+\overline{a_{n-2}v^{n-2}}+\cdots+\overline{a_1v}+\overline{a_0} \\ &= \overline{a_n}\,\overline{v^n}+\overline{a_{n-1}}\,\overline{v^{n-1}}+\overline{a_{n-2}}\,\overline{v^{n-2}}+\cdots+\overline{a_1}\,\overline{v}+a_0 \\ &= a_n\overline{v^n}+a_{n-1}\overline{v^{n-1}}+a_{n-2}\overline{v^{n-2}}+\cdots+a_1\overline{v}+a_0 \\ &= a_n(\overline{v})^n+a_{n-1}(\overline{v})^{n-1}+a_{n-2}(\overline{v})^{n-2}+\cdots+a_1(\overline{v})+a_0 \\ &= p(\overline{v}) \end{aligned}$$

Therefore since $p(\overline{v})=0$, $\overline{v}$ is a root. That is, both v and its conjugate $\overline{v}$ are roots.

As you read through the above proof try to work out why each line follows from the previous line. $\triangleleft$

2.6 Sine and cosine in terms of exponentials

Using Euler s formula, we have for all real θ,

(2.6a) $$e^{i\theta}=\cos\theta+i\sin\theta.$$

Replacing i with $-i$, gives

$$e^{-i\theta} = e^{i(-\theta)} = \cos(-\theta) + i\sin(-\theta) = \cos\theta - i\sin\theta. \tag{2.6b}$$

Adding equations (2.6a) and (2.6b) we obtain

$$e^{i\theta} + e^{-i\theta} = \cos\theta + i\sin\theta + \cos\theta - i\sin\theta = 2\cos\theta,$$

which rearranges to give $\quad \cos\theta = \dfrac{e^{i\theta} + e^{-i\theta}}{2}.$

Subtracting equations (2.6a) and (2.6b) gives

$$e^{i\theta} - e^{-i\theta} = \cos\theta + i\sin\theta - \cos\theta + i\sin\theta = 2i\sin\theta,$$

which rearranges to give $\quad \sin\theta = \dfrac{e^{i\theta} - e^{-i\theta}}{2i}.$

Sine and cosine in terms of exponentials

For all real θ, $\quad \cos\theta = \dfrac{e^{i\theta} + e^{-i\theta}}{2} \quad$ and $\quad \sin\theta = \dfrac{e^{i\theta} - e^{-i\theta}}{2i}.$

These expressions for $\cos\theta$ and $\sin\theta$ in terms of $e^{i\theta}$ will be used in Chapter 11 to derive trigonometric identities which can be used to solve integration problems involving powers of $\cos\theta$ and $\sin\theta$.

2.7 Complex exponential function

We have defined $e^{i\theta}$ as $\cos\theta + i\sin\theta$ for any $\theta \in \mathbb{R}$. This has given us a way of calculating complex exponentials where the exponent is a purely imaginary number. So, for example,

$$e^{3i} = \cos 3 + i\sin 3 \approx -0.989 + 0.141i.$$

Note that we are using radian measure, not degrees. The next step is to extend this so we can define exponentials of any complex number.

Complex exponential function

Let $z = x + iy$ be a complex number with x and y real, the complex exponential function is defined as

$$e^z = e^{x+iy} = e^x e^{iy} = e^x\big(\cos y + i\sin y\big).$$

Study this definition carefully, especially the last expression:

$$e^z = e^x(\cos y + i\sin y).$$

Notice that since x and y are real and e^x is real and positive, this displays e^z as a complex number in *polar form*. We can therefore read off the modulus and argument of e^z.

> **Modulus and argument of e^z**
> When $z = x + iy$ with x and y real, $|e^z| = e^x$ and $\arg(e^z) = y$.

The usual rules for multiplying, dividing and taking integer powers still apply. If $z = x + iy$ and $w = u + iv$ we know that $z + w = (x+u) + i(y+v)$, $z - w = (x-u) + i(y-v)$ and $nz = nx + iny$. Hence

$$e^z \times e^w = (e^x e^{iy})(e^u e^{iv}) = e^{x+u} e^{i(y+v)} = e^{z+w},$$

$$\frac{e^z}{e^w} = \frac{e^x e^{iy}}{e^u e^{iv}} = e^{x-u} e^{i(y-v)} = e^{z-w},$$

$$(e^z)^n = (e^x e^{iy})^n = (e^x)^n (e^{iy})^n = e^{nx} e^{iny} = e^{nz}, \text{ for any integer } n.$$

We can now calculate the value of e^z for any complex number z.

Example 2.7a Express the complex exponentials $e^0, e^{2+4i}, e^{-1+i\pi/4}, e^{-1+i17\pi/4}$ and e^x where x is real, as complex numbers in Cartesian form.

Using the definition of e^z, we obtain

$$e^0 = e^{0+0i} = e^0(\cos 0 + i\sin 0) = 1(1+i0) = 1,$$

$$e^{2+4i} = e^2(\cos 4 + i\sin 4) = e^2\cos 4 + ie^2\sin 4 \approx -4.83 - 5.59i,$$

$$e^{-1+i\pi/4} = e^{-1}(\cos\frac{\pi}{4} + i\sin\frac{\pi}{4}) = e^{-1}\cos\frac{\pi}{4} + ie^{-1}\sin\frac{\pi}{4} \approx 0.26 + 0.26i,$$

$$e^{-1+i17\pi/4} = e^{-1}(\cos\frac{17\pi}{4} + i\sin\frac{17\pi}{4}) = e^{-1}\cos\frac{17\pi}{4} + ie^{-1}\sin\frac{17\pi}{4} \approx 0.26 + 0.26i,$$

$$e^x = e^{x+i0} = e^x(\cos 0 + i\sin 0) = e^x(1+i0) = e^x.$$

That the third and fourth results are equal should be no surprise, since $\pi/4$ and $17\pi/4$ differ by an integer multiple of 2π. The last result shows that when z equals the real number x, the complex expression e^z agrees with the usual real exponential e^x. ◊

Example 2.7b Find a complex number z such that $e^z = 1+i$.

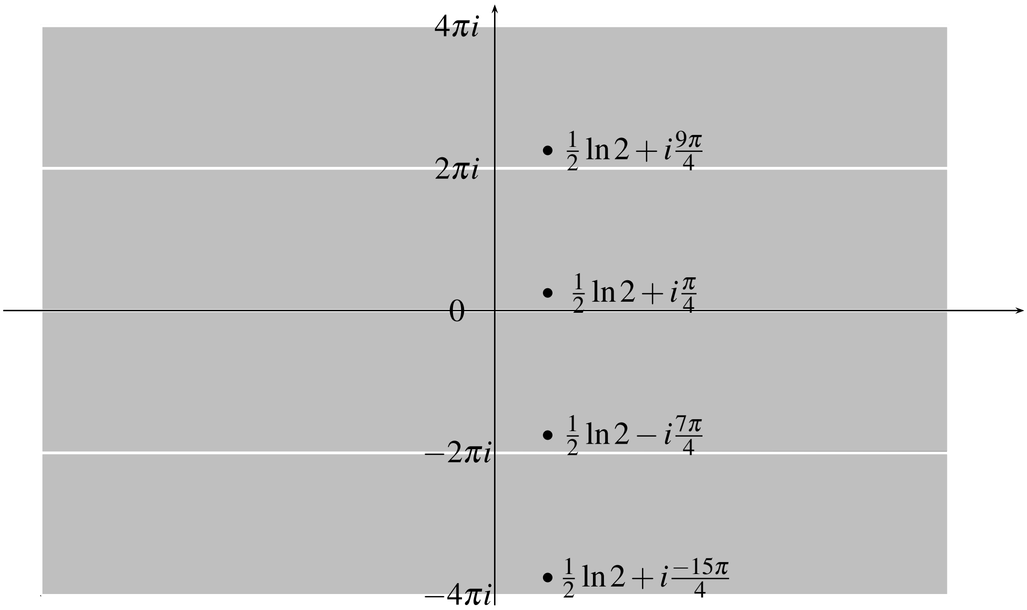
First write $1+i$ in polar exponential form. We have $1+i = \sqrt{2}e^{i\pi/4}$, and so we require $z = x+iy$ such that $e^z = e^x e^{iy} = \sqrt{2}e^{i\pi/4}$. This gives us $e^x = \sqrt{2}$, from which we find that $x = \frac{1}{2}\ln 2$, and $y = \pi/4 + 2k\pi$, where k is any integer.

We have in fact found infinitely many appropriate values of y to put together with a uniquely determined value of x. That is, there are infinitely many values of z satisfying $e^z = 1+i$. Here are some of them:

$$\left(\frac{1}{2}\ln 2\right) + i\,(\pi/4), \quad \left(\frac{1}{2}\ln 2\right) + i\,(9\pi/4), \quad \left(\frac{1}{2}\ln 2\right) + i\,(17\pi/4),$$

$$\left(\frac{1}{2}\ln 2\right) - i\,(7\pi/4), \quad \left(\frac{1}{2}\ln 2\right) - i\,(15\pi/4)$$

and so on. Some are plotted in the complex plane in the figure below. (Note that the scales on the horizontal and vertical axes are different).



◊

The previous example illustrates a most interesting property of the complex exponential function, namely its *periodicity*.

All points whose imaginary parts differ by integer multiples of 2π are mapped to the same point by the complex exponential function, because $e^{2k\pi i} = 1$ for every integer k, and hence for all $z \in \mathbb{C}$, $e^z = e^{z+2k\pi i}$ for any integer k. In particular, $e^z = e^{z+2\pi i}$ for all $z \in \mathbb{C}$, indicating that e^z is periodic with period $2\pi i$.

The complex exponential function is a periodic function, with period $2\pi i$.

An amazing formula

If we substitute $\theta = \pi$ into Euler's formula $e^{i\theta} = \cos\theta + i\,\sin\theta$, it gives

$$e^{i\pi} = \cos\pi + i\sin\pi = -1.$$

Rearranging we obtain

A remarkable formula

$$e^{i\pi} + 1 = 0.$$

This is a most remarkable expression. One equation of great simplicity contains five of the most important numbers in mathematics: e, π, i, 1 and 0. What is also remarkable is the fact that these five numbers come from very different areas of mathematics, like π from geometry, e from calculus and i from solving quadratic equations.

Summary of Chapter 2

- This chapter introduced three ways of expressing a complex number:
  a) **Cartesian form** $z = a + ib$
  b) **standard polar form** $z = r(\cos\theta + i\sin\theta)$
  c) **polar exponential form** $z = re^{i\theta}$
- **The Cartesian form** is convenient when adding or subtracting complex numbers.
- **The polar forms** greatly simplify multiplication, division and the calculation of powers and roots of complex numbers.
- **Euler's formula** $e^{i\theta} = \cos\theta + i\sin\theta$ allows trigonometric functions to be written in complex form and motivates the definition of the complex exponential function.
- **If one complex root** of a polynomial equation with real coefficients is known then its complex conjugate can immediately be written down to give another root.
- **The expressions** for $\cos\theta$ and $\sin\theta$ in terms of $e^{\pm i\theta}$ can be used to derive trigonometric identities which are helpful in solving certain types of integration problems.

Exercises

2.1 For each of the following numbers, give the numerical value of the real part x, the imaginary part y, the modulus r and the principal value of the argument θ. Plot the number as a point in the complex plane.

a) $1-i\sqrt{3}$

b) $1/(1-i)$

c) $(i+\sqrt{3})^2$

d) $2(\cos(\pi/6)+i\sin(\pi/6))$

e) $\left(\dfrac{1+i}{1-i}\right)^2$

f) $\dfrac{3+i}{2+i}$

2.2 Write each of the following complex numbers in polar form.

$$-4i, \qquad -2+2i, \qquad 1-i.$$

Use your results to perform the following operations in polar form.

a) $(-2+2i)(1-i)$

b) $-4i/(-2+2i)$

c) $(1-i)^6$

d) $(-2+2i)^{15}$

2.3 Use de Moivre's theorem to simplify

a) $\Big(\cos(2\pi/3)+i\sin(2\pi/3)\Big)^9$

b) $\Big(\cos(\pi/3)+i\sin(\pi/3)\Big)^4$

c) $\Big(\cos(2\pi/3)-i\sin(2\pi/3)\Big)^6$

d) $\Big(\sin(2\pi/3)+i\cos(2\pi/3)\Big)^9$.

2.4 Write the following complex numbers in the form $re^{i\theta}$:

a) $1+i$

b) $-\sqrt{3}-i$

c) $-2i$

d) -3

2.5 Calculate $(-\sqrt{3}-i)^{12}$.

2.6 Find $\displaystyle\int_0^{\pi/4} \sin^4\theta\, d\theta$ using the expression for $\sin^4\theta$ in terms of $\cos 4\theta$ and $\cos 2\theta$ in Example 12.1c.

2.7 Find a formula for $\cos 6\theta$ in terms of powers of $\sin\theta$ and $\cos\theta$.

2.8 Find all complex numbers z such that $e^z = i$.

2.9 Recall that if $p(z)$ is a polynomial with real coefficients and if $w \in \mathbb{C}$ is a root of $p(z)$ then so is $\overline{w}$.

Find the roots of the quadratic equation

$$q(z) = z^2 - 3(1+i)z - 2 + 6i = 0.$$

Verify that if w is a root of $q(z)$ then $\overline{w}$ is not a root. Explain why this does not contradict the statement at the start of this question.

2.10 Find all the roots of $f(z) = z^4 - 3z^3 + 7z^2 + 21z - 26$, given that $2 - 3i$ is a root.

2.11 Find all the roots of $z^4 - 5z^3 + 4z^2 + 2z - 8$, given that $1 - i$ is a root.

CHAPTER 3

Functions

The concept of function is fundamental to the study of all branches of mathematics. In this chapter we review the definition of a function and discuss some of their important properties.

3.1 Functions – definitions and examples

You would have often used the phrase "y is a function of x", or the equation $y = f(x)$, to indicate that y is a variable which depends on the variable x. If we denote the values over which x can vary as the set A, and if we observe that the values of y belong to a set B, then we say that f is a function from A to B, written $f : A \to B$.

Definition of function
If A and B are sets then a *function from A to B* (written $f : A \to B$) is a rule f which assigns to *each* element x in A *exactly one* element in B, denoted by $f(x)$.
We call $f(x)$ the *image of x under f* and we say *f maps x to $f(x)$* or *x is mapped to $f(x)$ by f*.

The rule f is sometimes a simple equation giving $f(x)$ in terms of x, such as $f(x) = x^2$. However, the rule may take other forms. In the case where A is a finite set, for example, the rule may take the form of a list of the values of $f(x)$, as x runs through A.

Domain, Codomain and Range of a function
Given a function f from A to B, $f : A \to B$, we call A the **domain** and B the **codomain** of the function.
The **range** of f is the set of values $\{f(x) \mid x \in A\}$. Thus the range is always the set B or a **proper subset** of B.

It is sometimes useful to think of a function f as a sort of machine which accepts an element from the domain as input, processes it, and produces an output. The set of all possible outputs forms the range.

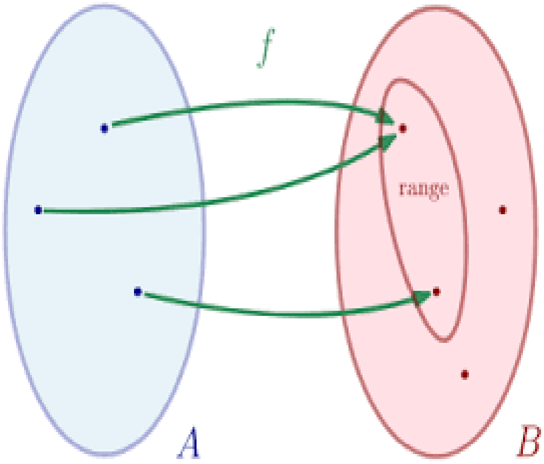
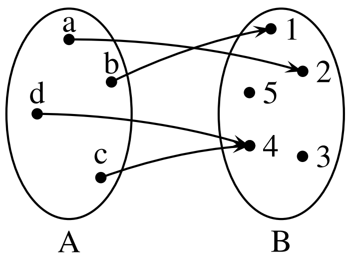


Figure 3.1: In the diagram A is the Domain and B the Codomain of $f(x)$. In this example the Range is a proper subset of the Codomain B.

Example 3.1a Let $A = \{a,b,c,d\}$, $B = \{1,2,3,4,5\}$ and consider the rule f given by the list of values

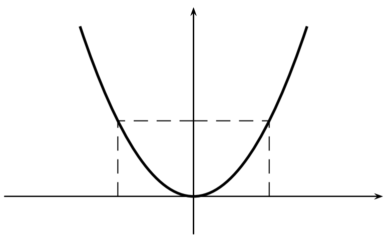
$$f(a) = 2,\ f(b) = 1,\ f(c) = 4,\ f(d) = 4.$$

Then f is a function from A to B with domain $A = \{a,b,c,d\}$, codomain $B = \{1,2,3,4,5\}$ and range $\{1,2,4\}$.



$\Diamond$

Example 3.1b Consider the function $f : \mathbb{R} \to \mathbb{R}$ given by the parabola $f(x) = x^2$. In this case the domain of f is $\mathbb{R}$, the codomain is also $\mathbb{R}$ and the range consists of the positive real numbers $f(x) \geq 0$.



$\Diamond$

Functions from $\mathbb{R}$ to $\mathbb{R}$

Many of the functions studied in elementary calculus have domain the set of real numbers $\mathbb{R}$ (or subsets of $\mathbb{R}$), and outputs which are also real numbers. These are the functions we study in this subject.

> **Natural domain**
> In circumstances when the domain is not specified explicitly, it should be assumed that the domain of f is the set of all real numbers x for which the defining formula gives a well defined, unique real number as the value for $f(x)$. Such a set of numbers is called the "natural domain" of f.

Examples 3.1c

i) Assuming that the domain is a subset of $\mathbb{R}$, the natural domain of the function given by the formula $f(x) = \dfrac{1}{x}$ is the set $\{\, x \in \mathbb{R} \mid x \neq 0 \,\}$. Zero cannot be included in the domain, since there is no real number $\dfrac{1}{0}$.

ii) In order to determine the natural domain of the function given by the formula $f(x) = \sqrt{x^2 - 4}$, where x is real, note that we must have $x^2 - 4 \geq 0$ (since we cannot take the square root of a negative number). That is, we must have $x \geq 2$ or $x \leq -2$. The natural domain is therefore $\{\, x \in \mathbb{R} \mid |x| \geq 2 \,\} = (-\infty, -2] \bigcup [2, \infty)$. $\Diamond$

Other types of functions

Not all functions have inputs and outputs in the set of real numbers $\mathbb{R}$. We have already seen some examples of functions where the domain is the set of complex numbers. Consider the following examples of functions, with their natural domains and their ranges.

Examples 3.1d

i) The function $f : \mathbb{N} \to \mathbb{N}$ given by $f(x) = x^2$ has domain the set of natural numbers. Its range is the set of all perfect squares, $\{0, 1, 4, 9, \ldots\}$.

ii) The function $f : \mathbb{C} \to \mathbb{C}$ with $f(z) = z^2$ has domain the set of complex numbers. The outputs are also complex numbers. What is the range? We're interested to find out if the set of outputs occupies the whole of $\mathbb{C}$ or if it is confined to a proper subset of $\mathbb{C}$. It turns out that the range is $\mathbb{C}$.

iii) The function $f : \mathbb{R} \to \mathbb{C}$ defined by $f(x) = \cos x + i \sin x$ is a function from the real numbers to the complex numbers. Note that the modulus of $\cos x + i \sin x$ is always equal to 1, whatever the value of x, and so the range of f is the set of complex numbers with modulus 1.

iv) The function $r : \mathbb{R} \to \mathbf{V}$ defined by $r(t) = \cos t\,\mathbf{i} + \sin t\,\mathbf{j} + t\,\mathbf{k}$ is a vector function, with domain the set of real numbers and outputs a subset of the set of vectors in 3-dimensional space. $\Diamond$

Modulus function

Considering the function $f : \mathbb{C} \to \mathbb{R}$ given by the formula $f(z) = |z|$ for all z in $\mathbb{C}$. It is easy to see that every output is a non-negative real number. The modulus of 0 is 0 and every other complex number has positive modulus, since the modulus of z measures the distance from the origin to z in the complex plane. For this function the range of f is equal to the set

$$\{x \in \mathbb{R} \mid x \geq 0\} = [0, \infty).$$

Natural domain and range of e^z

Let's now return to the complex exponential function which we defined earlier in the chapter. It is a function $f : \mathbb{C} \to \mathbb{C}$ given by the formula $f(z) = e^z$, or more usefully,

$$f(z) = f(x + iy) = e^x(\cos y + i \sin y),$$

where $z = x + iy$ in Cartesian form and e^x is just the ordinary real exponential function of the real variable x.

Recall that $|e^z| = e^x$. Therefore no matter which complex number z we select as an input, the complex number e^z has modulus which is *strictly positive, never zero*. This tells us that the complex number 0 is never an output for the complex eponential function, so the range is not equal to the whole of $\mathbb{C}$ but must be a proper subset of $\mathbb{C}$.

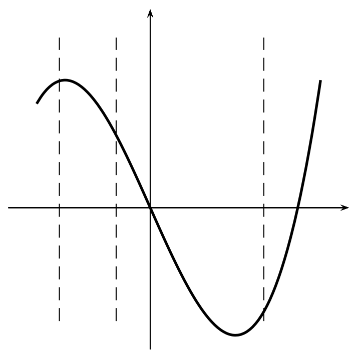
We now make use of another fact previously mentioned, namely that $\arg e^z = y$. Since y can be any real number, the outputs e^z can have any argument we please simply by selecting that value as the value of y.

Putting all this information together shows that e^z can have any positive modulus and, independently, any argument. That is, the complex exponential function has natural domain $\mathbb{C}$ and range $\{w \in \mathbb{C} \mid w \neq 0\}$.

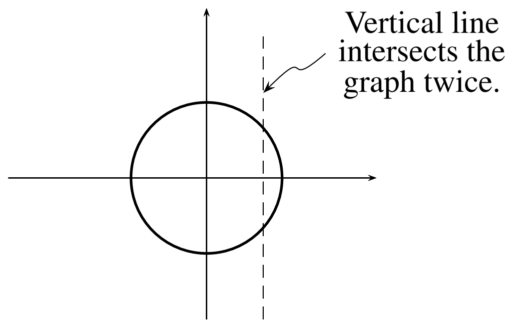
Pictures of functions

Recall that the "graph" of a function $f : \mathbb{R} \to \mathbb{R}$ is the set of points $\{\,\big(x, f(x)\big) \mid x \in \mathbb{R}\,\}$ in the ordinary Cartesian plane; we say that this is the graph $y = f(x)$. Graphical representations are possible because both the inputs and outputs of functions from $\mathbb{R}$ to $\mathbb{R}$ can be represented as points on a Cartesian plane.

> **Vertical line test**
> One special property of the graph of a functions is that every line $x = a$ parallel to the y-axis intersects the graph in exactly one point. The uniqueness of the point of intersection is a consequence of the requirement in the definition of a function that, given a in the domain of f, $f(a)$ must be unique.
> A curve which is such that some vertical line intersects it more than once is not the graph of a function.

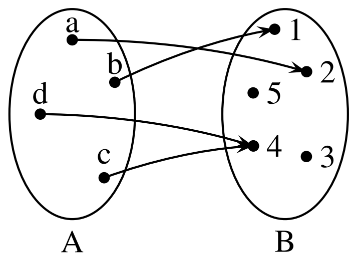


Graph of a function.



Not a function.

If A and B are finite sets, it is sometimes useful to represent a function $f : A \to B$ by an "arrow diagram". For example, the arrow diagram of the function f in Example 3.1a is:



Here, if x is in A and y is in B then we draw an arrow from the dot labelled x to the dot labelled y when $f(x) = y$. Note that there is exactly one arrow emanating from each of the dots representing the elements of A, and each arrow lands on a dot representing an element of B.

3.2 Combining functions

Suppose f and g are two functions whose ranges are contained in a set in which it is possible to add, subtract, multiply and divide. The functions f and g can then be combined in various ways to form new functions. In particular, they can be added, subtracted, multiplied or divided to form the functions $f+g$, $f-g$, fg or f/g respectively.

The functions $f+g$, $f-g$, fg have formulas $(f+g)(x) = f(x)+g(x)$, $(f-g)(x) = f(x)-g(x)$, and $(fg)(x) = f(x)\times g(x)$. These functions are defined for all values of x which lie in the domain of f and also in the domain of g. That is, if the domain of f is A and the domain of g is B, then the domain of each of $f+g$, $f-g$, fg is the intersection of A and B, namely $A\cap B$.

The function f/g is given by the formula $(f/g)(x) = \dfrac{f(x)}{g(x)}$ and is defined for all values of x which lie in the domain of f and also in the domain of g, such that $g(x) \neq 0$; that is, the domain of f/g is $\{x \in A\cap B \mid g(x) \neq 0\}$.

Composite functions

Another way to combine two functions is to allow one function to operate on the output of the other. The result is known as a composite function.

> **Composite functions**
> Given two functions f and g, the "composite function" $g\circ f$ is defined by the expression $(g\circ f)(x) = g(f(x))$ for all x in the domain of f such that $f(x)$ is in the domain of g.

Note that the output $u = f(x)$ of the function f is used as the input of the function g. If the range of f is a subset of the domain of g, then $(g\circ f)(x)$ will be defined for all x in the domain of f.

Examples 3.2a

i) Suppose $f(x) = \sqrt{x}$, and $g(x) = 1-x^2$. The natural domain of f is $\{x\in\mathbb{R} \mid x \geq 0\}$, and the natural domain of g is $\mathbb{R}$. The range of f is a subset of $\mathbb{R}$, and so the domain of the composite $g\circ f$ is equal to the domain of f.

$$(g\circ f)(x) = g(f(x)) = g(\sqrt{x}) = 1-(\sqrt{x})^2 = 1-x \text{ for } x \geq 0.$$

Note that the natural domain of the function $1-x$ is $\mathbb{R}$. Here, however, we are considering it as the composite function $g\circ f$ where $f(x) = \sqrt{x}$, and so the domain is $[0, \infty)$.

On the other hand, the range of g is $\{x\in\mathbb{R} \mid x \leq 1\}$, which is not a subset of the domain of f. The only values of x for which $f\circ g$ is defined are those such that $g(x) = 1-x^2 \geq 0$; that is, for $-1 \leq x \leq 1$.

$$(f\circ g)(x) = f(g(x)) = f(1-x^2) = \sqrt{1-x^2} \text{ for } |x| \leq 1.$$

Note that in general $f\circ g \neq g\circ f$. Note also that $g\circ f$ is quite different from the product gf of g and f. In this example, $(gf)(x) = g(x)\times f(x) = \sqrt{x}(1-x^2)$.

ii) If $f(x) = e^x$ and $g(x) = \cos x$, then $(g\circ f)(x) = \cos(e^x)$, $(f\circ g)(x) = e^{\cos x}$, and $(fg)(x) = (gf)(x) = e^x \cos x$. $\diamondsuit$

3.3 Injective and inverse functions

Here we revise the important concept of injective or ono-to-one functions. This is the crucial property of a function that guarantees the existence of an inverse.

Injective functions
A function $f : A \to B$ is said to be **injective** or **one-to-one** on the domain A if distinct elements in A are mapped to distinct elements in B; that is, if $x_1 \neq x_2$ implies that $f(x_1) \neq f(x_2)$, for all $x_1, x_2 \in A$.

Equivalently, f is injective if $f(x_1) = f(x_2)$ implies that $x_1 = x_2$, for all $x_1, x_2 \in A$.

Examples 3.3a

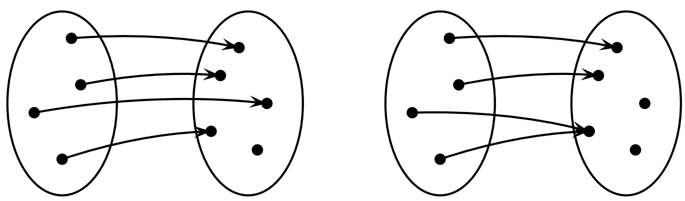
i) The function f with domain $\mathbb{R}$ defined by $f(x) = 3x+2$ is injective, since if $3x_1+2 = 3x_2+2$ then $x_1 = x_2$, for all pairs x_1, x_2.

ii) The function f with domain $\mathbb{R}$ defined by $f(x) = x^2$ is not injective since, for example, $f(-2) = f(2)$. This also shows that the function with domain $\mathbb{C}$ and formula $f(z) = z^2$ is not injective, since $\mathbb{R} \subset \mathbb{C}$.

iii) The sine function with domain $\mathbb{R}$ is not injective. If $\sin x_1 = \sin x_2$, then x_1 is not necessarily equal to x_2. However the sine function with domain $[-\frac{\pi}{2}, \frac{\pi}{2}]$ is indeed injective. (Draw the graph!) This example illustrates the fact that the property of injectivity depends on the domain we choose.

iv) The complex exponential function with domain $\mathbb{C}$ given by $f(z) = e^z$ is not injective because $e^z = e^{z+2\pi i}$.

v) The arrow diagram on the left represents an injective function, but that on the right does not.



$\Diamond$

Given the graph of a function from $\mathbb{R}$ to $\mathbb{R}$, the "horizontal line test" is a simple way to determine whether or not the function is injective.

> **The horizontal line test**
> A function $f : \mathbb{R} \to \mathbb{R}$ is injective if no horizontal line intersects its graph more than once.

Examples 3.3b

i) Any horizontal line intersects $y = x$ exactly once so this function is injective on $\mathbb{R}$.

ii) There are horizontal lines which intersect $y = x^2$ twice, and so the parabola is not an injective function on $\mathbb{R}$. $\Diamond$

The last example shows that in order to find an inverse of $y = x^2$ the domain must be restricted to some subset of $\mathbb{R}$ so that on the new domain the function is injective.

Inverse functions

Suppose that the function $f : A \to B$ has domain A and range B, and that on this domain, f is an injective function. Then we may define an inverse function $f^{-1} : B \to A$ with domain B and range A. The inverse function f^{-1} has the following property:

> **Property of the inverse function**
> $f^{-1}(f(x)) = x$ for all $x \in A$ and $f(f^{-1}(x)) = x$ for all $x \in B$. This means that
>
> $$f^{-1}(y) = x \text{ if and only if } y = f(x).$$
>
> The notation f^{-1} always means the inverse function. It does *not* denote the reciprocal $1/f(x)$.

The inverse function f^{-1} “undoes” what the function f does, and viceversa. That is, if $x \in A$ and $y \in B$, then $y = f(x)$ if and only if $x = f^{-1}(y)$. Thus to obtain a formula for the inverse function we must solve $y = f(x)$ for x, that is, make x the subject of the equation.

If f is a real valued invertible function of one real variable, the point (x,y) is on the graph of f if and only if (y,x) is on the graph of f^{-1}. This is because

$$\begin{aligned}
(x,y) \text{ is on the graph of } f \quad &\iff y = f(x)\\
&\iff x = f^{-1}(y)\\
&\iff (y,x) \text{ is on the graph of } f^{-1}.
\end{aligned}$$

Hence the graph of such a function and the graph of its inverse are always reflections of one another in the line $y = x$.

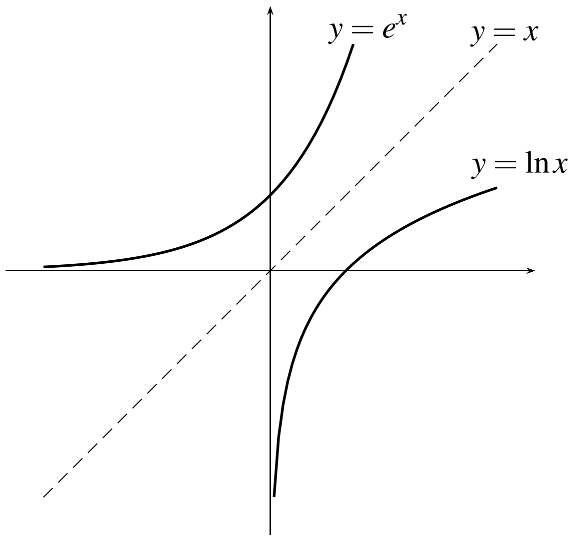
The following is a summary of the procedure for finding an inverse function formula for $f(x)$.

Calculating the inverse function of $f(x)$

- If f is not injective on its natural domain, we must restrict the domain to some subset A on which f is injective. This involves choosing a section of the graph of f where the function is strictly increasing or strictly decreasing.
- The formula for the inverse function $f^{-1} : B \to A$ is found by rearranging the equation $y = f(x)$ to make x the subject, $x = f^{-1}(y)$.
- Finally, we swap x and y to get the inverse in standard notation with x and y the independent and dependent variables, respectively.

Example 3.3c Possibly the most well-known pair of inverse functions are the exponential and logarithm functions.

The function $f : \mathbb{R} \to (0,\infty)$, given by $f(x) = e^x$, is clearly injective. Notice that with domain $\mathbb{R}$, the range of f is $(0,\infty)$. Its inverse, f^{-1} has domain $(0,\infty)$ and range $\mathbb{R}$ and is given by the rule $f^{-1}(y) = x$ if and only if $y = e^x$. That is, $f^{-1}(y) = \ln y$. The graph of the logarithm function is a reflection in the line $y = x$ of the graph of the exponential function. $\Diamond$



Example 3.3d Find the inverse of the function $f(x) = 4x - 1$.

First we note that f is injective on $\mathbb{R}$. For any x_1, x_2, if $f(x_1) = f(x_2)$ then $4x_1 - 1 = 4x_2 - 1$ and this implies that $x_1 = x_2$. The range of the function is clearly $\mathbb{R}$.

Now, make x the subject $x = \dfrac{y+1}{4}$ and therefore, $f^{-1}(y) = \dfrac{y+1}{4}$.

Now swap x and y, and the formula for f^{-1} is then written in the more customary notation using x to stand for the independent variable and y for the dependent variable,

$$y = f^{-1}(x) = \frac{x+1}{4}$$

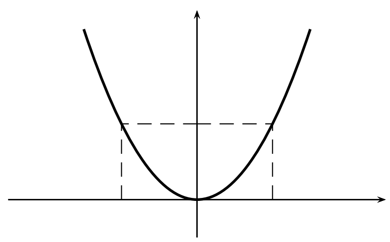
. ◊

Example 3.3e The natural domain of the function f given by $f(x) = \sqrt{e^x + 1}$ is $\mathbb{R}$, and the range is $(1, \infty)$. It is an increasing function (and hence injective). To see why, either note that it's the composite of two increasing functions, or apply the derivative test. Since $f'(x) = \dfrac{e^x}{2\sqrt{e^x+1}} > 0$ for all $x \in \mathbb{R}$, it is clear that f is increasing. So f has an inverse function $g : (1, \infty) \to \mathbb{R}$. To find the formula for g, we let $y = \sqrt{e^x + 1}$ and rearrange the equation to make x the subject, as follows:

$$\begin{aligned} y = \sqrt{e^x+1} \quad &\Rightarrow y^2 = e^x + 1 \\ &\Rightarrow e^x = y^2 - 1 \\ &\Rightarrow x = \ln\left(y^2 - 1\right). \end{aligned}$$

So $y = g(x) = \ln\left(x^2 - 1\right)$ for $x \in (1, \infty)$. ◊

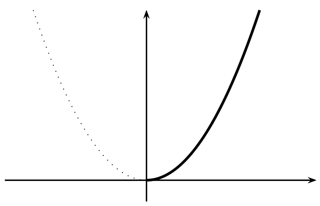
Example 3.3f The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ is not injective on its natural domain.



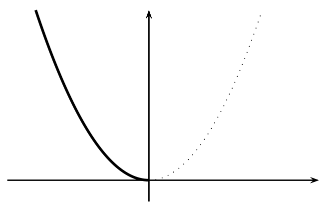
However, if we restrict the domain to some subset of $\mathbb{R}$ on which the function is injective, then we can find an inverse. For example, if $A = \{x \in \mathbb{R} \mid x \geq 0\}$, then the range of f is also A. In this case the function $f : A \to A$ has an inverse f^{-1}, also with domain and range A, defined by $f^{-1}(x) = \sqrt{x}$.

Note that it is possible to choose *any* domain over which the function is injective in order to be able to define an inverse. For example, we could take the set $S = \{x \in \mathbb{R} \mid x \leq 0\}$ as the new domain of f. The range of f is still the set A, and now the inverse function $f^{-1} : A \to S$ is defined by $f^{-1}(x) = -\sqrt{x}$.

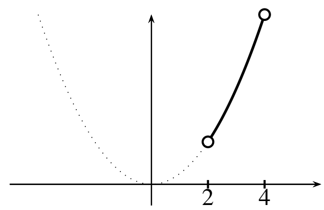
Or we could take, say, the interval $(2,4)$ as the new domain of f. This time the range is the interval $(4,16)$ and once again we would have an invertible function, namely the function $f^{-1}:(4,16)\to(2,4)$ given by $f^{-1}(x)=\sqrt{x}$.



$f{:}A\to A, f(x)=x^2$

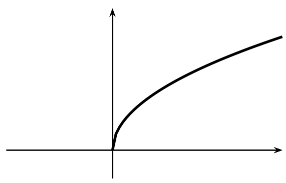


$f{:}S\to A, f(x)=x^2$

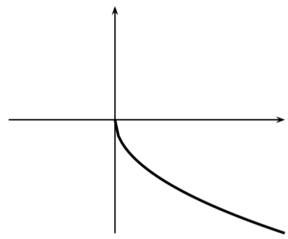


$f{:}(2,4)\to(4,16), f(x)=x^2$

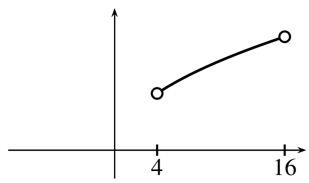
The corresponding inverse functions are shown below:



$f^{-1}{:}A\to A, f^{-1}(x)=\sqrt{x}$



$f^{-1}{:}A\to S, f^{-1}(x)=-\sqrt{x}$

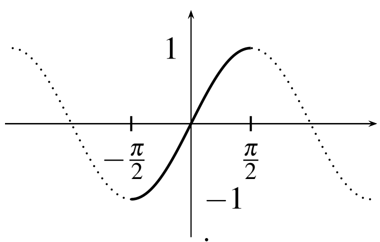


$f^{-1}{:}(4,16)\to(2,4), f^{-1}(x)=\sqrt{x}$

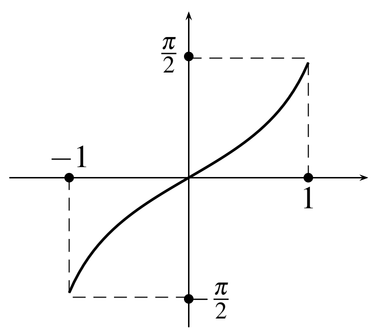
$\diamondsuit$

3.4 Inverse trigonometric functions

Example 3.4a The sine function has an inverse if we restrict its domain to the interval $[-\pi/2,\pi/2]$ over which the function is strictly increasing, the corresponding range being $[-1,1]$. Then the inverse function $\sin^{-1}$ has domain $[-1,1]$ and range $[-\pi/2,\pi/2]$. This is called the **principal value** range.



$y=\sin x$



$y=\sin^{-1}x$

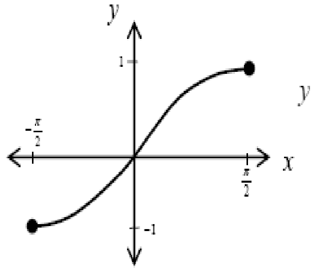
$\diamondsuit$

The table and figures below show the graphs of $\sin x$, $\cos x$ and $\tan x$ with restricted domains and their corresponding inverse functions.

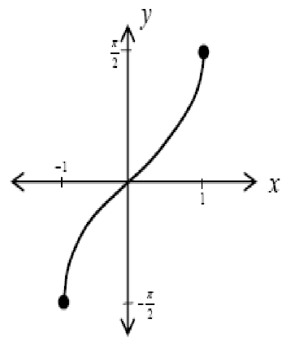
Trig function	Restricted domain	Inverse trig function	Principle value range
$y = \sin x$	$-\frac{\pi}{2} \le x \le \frac{\pi}{2}$	$y = \arcsin x$	$-\frac{\pi}{2} \le y \le \frac{\pi}{2}$
$y = \cos x$	$0 \le x \le \pi$	$y = \arccos x$	$0 \le y \le \pi$
$y = \tan x$	$-\frac{\pi}{2} < x < \frac{\pi}{2}$	$y = \arctan x$	$-\frac{\pi}{2} < y < \frac{\pi}{2}$

Graphs:

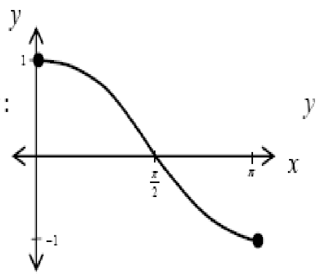
$y = \sin x$:



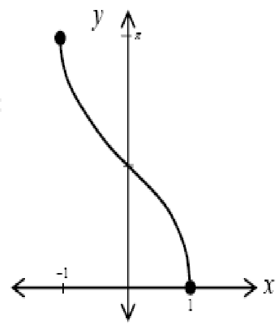
$y = \arcsin x = \sin^{-1} x$:



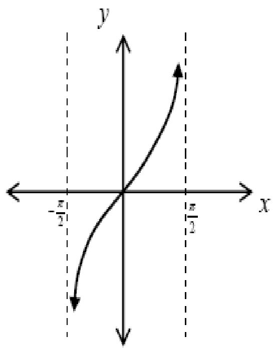
$y = \cos x$:



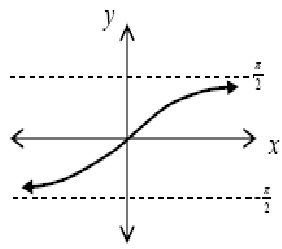
$y = \arccos x = \cos^{-1} x$:



$y = \tan x$:



$y = \arctan x = \tan^{-1} x$:



3.5 Hyperbolic functions and their inverses

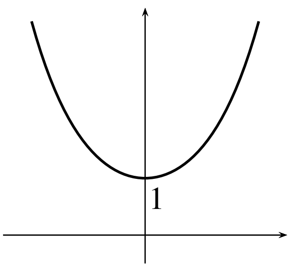
Certain combinations of the exponential functions e^x and e^{-x} occur so frequently in mathematical applications that they have been given special names. The two basic combinations are the hyperbolic sine, or "sinh" (pronounced “shine”), and the hyperbolic cosine, or "cosh" (pronounced "cosh"). They are defined on the domain $\mathbb{R}$ by the following formulas:

$$\boxed{\cosh x = \frac{e^x + e^{-x}}{2} \qquad \sinh x = \frac{e^x - e^{-x}}{2}}$$

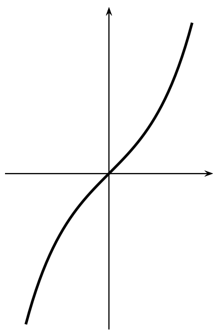
Although these functions are defined in terms of exponentials, they are called cosh and sinh because they have many properties which are similar to those of the trigonometric sine and cosine functions.

For example, note that $\cosh(-x) = \cosh x$ and therefore $\cosh x$ is an **even function** whose graph is symmetric about the y-axis as shown in the figure below, just as $\cos x$ is.

Similarly, $\sinh(-x) = -\sinh x$ and therefore $\sinh x$ is an **odd function** whose graph is skew-symmetric about the y-axis (symmetric about the origin), just as $\sin x$ is.



$y = \cosh x$



$y = \sinh x$

The graph of $\cosh x$ is called a *catenary* because it gives the shape of a chain hanging under gravity. The word catenary is derived from the Latin word for "chain." The German mathematician Joachim Jungius (1587-1657) showed that the shape of the hanging chain is $\cosh x$, disproving Galileo's claim that the curve was a parabola.

Some identities

The following identities are easily verified from the definitions.

$$\cosh^2 x - \sinh^2 x = 1$$

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

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

For example,

$$\begin{aligned}
\cosh^2 x - \sinh^2 x &= \left(\frac{e^x + e^{-x}}{2}\right)^2 - \left(\frac{e^x - e^{-x}}{2}\right)^2 \\
&= \left(\frac{e^{2x} + 2 + e^{-2x}}{4}\right) - \left(\frac{e^{2x} - 2 + e^{-2x}}{4}\right) \\
&= \frac{4}{4} = 1
\end{aligned}$$

The above identity explains the term “hyperbolic” functions and shows another similarity with sine and cosine (the “circular” functions). Squaring and adding the parametric equations of a circle $x = \cos t$ and $y = \sin t$, gives the Cartesian equation of the unit circle $x^2 + y^2 = 1$.

In contrast, squaring and adding the parametric equations $x = \cosh t$ and $y = \sinh t$, gives the Cartesian equation of the hyperbola $x^2 - y^2 = 1$.

Derivatives

Since we know that $\frac{d}{dx}(e^x) = e^x$, the derivatives of the hyperbolic functions are easily found. Again note the similarity with the trigonometric functions (note carefully the differences as well!).

$$\boxed{\frac{d}{dx}(\cosh x) = \sinh x \qquad \frac{d}{dx}(\sinh x) = \cosh x}$$

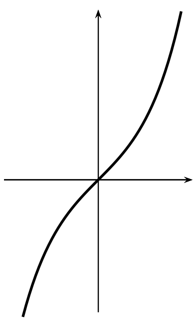
Technical aside Analogously with the trigonometric functions, we may also define

$$\tanh x = \frac{\sinh x}{\cosh x} \qquad \coth x = \frac{\cosh x}{\sinh x}$$
$$\operatorname{sech} x = \frac{1}{\cosh x} \qquad \operatorname{cosech} x = \frac{1}{\sinh x}$$

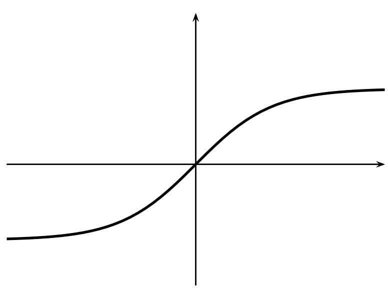
Note that $\coth x$ and $\operatorname{cosech} x$ are not defined when $x = 0$ since $\sinh 0 = 0$. $\triangleleft$

Inverse sinh function From the graph it is clear that the function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \sinh x$, is injective on $\mathbb{R}$. The function therefore has an inverse, called the "inverse hyperbolic sine function" usually written as $\sinh^{-1}$ and has domain and range equal to $\mathbb{R}$.

Note that $\sinh^{-1} y = x$ if and only if $y = \sinh x$. The graph of $\sinh^{-1} x$ is obtained by reflecting the graph of $\sinh x$ about the line $y = x$.



$y = \sinh x$



$y = \sinh^{-1} x$

To find the formula for $\sinh^{-1}(x)$, we note that if $y = \sinh x$, then

$$\begin{aligned} y = \frac{e^x - e^{-x}}{2} \quad &\Rightarrow 2y = e^x - e^{-x} \\ &\Rightarrow 2y = \frac{e^{2x} - 1}{e^x} \\ &\Rightarrow e^{2x} - 2ye^x - 1 = 0. \end{aligned}$$

This is a quadratic in e^x with solution

$$e^x = \frac{2y \pm \sqrt{4y^2 + 4}}{2} = y \pm \sqrt{y^2 + 1}.$$

Since e^x is always positive, we accept the $+$ sign and reject the $-$ sign, obtaining $e^x = y + \sqrt{y^2 + 1}$. Taking natural log on both sides gives $x = \ln(y + \sqrt{y^2 + 1})$. Finally, swapping $x \leftrightarrow y$ gives the expression

$$y = \ln(x + \sqrt{x^2 + 1}),$$

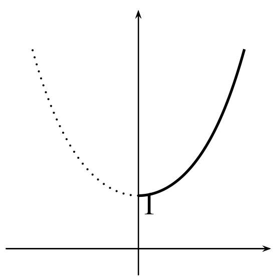
which is the formula for the inverse hyperbolic sine function

$$\sinh^{-1}(x) = \ln(x + \sqrt{x^2 + 1}).$$

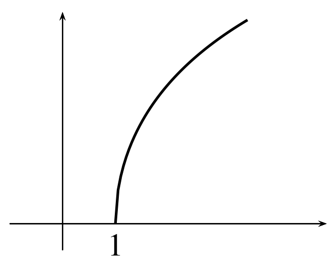
Inverse cosh function

Inspecting the graph of $\cosh x$ given earlier, we observe that the function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \cosh x$ is clearly not injective on $\mathbb{R}$. However, choosing just the right hand half of the curve (domain $[0, \infty)$) will give us an injective function whose range is $[1, \infty)$.

The inverse $f^{-1} : [1, \infty) \to [0, \infty)$ is called the "inverse hyperbolic cosine function", usually written as $\cosh^{-1}$. For $x \geq 0$, $\cosh^{-1} y = x$ if and only if $y = \cosh x$.



$y = \cosh x$



$y = \cosh^{-1} x$

As an exercise, show that with domain $[0, \infty)$ for the cosh function, the formula for $\cosh^{-1}$ is given by

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

for all $x \in [1, \infty)$. What would the formula be if we choose the domain of cosh to be $(-\infty, 0]$?

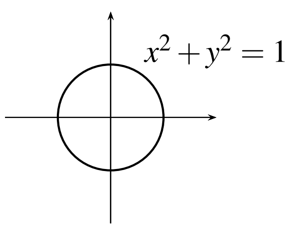
Summary of Chapter 3

- **Functions** are defined in this chapter and their **domain, codomain and range** are introduced.
- **Composite functions** (also known as **functions of a function**) are also studied and examples given to illustrate their application.
- **The natural domain** of a function f is the set of all real numbers x for which the defining formula gives a well defined, unique real number as the value for $f(x)$.
- **Injective or one-to-one functions** are defined and the horizontal line test is introduced to identify them.
- **The inverse trigonometric functions** $\sin^{-1}(x)$, $\cos^{-1}(x)$ and $\tan^{-1}(x)$ are defined by restricting the domain of $\sin x$, $\cos x$ and $\tan x$, respectively.
- **The hyperbolic functions** $\sinh x$ and $\cosh x$ are defined in terms of the exponential function and their inverses $\sinh^{-1}(x)$ and $\cosh^{-1}(x)$ are identified.

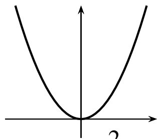
Exercises

3.1 Which of the following curves are the graphs of functions $f : A \to \mathbb{R}$?

a) $A = \{x \in \mathbb{R} \mid -1 \le x \le 1\}$

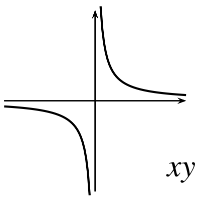


b) $A = \mathbb{R}$



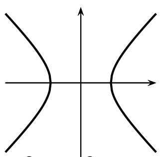
$y = x^2$

c) $A = \{x \in \mathbb{R} \mid x \neq 0\}$



$xy = 1$

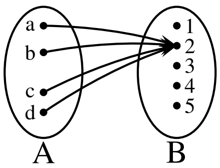
d) $A = \{x \in \mathbb{R} \mid |x| \geq 1\}$



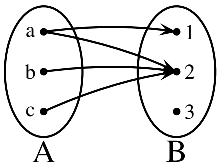
$x^2 - y^2 = 1$

3.2 Which of the arrow diagrams represent functions from A to B?

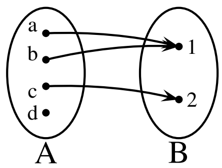
a)



b)



c)



3.3 Draw the graphs of the following functions, assuming that their domains are as large as possible.

a) $f(x) = x - 2$

b) $g(x) = \dfrac{x^2 - 4}{x + 2}$.

3.4 Find the largest subsets of $\mathbb{R}$ which are suitable domains of the following functions:

a) $f(x) = \sin(x + \pi)$

b) $f(x) = x^3 + x^{\frac{1}{3}}$

c) $f(x) = e^x$

d) $f(x) = -e^{-x}$

e) $f(x) = \sqrt{x - 2}$

f) $f(x) = \sqrt{x^2 + 4}$

g) $f(x) = \ln(x^2 + 1)$

h) $f(x) = \ln(x + 1)$

3.5 Find the ranges of the following functions $f : A \to \mathbb{R}$, where f and A are given below.

a) $f(x) = \sqrt{x+8}, \quad A = [-8, \infty).$

b) $f(x) = \sqrt{x+8}, \quad A = (0, \infty).$

c) $f(x) = \ln\left(x^2+2\right), \quad A = \mathbb{R}.$

d) $f(x) = |3\sin x|, \quad A = \mathbb{R}.$

e) $f(x) = \cos^2 x, \quad A = \left[0, \frac{\pi}{2}\right].$

f) $f(x) = x + \cos^2 x, \quad A = \mathbb{R}.$

g) $f(x) = \cos(x^2), \quad A = [-1, 1].$

h) $f(x) = \sqrt{e^x+3}, \quad A = \mathbb{R}.$

3.6 For what values of x is it possible to form the composite functions $f \circ g$ and $g \circ f$, in the following cases?

a) $f(x) = \sqrt{x}, \; g(x) = e^x$

b) $f(x) = \ln x, \; g(x) = -x^2.$

3.7 Check that $f : [-1, \infty) \to [0, \infty)$ with $f(x) = \sqrt{x+1}$ and $g : [0, \infty) \to [-1, \infty)$ with $g(x) = x^2 - 1$ are inverse to each other.

3.8 For each function $f(x)$ below, find a domain A and range B such that $f : A \to B$ has an inverse. In each case, find a formula for the inverse.

a) $f(x) = 5x + 1$

b) $f(x) = \dfrac{1}{x}$

c) $f(x) = \ln\left(e^x - 1\right)$

d) $f(x) = \sqrt{\ln x - 1}$

e) $f(x) = \sqrt{\left(\ln x\right)^2 + 5}$

f) $f(x) = e^{2x} + 1$

CHAPTER 4

Limits and Continuity

Limits are essential to the definition of derivative of a function which allows the calculation of the tangent to a curve or the instantaneous speed of a moving car. The concept of limit also appears in a natural way when attempting to calculate the area under a curve, which leads to the definition of the definite integral.

In this chapter we look at the concept of limit of a function from an intuitive point of view, what is called the **informal definition of limit,** and then study some of the basic laws that limits satisfy. The concept of limit is used to introduce the important definition of continuous functions.

An introduction to the **formal definition of limit** (not examinable here) is given in Appendix A for those interested in this topic.

4.1 Informal definition of limit

Suppose that f is a real-valued function defined at all points in an open interval $I \subseteq \mathbb{R}$ containing the point a, except possibly at a itself. In many cases we will observe that if we take values of x that are close to a but not equal to a then the values of $f(x)$ will become as close as we like to a point L. Informally, we say that $f(x)$ has limit L as x approaches a.

Informal definition of limit
We say "*the limit of a function $f(x)$, as x approaches a, is equal to L*" if we can make the values of $f(x)$ as close to L as we like by taking x sufficiently close to a, from both the left and right sides of a but **not equal** to a.
In this case we write

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

Referring to Figure 4.1, the notation $x \to a^-$ means that x approaches a from the **left** ($x < a$), and $x \to a^+$ means that x approaches a from the **right** ($x > a$).

For the limit to exist the function $f(x)$ must approach the same value L when x approaches a from the left and from the right. That is $\lim_{x\to a^-} f(x) = \lim_{x\to a^+} f(x) = L$.

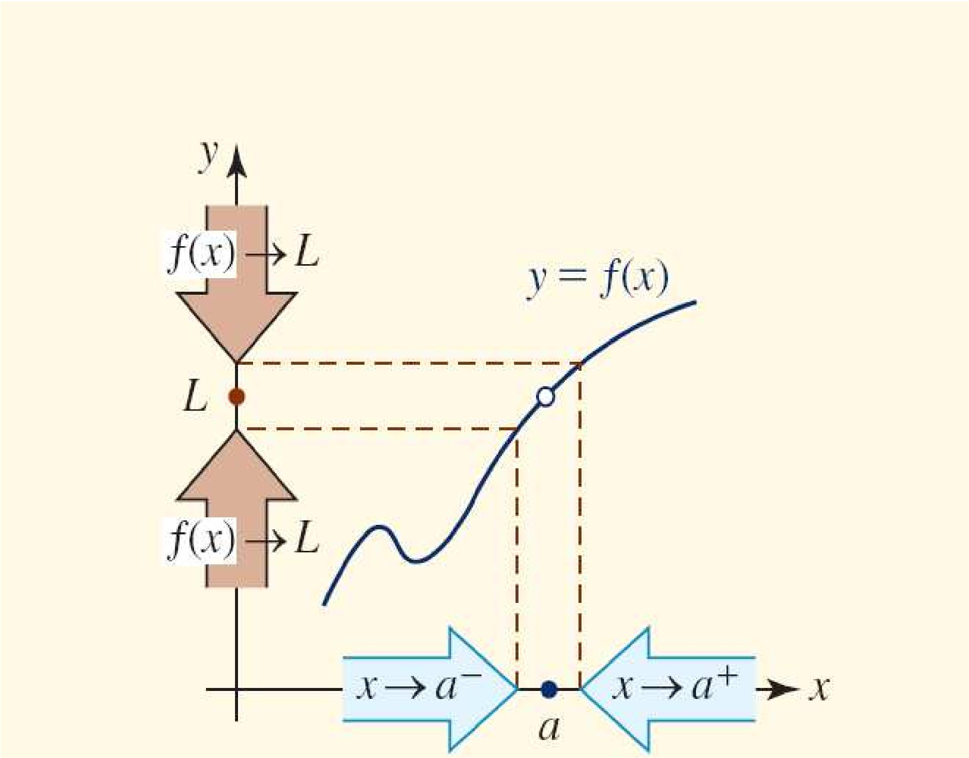


Figure 4.1: **Informal definition of limit –** The function need not be defined at $x = a$.

Example 4.1a Use this informal idea of limit to study the behaviour of the function f defined by the parabola $f(x) = x^2 + 1$ for values of x near 2.

We start by constructing two tables with values of $f(x)$ for x close to 2 from the left $x \to 2^-$ and close to 2 from the right $x \to 2^+$, but **not equal to** 2.

$x < 2$	1.5	1.8	1.95	1.99	1.995	1.999	1.9999
$f(x)$	3.3500	4.2400	4.8025	4.9601	4.9800	4.9960	4.9996

$x > 2$	2.5	2.2	2.05	2.01	2.005	2.001	2.0001
$f(x)$	7.2500	5.8400	5.2025	5.0401	5.0200	5.0040	5.0004

From the tables we see that when x is getting closer to 2 (on either side of 2) $f(x)$ is getting closer to 5. It is apparent that we can make the values of $f(x)$ as close as we like to 5 by taking x sufficiently close to 2. We express this behaviour of $f(x)$ by saying:

"The limit of the function $f(x) = x^2 + 1$ as x approaches 2 is equal to 5 and write"

$$\lim_{x \to 2} \left(x^2 + 1\right) = 5.$$

$\Diamond$

Example 4.1b Using a calculator in radian mode construct a table of values to six decimal places to guess the value of $\lim_{x\to 0} \frac{\sin x}{x}$.

The function $\frac{\sin x}{x}$ is not defined at $x = 0$. However, we only need values of $f(x)$ near $x = 0$ and not at $x = 0$ itself.

Since the values of $f(x)$ do not change by replacing x with $-x$, only one table is necessary to tabulate vales of $f(x)$ for x on either side of zero.

x	± 0.5	± 0.3	± 0.1	± 0.01
$\frac{\sin x}{x}$	0.958851	0.985067	0.998334	0.999983

From the values on the table we guess that $\lim_{x\to 0} \frac{\sin x}{x} = 1$, which is the correct answer as confirmed by the graph of $f(x)$ in Figure 4.2 below.

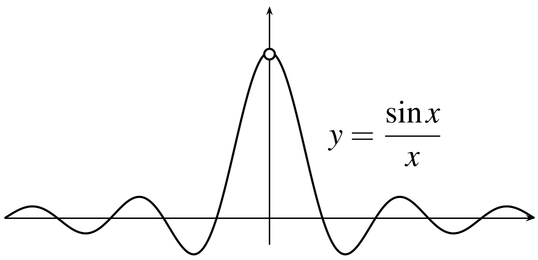


Figure 4.2:

In Appendix B we prove that $\lim_{x\to 0} \frac{\sin x}{x} = 1$ using a geometric argument. $\Diamond$

Example 4.1c Investigate $\lim_{x\to 0} \sin \frac{\pi}{x}$. Calculating the function for some small values of x, we get $f(1) = \sin \pi = 0$, $f(0.1) = \sin 10\pi = 0$, $f(0.01) = \sin 100\pi = 0$, and so on. On the basis of these results we might be tempted to guess that

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

However, this limit is wrong because the values of $\sin \frac{\pi}{x}$ oscillate between 1 and -1 infinitely many times as x approaches 0 as shown in the graph in Figure 4.3. Since the values of the function do not tend to a fixed number, the limit does not exist. $\Diamond$

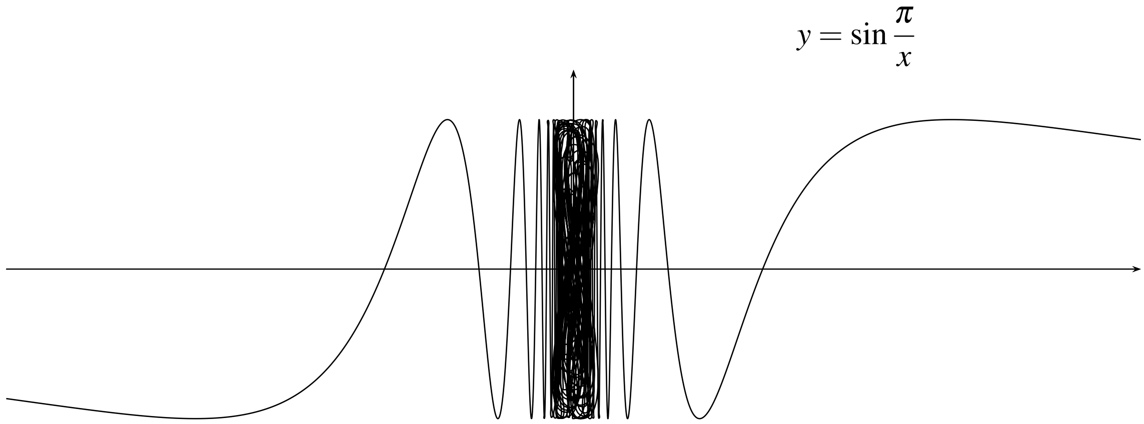
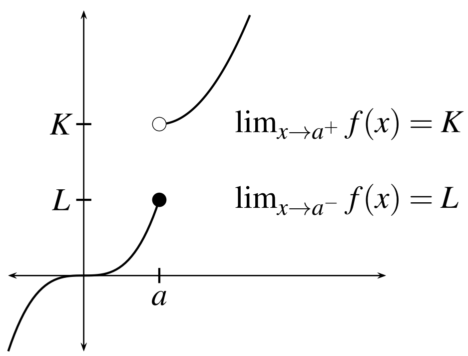


Figure 4.3:

4.2 One-sided limits

Consider the function $f(x)$ which has the following graph:



and suppose that we wanted to calculate $\lim_{x\to a} f(x)$. As we mentioned earlier we use a^- to indicate that x approaches a from the left, that is we approach a through values of x that are less than a. Similarly, a^+ is used to indicate that x approaches a from the right, that is, through values of x that are greater than a.

It is easy to see that if x approaches a along the x–axis from the *left* then $\lim_{x\to a^-} f(x) = L$; whereas, if x approaches a along the x–axis from the *right* then $\lim_{x\to a^+} f(x) = K$.

Since $L \neq K$ it follows that the ordinary limit, $\lim_{x\to a} f(x)$, does not exist. However, we can still define *one-sided limits* as follows,

Left-hand limit
We say that L is the limit of $f(x)$ as x approaches a from the left if we can make the values of $f(x)$ as close to L as we like by taking x sufficiently close to a and such that x **is less than** a. We write

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

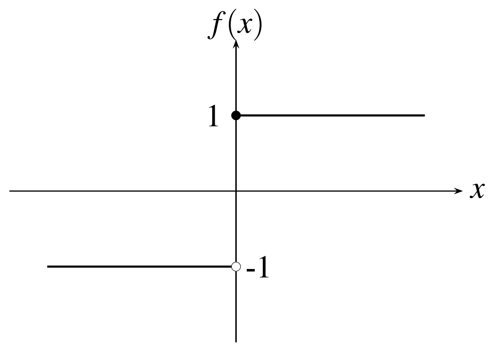
Right-hand limit
We say that K is the limit of $f(x)$ as x approaches a from the right if we can make the values of $f(x)$ as close to K as we like by taking x sufficiently close to a and such that x **is greater than** a. We write

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

Example 4.2a Consider the function

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

which has the following graph:



At the point $x = 0$ the limit of $f(x)$ as x appproaches 0 depends on the direction of approach.

As $x \to 0^+$ (through values greater than 0), the limit is 1. However, as $x \to 0^-$ (through values less than 0), the limit is -1.

Conclusion
Since the value of the limit cannot be both $+1$ and -1 we say that the limit does not exist. However, the one sided limits do exist and

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

◊

4.3 The basic limit laws

The limit laws that we introduce in this section consist of a set of rules for calculating the limit of one function in terms of the limits of simpler functions.

The sum law
Suppose that $\lim_{x\to a} f(x)$ and $\lim_{x\to a} g(x)$ both exist. Then

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

The difference law
Suppose that $\lim_{x\to a} f(x)$ and $\lim_{x\to a} g(x)$ both exist. Then

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

The product law
Suppose that $\lim_{x\to a} f(x)$ and $\lim_{x\to a} g(x)$ both exist. Then

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

The quotient law
Suppose that $\lim_{x\to a} f(x)$ and $\lim_{x\to a} g(x)$ both exist and that $\lim_{x\to a} g(x) \neq 0$. Then

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

The power law
Suppose that $\lim_{x\to a} f(x)$ exists. Then

$$\lim_{x\to a}\left[f(x)\right]^n = \left[\lim_{x\to a} f(x)\right]^n \quad \text{where } n \text{ is a positive integer.}$$

The root law
Suppose that $\lim_{x\to a} f(x)$ exists. Then

$$\lim_{x\to a}\sqrt[n]{f(x)} = \sqrt[n]{\lim_{x\to a} f(x)} \quad \text{where } n \text{ is a positive integer.}$$

If n is even, we assume that $\lim_{x\to a} f(x) > 0$.

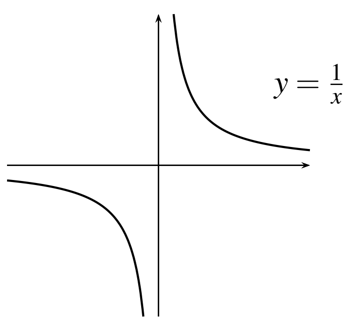
4.4 Limits at infinity – Horizontal asymptotes

The definition of $\lim_{x\to a} f(x)$ given in Section 4.1 tells us about the expected behaviour of $f(x)$ as x approaches the **finite** number a. We can also ask how $f(x)$ behaves as x becomes arbitrarily **large and positive** ($x\to\infty$) and arbitrarily **large and negative** ($x\to-\infty$).

Looking at the graph of the function $f(x)=\dfrac{1}{x}$ below, it is intuitively clear that

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

.



We observe that the curve gets closer to the x-axis (the line $y=0$) when $x\to\pm\infty$. In this case we say that the line $y=0$ is a horizontal asymptote of the curve $y=\dfrac{1}{x}$. In general, we have the following definition:

Horizontal asymptotes
The line $y = L$ is called a horizontal asymptote of the curve $y = f(x)$ if either

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

Important results

We will always assume the following basic results when calculating limits as $x \to \pm\infty$:

$$\lim_{x\to-\infty} e^x = 0, \qquad \lim_{x\to\infty} e^{-x} = 0, \qquad \lim_{x\to\infty} \frac{1}{x} = 0, \qquad \lim_{x\to-\infty} \frac{1}{x} = 0.$$

Example 4.4a Note that $\lim_{x\to\pm\infty} \sin x$ and $\lim_{x\to\pm\infty} \cos x$ do not exist. These functions oscillate between ± 1 and never settle down to a finite limit as $x \to \pm\infty$. $\diamondsuit$

Limits at infinity of rational functions

Rational functions are expressions of the form $f(x) = \dfrac{p(x)}{q(x)}$, where $p(x)$ and $q(x)$ are both polynomials. When calculating $\lim_{x\to\pm\infty} \dfrac{p(x)}{q(x)}$, the numerator and denominator both tend to infinity and we obtain what is known as an "indeterminate form" of type $\dfrac{\infty}{\infty}$. In this case we may use the following technique:

To find the limits of rational functions as $x \to \pm\infty$,
divide top and bottom by the largest power of x appearing in the denominator.

The following examples illustrate this idea.

Example 4.4b Find $\lim_{x\to-\infty} \dfrac{3x^3 - 2}{x^4 + 4x^2 - 1}$. The numerator approaches $-\infty$ and the denominator approaches ∞. Since the largest power of x appearing in the denominator is x^4, we divide top and bottom by x^4,

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

$\diamondsuit$

Example 4.4c Now consider $\lim_{x\to\infty} \dfrac{x^2+1}{2x^2-1}$. The numerator and denominator both approach ∞. Dividing top and bottom by x^2 gives

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

$\Diamond$

Example 4.4d Next, consider $\lim_{x\to-\infty} \dfrac{x^4+4x^2-1}{3x^3-2}$. The numerator approaches ∞ and the denominator approaches $-\infty$. We divide top and bottom by the largest power of x in the denominator, which is x^3.

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

$\Diamond$

4.5 Infinite limits – Vertical asymptotes

The term "infinite limit" is used to describe the situation where function values become arbitrarily large in magnitude (either positive or negative) as x approaches either a fixed point a or as $x \to \pm\infty$. Loosely speaking, we say the function "approaches infinity" or "approaches minus infinity".

Polynomial functions such as $f(x) = x^4$ or $g(x) = x^3$ approach infinity or minus infinity as $x \to \pm\infty$. Strictly speaking, limits such as

$$\lim_{x\to 0^+} \frac{1}{x}, \qquad \lim_{x\to 0^-} \frac{1}{x}, \qquad \lim_{x\to\infty} x^3$$

and so on *do not exist*, since they are not finite numbers. Nevertheless, to know whether such functions are becoming arbitrarily large in either a positive or negative direction is often extremely useful information and so we adopt an "abuse of notation" and use the symbol ∞ as if it were a number, as follows.

Infinite limits
We write $\lim_{x\to a} f(x) = \infty$ if $f(x)$ can be made to exceed any positive number we please by choosing x sufficiently close to a .

We write $\lim_{x\to\infty} f(x) = \infty$ if $f(x)$ can be made to exceed any positive number we please by choosing x sufficiently large and positive.

We write $\lim_{x\to-\infty} f(x) = \infty$ if $f(x)$ can be made to exceed any positive number we please by choosing x sufficiently large and negative.

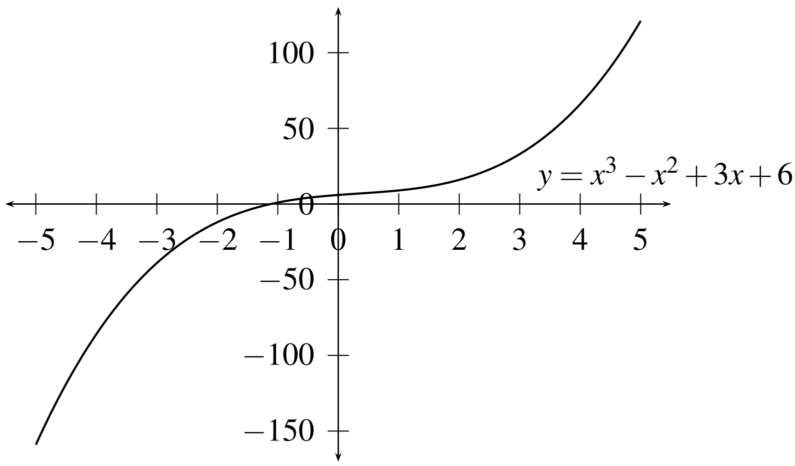
Similar statements hold when the infinite limit is $-\infty$.

A word of warning: the limit laws requires that $\lim_{x\to c} f(x)$ and $\lim_{x\to\pm\infty} f(x)$ exist, in other words, they must be finite. Consequently, *the limit laws do not apply to infinite limits*. This means that when infinite limits are involved we must exercise some care.

Example 4.5a Consider $\lim_{x\to\infty}\left(x^3 - x^2 + 3x + 6\right)$. If the sum and product rules did actually apply here then this limit would be $\infty^3 - \infty^2 + 3\infty + 6$. What is the meaning of sums and differences of infinity? Expressions such as this need careful scrutiny. The main point here is that if x is very large and positive then x^3 is much bigger than $-x^2 + 3x + 6$. Therefore for very large x, $x^3 - x^2 + 3x + 6 \approx x^3$; consequently,

$$\lim_{x\to\infty}\left(x^3 - x^2 + 3x + 6\right) = \lim_{x\to\infty} x^3 = \infty.$$

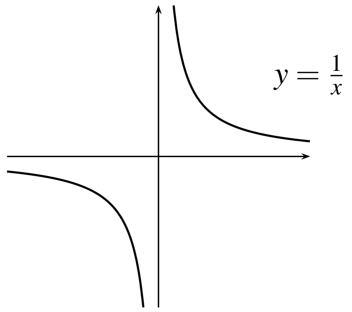
This calculation agrees with the graph of $y = x^3 - x^2 + 3x + 6$.



Similarly, $\lim_{x\to-\infty}\left(x^3 - x^2 + 3x + 6\right) = -\infty$. The term x^3 dominates all other terms and the limit as $x \to -\infty$ depends on its behaviour. $\Diamond$

Vertical asymptotes

Going back to the function $f(x) = \dfrac{1}{x}$ and its graph (reproduced below again for convenience), we see that the curve approaches $+\infty$ as x approaches 0 through positive values and $-\infty$ as x approaches 0 through negative values.



This behaviour motivates the following definition,

> **Vertical asymptotes**
> The line $x = a$ is called a vertical asymptote of the curve $y = f(x)$ if at least one of the following statements is true:
>
> $$\lim_{x\to a} f(x) = \infty \qquad \lim_{x\to a^-} f(x) = \infty \qquad \lim_{x\to a^+} f(x) = \infty$$
>
> $$\lim_{x\to a} f(x) = -\infty \qquad \lim_{x\to a^-} f(x) = -\infty \qquad \lim_{x\to a^+} f(x) = -\infty$$

Example 4.5b Consider the function $f(x) = \dfrac{x}{x-2}$. Intutively we can see that

$$\lim_{x\to 2^+} = \infty \qquad \text{and} \qquad \lim_{x\to 2^-} = -\infty,$$

and therefore the line $x = 2$ is a vertical asymptote. ♢

4.6 The squeeze law

> **The squeeze law**
> Suppose that $g(x) \le f(x) \le h(x)$, for all x near a, and that $\lim_{x\to a} g(x) = L$ and $\lim_{x\to a} h(x) = L$. Then $\lim_{x\to a} f(x)$ exists and, moreover,
>
> $$\lim_{x\to a} f(x) = L.$$

The next examples show just how useful the squeeze law is.

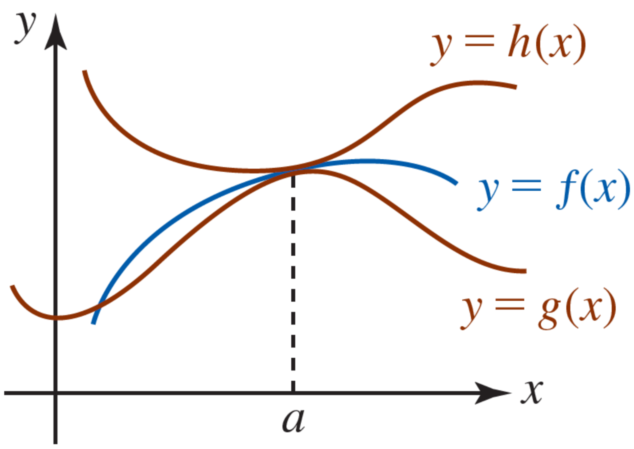


Figure 4.4: The function $f(x)$ is squeezed or sandwiched between $h(x)$ and $g(x)$. and so $\lim_{x\to a} f(x) = \lim_{x\to a} g(x) = \lim_{x\to a} h(x)$

Example 4.6a As our first application of the squeeze law, we calculate $\lim\limits_{x\to 0} x\sin\dfrac{1}{x}$. Notice that $x\sin\dfrac{1}{x}$ is not defined at $x = 0$ and therefore 0 does not belong to the domain.

Before we calculate the limit, look at the graph of $y = x\sin\dfrac{1}{x}$ below and guess what the value of $\lim\limits_{x\to 0} x\sin\dfrac{1}{x}$ should be. In order to apply the squeeze law notice that for all $x \neq 0$,

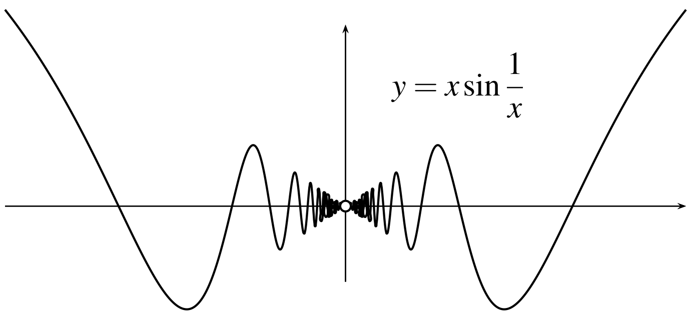


Figure 4.5:

$$\left|\sin\frac{1}{x}\right| \leq 1.$$

This inequality is our starting point. Multiplying both sides by $|x|$ gives $\left|x\sin\dfrac{1}{x}\right| \leq |x|$ or alternatively

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

Referring to Figure 4.4 we can see that in this example $f(x) = x\sin\frac{1}{x}$, $g(x) = -|x|$ and $h(x) = |x|$. Since $\lim_{x\to 0}|x| = 0$, by the squeeze law

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

which is consistent with the graph in Figure 4.5. $\Diamond$

Example 4.6b Show that $\lim\limits_{x\to\infty} \sin(e^{-x^2}) = 0$.

First we use the squeeze law to show that $\lim_{x\to\infty} e^{-x^2} = 0$. Since $x < x^2$ for all $x > 1$, it follows that $e^x < e^{x^2}$ because the exponential is an increasing function.

Therefore, taking reciprocals gives $e^{-x} > e^{-x^2} > 0$, which is valid for $x > 1$. Taking limits on this expression gives

$$\lim_{x\to\infty} e^{-x} > \lim_{x\to\infty} e^{-x^2} > 0$$

and since $\lim\limits_{x\to\infty} e^{-x} = 0$ by our previous assumptions, it follows by the squeeze law that $\lim\limits_{x\to\infty} e^{-x^2} = 0$ also.

Finally, by the composition law,

$$\lim_{x\to\infty} \sin(e^{-x^2}) = \sin(\lim_{x\to\infty} e^{-x^2}) = \sin 0 = 0.$$

$\Diamond$

4.7 Continuous and discontinuous functions

We have seen that the intuitive notion of limit of a function at a point a is the unique number L such that the function values can be made as close as we please to L by selecting domain elements sufficiently close to a, in both directions, from the right and from the left.

It is important to stress that when taking limits of a function at a point, the point in question *need not be in the domain of the function*. We are interested only in the values of the function *near the point*.

However, when the point in question *is* in the domain of the function, then the obvious question arises: if $\lim_{x\to a} f(x)$ exists, does $\lim_{x\to a} f(x) = f(a)$? In other words, is the limit equal to the value of the function at that point?

If the answer to this question is yes, we say the function is *"continuous at the point"*.

Continuous functions
A real valued function f of one real variable is said to be "continuous at the point" a in its domain if $\lim_{x\to a} f(x)$ exists and equals $f(a)$.

Functions which are continuous at every point of their domains are said to be "continuous functions".

If f is continuous at a point a this definition implies that the next three conditions are automatically satisfied:

a) $f(a)$ is defined, that is, the point a is in the domain of f.

b) $\lim_{x\to a} f(x) = L$ exists, that is, L is finite.

c) $\lim_{x\to a} f(x) = f(a)$, that is, $L = f(a)$.

If any of these conditions is not satisfied we say that f is **discontinuous** at a, or that f has a **discontinuity** at $x = a$.

Continuity of Elementary Functions

We will always assume that the elementary functions e^x, $\ln x$, $\sin x$, $\cos x$, $\tan x, x^a$ and polynomials are continuous on their natural domains. Similarly, any functions formed by adding, multiplying, dividing and composing such functions are continuous, provided we exclude division by zero.

One-sided continuity

It is also possible to define

One-sided continuity
A function f is **continuous from the right at a point** a if

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

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

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

Example 4.7d below shows a function that is continuous from the right and discontinuous from the left.

Continuity on a closed interval

Continuity on a closed interval
A function f is **continuous on the closed interval** $[a,b]$ if it is continuous at every point inside the interval and is continuous from the right at a and continuous from the left at b.

Limits of continuous functions

To calculate the limit of a continuous function we use the following property:

Direct substitution rule
Suppose we know that a function $f(x)$ is continuous at a point a and we wish to calculate $\lim_{x\to a} f(x)$. In this case all we have to do is substitute $x = a$ into the function to obtain the value $f(a)$. Then

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

Example 4.7a Calculate $\lim_{x\to 2} f(x)$, where $f(x) = \left(x^3 + 2x^2 - 4x + 1\right)$.
Since we know all polynomials are continuous functions for $x \in \mathbb{R}$, using the direct substitution rule, gives

$$\begin{aligned}\lim_{x\to 2} f(x) &= f(2)\\ &= \left(2^3 + 2\times 2^2 - 4\times 2 + 1\right)\\ &= 9.\end{aligned}$$

$\Diamond$

To calculate the limit of the composition of continuous functions we use the following law:

The composition law
Suppose that $f(x)$ is a continuous function and that $\lim_{x\to a} g(x) = L$. Then

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

Consequently, if $f(x)$ and $g(x)$ are both continuous functions then $f(g(x))$ is also continuous.

Example 4.7b Calculate $\lim_{x\to\pi/2} e^{\sin(x)}$.

Since e^x and $\sin x$ are continuous functions in $\mathbb{R}$, by the composition law, the composite function $e^{\sin(x)}$ is also continuous in $\mathbb{R}$. Therefore $\lim_{x\to\pi/2} e^{\sin(x)} = e^{\sin(\pi/2)} = e^1 = e \approx 2.718$. ♢

Example 4.7c Consider the function f given by $f(x) = \dfrac{2-\sin(\ln x)}{x^2+2}$. What is its limit as $x \to 4$?

We notice that the function is continuous at $x = 4$. Therefore, we may use the direct substitution rule:

$$\lim_{x\to 4} f(x) = f(4) = \frac{2-\sin(\ln 4)}{18}.$$

♢

Discontinuous functions

Most of the functions we will come across in this course are continuous. To see some discontinuous functions we generally have to consider functions which have gaps or jumps in their graphs. Here are a few examples to illustrate functions of one variable which are discontinuous at a point in their domain.

Example 4.7d Jump discontinuity

Consider the function

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

which has the following graph: The function has domain $\mathbb{R}$ and is continuous at each $x \in \mathbb{R}$

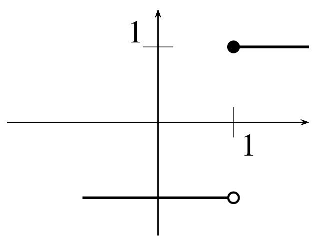


Figure 4.6:

except (intuitively) at the point $x = 1$. At this point, the limit of $f(x)$ as x appproaches 1 depends on the direction of approach. As $x \to 1^+$ (through values greater than 1), the limit is 1. As $x \to 1^-$ (through values less than 1), the limit is -1. However, the value of the limit cannot be both $+1$ and -1. Hence this limit cannot exist and so $f(x)$ is *not* continuous at $x=1$. It has a **jump discontinuity** at $x=1$.

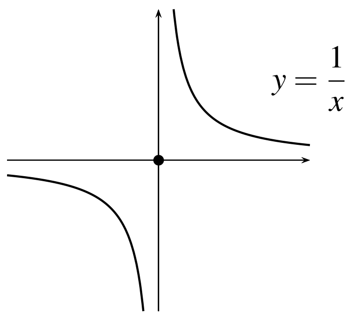
It should be clear that there is no way in which we could re-define $f(1)$ in order to make this function continuous on $\mathbb{R}$. However, it can be said that f is continuous at $x=1$ **from the right**. $\Diamond$

Example 4.7e Infinite discontinuity

Next let $f(x)$ be the function with domain $\mathbb{R}$ defined by

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

Then $y=f(x)$ has the following graph (Note the dot drawn at the origin indicating the value of $f(x)$ when $x=0$)



As x approaches 0 through positive values, $1/x$ becomes increasingly large and positive and approaches infinity. As x approaches 0 through negative values, $1/x$ become increasingly large and negative and approaches minus infinity. Thus $\lim_{x\to 0}\dfrac{1}{x}$ does not exist and $f(x)$ is said to have an **infinite discontinuity** at $x=0$.

Setting $f(0)=0$ is quite arbitrary; if we define $f(0)=a$ for any $a\in\mathbb{R}$, this function would still fail to be continuous at $x=0$ because $\lim_{x\to 0}\dfrac{1}{x}$ does not exist. $\Diamond$

Example 4.7f Removable discontinuity

Consider once again the function $f(x)=\dfrac{\sin x}{x}$ that is undefined at $x=0$ and therefore is discontinuous at $x=0$. As a matter of fact, it is an *indeterminate form* of type $\dfrac{0}{0}$. However, in Example 4.1b we showed that $\lim_{x\to 0}\dfrac{\sin x}{x}=1$. Here again is the graph of $y=\dfrac{\sin x}{x}$ showing the gap at $x=0$ where $f(x)$ is undefined. The fact that the limit exists and is finite, means that it is possible to define a new function

$$g(x)=\begin{cases}\dfrac{\sin x}{x}, & \text{if } x\neq 0,\\ 1, & \text{if } x=0,\end{cases}$$

which is continuous for all real values of x. Because we were able to eliminate the discontinuity at $x=0$, $f(x)$ is said to have a **removable discontinuity** at that point. $\Diamond$

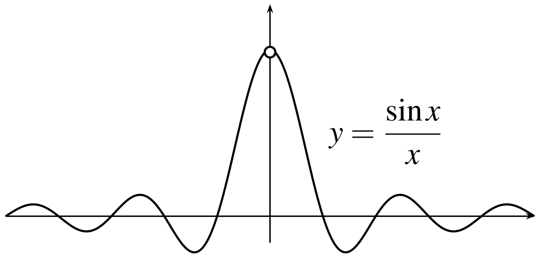


Figure 4.7:

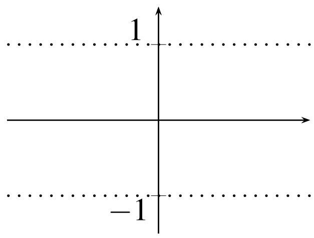
We now turn to an example of a function with domain $\mathbb{R}$ which is discontinuous at every point $x \in \mathbb{R}$.

Example 4.7g Discontinuous at every point $x \in \mathbb{R}$

Recall that $\mathbb{Q}$ is the set of rational numbers. Let

$$f(x) = \begin{cases} 1, & \text{if } x \in \mathbb{Q}, \\ -1, & \text{if } x \notin \mathbb{Q}. \end{cases}$$

The graph of $f(x)$ is a bit hard to draw because the function jumps too quickly between ± 1; however, it looks something like the following:



◊

Summary of Chapter 4

- **Limits** were introduced in a heuristic, informal manner.
  When we say that $f(x)$ has a limit L as x approaches a we are excluding the point a itself and considering only the behaviour of $f(x)$ near a, irrespective of whether the function is defined at a or not.
- **A function is continuous** at $x = a$ if $\lim_{x\to a} f(x)$ exists and is equal to the value of the function $f(a)$.
- **The direct substitution rule** says that if we want $\lim_{x\to a} f(x)$ and we know that $f(x)$ is continuous at $x = a$, then we simply substitute $x = a$ and find $f(a)$.
- **Discontinuous functions –** We looked at examples of the following types of discontinuities:
  - Jump discontinuities,
  - Infinite discontinuities,
  - Removable discontinuities, and
  - Functions that are discontinuous everywhere.

Exercises

4.1 Use continuity to calculate:

a) $\displaystyle\lim_{x\to -1} (6x^4 - 2x^3 + x^2 + x - 1)$

b) $\displaystyle\lim_{x\to -3} \frac{x^4 + x^2 - 6}{x+9}$

c) $\displaystyle\lim_{x\to 2} \frac{x^2 - 4}{x+2}$

d) $\displaystyle\lim_{x\to 0} \frac{x^3 - 2x + 1}{2x^4 - 5x + 2}$.

4.2 Calculate the following limits by cancelling out the common factor:

a) $\displaystyle\lim_{x\to 1} \frac{x^3 - 1}{x-1}$

b) $\displaystyle\lim_{x\to 3} \frac{x^3 - 27}{x-3}$

c) $\displaystyle\lim_{x\to -1} 3x + \frac{4x^2 - x}{x+2}$

d) $\displaystyle\lim_{x\to 5} \left(\frac{2x^2 - 10x}{x-5}\right)^2$

4.3 Calculate the limit of the following rational functions:

a) $\lim_{x\to\infty} \dfrac{3x^3-2x^2+4x-6}{4x^3+17x-24}$.

b) $\lim_{x\to-\infty} \dfrac{17x^2-18x+1}{4x^4-15x^3+2x^2+1}$.

c) $\lim_{x\to-\infty} \dfrac{1+x^2}{1-x^2}$.

d) $\lim_{x\to\infty} -7x^3+4x^2+2x-\dfrac{1}{x}$.

e) $\lim_{x\to 4} \dfrac{1}{x-4}$.

4.4 Use the squeeze law to calculate

a) $\lim_{x\to 0} x^2 \sin\left(\frac{1}{x}\right)$

b) $\lim_{x\to 1} |x-1| \sin\left(\dfrac{1}{x-1}\right)$

c) $\lim_{x\to 0} x\cos\left(\dfrac{2}{x}\right)$

4.5 Compute the following limits. **Hint:** let $x=\dfrac{1}{t}$.

a) $\lim_{t\to\infty} \dfrac{\sin t}{t}$.

b) $\lim_{t\to\infty} t\sin\left(\dfrac{1}{t}\right)$.

4.6 Sketch the graphs of the functions specified by the formulas. For each function, state whether $\lim_{x\to 1} f(x)$ exists, and if it does, find its value.

a) $f(x)=|x-1|$

b) $f(x)=\begin{cases} 0, & \text{when } x=1, \\ x, & \text{when } x\neq 1. \end{cases}$

4.7 Use continuity to calculate the limits:

a) $\lim_{x\to 1} (x^2+3)\cos(\pi x)$

b) $\lim_{x\to 2} e^{\sin(\pi x)}$

c) $\lim_{x\to 0} \left(e^x+3-\sqrt{x^2+1}\right)$.

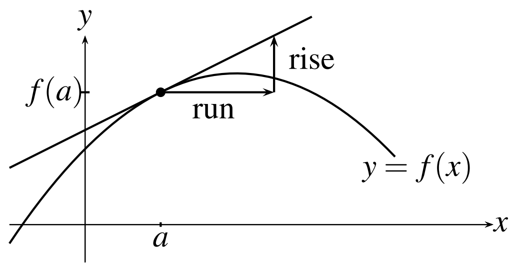
CHAPTER 5

Differentiation

In this chapter we use the concept of limit developed in Chapter 4 to define the derivative of a function. The basic limit laws studied in that chapter can be used to derive the basic rules of differentiation including the chain rule for differentiating composite functions. We also look at "implicit differentiation" which allows us to calculate the tangent line to a curve given by an implicit function.

5.1 The derivative at a point

Recall that the derivative of a function f at a point $x = a$ in its domain can be interpreted as the slope of the tangent line to the graph $y = f(x)$ at the point $(a, f(a))$. In the Cartesian plane the slope is often said to be given by "rise over run", where the run is the distance measured in the positive x direction and the rise is the distance measured in the positive y direction. This is illustrated in the diagram below.



Consider points P and Q on the graph below with coordinates $(a, f(a)$ and $(a+h, f(a+h))$, respectively. Note that h can be positive or negative – in Figure 5.1 below, we show h positive. Intuitively, the tangent line to the curve at P can be considered as the limiting position of the secant line PQ as h approaches 0. We can write down the slope of the secant PQ using the "rise over run" idea, obtaining

$$\text{Slope of } PQ = \frac{f(a+h)-f(a)}{h}.$$

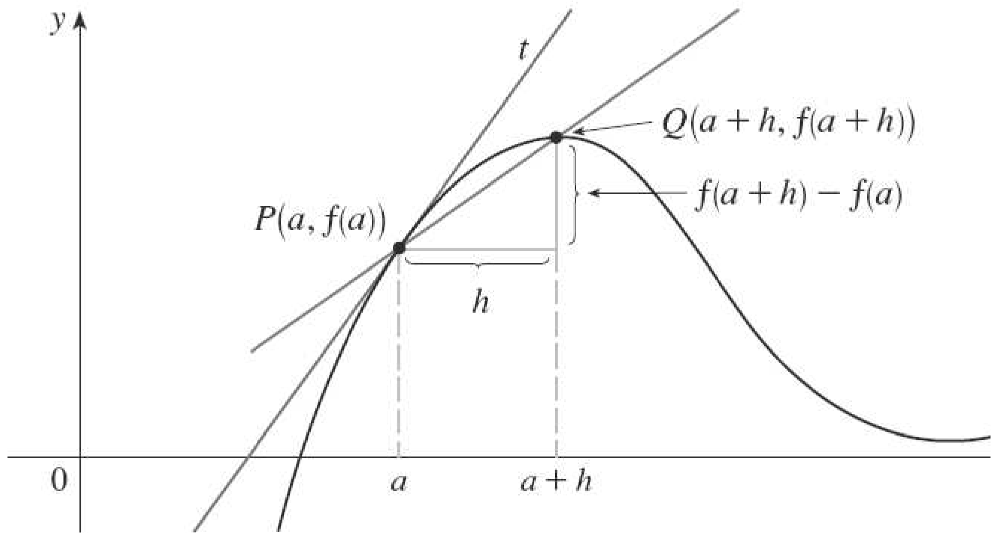


Figure 5.1:

Then the derivative of f at the point $x = a$ is defined as follows:

The derivative of a function f at a point a is

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

provided this limit exists. Note that in this definition $f'(a)$ is just a number. An alternative notation for $f'(a)$ is $\left.\frac{df}{dx}\right|_{x=a}$.

It is important to emphasize that the limit must exist. The reason is that there are functions which do not have a derivative at one or more points of their domains.

Example 5.1a The classic example of a function that does not have a derivative at a point, is the absolute value function $f(x) = |x|$ defined by the expression

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

whose graph has a V-shape with its point at the origin as shown in Figure 5.2.

This function does not have a derivative at $x = 0$, for

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

and this limit does not exist as the next example shows. $\Diamond$

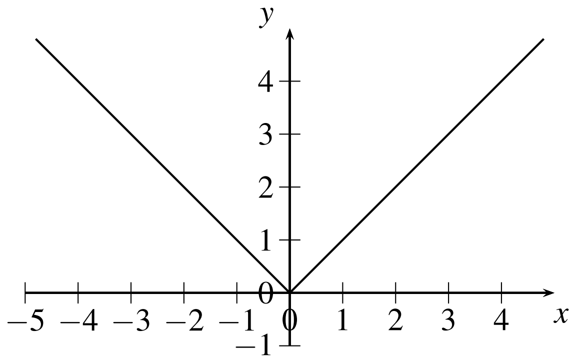


Figure 5.2:

Example 5.1b Explain informally why $\lim_{h\to 0} \dfrac{|h|}{h}$ does not exist, and hence why the absolute value function $f(x) = |x|$ does not have a derivative at $x = 0$.

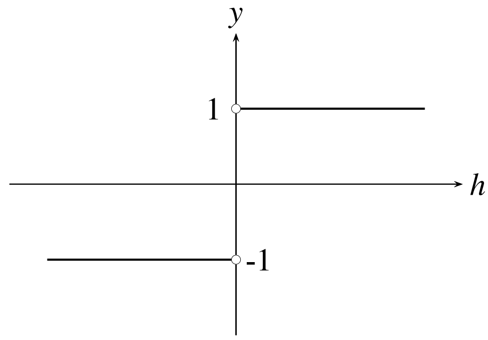
Since

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

it follows that

$$\frac{|h|}{h} = \begin{cases} 1 & \text{if } h \geq 0, \\ -1 & \text{if } h < 0, \end{cases}$$

and of course $y = \dfrac{|h|}{h}$ is not defined when $h = 0$. Its graph is shown in the diagram below, and demonstrates that there is no tendency for the function values to become close to a single number which could be regarded as the (unique) limit as h approaches 0 through both positive and negative values.



Observe that if we get closer to $x = 0$ from the right, the limit is 1, and if we approach $x = 0$ from the left, the limit is -1. Since $\lim_{x\to 0^-} \neq \lim_{x\to 0^+}$, the derivative of $f(x) = |x|$ does not exist at $x = 0$. $\diamondsuit$

5.2 The derivative as a function

In the previous section we studied the derivative of a function f at a fixed point $x = a$. The answer is a real number that represents the slope of the tangent at that point. We now change our emphasis and replace the number a by the variable x in the definition of derivative at a point:

The derivative of f as a function is

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

provided the limit exists. Note that in this definition $f'(x)$ is an actual function.

The new function f' is called the **derivative** of the function f. Its value at x can be interpreted geometrically as the slope of the tangent at the point $(x, f(x))$.

Differentiable function
If $f : I \to \mathbb{R}$ is a real-valued function of one variable and its derivative $f'(x)$ is defined for all x in the open interval I, we say f is *differentiable on I*.

Alternative notation

It is common practice to use Δx instead of h and define $\Delta y = f(x + \Delta x) - f(x)$. With this new notation the definition of the derivative of f becomes

$$f'(x) = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x},$$

provided the limit exists. The numbers Δx and Δy are called **the increments** of the variables x and y, respectively.

Example 5.2a Use the definition of derivative as a function to calculate the derivative $f'(x)$ of the function $f(x) = x^2 - x$.

Applying the formula we obtain

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

This is the same answer we would get if we used the basic rules of differentiation that we are revising below. ♢

The following important result shows how the properties of **continuity** and **differentiability** are related:

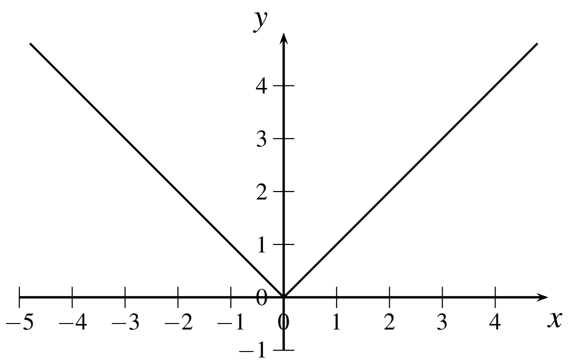
Theorem
If a function f is differentiable at a point a, then f is also continuous at a.

The converse of this theorem is not true; there are functions that are continuous at a point but not differentiable at that point.

Example 5.2b For example, the function $f(x)=|x|$ is continuous at $x=0$ because

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

However, we showed in Example 5.1b that f is not differentiable at $x=0$, showing that the converse theorem does not hold. This behaviour is shown in the plot of $f(x)=|x|$ (reproduced



here again) where we can see the cusp or corner of the curve at $x=0$. ♢

Second and higher order derivatives

If a function f is differentiable its derivative f' is also a function which may have its own derivative, denoted by $\left(f'\right)' = f''$. This new function f'' is called the **second derivative** of the original function f. An alternative notation for the second derivative of $y = f(x)$ is

$$\frac{d}{dx}\left(\frac{dy}{dx}\right) = \frac{d^2y}{dx^2}.$$

Similarly, the **third derivative** f''' is the derivative of the second derivative. In general, the n^{th} derivative of f is denoted by $f^{(n)}$ and is obtained from f by differentiating n times. If $y = f(x)$ then we write the n^{th} derivative as

$$y^{(n)} = f^{(n)}(x) = \frac{d^ny}{dx^n}.$$

Example 5.2c Find the third derivatives of the functions **a)** $f(x) = 3x^4 + x^2$ and **b)** $g(x) = \sin x$.

a) $f(x) = 3x^4 + x^2, \quad f'(x) = 12x^3 + 2x, \quad f''(x) = 36x^2 + 2, \quad f'''(x) = 72x.$

b) $g(x) = \sin x, \quad g'(x) = \cos x, \quad g''(x) = -\sin x, \quad g'''(x) = -\cos x.$ $\Diamond$

5.3 Basic rules of differentiation

Suppose that f and g are differentiable functions of one variable on the interval I. The following well–known formulas give the derivatives of kf, $f+g$, $f-g$, fg and f/g.

The constant multiple rule

$$(kf)' = kf', \quad k \text{ constant}$$

The sum/difference rule

$$(f \pm g)' = f' \pm g'$$

The product rule

$$(fg)' = f'g + fg'$$

The quotient rule

$$\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2}$$

The power rule

$$(x^a)' = a x^{a-1} \quad (a \text{ constant})$$

Even though we are not proving these results here, we should emphasize that the rules are a direct consequence of the limit laws introduced in Chapter 4.

Examples 5.3a Use the basic rules of differentiation to check the following derivatives:

i) $\dfrac{d}{dx}\left(4x^3 - 2x + 4\cos x - \dfrac{1}{e^x + x}\right) = 12x^2 - 2 - 4\sin x + \dfrac{e^x + 1}{(e^x + x)^2}$.

ii) $\dfrac{d}{dx}\left(4xe^{x^2}\right) = 4e^{x^2} + 8x^2 e^{x^2}$.

iii) $\dfrac{d}{dx}\Big(3\sin(2x+1)\Big) = 6\cos(2x+1)$.

iv) $\dfrac{d}{dx}\left(\dfrac{7x+1}{3x^2-1}\right) = -\dfrac{21x^2 + 6x + 7}{(3x^2-1)^2}$. $\diamondsuit$

5.4 The chain rule

The chain rule is used to find the derivative of a "composite function", also known as a "function of a function". Recall the definition of composite function given in Section 3.2.

Composite function
Given two functions f and g, the "composite function" $f \circ g$ is defined by the expression $(f \circ g)(x) = f(g(x))$ for all x in the domain of g such that $g(x)$ is in the domain of f.

When dealing with a composite function $f \circ g$ it is often useful to think of g as the "inside" function, and of f as the "outside" function. Using an intermediate variable u we can write a function of a function in the form

$$(f \circ g)(x) = f(g(x)) \implies y = f(u) \text{ where } u = g(x).$$

For example, the function $y = (\cos x)^3$ may be expressed in the form $y = u^3$ where $u = \cos x$. Using this notation the Chain Rule for differentiation may be written as follows:

Chain Rule

If $y = f(u)$ where $u = g(x)$ then

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

Example 5.4a Use the chain rule to find the derivatives of the functions:

a) $y = \sin(e^x)$. In this case we let $y = \sin u$ where $u = e^x$. Therefore

$$\frac{d}{dx}\sin(e^x) = \frac{dy}{du}\frac{du}{dx} = \cos(u) \times e^x = \cos(e^x) \times e^x = e^x \cos(e^x).$$

b) $y = e^{\cos x}$. In this case we let $y = e^u$ where $u = \cos x$. Therefore

$$\frac{d}{dx}e^{\cos x} = \frac{dy}{du}\frac{du}{dx} = e^u \times (-\sin x) = e^{\cos x} \times (-\sin x) = -\sin x e^{\cos x}.$$

c) $y = (x^3 - 1)^{100}$. In this case we let $y = u^{100}$ where $u = x^3 - 1$. Therefore

$$\frac{d}{dx}(x^3 - 1)^{100} = \frac{dy}{du}\frac{du}{dx} = 100u^{99} \times (3x^2) = 300x^2(x^3 - 1)^{99}$$

. $\Diamond$

Basic table of derivatives

$f(x)$	$f'(x)$
$\sin x$	$\cos x$
$\cos x$	$-\sin x$
$\tan x$	$\sec^2 x$
e^x	e^x
$\ln x$	$1/x \quad (x > 0)$
x^a	$ax^{a-1} \quad (a \text{ constant})$

5.5 Implicit differentiation

Implicit differentiation allows us to calculate the derivative of functions that are defined implicitly by a relation of the form $F(x,y) = 0$. Sometimes it is possible to solve for y and obtain an explicit function of x in the usual form $y = f(x)$ but this is not always the case. For example, consider the relations given by

$$x^2 + y^2 - 9 = 0 \qquad \text{or} \qquad x^2y^2 + x\sin y - 4 = 0.$$

The first equation represents a circle and it is possible to make y the subject to obtain the two explicit functions

$$y = \sqrt{9 - x^2} \quad \text{and} \quad y = -\sqrt{9 - x^2},$$

corresponding to the upper and lower semicircles, respectively. It is now possible to find the slope of the tangent $\frac{dy}{dx}$ using ordinary differentiation.

However, the functions defined by the second relation cannot be expressed in explicit form $y = f(x)$. You may convince yourself by trying to make y the subject as an exercise.

Nevertheless, it is still possible to calculate the derivative using the method of implicit differentiation:

> **Implicit differentiation** consists of differentiating both sides of the relation $F(x,y) = 0$ with respect to x and then solving the resulting equation for $\frac{dy}{dx}$.
> **Important**
> We must remember to use the **chain rule** everytime we differentiate a function of y with respect to x.

Example 5.5a Use implicit differentiation to find the derivative $\frac{dy}{dx}$ if $F(x,y) = 0$ is the circle $x^2 + y^2 - 9 = 0$. Then find the equation of the tangent to the circle at the point $(x,y) = (\sqrt{5}, 2)$.

Differentiate both sides of $x^2 + y^2 - 9 = 0$ with respect to x to give

$$\begin{aligned}
&\frac{d}{dx}\left(x^2 + y^2 - 9\right) = \frac{d}{dx}(0), \\
&\frac{d}{dx}\left(x^2\right) + \frac{d}{dx}\left(y^2\right) - \frac{d}{dx}(9) = \frac{d}{dx}(0), \\
&2x + 2y\frac{dy}{dx} - 0 = 0, \\
&x + y\frac{dy}{dx} = 0.
\end{aligned}$$

Making $\frac{dy}{dx}$ the subject in the last equation we obtain $\frac{dy}{dx} = -\frac{x}{y}$. Note that the derivative is both, function of x and function of y also.

The equation of the tangent line is given by $y - y_0 = m(x - x_0)$, where $(x_0, y_0) = (\sqrt{5}, 2)$ and the slope $m = \dfrac{dy}{dx} = -\dfrac{x}{y} = -\dfrac{\sqrt{5}}{2}$. Therefore the equation of the tangent is

$$y - 2 = -\frac{\sqrt{5}}{2}(x - \sqrt{5}).$$

◊

In the next example we simplify the notation and use y' instead of $\dfrac{dy}{dx}$ for the derivative.

Example 5.5b Calculate y' if $\sin\big(x+y\big) = y^2 \cos x$.

Differentiating implicitly both sides with respect to x and remembering to use the product and chain rules, we have

$$\cos\big(x+y\big) \times \big(x+y\big)' = 2yy' \cos x - y^2 \sin x,$$
$$\cos\big(x+y\big) \times \big(1+y'\big) = 2yy' \cos x - y^2 \sin x.$$

Expanding and solving for y', gives

$$y' = \frac{y^2 \sin x + \cos\big(x+y\big)}{2y\cos x - \cos\big(x+y\big)}.$$

◊

Logarithmic differentiation

The calculation of derivatives of complicated functions involving products, quotients or powers can sometimes be simplified by taking the natural logarithm of both sides of the equation and then using implicit differentiation.

Example 5.5c Differentiate the function $y = \left(\dfrac{x^3+1}{x^{7/9}}\right)^{1/4}$.

First take the natural logarithm of both sides:

$$\ln y = \ln\left(\frac{x^3+1}{x^{7/9}}\right)^{1/4} = \frac{1}{4}\ln\left(\frac{x^3+1}{x^{7/9}}\right)$$
$$= \frac{1}{4}\ln(x^3+1) - \frac{7}{36}\ln x.$$

We now differentiate implicitly:

$$\frac{1}{y}\frac{dy}{dx} = \frac{3x^2}{4(x^3+1)} - \frac{7}{36x}$$

or

$$\frac{dy}{dx} = y\left[\frac{3x^2}{4(x^3+1)} - \frac{7}{36x}\right] = \left(\frac{x^3+1}{x^{7/9}}\right)^{1/4}\left[\frac{3x^2}{4(x^3+1)} - \frac{7}{36x}\right].$$

◊

Example 5.5d Differentiate the function $y = x^x$.

Taking natural logarithm of both sides gives

$$\ln y = x \ln x$$

and differentiating implicitly:

$$\frac{1}{y}\frac{dy}{dx} = x \times \frac{1}{x} + \ln x = 1 + \ln x \quad \Longrightarrow \quad \frac{dy}{dx} = y(1 + \ln x) = x^x(1 + \ln x).$$

$\Diamond$

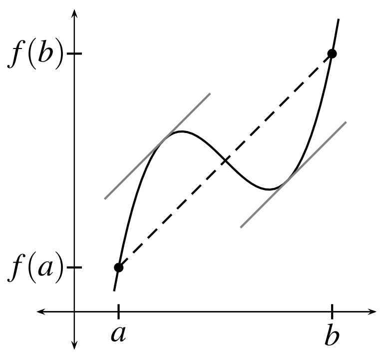
5.6 The Mean Value Theorem

The Mean Value Theorem is not examinable but we will use it to prove the Lagrange form of the remainder of Taylor series in Chapter 7 and to prove the Fundamental Theorem of Calculus in Chapter 10.

Theorem 5.6a (The Mean Value Theorem) *Suppose that f is differentiable on some interval I containing the points a and b. Then there exists a number c in the open interval (a,b) such that*

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

In other words, the Mean Value Theorem guarantees that there is at least one point $c \in (a,b)$ for which the tangent line to f at $x = c$ has the same slope as the slope of the secant line joining points $(a, f(a))$ and $(b, f(b))$. The following picture illustrates this idea.



The "dashed" line is the secant line joining points $(a, f(a))$ and $(b, f(b))$. For the particular function drawn above, there are two points c between a and b where the tangent line has the same slope as the secant – these tangent lines are also drawn. If you think about it a bit you'll understand why we need the function to be differentiable and hence continuous.

Summary of Chapter 5

- **The derivative** of a function is defined in terms of limits and an example of calculation from first principles is given for a simple function to illustrate the definition.
- **The basic rules of differentiation** were introduced along with examples that show that the calculation of derivatives is greately simplified by these rules.
- **The chain rule** is used to calculate derivatives of composite functions (function of a function).
- **Implicit differentiation** makes use of the chain rule to calculate derivatives of implicit functions.
- **Logarithmic differentiation** is an application of implicit differentiation to functions involving products, quotients and powers.
- **The Mean Value Theorem** introduced in this chapter will be used later on in the proof of the Fundamental Theorem of Calculus.

Exercises

5.1 Differentiate the following functions:

a) $f(x) = e^{x+5}$

b) $f(x) = (\ln 4)e^x$

c) $f(x) = xe^x$

d) $f(x) = \dfrac{x^2+5x+2}{x+3}$

e) $f(x) = (x+1)^{99}$

f) $f(x) = xe^{-x^2}$

g) $f(t) = e^{\cos t}$

h) $f(t) = e^{t\cos 3t}$

i) $f(t) = \ln\big(\cos(1-t^2)\big)$

j) $f(x) = (x+\sin^5 x)^6$

k) $f(x) = \sin(\sin(\sin x))$

l) $f(x) = \sin(6\cos(6\sin x))$

5.2 For each of the following functions f, find $f(f'(x))$ and $f'(f(x))$.

a) $f(x) = \dfrac{1}{x}$,

b) $f(x) = x^2$,

c) $f(x) = 2$,

d) $f(x) = 2x$.

5.3 Use implicit differentiation to find $\dfrac{dy}{dx}$ if:

a) $x^3 + y^3 = 1$

b) $x^2 + xy - y^2 = 4$

c) $4\cos x \sin y = 1$

d) $\sqrt{xy} = 1 + x^2 y$

e) $2\sqrt{x} + \sqrt{y} = 3$

5.4 Use implicit differentiation to find the equation of the tangent line to the following curves at the given point:

a) $x^2 + xy + y^2 = 3$ at the point (1,1) – This is an ellipse.

b) $x^2 + 2xy - y^2 + x = 2$ at the point (1,2) – This is a hyperbola.

CHAPTER 6

Applications of Differentiation

Optimizing functions of one variable is an extremely useful application of calculus. In this chapter we develop various procedures such as the first and second derivative tests to identify local and global extrema of such functions. We also study methods to find points of inflection and identify concavity that will allow us to sketch functions of one variable. Other important applications of differentiation are investigated including L'Hopital's rule for finding limits of indeterminate forms.

6.1 Optimizing functions of one variable

Problem optimization means finding the optimal or best way of solving a given problem. Examples are available in the physical, biological, behavioral sciences as well as in finance and economics. We might be interested in minimizing the cost of manufacturing a certain item or in maximizing the profit in performing certain transactions, and so on.

Mathematically, these problems can be reduced to finding the **extrema** of a function $f(x)$, that is, identifying the **maximum** or **minimum** values of the function in some part of its domain. There are two types of extrema, local and global:

Local extrema
We say that f has a **local maximum** (or **relative maximum**) at a point a if $f(a) \geq f(x)$ when x is **near** the point a.

Similarly, we say that f has a **local minimum** (or **relative minimum**) at a point a if $f(a) \leq f(x)$ when x is **near** the point a.

Global extrema
A function f has a **global maximum** (or **absolute maximum**) at a point a if $f(a) \geq f(x)$ for all x in its domain.

Similarly, f has a **global minimum** (or **absolute minimum**) at a point a if $f(a) \leq f(x)$ for all x in its domain.

The following is an important result that guarantees the existence of global extrema for functions defined on closed intervals:

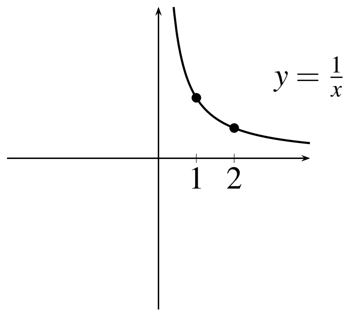
> **The Extreme Value Theorem**
> If f is continuous on a closed interval $[a,b]$, then f attains a **global maximum value** $f(c)$ and a **global minimum value** $f(d)$, where c and d are some real numbers on $[a,b]$.

If a function is defined on an open interval of the form (a,b) where the end-points are excluded, it need not have global maxima and minima.

Example 6.1a Functions without local or global maxima or minima

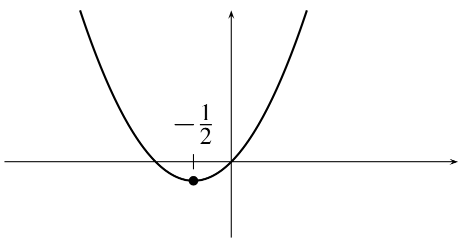
Consider the function $f : (0,\infty) \to (0,\infty)$ given by $f(x) = \dfrac{1}{x}$. There is no global maximum because the value of $\dfrac{1}{x}$ can be made as large as we like simply by choosing the positive number x sufficiently close to (but not equal to) 0, and there is no global minimum since $\dfrac{1}{x}$ can be made as small as we like by choosing x sufficiently large and positive.

However if we change the domain from $(0,\infty)$ to $[1,2]$, we see that the global maximum is 1 and the global minimum is $\dfrac{1}{2}$. There are no local maxima or minima in either case.



In the next example the function (whose domain is $\mathbb{R}$) has a local minimum, a global minimum and no local or global maxima. $\Diamond$

Example 6.1b Suppose that $f(x) = x^2 + x$.



Completing the square in the form $f(x) = (x+\frac{1}{2})^2 - \frac{1}{4}$ shows that the smallest value of f occurs when $x = -\frac{1}{2}$, that is, $f(-\frac{1}{2}) = -\frac{1}{4}$. Therefore $f(x) \geq f(-\frac{1}{2}) = -\frac{1}{4}$ for every x in any open interval about $-\frac{1}{2}$ and hence f has a local minimum at $x = -\frac{1}{2}$.

As the domain of f is $\mathbb{R}$, it is clear that f has no global maximum since $f(x)$ may be made as large as we please by choosing x sufficiently large. The global minimum is the same as the local minimum, namely $-\frac{1}{4}$ at the point $x = -\frac{1}{2}$. ◊

Critical points

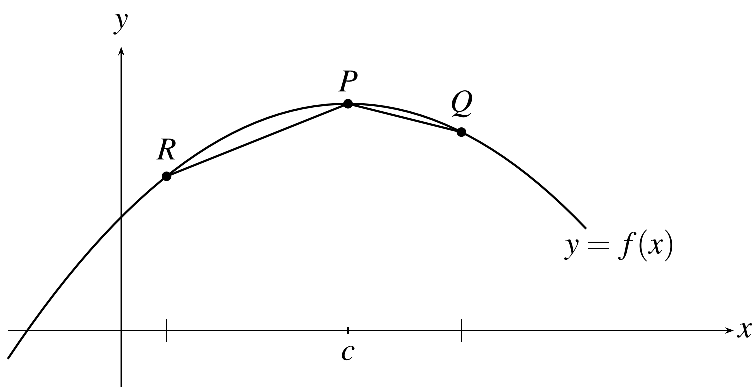
Critical points are important because they are potential candidates to become local extrema and therefore also potential global extrema.

> **A critical point** of $f(x)$ is a number c in the domain of f, where either $f'(c) = 0$ or where $f'(c)$ does not exist.

The following result allows us to identify where local maxima and local minima are located:

> **Property of critical points**
> If f is differentiable at the point c and has a local maximum or minimum at c, then $f'(c) = 0$.

This result will not be proved formally here, but it is easy to convince yourself that it is true by looking at the following diagram, which illustrates the case of a maximum point.

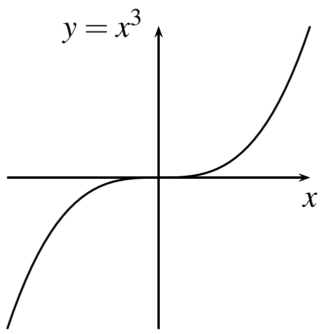


The plot shows that the point P at $x = c$ is a maximum point. Therefore, there is an interval I about c such that $f(x) \leq f(c)$ for all $x \in I$. Select two values of x in I, one to the left of c and the other to the right. Let R and Q be the corresponding points on the graph, and let PR, PQ be the secants as shown.

The slope of secant PQ is negative and the slope of secant PR is positive. The derivative $f'(c)$ is the limiting value of each of these slopes, as the points Q and R move along the curve towards P. Since the secants PR have positive slope their limit cannot be negative. Similarly, the secants PQ have negative slope and so their limit cannot be positive.

As the derivative of f exists at c, these two slopes must ultimately tend towards a common value, which must be zero. Thus $f'(c) = 0$. A similar situation occurs for the case of a local minimum.

Unfortunately the converse of this property is not always true: if $f'(c) = 0$ it does not always follow that c is a maximum or minimum point. The standard counter example is the cubic function $f(x) = x^3$ whose graph is shown below.



The point $x = 0$ is a critical point because $f'(0) = 0$ and yet the origin is neither a minimum nor a maximum. It is what we call a point of inflection to be introduced in the next sections. Nevertheless, finding the critical points gives us a list of possible places where maxima or minima may exist.

Calculation of global extrema

If the function f is continuous in a closed interval $[a,b]$, the global maximum and global minimum values occur either at the critical points or at the endpoints of the interval.

Calculation of global extrema

a) Find the values of f at the critical points on $[a,b]$.

b) Calculate the values of f at the endpoints of the interval. That is, calculate $f(a)$ and $f(b)$.

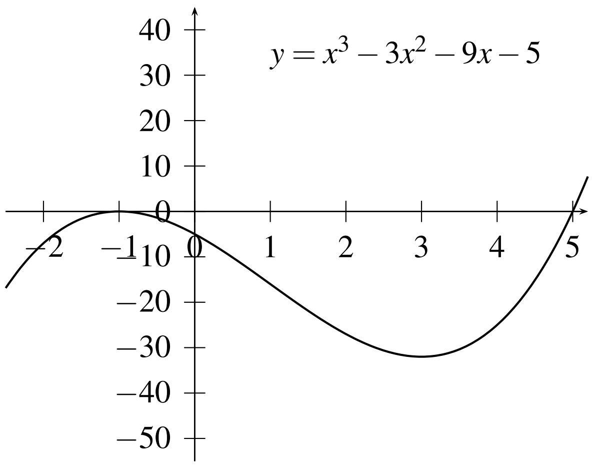
c) Compare all the numbers obtained in the first two steps. The largest of all the values is the **global maximum** value and the smallest of these values is the **global minimum** value.

Example 6.1c Find the global maximum and minimum values of the function
$f(x) = x^3 - 3x^2 - 9x - 5$, 1) on the interval $[0,4]$ and 2) on the interval $[-2,4]$.

The derivative is $f'(x) = 3x^2 - 6x - 9 = 3(x+1)(x-3)$, therefore setting $f'(x) = 0$ gives the critical points $x = -1$ or $x = 3$.

1) In this case we do not have to consider $x = -1$ because it is outside the interval $[0,4]$. Using the three step process defined above to calculate global extrema, we have

a) The value of f at the critical point $x = 3$ inside $[0,4]$ is $f(3) = -32$.
b) The values of f at the endpoints of the interval are $f(0) = -5$ and $f(4) = -25$.
c) Comparing these three numbers, shows that the global maximum is $f(0) = -5$ and the global minimum $f(3) = -32$.



2) Now consider f on the interval $[-2,4]$. Again, using the same method,

a) Both critical points are now inside the interval $[-2,4]$ so we get $f(-1) = 0$ and $f(3) = -32$.
b) The values of f at the endpoints of the interval are $f(-2) = -7$ and $f(4) = -25$.
c) Comparing these four numbers shows that in this case the global maximum is $f(-1) = 0$ and the global minimum is still $f(3) = -32$.

Note that when the domain is $\mathbb{R}$, f has no global maximum or minimum because
$\lim_{x\to-\infty} f(x) = -\infty$ and $\lim_{x\to\infty} f(x) = \infty$. ◊

6.2 Increasing and decreasing functions

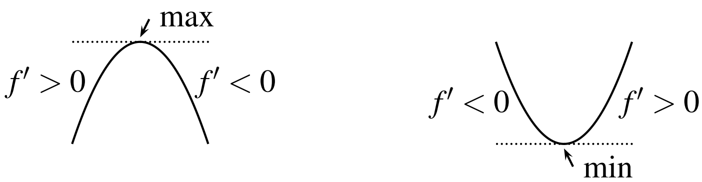
Recall that $f'(x)$ represents the slope of the curve $y = f(x)$ at the point x, which by definition is the slope of the tangent. Therefore, $f'(x)$ tells us the direction of the curve at the point x.

In particular, it can tell us whether it increases or decreases near the point.

Increasing/Decreasing test

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

These tests also help us identify the nature of a critical point at $x = c$. For example, if f has a **local maximum** at a point, it must first increase to that point and then decrease. Similarly, if f has a **local minimum** at a point, it must decrease to the point and then increase. The diagram below illustrates maximum and minimum critical points with $f'(c) = 0$.



These observations form the basis of the **first derivative test** to identify the nature of the critical points:

First Derivative Test

Suppose that c is a critical point of a continuous function f.

a) If f' changes from positive to negative at c, then f has a local maximum at c.
b) If f' changes from negative to positive at c, then f has a local minimum at c.
c) If f' does not change sign at c, then f has no local maximum or minimum at c.

Note that the First Derivative Test is a direct consequence of the Increasing/Decreasing Test.

Example 6.2a Consider the function $f(x) = x^2 e^x$. Find the critical points and draw a *sign diagram* to indicate the regions where $f(x)$ increases or decreases, identifying the critical points as local maxima or local minima. Use this information to draw a sketck of $f(x)$ showing the most salient features.

The natural domain of f is the whole of the real line $-\infty < x < \infty$ so there are no singularities. Differentiating $f(x) = x^2e^x$ using the product rule, gives

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

and hence there are two critical points at $x = 0$ and $x = -2$ such that $f'(0) = f'(-2) = 0$. The values of the function at those two points are $f(0) = 0$ and $f(-2) = 4e^{-2} \approx 0.5$.

Now to determine the nature of the critical points $(0,0)$ and $(-2, 4e^{-2})$ we look at the sign of $f'(x)$ on either side of $x = 0$ and $x = -2$ and use the first derivative test.

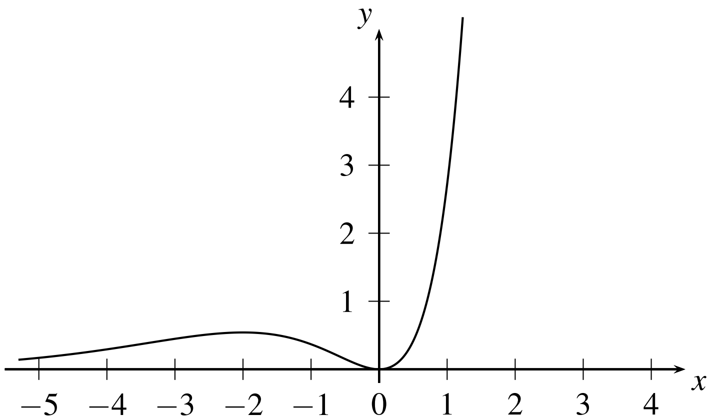
It is helpful to draw a *sign diagram* for $f'(x)$ as follows:

	$x < -2$	$x = -2$	$-2 < x < 0$	$x = 0$	$x > 0$
$f'(x)$	+ve	0	$-$ve	0	+ve
slope	$\nearrow$	$\longrightarrow$	$\searrow$	$\longrightarrow$	$\nearrow$

Table 1

Thus the graph slopes up to a local maximum at $(-2, 4e^{-2})$, then down to a local minimum at $x = 0$, then up again after that. We note also that $f(x) = x^2 e^x \geq 0$ for all x and that $\lim_{x \to -\infty} = 0$ and $\lim_{x \to \infty} = \infty$.

We now have enough information at our disposal to be able to sketch the graph. It is immediately clear that for large



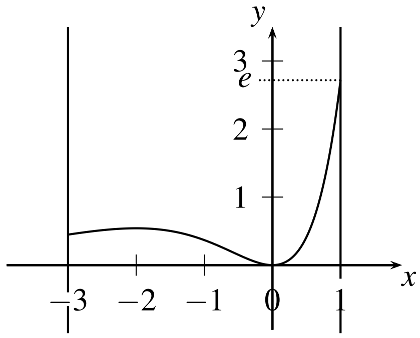
positive x the function takes on values much bigger than $4e^{-2}$; so $(-2, 4e^{-2})$ is not a global maximum, but only a local maximum. The function value at $x = -2$ is bigger than the values immediately around, but by no means the biggest value the function attains. On the other hand, the point $(0,0)$ is a **global minimum**, since $x^2 e^x \geq 0$ for all x. This function **does not have a global maximum** since it goes all the way up to $+\infty$. $\diamondsuit$

Example 6.2b In the previous example, we saw that $f(x) = x^2 e^x$ does not have a global maximum over its natural domain $-\infty < x < \infty$. In this example we *restrict the domain* to the closed interval $[-3, 1]$ and so the Extreme Value Theorem asserts that there will be both, global maximum and global minimum. Now use the three step process to calculate global extrema.

a) Both critical points $x = 0$ and $x = -2$ are inside the interval $[-3, 1]$. The function values are $f(0) = 0$ and $f(-2) = 4e^{-2} \approx 0.5$.

b) The values of f at the endpoints of the interval are $f(-3) = 9e^{-3} \approx 0.45$ and $f(1) = e \approx 2.72$.

c) Comparing these four values, shows that the global maximum is $f(1) = e \approx 2.72$ and the global minimum $f(0) = 0$.



◊

Maxima and minima when the derivative is undefined

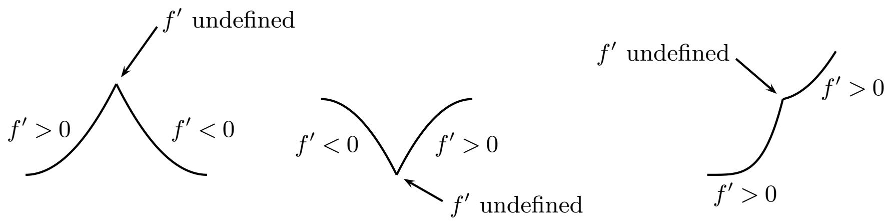
The maximum and minimum points we have encountered so far have all occurred at points where the first derivative is zero (that is, at points where the curve has a horizontal tangent). We now turn to cases in which the first derivative is undefined.

We have already shown in Example 5.1b that the absolute value function

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

does not have a derivative at $x = 0$. In fact, its graph has a V-shape with a cusp at the origin. Because there is no tangent at 0, we say that the graph is not smooth at the origin, even though it is continuous there. Indeed, it is certainly defined at $x = 0$ and $|0| = 0$. As this example shows, very frequently, cusps will be local maximum or minimum.

As for critical numbers with $f'(x) = 0$, the determining feature of a maximum at a cusp is that $f'(x)$ changes from positive to negative as x increases through the value x_0. Similarly, a minimum at a cusp is characterized by $f'(x)$ changing from negative to positive at the critical number.



The diagrams illustrate a cusp that is a local maximum, a cusp that is a local minimum and a cusp that is neither a local maximum nor a local minimum.

Example 6.2c The first derivative of $f(x) = (x-1)^{2/5}$ is

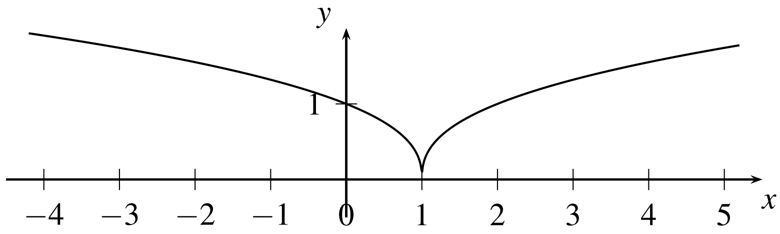
$$\begin{aligned} f'(x) &= \frac{2}{5}(x-1)^{-(3/5)} \\ &= \frac{2}{5(x-1)^{3/5}}, \end{aligned}$$

which is undefined at $x = 1$. There are no values of x that make $f'(x)$ zero; so $x = 1$ is the only critical number. The corresponding critical point is $(1,0)$. The sign diagram for the slope is shown in the table below.

	$x < 1$	$x = 1$	$x > 1$
$f'(x)$	$-$ve	undefined	+ve
slope	$\searrow$	undefined	$\nearrow$

Table 2

We can now draw the graph of $f(x)$ clearly showing that the cusp at $(1,0)$ is also a global minimum.



$\Diamond$

6.3 Concavity and points of inflection

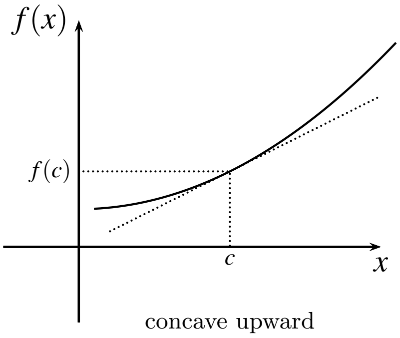
Let $x = c$ be a point in the domain of a function f at which the derivative $f'(c)$ is defined. The tangent to the graph at the point $(c, f(c))$ can then be drawn.

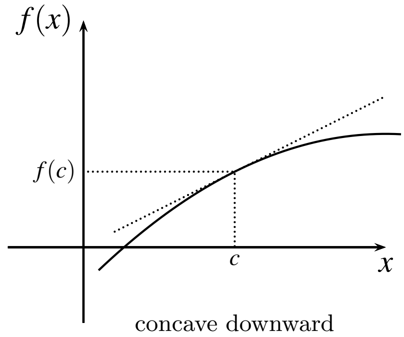
Concavity

a) If the graph of f lies above the tangent line for all points close to $(c, f(c))$ then we say that the graph is **concave upward** at this point.

b) If the graph lies below the tangent line at all points close to $(c, f(c))$ then the graph is said to be **concave downward** at this point.

Note from the diagram that if the concavity is upwards, going from left to right, the slope of the tangent increases. This means that the derivative f' is an increasing function and therefore its derivative f'' is positive.





Similarly, if the concavity is downwards, going from left to right, the slope of the tangent decreases. This means that the derivative f' is a decreasing function and therefore its derivative f'' is negative.

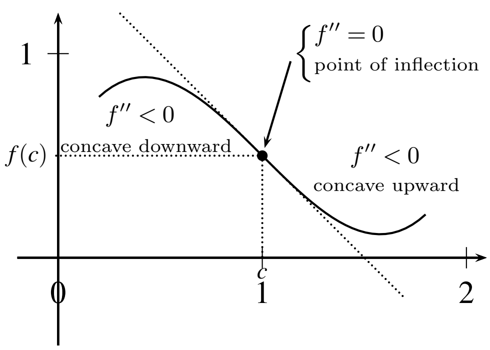
These observations lead to the following test, which may be proved using the Mean Value Theorem:

Concavity Test

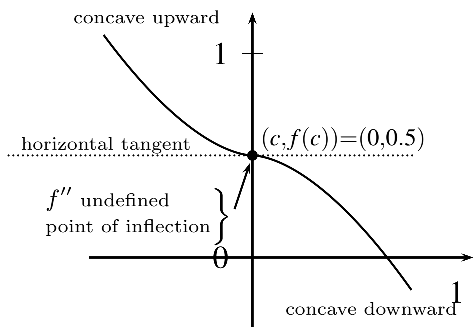
a) If $f''(x) > 0$ for all x in an interval I, then the graph of f is **concave upwards** on I.

b) if $f''(x) < 0$ for all x in an interval I, then the graph of f is **concave downwards** on I.

A point at which the concavity changes from upward to downward (or vice versa) is called a *point of inflection*. Since the sign of the second derivative changes as we pass through a point of inflection, at the point of inflection itself the second derivative must either be zero or undefined.



$f(x) = x^3 - 3x^2 + 2x + 0.5$



$f(x) = 0.5 - \sqrt[3]{x^5}$

Observe that although the graph and its tangent line at the point $(c, f(c))$ have the same slope, as they must by the definition of "tangent", the fact that $(c, f(c))$ is a point of inflection means that the graph actually crosses the tangent at $(c, f(c))$ as shown in the diagrams.

It is helpful to locate points of inflection when sketching graphs, since changes in concavity are significant aspects of a curve's shape.

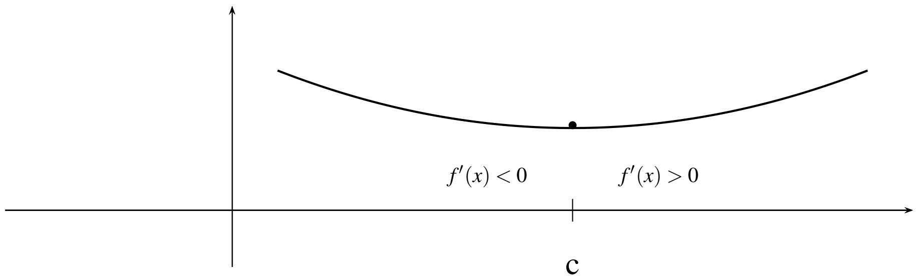
Second derivative test

Since we are discussing applications of second derivatives, we should mention the Second Derivative test for identifying the nature of local extrema. The method assumes that the second derivative of a function f exists.

> **Second Derivative Test**
> If f and f' are differentiable functions and $f'(c) = 0$ (that is, $x = c$ is a critical point of f), then
> a) If $f''(c) > 0$, there is a local minimum of f at $x = c$,
> b) if $f''(c) < 0$, there is a local maximum of f at $x = c$, and
> c) if $f''(c) = 0$, we cannot draw any conclusions without further work.

To understand why this works, consider the following diagram which illustrates case (a). If $f''(c) > 0$ then we also have $f''(x) > 0$ for all x sufficiently close to c, and hence $f'(x)$ must be an increasing function in some interval containing c.

Since $f'(c) = 0$, this means that a little to the left of c we must have $f'(x)$ negative and a little to the right we must have $f'(x)$ positive. In other words, f has a local minimum at $x = c$.



Similar reasoning holds for case b). The examples below show why we can make no conclusion when $f''(c) = 0$, in case c).

Critical points and concavity

Note that if $f'(c) = 0$ and f is concave upward at $(c, f(c))$ then the critical point $(c, f(c))$ is a local minimum of f. Similarly, if f is concave downward at a critical point then the point is a local maximum. However, if $f'(c) = 0$ and $(c, f(c))$ is a point of inflection (as well as being a critical point) then it is neither a local maximum nor a local minimum.

This is because the tangent at $(c, f(c))$ must be horizontal (since $f'(c) = 0$) and the graph must cross the tangent at $(c, f(c))$ (since we have a point of inflection); so $f(x) > f(c)$ on one side of $x = c$ and $f(x) < f(c)$ on the other side.

The simplest example of this is the graph of $y = x^3$ at the origin where both $f'(0) = 0$ and $f''(0) = 0$. Hence it must be a point of inflection as well as a critical point.

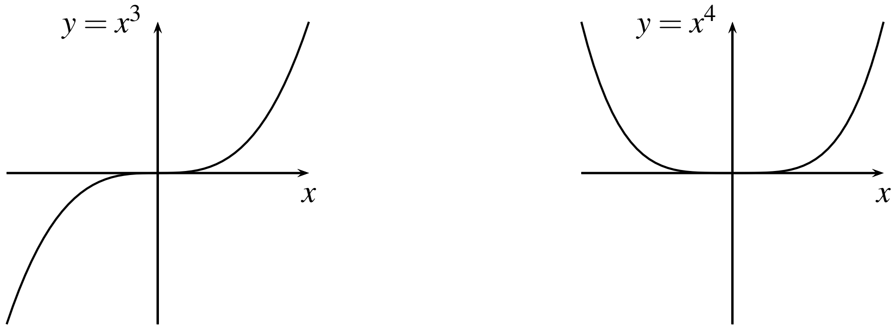
By contrast, $y = x^4$ also has the property that $f'(0) = f''(0) = 0$, however, in this case $(0,0)$ is a local minimum. This example shows that the condition $f''(x) = 0$ is no guarantee that we have a point of inflection.

Examples 6.3a

i) Let $f(x) = x^3 - 3x^2 - 9x - 5$. Then $f'(x) = 3x^2 - 6x - 9 = 3(x-3)(x+1)$; so $x = -1$ and $x = 3$ are the only critical points of f. The second derivative is $f''(x) = 6(x-1)$, so $f''(-1) = -12 < 0$ and $f''(3) = 12 > 0$; therefore, $x = -1$ is a local maximum for f and $x = 3$ is a local minimum.

ii) Next, consider the function $g(x) = x^3$. The first and second derivatives are $g'(x) = 3x^2$ and $g''(x) = 6x$ so $x = 0$ is the only critical point of $y = x^3$. For this function the second derivative test is of no help in deciding the nature of the critical point at $x = 0$. However, as $g'(x) = 3x^2$ we see that $g'(x) \geq 0$ for all x; hence, there is no change in the sign of $g'(x)$ around $x = 0$ and we have a point of (horizontal) inflection.

iii) Finally, consider $h(x) = x^4$. This time, $h'(x) = 4x^3$ and $h''(x) = 12x^2$, so $x = 0$ is the only critical point and $h''(0) = 0$. Unlike the last example, the sign of $h'(x)$ changes around $x = 0$: more precisely, $h'(x) < 0$ if $x < 0$, and $h'(x) > 0$ if $x > 0$. Hence, h is decreasing on $(-\infty, 0)$ and increasing on $(0, \infty)$, and therefore, $x = 0$ is a local minimum. $\Diamond$



The graphs of $y = x^3$ and $y = x^4$ both satisfy $f'(0) = f''(0) = 0$. In both cases the x-axis is the tangent to the graph at the origin. For $y = x^3$ the graph crosses its tangent at the origin, indicating a point of inflection. The graph of $y = x^4$ is concave upward at the origin, indicating a local minimum.

6.4 Curve sketching

The following list provides the steps necessary to sketch a curve and is intended as a guide only. Not every item is relevant to every function.

CURVE SKETCHING SUMMARY

a) **Domain** – Find the set of values of x for which the function $f(x)$ is defined. This could be the natural domain or some specified subset of the natural domain (Section 3.1).

b) **Intercepts** – Find the y-intercept from $y = f(0)$ and the x-intercepts by solving the equation $f(x) = 0$.

c) **Horizontal asymptotes** – Calculate $\lim_{x\to\infty} f(x)$ and $\lim_{x\to-\infty} f(x)$. If either limit is finite equal to L, then the line $y = L$ is a horizontal asymptote (Section 4.4)

d) **Vertical asymptotes** – If there is a value $x = a$ such that $\lim_{x\to a^+} f(x) = \pm\infty$ or $\lim_{x\to a^-} f(x) = \pm\infty$ then there is a vertical asymptote at $x = a$ (Section 4.5).

e) **Critical points** – Find the critical points c of f where $f'(c) = 0$ or $f'(c)$ does not exist. (Section 6.1).

f) **Intervals of increase or decrease** – Using the critical points found above, draw a *sign diagram* to find the intervals on which $f'(x) > 0$ (f increasing) and the intervals on which $f'(x) < 0$ (f decreasing) (Section 6.2). This process will also identify the nature of the critical points (maximum, minimum or neither) using the First Derivative Test.

g) **Concavity and points of inflection** – Calculate $f''(x)$ and use the concavity test. Find the points of inflection by solving $f''(x) = 0$ (Section 6.3).
You may draw another *sign diagram* (or combine with previous one) to find the intervals on which $f''(x) > 0$ (f concave upwards) and the intervals on which $f''(x) < 0$ (f concave downwards)

h) **Sketch the curve** – Put together the information obtained above to draw the graph of f (See the example below).

Example 6.4a Sketch the graph of $f(x) = x^{2/3}(x-5)$ indicating the most important features according to the list of steps above.

a) **Domain** – The natural domain of the function $f(x) = x^{2/3}(x-5)$ consists of all x such that $x \in \mathbb{R}$.

b) **Intercepts** – The y-intercept is $y = f(0) = 0$, and since $f(0) = f(5) = 0$ the x-intercepts occur when $x = 0$ and $x = 5$.

c) **Horizontal asymptotes** – Since $x^{2/3}$ is the cube root of x^2, then it is zero at $x = 0$ and positive elsewhere. Thus $x^{2/3}(x-5)$ has the same sign as $x-5$ for all nonzero values of x. This means that $f(x) < 0$ for $x < 5$ and $f(x) > 0$ for $x > 5$. This also implies that $\lim_{x\to\infty} x^{2/3}(x-5) = \infty$ and $\lim_{x\to-\infty} x^{2/3}(x-5) = -\infty$. Therefore there are no horizontal asymptotes.

d) **Vertical asymptotes** – Because $f(x) = x^{2/3}(x-5)$ is defined for all values of x, there are no points $x = a$ such that $\lim_{x\to a^+} f(x) = \pm\infty$ or $\lim_{x\to a^-} f(x) = \pm\infty$. That is, there are no vertical asymptotes.

e) **Critical points** – To find the critical points we calculate the derivative

$$\begin{aligned} f'(x) &= x^{2/3} + (x-5)\frac{2}{3}x^{-1/3} \\ &= x^{2/3} + \frac{2(x-5)}{3x^{1/3}} \\ &= \frac{3x + 2(x-5)}{3x^{1/3}} \\ &= \frac{5x-10}{3x^{1/3}} \\ &= \frac{5}{3}\frac{(x-2)}{x^{1/3}} \end{aligned}$$

So $f'(2) = 0$ and is undefined when $x = 0$. Therefore, there are two critical points, one at $(0,0)$ and the other at $(2, -3\sqrt[3]{4}) \approx (2, -4.8)$.
Regarding the critical point $(0,0)$, we had already found that $f(0) = 0$ and that $f(x) < 0$ for nearby values of x on either side of $x = 0$. This shows that $(0,0)$ is a local maximum, and it might have led us to expect that the derivative would be zero at $x = 0$.
In fact, there is a cusp at $x = 0$, as can be seen by inspecting the values of the derivative for points close to zero. If x is small and positive then $x^{1/3}$ (the cube root of x) is also small and positive, whereas $x - 2$ is close to -2. So $\dfrac{(x-2)}{x^{1/3}}$ is a negative number of large magnitude.
On the other hand, when x is negative and close to zero, $x^{1/3}$ is also negative and close to zero, while $x-2$ is still close to -2; so in this case $\dfrac{(x-2)}{x^{1/3}}$ is positive and of large magnitude. So we have an extremely sharp cusp at the origin, the graph being nearly vertical on either side of the cusp.
Since $f(0) = f(5) = 0$ and $f(x) < 0$ for $0 < x < 5$, it is clear that the critical point $(2, -3\sqrt[3]{4})$ must be a local minimum.

f) **Intervals of increase or decrease** – Look at the Table below that shows the sign diagram for $f'(x)$.

g) **Concavity and points of inflection** – Calculate $f''(x)$:

$$\begin{aligned} f''(x) &= \frac{5}{3}\left(\frac{x^{1/3}-(x-2)\,\frac{1}{3}\,x^{-(2/3)}}{x^{2/3}}\right) \\ &= \frac{5}{3}\,\frac{(3x^{1/3}-(x-2)x^{-2/3})}{3x^{2/3}} \\ &= \frac{5}{3}\,\frac{(3x-(x-2))}{3x^{4/3}} \\ &= \frac{5}{9}\,\frac{(2x+2)}{3x^{4/3}} \\ &= \frac{10}{9}\,\frac{(x+1)}{x^{4/3}}. \end{aligned}$$

In the first line of the above equation we applied the quotient rule for differentiation and in the third line we multiplied both the numerator and the denominator by $x^{2/3}$.

So $f''(-1)=0$ and at that point $f(-1)=(-1)^{2/3}(-1-5)=-6$. Since $x^{4/3}$ is never negative we see that $f''<0$ when $x<-1$, and $f''>0$ for $x>-1$ (excluding $x=0$).

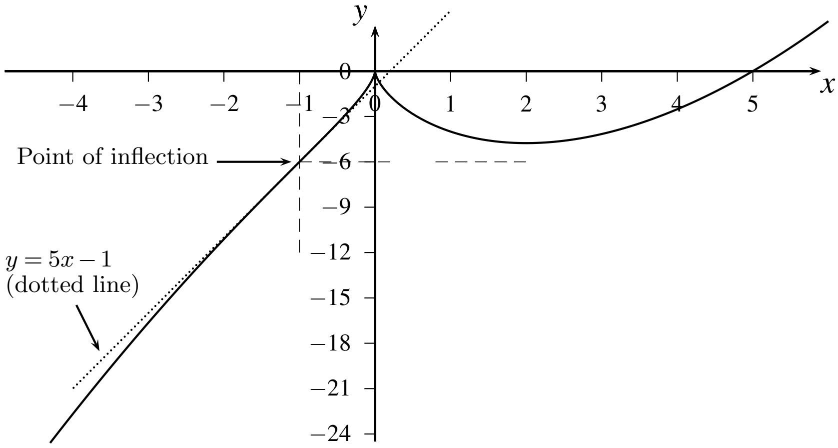
In summary, the concavity is downward for $x<-1$, the point $(-1,-6)$ is a point of inflection, and the concavity is upward for $-1<x<0$ and for $x>0$.

h) **Sketch the curve** – We now put together all the information obtained above to draw the following sign diagram:

	$x<-1$	$x=-1$	$-1<x<0$	$x=0$	$0<x<2$	$x=2$	$x>2$
dy/dx	+ve	+ve	+ve	undef.	−ve	0	+ve
slope	$\nearrow$	$\nearrow$	$\nearrow$	$\uparrow\downarrow$	$\searrow$	$\longrightarrow$	$\nearrow$
y	$y<-6$	$y=-6$	$-6<y<0$	$y=0$	$0>y>-3\sqrt[3]{4}$	$y=-3\sqrt[3]{4}$	$-3\sqrt[3]{4}<y$
d^2y/dx^2	−ve	0	+ve	undef.	+ve	+ve	+ve
concavity	downward	inflection	upward	undef.	upward	upward	upward

Table 3

Finally, after calculating that $f(-4) \approx -22.6$ and $f(6) \approx 3.3$, we are able to draw the graph quite accurately. In order to exaggerate the change in concavity at $(-1,-6)$ we have used different scales for x and y.



The graph of $f(x) = x^{2/3}(x-5)$. The line $y = 5x - 1$ is the tangent at $(-1,-6)$; it crosses the graph at the point of inflection $(-1,-6)$.

$\Diamond$

6.5 L'Hôpital's rule

In Chapter 4 we mentioned indeterminate forms in the context of the quotient law for calculating limits. The quotient law states that

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

If both limits are 0 or both are ∞, then the limit of the ratio f/g can be 0, ∞ or any real number, in other words, the limit of the ratio is **indeterminate**. These two cases lead to two types of indeterminations:

Case 1 – When both, the numerator and denominator tend to infinity, we obtain what is known as an "indeterminate form" of type $\dfrac{\infty}{\infty}$.

Case 2 – Similarly, if both, the numerator and denominator tend to zero, we obtain what is called an "indeterminate form" of type $\dfrac{0}{0}$.

l'Hôpital's rule provides a powerful method for calculating the limits of these two types of indeterminate forms.

L'Hôpital's rule $\left(\frac{0}{0}\text{ form}\right)$

Suppose that $\lim_{x\to c} f(x) = 0$, $\lim_{x\to c} g(x) = 0$ and that $\lim_{x\to c} \dfrac{f'(x)}{g'(x)}$ exists. Then
$$\lim_{x\to c} \frac{f(x)}{g(x)} = \lim_{x\to c} \frac{f'(x)}{g'(x)}.$$

L'Hôpital's rule $\left(\frac{\pm\infty}{\pm\infty}\text{ form}\right)$

Suppose that $\lim_{x\to c} f(x) = \pm\infty$, $\lim_{x\to c} g(x) = \pm\infty$ and that $\lim_{x\to c} \dfrac{f'(x)}{g'(x)}$ exists. Then
$$\lim_{x\to c} \frac{f(x)}{g(x)} = \lim_{x\to c} \frac{f'(x)}{g'(x)}.$$

In both the $\dfrac{0}{0}$ form and $\dfrac{\pm\infty}{\pm\infty}$ form of L'Hôpital's rule, we also allow c to be $\pm\infty$.

Important

When applying the rule more than once to the same expression, is important to ensure that the conditions are satisfied at each step or we will get the wrong answer.

Examples 6.5a

i) We calculate $\lim_{x\to\infty} \dfrac{e^x}{x}$.

   As $x \to \infty$ this expression has the form $\dfrac{\infty}{\infty}$ and so we may apply l'Hôpital's rule. This gives
   $$\lim_{x\to\infty} \frac{e^x}{x} = \lim_{x\to\infty} \frac{e^x}{1} = \infty.$$
   What this is saying is that the function $y = e^x$ grows much more quickly than the function $y = x$; that is, if x is large enough then e^x dominates x.

ii) Find $\lim_{x\to\infty} \dfrac{e^x}{x^2}$.

   This time we must use l'Hôpital's rule twice; after the first application the limit still cannot be evaluated, but the expression remains in the form appropriate for l'Hôpital's rule to be applied again,
   $$\lim_{x\to\infty} \frac{e^x}{x^2} = \lim_{x\to\infty} \frac{e^x}{2x} = \lim_{x\to\infty} \frac{e^x}{2} = \infty.$$

Therefore e^x dominates x^2 as x becomes very large. More generally, it can be shown that e^x dominates x^n, for any $n > 0$. (Convince yourself that $\lim_{x\to\infty} \frac{e^x}{x^n} = \infty$ for any integer $n > 0$.)

iii) Consider $\lim_{x\to\infty} \frac{6x^2+2x-6}{2x^2-7x+4}$. Applying l'Hôpital's rule twice we find that

$$\lim_{x\to\infty} \frac{6x^2+2x-6}{2x^2-7x+4} = \lim_{x\to\infty} \frac{12x+2}{4x-7} = \lim_{x\to\infty} \frac{12}{4} = 3.$$

Check that the two applications of l'Hôpital's rule are allowed. Of course, we could also evaluate this limit by dividing both the numerator and the denominator by x^2.

iv) Now for a slightly harder example. Consider $\lim_{x\to 0^+} x \ln x$. Recall that $x \to 0^+$ indicates that we consider the limit only as x approaches 0 from the right. Happily, l'Hôpital's rule can also be used to evaluate such one-sided limits. However on the face of it, l'Hôpital's rule does not seem to apply in this particular problem because $x \ln x$ is not in the form of a fraction $f(x)/g(x)$.

However, it is possible to rewrite the expression as follows

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

Since $\lim\limits_{x\to 0^+} \ln x = -\infty$ and $\lim\limits_{x\to 0^+} \frac{1}{x} = \infty$ we see that the new expression is an indeterminate form and so l'Hôpital's rule can be used to compute this limit. We have

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

v) Finally, *the following calculation contains a mistake*; find it.

$$\lim_{x\to 1} \frac{x^3-x^2+2x-1}{x^2-2x-2} = \lim_{x\to 1} \frac{3x^2-2x+2}{2x-2} = \lim_{x\to 1} \frac{6x-2}{2} = 2.$$

What is the actual value of this limit? ◊

Example 6.5b To conclude this section we give the harder example

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

Having x appear as an exponent creates some problems; we can get rid of the exponent by using the formula

$$f(x) = \exp\left[\ln f(x)\right]$$

which is a direct consequence of fact that the exponential and the natural log are inverse functions. Therefore,

$$\left(1+\frac{1}{x}\right)^x = \exp\left[\ln\left(1+\frac{1}{x}\right)^x\right] = \exp\left[x \ln\left(1+\frac{1}{x}\right)\right] = \exp\left[\frac{\ln\left(1+\frac{1}{x}\right)}{1/x}\right].$$

Hence,

$$\begin{aligned}
\lim_{x\to\infty}\left(1+\frac{1}{x}\right)^x &= \exp\left[\lim_{x\to\infty}\frac{\ln\left(1+\frac{1}{x}\right)}{1/x}\right] \\
&= \exp\left[\lim_{x\to\infty}\frac{\frac{1}{1+\frac{1}{x}}\times\left(-\frac{1}{x^2}\right)}{-\frac{1}{x^2}}\right] \\
&= \exp\left[\lim_{x\to\infty}\frac{1}{1+\frac{1}{x}}\right] \\
&= \exp\left[\frac{1}{1+0}\right] = \exp[1] = e.
\end{aligned}$$

This limit is both unexpected and amazing. Our working shows that if x is very large then

$$e \approx \left(1+\frac{1}{x}\right)^x.$$

For example, taking $x = 1000$ we find that

$$\left(1+\frac{1}{1000}\right)^{1000} = 2.7169\ldots$$

and $e = 2.7182818\ldots$. You might like to experiment with your calculator to see what happens with larger values of x. $\Diamond$

Summary of Chapter 6

- **Optimization of functions** is related to the problem of identifying their global extrema. The **Extreme Value Theorem** guarantees the existance of global maxima and minima for functions that are continuous on a closed interval $[a,b]$.
- **Curve sketching** is another important application of differentiation. A summary of the most important tools needed for sketching is provided in Section 6.4.
- **L'Hôpital's rule** is concerned with the application of differentiation to the calculation of limits of indeterminate forms.

Exercises

6.1 Find the derivatives of the following functions:

a) $f(x) = 7x^3 - 2x + \dfrac{1}{x}$,
b) $f(x) = (x + \cos x)^3$,
c) $f(x) = \sin\left(\cos(x^2)\right)$,
d) $f(x) = x^3 \sin x + e^{x\cos x}$,
e) $f(x) = 7x^8 + \dfrac{1}{1+\cos x^2}$.

6.2 Find the equation of the tangent line to the function $f(x) = 4xe^x + \cos x$ at the point $x = 0$.

6.3 Find the global minimum and maximum values of the following functions on the closed interval $[-3, 3]$.

a) $f(x) = x^2 e^x$
b) $f(x) = 3x^2 - 6$
c) $f(x) = |x - 1| + 2$
d) $f(x) = \cos\left(\dfrac{x}{4}\right)$

6.4 Show that the polynomial

$$f(x) = 1 + x + x^3 + x^5 + x^7$$

is strictly increasing over the whole real line.

6.5 Find the value a such that the function

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

is strictly decreasing for $x \in (0, a)$ and strictly increasing when $x > a$.

6.6 Use l'Hôpital's rule to find $\lim\limits_{x\to\infty} \dfrac{\sinh x}{\cosh x}$. Now find the limit using the definition of sinh and cosh in terms of exponentials and compare the results.

6.7 Use l'Hôpital's rule to find the following limits:

a) $\lim\limits_{x\to 0} \dfrac{x\cosh x}{\sin x}$
b) $\lim\limits_{r\to 0} \dfrac{\sin r^2}{r^2}$
c) $\lim\limits_{x\to 1} \dfrac{\ln x}{(x-1)}$
d) $\lim\limits_{x\to 0^+} x \ln x$
e) $\lim\limits_{x\to 0} \dfrac{e^x - 1}{\sin x}$
f) $\lim\limits_{x\to 0} \dfrac{1 - \cos x}{x^2}$
g) $\lim\limits_{x\to 0} \dfrac{e^x - x - 1}{\cos x - 1}$
h) $\lim\limits_{x\to 1} \dfrac{x^2 - 1}{\ln x}$
i) $\lim\limits_{x\to 0} \dfrac{\sin x}{1 - e^x}$
j) $\lim\limits_{x\to 0} \dfrac{\cos x - 1}{x \sin x}$

CHAPTER 7

Taylor Polynomials

As we saw in Chapter 2, a polynomial is a function of the form

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n,$$

where the a_k are constants and n is a non-negative integer. "Taylor polynomials" are special polynomials which are used to approximate other types of functions near particular points. The reason for this is that polynomials are easy to evaluate, requiring only addition and multiplication.

Other types of functions, for example the trigonometric functions or logarithmic functions, are very difficult to evaluate exactly. (Have you ever thought about the process used by your calculator in finding, say, $\sin 2$ or $\ln 10$?) In this chapter we also learn to determining how good an approximation it provides.

7.1 An approximation for e^x

Let's start by thinking about how we might construct a cubic (polynomial of degree 3) to approximate e^x near the point $x = 0$,

$$p(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3.$$

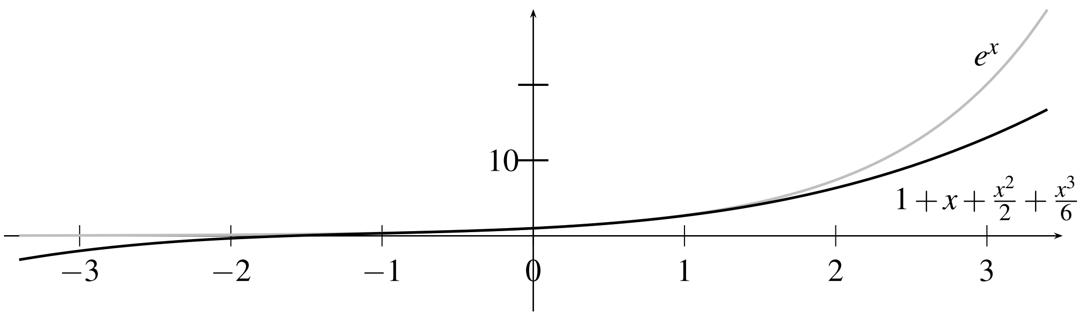
a) Firstly, we would certainly want the value of $p(x)$ to be the same as the value of e^x at $x = 0$. So let's take $p(0) = a_0 = e^0 = 1$, and thus $p(x) = 1 + a_1 x + a_2 x^2 + a_3 x^3$.

b) Next, perhaps it would be a good idea to make the first derivatives the same, so that the graphs of $p(x)$ and e^x have the same slope at $x = 0$. The derivative of e^x is e^x, while $p'(x) = a_1 + 2a_2 x + 3a_3 x^2$. At $x = 0$ we have $p'(0) = a_1 = e^0 = 1$ and hence $p(x) = 1 + x + a_2 x^2 + a_3 x^3$.

c) If we make the second derivatives the same as well, the graphs would then have the same concavity at $x = 0$. The second derivative of e^x is e^x. (Indeed all derivatives of e^x are e^x.) Now, $p''(x) = 2a_2 + 6a_3 x$ and $p''(0) = 2a_2$. Since we want this to equal $e^0 = 1$, we'll let $a_2 = \frac{1}{2}$, and $p(x) = 1 + x + \frac{x^2}{2} + a_3 x^3$.

d) Now let's match up the third derivatives also. Since $p'''(0) = 6a_3$, we should take $6a_3 = 1$, $a_3 = \frac{1}{6}$ and $p(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{6}$.

The graph below shows that $p(x)$ does indeed match e^x very closely near $x=0$.



Suppose we use the same idea to construct a polynomial of degree n (with $n>3$). That is, we look for values of $a_0, a_1, a_2, \ldots, a_n$ such that the polynomial

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

is equal to e^x at $x=0$ and has each of its first n derivatives at $x=0$ equal to the corresponding derivative of e^x at 0. (Note that for a polynomial of degree n all derivatives of order higher than n are 0.) Since every derivative of e^x is equal to 1 at $x=0$, we want to choose the coefficients of $P_n(x)$ so that the first, second, third, …, n^{th} derivatives of $P_n(x)$ are all equal to 1.

Notation The abbreviations that we commonly use for first, second or third derivatives of a function f, namely $f' = \dfrac{df}{dx}$, $f'' = \dfrac{d^2f}{dx^2}$, $f''' = \dfrac{d^3f}{dx^3}$ are inconvenient for higher derivatives. For this reason we introduce the notation $f^{(k)}$ to mean the k^{th} derivative of f. That is, $f^{(k)} = \dfrac{d^kf}{dx^k}$. Sometimes it is convenient to allow k to equal 0 in this notation, so that we have $f^{(0)}$, by which we mean simply f.

Using this notation, we want to find $P_n(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$ such that $P_n^{(k)}(0) = 1$ for $0 \le k \le n$. The first few derivatives of $P_n(x)$ are as follows:

$$\begin{aligned}
P_n(x) &= a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \\
P_n'(x) &= a_1 + 2a_2x + 3a_3x^2 + \cdots + na_nx^{n-1} \\
P_n''(x) &= 2a_2 + 2\cdot 3a_3x + 3\cdot 4a_4x^2 + \cdots + (n-1)na_nx^{n-2} \\
P_n'''(x) &= 2\cdot 3a_3 + 2\cdot 3\cdot 4a_4x + 3\cdot 4\cdot 5a_5x^2 + \cdots + (n-2)(n-1)na_nx^{n-3} \\
P_n^{(4)}(x) &= 2\cdot 3\cdot 4a_4 + 2\cdot 3\cdot 4\cdot 5a_5x + \cdots + (n-3)(n-2)(n-1)na_nx^{n-4}
\end{aligned}$$

Now think about what happens to $P_n(x)$ once it has been differentiated k times ($0 \le k \le n$). All terms in powers of x smaller than k have been differentiated away to zero. The term in x^k will have been multiplied by k on the first differentiation, then by $k-1$, then by $k-2$, and so on. After k differentiations, the term in x^k will have been multiplied by $k(k-1)(k-2)\ldots 2.1 = k!$. All terms in $P_n(x)$ with powers higher than k will still contain a power of x after k differentiations. We therefore have the following:

$$\begin{aligned}
P_n^{(k)}(x) &= k!a_k + \text{ terms containing } x \qquad (k<n)\\
\vdots \quad & \quad \vdots\\
P_n^{(n)}(x) &= n!a_n
\end{aligned}$$

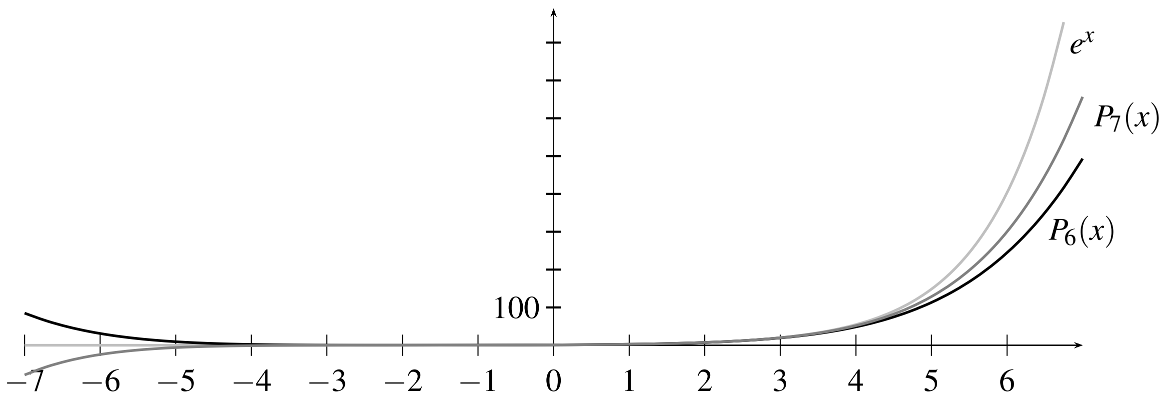
Substituting $x=0$ into the equations above, we find:

$$\begin{aligned}
P_n(0) &= a_0 = 1\\
P_n'(0) &= a_1 = 1\\
P_n''(0) &= 2a_2 = 1, \quad \Longrightarrow \quad a_2 = \frac{1}{2}\\
P_n^{(3)}(0) &= 3!a_3 = 1, \quad \Longrightarrow \quad a_3 = \frac{1}{3!}\\
\vdots \quad & \quad \vdots \qquad\qquad\qquad \vdots\\
P_n^{(k)}(0) &= k!a_k = 1, \quad \Longrightarrow \quad a_k = \frac{1}{k!}\\
\vdots \quad & \quad \vdots \qquad\qquad\qquad \vdots\\
P_n^{(n)}(0) &= n!a_n = 1, \quad \Longrightarrow \quad a_n = \frac{1}{n!}
\end{aligned}$$

Hence the polynomial

$$P_n(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!}$$

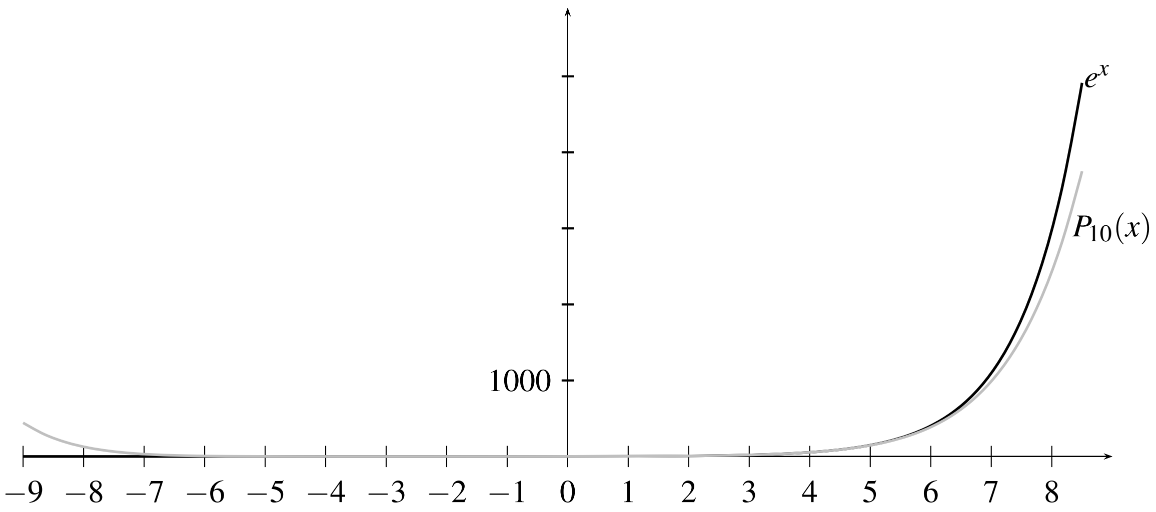
is equal to e^x at $x=0$, and each of its first n derivatives is equal to the corresponding derivative of e^x at $x=0$. Let's see how well this polynomial approximates e^x. The following diagrams show the graphs of $P_6(x)$ and $P_7(x)$ superimposed on the graph of e^x.



We see that the polynomials of degree 6 and degree 7 look virtually identical with e^x for $-2.5 \leq x \leq 2.5$. Taking higher values of n makes

$$P_n(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!}$$

a better approximation to e^x in the sense that it matches the exponential function over a larger interval. Here is the graph of $P_{10}(x)$, superimposed on the graph of e^x.



It is somewhat surprising that the polynomial is a good match for e^x for $-6 \leq x \leq 6$ since we set out to find an approximation to e^x around $x = 0$. What is even more surprising is that if we let n become infinitely big, then the "infinite polynomial"

$$P_n(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

is *exactly* equal to e^x for *all* values of x! (This "infinite polynomial" is more correctly known as a "power series". We will deal with series in the next chapter.)

7.2 Taylor polynomials about $x = 0$

The polynomial we found in the previous section,

$$P_n(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!}$$

is called the nth order "Taylor polynomial" of e^x about $x = 0$.

The process we used for constructing this polynomial is precisely the one that we use for constructing Taylor polynomials for other functions. That is, given a function $f(x)$, we find a polynomial $P_n(x)$ such that $P_n(0) = f(0)$, $P_n'(0) = f'(0)$, $P_n''(0) = f''(0)$, …, $P_n^{(n)}(0) = f^{(n)}(0)$. Of course, the function f must be differentiable n times at $x = 0$.

As we saw in the previous section, if $P_n(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$, then the k^{th} derivative of P_n at $x = 0$, $P_n^{(k)}(0) = k!a_k$, for $0 \le k \le n$.

Since we want $P_n^{(k)}(0)$ to equal $f^{(k)}(0)$, we choose $a_k = \dfrac{f^{(k)}(0)}{k!}$ for each k, $0 \le k \le n$. (Note that $0! = 1$, and $f^{(0)} = f$.)

Taylor polynomial about $x = 0$
The n^{th} order "Taylor polynomial" of a function f about $x = 0$ is

$$P_n(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f^{(3)}(0)}{3!}x^3 + \cdots + \frac{f^{(n)}(0)}{n!}x^n.$$

In sigma notation: $\quad P_n(x) = \sum_{k=0}^{n} \dfrac{f^{(k)}(0)}{k!}x^k$

Taylor polynomials about $x = 0$ are known as **Maclaurin polynomials**

You should memorise this formula. Constructing a Taylor polynomial for a function f is simply a matter of finding the values of the derivatives of f at zero, and substituting into this formula. For example, since every derivative of $f(x) = e^x$ is e^x, and $e^0 = 1$, it is immediately obvious that the n^{th} order Taylor polynomial for the exponential function e^x is

$$P_n(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!}$$

.

The sine function is also easy to deal with. Let $f(x) = \sin x$, and we have the following:

$$\begin{array}{rcl rcl}
f(x) & = & \sin x & f(0) & = & 0 \\
f'(x) & = & \cos x & f'(0) & = & 1 \\
f''(x) & = & -\sin x & f''(0) & = & 0 \\
f^{(3)}(x) & = & -\cos x & f^{(3)}(0) & = & -1 \\
f^{(4)}(x) & = & \sin x & f^{(4)}(0) & = & 0 \\
f^{(5)}(x) & = & \cos x & f^{(5)}(0) & = & 1 \\
f^{(6)}(x) & = & -\sin x & f^{(6)}(0) & = & 0 \\
f^{(7)}(x) & = & -\cos x & f^{(7)}(0) & = & -1
\end{array}$$

Therefore the Taylor polynomial of order 7 for $f(x) = \sin x$, about $x = 0$, is

$$P_7(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!}.$$

Note that $P_8(x)$ is the same as $P_7(x)$, since $f^{(8)}(0) = \sin 0 = 0$.

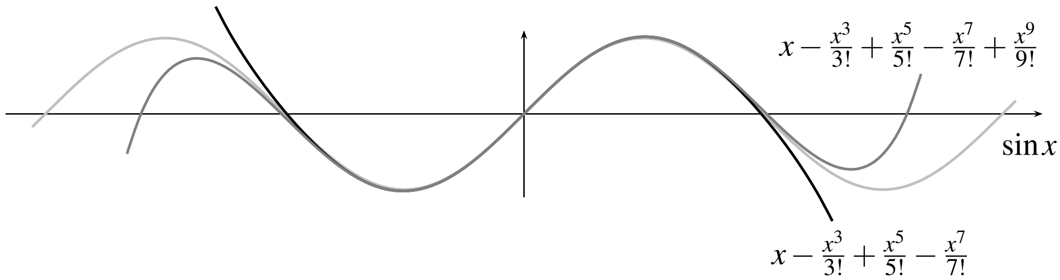
It is clear that successive derivatives of $\sin x$, evaluated at 0, will continue to follow the pattern we can see in the table above. (That is, starting with the 0^{th} derivative, the pattern is $0, 1, 0, -1, 0, 1, 0, -1, \ldots$.) Hence, successive terms in the Taylor polynomial for $\sin x$ will follow the pattern that is obvious in $P_7(x)$, and the general polynomial, of order $2n+1$, $(n = 0, 1, 2, \ldots)$ for the sine function, is

$$x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots + (-1)^n \frac{x^{2n+1}}{(2n+1)!}.$$

It is always worthwhile attempting to find a pattern in the values of derivatives when calculating Taylor polynomials, since finding a pattern allows us to write down an expression for the general polynomial.

Normally we would look for the general polynomial of order n, but in the case of $\sin x$ it is easier to write one in terms of odd numbers $2n+1$ only, since all the coefficients of even powers are zero, and the polynomial consists of terms in odd powers of x only.

The following diagram compares the graph of $\sin x$ with those of the Taylor polynomials of order 7 and order 9.



Both polynomials are very good approximations over the interval $[-\pi, \pi]$, although they both eventually diverge dramatically from $\sin x$. Of course, if we know $\sin x$ for $-\pi \le x \le \pi$, we can calculate it for any value of x using the periodicity of the sine function.

Examples 7.2a

i) If $f(x) = \cos x$, then $f(0) = 1$, and the values of the first 6 derivatives at $x = 0$ are $1, 0, -1, 0, 1, 0$. The Taylor polynomial of order 6 for $\cos x$ is therefore

$$1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!}.$$

This time we have an even function, and its Taylor polynomial contains even powers only. As with the sine function, the pattern continues and the polynomial of order $2n$ for $\cos x$ is

$$1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots + (-1)^n \frac{x^{2n}}{(2n)!}.$$

ii) Find the Taylor polynomial of order 3 for the function $f(x) = x^3 - 7x^2 + 4x - 2$. You should find that it is precisely $x^3 - 7x^2 + 4x - 2$. What about the Taylor polynomial of order 4? The 4^{th} derivative of f is 0, so the 4^{th} order Taylor polynomial is also $x^3 - 7x^2 + 4x - 2$ (as, indeed, is any Taylor polynomial of higher degree).

In general, if f is a polynomial of degree n, then its Taylor polynomial of order n or higher is equal to f. Can you verify this statement?

iii) Let $f(x) = \dfrac{1}{1-x}$. Then:

$$\begin{aligned}
f(x) &= \frac{1}{1-x} & f(0) &= 1\\
f'(x) &= \frac{1}{(1-x)^2} & f'(0) &= 1\\
f''(x) &= \frac{2}{(1-x)^3} & f''(0) &= 2\\
f^{(3)}(x) &= \frac{3\cdot 2}{(1-x)^4} & f^{(3)}(0) &= 3!\\
f^{(4)}(x) &= \frac{4\cdot 3\cdot 2}{(1-x)^5} & f^{(4)}(0) &= 4!\\
f^{(5)}(x) &= \frac{5\cdot 4\cdot 3\cdot 2}{(1-x)^6} & f^{(5)}(0) &= 5!
\end{aligned}$$

The pattern is clear; $f^{(n)}(x) = \dfrac{n!}{(1-x)^{n+1}}$, and $f^{(n)}(0) = n!$. The Taylor polynomial of order n for $\dfrac{1}{1-x}$ is therefore

$$\begin{aligned}
P_n(x) &= 1 + x + \tfrac{2}{2!}x^2 + \tfrac{3!}{3!}x^3 + \tfrac{4!}{4!}x^4 + \cdots + \tfrac{n!}{n!}x^n\\
&= 1 + x + x^2 + x^3 + x^4 + \cdots + x^n.
\end{aligned}$$

iv) Show that the Taylor polynomials for $\cosh x$ and $\sinh x$, about $x = 0$, are $\displaystyle\sum_{k=0}^{n} \frac{x^{2k}}{(2k)!}$ and $\displaystyle\sum_{k=0}^{n} \frac{x^{2k+1}}{(2k+1)!}$ respectively. Compare these polynomials with those for $\cos x$ and $\sin x$.

$\Diamond$

7.3 Taylor polynomials about $x = a$

In order to be able to use the formula given in the previous section for a Taylor polynomial, we must have a function $f(x)$ which not only exists at $x = 0$, but is also differentiable n times at $x = 0$. Not all functions satisfy this requirement. The function $f(x) = \ln x$ is an obvious example. It is not possible, therefore to approximate $\ln x$ with a polynomial using the formula in the previous section. It is possible, however, to find Taylor polynomials about values other than zero.

Taylor polynomial about $x = a$
The n^{th} order "Taylor polynomial" of a function f about $x = a$ is

$$
\begin{aligned}
f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f^{(3)}(a)}{3!}(x-a)^3 + \cdots \\
\cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n \\
= \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k.
\end{aligned}
$$

In order to use this formula the function f must exist at a and be differentiable n times at a. Note that the formula given in the previous section, for a Taylor polynomial about 0, is just a special case of this formula (with $a = 0$).

Example 7.3a Let us find a Taylor polynomial for $f(x) = \ln x$. As mentioned above, we cannot find one about $x = 0$. We must choose a value for a at which $\ln x$ exists, and is differentiable n times. Suppose we choose $a = 1$, which should make evaluation of the derivatives relatively simple. (There is no point in making things hard for ourselves!) So we have:

$$
\begin{array}{rlrl}
f(x) & = \ln x & f(1) & = 0 \\
f'(x) & = 1/x & f'(1) & = 1 \\
f''(x) & = -1/x^2 & f''(1) & = -1 \\
f^{(3)}(x) & = 2/x^3 & f^{(3)}(1) & = 2 \\
f^{(4)}(x) & = -3!/x^4 & f^{(4)}(1) & = -3! \\
f^{(5)}(x) & = 4!/x^5 & f^{(5)}(1) & = 4! \\
\vdots & \quad\vdots & \vdots & \quad\vdots \\
f^{(n)}(x) & = (-1)^{n+1}(n-1)!/x^n & f^{(n)}(1) & = (-1)^{n+1}(n-1)!
\end{array}
$$

The Taylor polynomial of order n for $\ln x$, about $x = 1$, is:

$$
\begin{aligned}
P_n(x) &= (x-1) + \frac{-1}{2}(x-1)^2 + \frac{2}{3!}(x-1)^3 + \frac{-3!}{4!}(x-1)^4 + \cdots + \frac{(-1)^{n+1}(n-1)!}{n!}(x-1)^n \\
&= (x-1) - \frac{(x-1)^2}{2} + \frac{(x-1)^3}{3} - \frac{(x-1)^4}{4} + \cdots + (-1)^{n+1}\frac{(x-1)^n}{n}
\end{aligned}
$$

A somewhat neater and more user-friendly polynomial for approximating the logarithm function is obtained by letting $t = x - 1$. Then $x = 1 + t$, and we have the following Taylor polynomial for $\ln(1+t)$, about $t = 0$:

$$
t - \frac{t^2}{2} + \frac{t^3}{3} - \frac{t^4}{4} + \cdots + (-1)^{n+1}\frac{t^n}{n}.
$$

◊

7.4 Taylor's formula – The remainder term

An approximation is not particularly useful if we do not have some idea of how good the approximation is. If we approximate the value of a function by a value of its Taylor polynomial, then it is important that we are able to say what the error in the approximation might be. We do this by looking at the "remainder term", which is simply the difference between a function and its Taylor polynomial. In the next chapter we will see that the remainder term plays another important role as well.

Definition of remainder term
Given a function $f(x)$ and its Taylor polynomial $P_n(x)$ of order n, the "remainder term", $R_n(x)$, is the difference between $f(x)$ and $P_n(x)$.

$$R_n(x) = f(x) - P_n(x).$$

In our discussion of the remainder term we will restrict our attention to Taylor polynomials about $x = 0$.

Note – The fact that the remainder term is denoted by $R_n(x)$ indicates that the difference between a function and its Taylor polynomial depends on both n and x. The following expression for $R_n(x)$ is called the "Lagrange form" of the remainder. (There are other forms of the remainder which will not concern us here.)

Lagrange form of the remainder

$$R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}x^{n+1} \qquad \text{for some } c \text{ between 0 and } x.$$

Note that the formula for the remainder term $R_n(x)$ is easy to remember, since it is the same as the last term of $P_{n+1}(x)$, with 0 replaced by c. You will need to memorise this formula and be able to use it to estimate errors in various Taylor approximations. If we substitute the expression for $R_n(x)$ into the equation

$$f(x) = P_n(x) + R_n(x)$$

we obtain the formula known as "Taylor's formula":

Taylor's formula

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f^{(3)}(0)}{3!}x^3 + \cdots + \frac{f^{(n)}(0)}{n!}x^n + \frac{f^{(n+1)}(c)}{(n+1)!}x^{n+1}$$

for some c between 0 and x.

If we let $n = 0$ in Taylor's formula, we have

$$f(x) = f(0) + f'(c)x, \text{ or } f'(c) = \frac{f(x) - f(0)}{x - 0} \quad \text{for some } c \text{ between 0 and } x.$$

This is precisely the Mean Value Theorem (see Section 5.6). In fact, Taylor's formula can be considered as a generalised form of the Mean Value Theorem.

7.5 How good is the Taylor polynomial approximation?

Note that the Mean Value Theorem **does not** tell us how to find the value of c, and it is almost always impossible to do so. In other words, we cannot expect to find an exact value for the remainder. Often, however, we can find a *bound* on the value of the remainder that will tell us how good (or bad) the Taylor polynomial approximation is.

Example 7.5a Suppose we want to estimate the value of $\cos(1)$ using $P_4(1)$, where $P_4(x)$ is the Taylor polynomial of order 4 for $\cos(x)$ about $x = 0$. Find an upper bound for the error, ie., find a bound for the remainder term.

Using Taylor's formula we obtain $\cos(x) = P_4(x) + R_4(x)$ where $P_4(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!}$ and

$$R_4(x) = f^{(5)}(c)\frac{x^5}{5!} = -\sin(c)\frac{x^5}{5!} \quad \text{for} \quad 0 < c < x.$$

Since we know that $|\sin(c)| \leq 1$ for any value of c, taking absolute values gives

$$|R_4(x)| = \left|\sin(c)\frac{x^5}{5!}\right| \leq \left|\frac{x^5}{5!}\right| = \left|\frac{x^5}{120}\right|.$$

Letting $x = 1$ gives

$$|R_4(1)| \leq \frac{1}{120} \approx 0.0083$$

and therefore

$$\cos(1) = P_4(1) + R_4(1) = 1 - \frac{1}{2!} + \frac{1}{4!} + R_4(1) \approx 0.54166 \pm 0.00833,$$

or equivalently

$$0.53333 < \cos(1) < 0.54999.$$

The exact value from a calculator is $\cos(1) = 0.540302$. $\Diamond$

Taylor polynomials are also useful in calculating approximate values of definite integrals in cases where standard methods of finding anti-derivatives fail. Polynomials are easy to integrate, and we can use the remainder term to estimate the maximum possible error in the approximation.

Example 7.5b In this example we estimate the value of the definite integral $\int_0^1 e^{x^2}\,dx$ using a Taylor polynomial. Note that there is no simple function which is an anti-derivative of e^{x^2}, and so it is not possible to evaluate this integral by anti-differentiation.

The Taylor polynomial of degree 5 for e^x is $1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\frac{x^4}{4!}+\frac{x^5}{5!}$, so Taylor's Formula gives

$$e^x = 1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\frac{x^4}{4!}+\frac{x^5}{5!}+R_5(x),$$

where $R_5(x) = e^c\,\dfrac{x^6}{6!}$ for some c between 0 and x.

This is an identity, true for all values of x, and so we can replace x by x^2 to obtain

$$e^{x^2} = 1+x^2+\frac{x^4}{2!}+\frac{x^6}{3!}+\frac{x^8}{4!}+\frac{x^{10}}{5!}+R_5(x^2),$$

and $R_5(x^2) = e^c\,\dfrac{x^{12}}{6!}$ for some c between 0 and x^2.

Integrating both sides of this equation we have

$$\begin{aligned}
\int_0^1 e^{x^2}\,dx &= \int_0^1 \left(1+x^2+\frac{x^4}{2!}+\frac{x^6}{3!}+\frac{x^8}{4!}+\frac{x^{10}}{5!}+R_5(x^2)\right)dx \\
&= \left[x+\frac{x^3}{3}+\frac{x^5}{5\times 2!}+\frac{x^7}{7\times 3!}+\frac{x^9}{9\times 4!}+\frac{x^{11}}{11\times 5!}\right]_0^1 + \int_0^1 R_5(x^2)\,dx \\
&= 1+\frac{1}{3}+\frac{1}{10}+\frac{1}{42}+\frac{1}{216}+\frac{1}{1320}+\int_0^1 R_5(x^2)\,dx \\
&= 1.46253\ldots + \int_0^1 R_5(x^2)\,dx.
\end{aligned}$$

Now, what can be said about the value of $\int_0^1 R_5(x^2)\,dx$? We know that $R_5(x^2) = e^c\,\dfrac{x^{12}}{6!}$ for some c between 0 and x^2, and since we are integrating from 0 to 1, x lies between 0 and 1, as does x^2. So c also lies between 0 and 1, and hence $1 = e^0 < e^c < e^1 < 3$. Therefore

$$\int_0^1 \frac{x^{12}}{6!}\,dx < \int_0^1 R_5(x^2)\,dx < 3\int_0^1 \frac{x^{12}}{6!}\,dx,$$

and so

$$0.0001 < \left[\frac{x^{13}}{13\times 6!}\right]_0^1 < \int_0^1 R_5(x^2)\,dx < 3\left[\frac{x^{13}}{13\times 6!}\right]_0^1 < 0.0004.$$

We may therefore conclude that $\int_0^1 e^{x^2}\,dx = 1.463$ correct to 3 decimal places. $\Diamond$

7.6 Proof of the remainder formula

The proof of the Lagrange form of the remainder relies on the Mean Value Theorem given in Chapter 4. It is not part of the examinable material of the course but it is given here because it is a very important application of the MVT.

We define an unlikely looking function $g(t)$ as follows:

$$g(t) = f(x) - f(t) - f'(t)(x-t) - \frac{f''(t)}{2!}(x-t)^2 - \cdots\cdots$$
$$\cdots\cdots - \frac{f^{(n)}(t)}{n!}(x-t)^n - R_n(x)\frac{(x-t)^{n+1}}{x^{n+1}}.$$

The introduction of this particular function g in the proof probably seems somewhat mysterious. However, it was carefully chosen so that after applying the Mean Value Theorem, we will obtain the formula we want.

First, we regard x as fixed at some non-zero value, and show that $g(x) = 0$ and $g(0) = 0$. Substituting x for t into the formula for $g(t)$ gives

$$g(x) = f(x) - f(x) - 0 - 0 - \cdots\cdots - 0 = 0.$$

Substituting 0 for t into the formula for $g(t)$ gives

$$g(0) = f(x) - f(0) - f'(0)x - \frac{f''(0)}{2!}x^2 - \cdots\cdots - \frac{f^{(n)}(0)}{n!}x^n - R_n(x)\frac{x^{n+1}}{x^{n+1}},$$

that is,

$$g(0) = f(x) - \left(f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \cdots\cdots + \frac{f^{(n)}(0)}{n!}x^n \right) - R_n(x),$$

in other words,

$$g(0) = f(x) - P_n(x) - R_n(x) = 0.$$

Now differentiate $g(t)$ with respect to t (treating x as a constant):

$$\begin{aligned} g'(t) =& 0 - f'(t) - \big(f''(t)(x-t) - f'(t)\big) \\ & - \left(\frac{f'''(t)}{2!}(x-t)^2 - \frac{f''(t)}{2!}2(x-t)\right) \\ & - \left(\frac{f^{(4)}(t)}{3!}(x-t)^3 - \frac{f'''(t)}{3!}3(x-t)^2\right) \\ & - \cdots - \left(\frac{f^{(n+1)}(t)}{n!}(x-t)^n - \frac{f^{(n)}(t)}{n!}n(x-t)^{n-1}\right) \\ & + R_n(x)\frac{(n+1)(x-t)^n}{x^{n+1}} \end{aligned}$$

Most terms in this expression cancel out in pairs, and we are left with

$$g'(t) = R_n(x)\frac{(n+1)(x-t)^n}{x^{n+1}} - \frac{f^{(n+1)}(t)}{n!}(x-t)^n.$$

Now, since $g(x) = 0$ and $g(0) = 0$ we can apply the Mean Value Theorem. This guarantees that there exists some number c between 0 and x such that $g'(c) = 0$. That is,

$$0 = R_n(x)\frac{(n+1)(x-c)^n}{x^{n+1}} - \frac{f^{(n+1)}(c)}{n!}(x-c)^n$$

and solving for $R_n(x)$ gives

$$R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}x^{n+1}.$$

This completes the proof of the formula for the remainder term.

Summary of Chapter 7

- **Taylor polynomials** are introduced to approximate functions **near a point** $x = a$ in terms of the values of the function and its derivatives evaluated at that point.
- **The main reason** for approximating functions by polynomials is that polynomials are easy to calculate, requiring only addition and multiplication.
- **The remainder term** is derived to be able to give an idea of the error occured in the approximation. In other words, it will be able to tell us **how good the approximation is**.

Exercises

7.1 Suppose $f(x) = \sin x$ and we want a straight-line approximation to $f(x)$ accurate to within 0.05 on an interval $[0,h]$. Use a calculator to find how big h can be (to within 0.01 say) for:

a) The constant approximation (that is, the Taylor polynomial of order zero).
b) The approximation by the tangent line at $x = 0$ (that is, the Taylor polynomial of order one).

7.2 Calculate the first 5 nonzero terms of the Taylor series for $\tan x$ at $x = 0$.

7.3 Calculate the Taylor polynomial of order 4 at $x = 0$ for the function $f(x) = e^{-x^2}$.

7.4 What happens when you calculate the Taylor polynomial at $x = 0$ of $f(x)$, where f is already a polynomial in x? Experiment to see what happens in the following examples:

a) $(1+x)^2$
b) $(1+x)^3$
c) $(1+x)^n$, for n a positive integer.

7.5 Compute the first 4 nonzero terms of the Taylor polynomials for the following functions

a) $\ln(1+x)$
b) $\sec x$
c) $\sinh x$
d) $\cosh x$
e) $\sqrt{1+x}$

7.6 Find the Taylor polynomial of order 3, about $x = 1$, for $x^3 - 3x^2 + 3x - 1$.

7.7 Show that e^x can be approximated to within 0.025 by the polynomial

$$1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!}$$

for all x in the interval $[-1, 1]$. (You may assume $e < 3$.)

7.8 Show that $\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + R(x)$, where $|R(x)| \le \frac{|x|^6}{6!}$.

The next four exercises concern the **Mean Value Theorem (MVT)**, which is not examinable; nevertheless we encourage you to try these problems.

7.9 Use the Mean Value Theorem to show that if $f(0) = 0$ and $f'(x) \ge 7$ for all $x \ge 0$, then $f(x) \ge 7x$ for all $x \ge 0$.

7.10 Suppose that f is a differentiable function with $f(0) = 200$ and $f'(x) < \frac{1}{2}$ for all x. Use the Mean Value Theorem to show that $f(1000) < 700$.

7.11 Use the Mean Value Theorem to show that there is at least one point on the graph of

$$y = x^7 + 3x^4 - 4x + 5$$

between $x = 0$ and $x = 1$ where the tangent to the curve is horizontal.

7.12 Use the Mean Value Theorem to prove the following tests for increasing and decreasing functions that were referred to in Section 6.2.

a) Suppose f is a differentiable function such that $f'(x) > 0$ for all x in some open interval I. Prove that f is an increasing function on I. (That is, for any two points $a, b \in I$ such that $a < b$, use the MVT to prove that $f(a) < f(b)$.)

b) Similarly, prove that if $f'(x) < 0$ for all x in some open interval I then f is a decreasing function on I.

CHAPTER 8

Taylor Series

This chapter begins with an extremely brief introduction to the theory of infinite series. Infinite series are important in many areas of mathematics, and the associated theory is extensive. Our introduction to the theory is just sufficient to allow us to extend the idea of a Taylor polynomial to “infinite polynomials”, or Taylor series, that were mentioned in the previous chapter. Using the remainder term we are able to show that some functions $f(x)$ are equal to their Taylor series for all values of x. The complex exponential function is defined as a series, and this leads us to revisit Euler’s formula which we first met in Chapter 2.

8.1 Infinite series

An "infinite sequence" is simply an infinite list of numbers $a_0, a_1, a_2, \ldots, a_n, \ldots$.
If we add up the terms in such a sequence we obtain an expression of the form

$$a_0 + a_1 + a_2 + \ldots + a_n + \ldots = \sum_{n=0}^{\infty} a_n.$$

We call such an expression an "infinite series" (or simply a "series"). It is not immediately clear, however, what the sum of infinitely many numbers might mean.

A familiar example is the geometric series

$$1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots.$$

You will have learnt in high school that this series is equal to 2, and if you start adding up the terms $1 + \frac{1}{2} + \frac{1}{4} + \cdots$ you will see that you quite quickly obtain a number close to 2.

However, it is not hard to see that each successive term is exactly half the difference between 2 and the sum of all the preceding terms, so that the sum will never actually reach 2. No matter how many terms you have managed to add together, there are still infinitely many left! So what do we mean, exactly, when we say

$$1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots = 2?$$

We need a definition, and for this purpose we introduce the idea of partial sums.

Partial sum of an infinite series
The kth "partial sum", S_k, of the series $\sum_{n=0}^{\infty} a_n$ is the sum of all the terms from a_0 to a_k.

$$S_k = a_0 + a_1 + a_2 + \ldots + a_k = \sum_{n=0}^{k} a_n.$$

For any series we therefore have a sequence of partial sums,

$$\begin{aligned} S_0 &= a_0 \\ S_1 &= a_0 + a_1 \\ S_2 &= a_0 + a_1 + a_2 \\ &\vdots \\ S_k &= a_0 + a_1 + a_2 + \ldots + a_k, \end{aligned}$$

and we define the sum $\sum_{n=0}^{\infty} a_n$ as the limit of these partial sums S_k as $k \to \infty$, if it exists.

The sum of an infinite series
The "sum" of the infinite series $\sum_{n=0}^{\infty} a_n$ is the limit, as $k \to \infty$, of the partial sums $S_k = \sum_{n=0}^{k} a_n$, provided this limit exists,

$$\sum_{n=0}^{\infty} a_n = \lim_{k\to\infty} S_k = \lim_{k\to\infty} \sum_{n=0}^{k} a_n.$$

If the sequence of partial sums fails to converge to a limit we say that the series **diverges.**

When the limit of the partial sums S_k exists and is equal to L, we say that the series "converges" to L and we write

$$\sum_{n=0}^{\infty} a_n = L.$$

For example, we show below that the infinite series $1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots$ converges to 2.

The series $1 + 1 + 1 + 1 + \cdots$, for example, clearly increases indefinitely and hence diverges.

Geometric series

Consider the geometric series $1+r+r^2+r^3+\cdots$ with common ratio r.

The kth partial sum of this series is the geometric progression

$$S_k = \sum_{n=0}^{k} r^n = 1+r+r^2+r^3+\cdots+r^k. \tag{8.1a}$$

Now, multiply both sides by r:

$$r\,S_k = r+r^2+r^3+\cdots+r^k+r^{k+1}, \tag{8.1b}$$

and subtract (8.1b) from (8.1a), noting that all terms except the first and the last cancel,

$$S_k - r\,S_k = 1 - r^{k+1}.$$

Finally, solving for S_k gives the expression for the sum of the first k terms as

$$S_k = \frac{1-r^{k+1}}{1-r}.$$

When r is a real number such that $-1 < r < 1$, $r^{k+1} \to 0$ as $k \to \infty$, and so

$$\begin{aligned} 1+r+r^2+r^3+\cdots &= \lim_{k\to\infty}(1+r+r^2+r^3+\cdots+r^k) \\ &= \lim_{k\to\infty}\frac{1-r^{k+1}}{1-r} \\ &= \frac{1}{1-r}. \end{aligned}$$

However, if $|r| > 1$, the magnitude of r^{k+1} increases without bound as k increases, and $\lim\limits_{k\to\infty}\dfrac{1-r^{k+1}}{1-r}$ does not exist.

When $r=1$, the series is the divergent series $1+1+1+\cdots$.

When $r=-1$, we have the series $1-1+1-1+1-1+\cdots$. Consideration of the partial sums in this case should convince you that this series diverges.

So the geometric series $1+r+r^2+r^3+\cdots$ converges to $\dfrac{1}{1-r}$ if $|r| < 1$, and diverges otherwise.

Example 8.1c Show that the series $1+\dfrac{1}{2}+\dfrac{1}{4}+\dfrac{1}{8}+\ldots$ converges to 2.

The series may be written in the form $1+\dfrac{1}{2}+\left(\dfrac{1}{2}\right)^2+\left(\dfrac{1}{2}\right)^3+\ldots$ which shows that it is a geometric series with common ratio $r=\dfrac{1}{2}$ and therefore its sum is $S=\dfrac{1}{1-r}=\dfrac{1}{1-1/2}=2$.

◊

Warning: It is tempting to think that a series will converge if the terms in the series approach zero. In general, this is *not* true. For example, the infinite series

$$\sum_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots$$

diverges, that is, we can make the sum as large as we like by taking a sufficiently large number of terms.

8.2 Taylor series

We have been considering series of constant terms which, when they converge, sum to a real number. In this section we look at series with variable terms which, if they converge, sum to a function. The importance of the role of infinite series in mathematics is largely due to the fact that many functions have a representation as a series.

We have, in fact, already seen such a representation. In Section 8.1 we saw the formula for the sum of a geometric series:

$$1 + r + r^2 + r^3 + r^4 \cdots = \frac{1}{1-r} \text{ for } |r| < 1.$$

Now consider this result from a slightly different point of view, by thinking of r as a variable. Let's replace r by x to reinforce this view, and turn the formula around, so we have

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + x^4 \cdots \qquad \text{for } |x| < 1.$$

Now we have a "power series", $1 + x + x^2 + x^3 + x^4 \cdots$, which is equal to the function $\dfrac{1}{1-x}$ for $|x| < 1$. Notice that this power series is precisely the series we would obtain by extending indefinitely the Taylor polynomial of $\dfrac{1}{1-x}$.

We can find the Taylor series of any function simply by finding the Taylor polynomial, and extending it to an infinite number of terms. The function must, of course, have derivatives of any order at the point about which the polynomial is found.

Taylor series
The "Taylor series" for a function f about $x = 0$ is

$$f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \cdots\cdots = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!}x^k$$

Taylor series about $x = 0$ are known as **Maclaurin series**

Examples 8.2a These examples use the polynomials we found in the previous chapter. All the series are about the point $x = 0$.

i) The Taylor series for $\cos x$: $\quad 1 - \dfrac{x^2}{2!} + \dfrac{x^4}{4!} - \dfrac{x^6}{6!} + \cdots.$

ii) The Taylor series for $\sin x$: $\quad x - \dfrac{x^3}{3!} + \dfrac{x^5}{5!} - \dfrac{x^7}{7!} + \cdots.$

iii) The Taylor series for e^x: $\quad 1 + x + \dfrac{x^2}{2!} + \dfrac{x^3}{3!} + \dfrac{x^4}{4!} + \cdots.$

Try differentiating, term-by-term, each of the series above. What do you notice?

iv) The Taylor series for $\dfrac{1}{1-x}$: $\quad 1 + x + x^2 + x^3 + x^4 + \cdots.$

Try multiplying $(1 - x)$ by $1 + x + x^2 + x^3 + x^4 + \cdots$. What do you find?

v) The Taylor series for $\ln(1 + x)$: $\quad x - \dfrac{x^2}{2} + \dfrac{x^3}{3} - \dfrac{x^4}{4} + \cdots.$

vi) The Taylor series for $\cosh x$: $\quad 1 + \dfrac{x^2}{2!} + \dfrac{x^4}{4!} + \dfrac{x^6}{6!} + \cdots.$

vii) The Taylor series for $\sinh x$: $\quad x + \dfrac{x^3}{3!} + \dfrac{x^5}{5!} + \dfrac{x^7}{7!} + \cdots.$ ◊

Simply knowing the Taylor series of a function f is not particularly useful. We need to know the values of x for which the series converges to $f(x)$. The answer to this question depends on the function.

For example, we have seen above that the Taylor series for $\dfrac{1}{1-x}$ converges to $\dfrac{1}{1-x}$ for values of x such that $|x| < 1$.

For some well-behaved functions f the Taylor series converges to $f(x)$ for every value of x. The functions $\sin x$, $\cos x$ and e^x all fall into this category. In order to see this, we consider the behaviour of the remainder term, $R_n(x)$, as $n \to \infty$.

The remainder term

Recall that the remainder term $R_n(x) = f(x) - P_n(x)$, where $P_n(x)$ is the n^{th} order Taylor polynomial of f. Rewrite the equation as $f(x) = P_n(x) + R_n(x)$, consider x as fixed at some particular value, and take the limit as $n \to \infty$. We therefore have

$$f(x) = \lim_{n\to\infty} P_n(x) + \lim_{n\to\infty} R_n(x).$$

Hence, if $\lim_{n\to\infty} R_n(x) = 0$, then $f(x) = \lim_{n\to\infty} P_n(x)$. Now, $\lim_{n\to\infty} P_n(x)$ is the Taylor series of f, and so we have the following result:

Convergence of Taylor series
If $\lim_{n\to\infty} R_n(x) = 0$ for a particular value of x, then the Taylor series of f converges to f at that point.

For the functions $\sin x$, $\cos x$ and e^x we are able to show that $\lim_{n\to\infty} R_n(x) = 0$ for all x.

Example 8.2b Recall that the remainder term $R_n(x)$ is equal to $\dfrac{f^{(n+1)}(c)}{(n+1)!}x^{n+1}$ for some c between 0 and x.

For the function $f(x) = \sin x$, every derivative is one of $\sin x$, $\cos x$, $-\sin x$ or $-\cos x$. Therefore, no matter what the value of c, or of n, $-1 \le f^{(n+1)}(c) \le 1$, or $|f^{(n+1)}(c)| \le 1$. Hence, the remainder term for $f(x) = \sin x$ is such that

$$|R_n(x)| \le \left|\frac{x^{n+1}}{(n+1)!}\right|.$$

Now, $\lim_{n\to\infty} \dfrac{x^{n+1}}{(n+1)!} = 0$ for any real number x (see the technical aside below). It follows (by the squeeze law) that $\lim_{n\to\infty} R_n(x) = 0$, and so

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots \text{ for all } x \in \mathbb{R}.$$

Precisely the same argument applies to the function $\cos x$, and

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots \text{ for all } x \in \mathbb{R}.$$

$\diamondsuit$

Technical aside We want to show here that $\lim_{n\to\infty} \dfrac{x^n}{n!} = 0$ for any real number x. Since x can be positive or negative, it will be convenient to consider $\left|\dfrac{x^n}{n!}\right| = \dfrac{|x|^n}{n!}$.

Choose an integer k such that $k > |x|$. Then for $n > k$ we can write

$$\frac{|x|^n}{n!} = \frac{|x|}{1}\cdot\frac{|x|}{2}\cdots\frac{|x|}{k}\cdot\frac{|x|}{(k+1)}\cdot\frac{|x|}{(k+2)}\cdots\frac{|x|}{(n-1)}\cdot\frac{|x|}{n}.$$

The product of the first k terms is just some finite number; let's call it K. Since $k > |x|$, each of the terms

$$\frac{|x|}{(k+1)}, \frac{|x|}{(k+2)}, \ldots, \frac{|x|}{(n-1)}$$

is less than 1. Therefore,

$$0 \leq \frac{|x|^n}{n!} < \frac{K|x|}{n}.$$

Since $\lim_{n\to\infty} \frac{K|x|}{n} = K|x| \lim_{n\to\infty} \frac{1}{n} = 0$, it follows (by the squeeze law) that $\lim_{n\to\infty} \frac{|x|^n}{n!} = 0$ and hence

$$\lim_{n\to\infty} \frac{x^n}{n!} = 0 \text{ for any real number } x.$$

◁

Example 8.2c Now consider the remainder term for the function $f(x) = e^x$. Since any derivative of e^x is e^x, we have $f^{(n+1)}(c) = e^c$, and

$$R_n(x) = \frac{e^c x^{n+1}}{(n+1)!} \qquad \text{for some } c \text{ between 0 and } x.$$

If x is negative, and c is between 0 and x, then $0 < e^c < 1$ and $0 < |R_n(x)| < \frac{|x|^{n+1}}{(n+1)!}$.

Since $\lim_{n\to\infty} \frac{|x|^{n+1}}{(n+1)!} = 0 \implies \lim_{n\to\infty} |R_n(x)| = 0$.

If x is positive, $1 < e^c < e^x$ (since e^x is an increasing function) and $0 < |R_n(x)| < \frac{e^x|x|^{n+1}}{(n+1)!}$.

Now, $\lim_{n\to\infty} \frac{e^x|x|^{n+1}}{(n+1)!} = e^x \lim_{n\to\infty} \frac{|x|^{n+1}}{(n+1)!} = 0$, and once again $\lim_{n\to\infty} |R_n(x)| = 0$.

So the remainder term tends to zero as $n \to \infty$, and hence the Taylor series for e^x converges to the function e^x for all x. That is,

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!} \text{ for all real } x.$$

Substituting the number 1 for x we obtain a series of constant terms for the number e itself:

Series for the number e

$$e = 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \cdots = \sum_{n=0}^{\infty} \frac{1}{n!}$$

The 10th partial sum gives

$$e \approx 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{10!} \approx 2.7182818,$$

which is accurate to the number of decimal places shown. ◊

The series representations of the functions e^x, $\sin x$ and $\cos x$ should be remembered. Here they are again:

Series of elementary functions

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots$$
$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots$$
$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots$$

It is important to realise that these equations are *identities* that hold for all values of x. In other words, for each of the equations in the box above, the function on the left is *exactly* equal to the sum on the right.

8.3 Euler's formula

In this section we will briefly discuss series of complex terms. Our treatment of complex series will not be rigorous because our sole aim is to show how Euler's formula (which expresses the exponential function in terms of the sine and cosine functions) can be discovered using infinite series. If you have forgotten about Euler's formula it would be a good idea to revise Chapter 2 before reading this section.

The infinite series of complex numbers $c_n = a_n + ib_n$ is defined by

$$\sum_{n=0}^{\infty} c_n = \sum_{n=0}^{\infty} a_n + i \sum_{n=0}^{\infty} b_n.$$

Provided the two series of real terms on the right-hand side of this equation converge, we say that the series of complex terms converges.

It can be shown that the exponential series

$$1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots$$

converges whenever the real number x is replaced by a complex number $x + iy$. (We are not going to prove this here; it is beyond the scope of this course and you will just have to accept it as a fact.) Consequently, we can define the exponential function e^z for complex as well as real numbers, by the series:

The complex exponential function

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!} = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \frac{z^4}{4!} + \cdots$$

Substituting the purely imaginary number $z = i\theta$ (where θ is real) into this infinite series formula for e^z, and using the fact that $i^2 = -1$, we obtain

$$\begin{aligned} e^{i\theta} &= 1 + i\theta + \frac{(i\theta)^2}{2!} + \frac{(i\theta)^3}{3!} + \frac{(i\theta)^4}{4!} + \cdots \\ &= 1 + i\theta - \frac{\theta^2}{2!} - \frac{i\theta^3}{3!} + \frac{\theta^4}{4!} + \frac{i\theta^5}{5!} + \cdots \\ &= \left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \cdots\right) + i\left(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \frac{\theta^7}{7!} + \cdots\right). \end{aligned}$$

But for any real θ, we have seen that

$$\cos\theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \cdots \quad \text{and} \quad \sin\theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \frac{\theta^7}{7!} + \cdots$$

Hence the expression for $e^{i\theta}$ given above reduces to the famous "Euler's formula".

Euler's formula $\quad e^{i\theta} = \cos\theta + i\sin\theta \quad$ for all real θ.

We now have two ways to think about the complex exponential function e^z: as an infinite series of powers of z and as a concise formula in terms of the real and imaginary parts of z. If $z = x + iy$ then

$$e^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \frac{z^4}{4!} + \cdots = e^x(\cos y + i\sin y).$$

8.4 The binomial series

One of the most useful series, particularly in applied mathematics, is the "binomial series". It is the Taylor series of the function $(1+x)^p$, where p is any real number, and is a generalisation of the binomial theorem that we revise in Chapter 12.

To find the Taylor series, about 0, of $(1+x)^p$, we first list the derivatives. and evaluate them at 0:

$$\begin{aligned}
f(x) &= (1+x)^p \\
f'(x) &= p(1+x)^{p-1} \\
f''(x) &= p(p-1)(1+x)^{p-2} \\
f'''(x) &= p(p-1)(p-2)(1+x)^{p-3} \\
&\vdots \\
f^{(n)}(x) &= p(p-1)(p-2)\cdots(p-n+1)(1+x)^{p-n}
\end{aligned}$$

Now evaluate the derivatives at $x = 0$.

$$\begin{aligned}
f(0) &= 1 \\
f'(0) &= p \\
f''(0) &= p(p-1) \\
f'''(0) &= p(p-1)(p-2) \\
&\vdots \\
f^{(n)}(0) &= p(p-1)(p-2)\cdots(p-n+1)
\end{aligned}$$

The Taylor series for $(1+x)^p$, also known as the "binomial series" is, therefore:

Binomial series

$$\begin{aligned}
1 + px + \frac{p(p-1)}{2!}x^2 + \frac{p(p-1)(p-2)}{3!}x^3 + \cdots\cdots \\
= 1 + \sum_{n=1}^{\infty} \frac{p(p-1)(p-2)\cdots(p-n+1)}{n!}x^n
\end{aligned}$$

$\triangleright$ **Aside** The above series is *not* always equal to $(1+x)^p$. In fact, for all values of p, the series converges to $(1+x)^p$ whenever $|x| < 1$, and diverges if $|x| > 1$. The proof of this is beyond us at this stage. Whether or not the series converges for $x = \pm 1$ depends on the value of p. $\triangleleft$

Examples 8.4a

i) When $p=-1$, we obtain the following series for $(1+x)^{-1}=\dfrac{1}{1+x}$:

$$\begin{aligned}&1+(-1)x+\frac{(-1)(-2)}{2!}x^2+\frac{(-1)(-2)(-3)}{3!}x^3+\frac{(-1)(-2)(-3)(-4)}{4!}x^4+\cdots\\&=1-x+x^2-x^3+x^4+\cdots\end{aligned}$$

Compare this series with that for $(1-x)^{-1}$ in Example 8.2a.

ii) A special case occurs when p is a positive integer. In this case only the first p derivatives of $f(x)$ are non-zero, since the factor $(p-p)=0$ appears in the Taylor series coefficient of x^{p+1} and subsequent terms.

For example, with $p=3$ the binomial series gives

$$(1+x)^3=1+3x+\frac{3\cdot 2}{2!}x^2+\frac{3\cdot 2\cdot 1}{3!}x^3+\frac{3\cdot 2\cdot 1\cdot 0}{4!}x^4+\cdots.$$

The coefficient of x^4 is zero, as are the coefficients of all higher powers of x. So we have the familiar binomial formula

$$(1+x)^3=1+3x+3x^2+x^3$$

in this case.

In general, the coefficient of x^n in the Taylor series for $(1+x)^p$ is

$$\frac{p(p-1)(p-2)\cdots(p-n+1)}{n!}$$

and in the case that p is a positive integer this can also be written as the binomial coefficient

$$\frac{p!}{n!\times(p-n)!}=\binom{p}{n},$$

and the binomial series reduces to the binomial theorem:

$$(1+x)^p=\sum_{n=0}^{p}\binom{p}{n}x^n.$$

iii) Show that the series for $\dfrac{1}{\sqrt{1-x}}=(1+(-x))^{-\frac{1}{2}}$ is

$$1+\sum_{n=1}^{\infty}\frac{1\cdot 3\cdot 5\cdots(2n-1)}{2\cdot 4\cdot 6\cdots(2n)}x^n.$$

◊

8.5 A series for the inverse tan function

Sometimes it is not convenient to find the higher order derivatives of a function in order to find its Taylor series. For example, try finding the first few derivatives of $\tan^{-1} x$.

You should see that the derivatives are going to be very messy after the third derivative. In this section we see an alternative method of finding the Taylor series of $\tan^{-1} x$.

We note that
$$\frac{d}{dx}\tan^{-1} x = \frac{1}{1+x^2} \implies \int \frac{1}{1+x^2}\,dx = \tan^{-1} x + C,$$
where C is an arbitrary constant.

Now, in Example 8.4a, we saw that the power series for $\dfrac{1}{1+x}$ is

$$1 - x + x^2 - x^3 + x^4 - x^5 + \cdots .$$

Indeed, this is just a geometric series, and we know that

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + x^4 - x^5 + \cdots \quad \text{for } |x| < 1.$$

Replacing x by x^2, we have

$$\frac{1}{1+x^2} = 1 - x^2 + x^4 - x^6 + x^8 - x^{10} + \cdots \quad \text{for } |x| < 1.$$

Now integrate this series term-by-term to obtain the following series for $\tan^{-1} x$. We have

$$\tan^{-1} x + C = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \frac{x^9}{9} - \frac{x^{11}}{11} + \cdots \quad \text{for } |x| < 1.$$

Substituting $x = 0$ into both sides gives $C = 0$, and so

A series for $\tan^{-1} x$

$$\tan^{-1} x = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \frac{x^9}{9} - \frac{x^{11}}{11} + \cdots \quad \text{for } |x| < 1.$$

What we have done here is to take the infinite series which converges to $\dfrac{1}{1+x^2}$ for $|x| < 1$, integrate it term-by-term, and claim that the resulting series converges to the integral of $\dfrac{1}{1+x^2}$ for $|x| < 1$. It is by no means obvious that this is a legitimate thing to do. It is indeed true, however, that the series shown above does converge to $\tan^{-1}$ for $|x| < 1$. What's more, it converges for $|x| \leq 1$.

Substituting $x = 1$ into the equation $\tan^{-1} x = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \frac{x^9}{9} - \frac{x^{11}}{11} + \cdots$ gives a quite remarkable formula:

A series for $\pi/4$

$$\tan^{-1} 1 = \frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \cdots\cdots$$

While this is very nice series for $\pi/4$, it is not an efficient way to compute π, since one needs to take a very large number of terms before π is calculated with any degree of accuracy.

Summary of Chapter 8

- **Infinite sequences** and **infinite series** of numbers are briefly dealt with, the aim being to apply these ideas to extend the definition of Taylor polynomials to **infinite polynomials** or **Taylor series**.
- **Taylor series** are used to expand the exponential, trigonometric and other elementary functions.
- **Taylor series** are then used to prove **Euler's formula**

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

- **Term-by-term** integration of the convergent Taylor series of a function $f(x)$ sometimes leads to the series of the integral of $f(x)$. We used this result to calculate the Taylor series of $\tan^{-1} x$ from the series of its derivative: $(1+x)^{-1} = 1 - x + x^2 - x^3 + x^4 - x^5 + \cdots \quad$ for $|x| < 1$.

Exercises

8.1 Show that the Taylor series for $\sinh x$ converges to $\sinh x$ for all real x, by showing that the remainder term tends to zero as $n \to \infty$.

8.2 Write the repeating decimal $0.555555\ldots$ as the sum of a geometric series, and hence as a rational number m/n.

8.3 Write the repeating decimal $0.525252\ldots$ as the sum of a geometric series, and hence as a rational number m/n.

8.4 Write the following complex numbers in the form $re^{i\theta}$:

a) $w = 1 - i$

b) $z = \sqrt{3} + i$

8.5 Using the complex numbers z and w from the previous exercise, calculate

a) zw and $(zw)^{12}$

b) e^w in Cartesian form.

c) e^z in Cartesian form.

8.6 a) Calculate the binomial series for the function $f(x) = (1-x)^{-1/2}$ about $x = 0$.

b) Hence find a series for $(1-x^2)^{-1/2}$ and then a series for $\sin^{-1} x$ by term-by-term integration.

CHAPTER 9

The Riemann Integral

In this and the next few chapters we develop the theory of *integration* along with some of its applications. One of the main reasons for getting a good understanding of integration is that a major practical use of mathematics is concerned with building 'mathematical models' of physical, biological or financial systems. Typically these models are written in terms of *differential equations*, which express relationships between the *derivatives* of the different quantities in the system and their solution involves applications of the theory of integration.

9.1 Riemann sums – The area problem

Here we study the area problem which is to calculate the area of the region bounded by a curve $y = f(x)$ and the x-axis between two points $x = a$ and $x = b$. The distance problem is considered in Appendix D.

a) Start with a continuous function $f(x)$ defined on a closed interval $[a,b]$ and for simplicity assume the function is *n*on-negative, so $f(x) \geq 0$ for all x in the interval.

b) Fix an integer $N \geq 1$ and divide the interval $[a,b]$ into N subintervals of equal length

$$[x_0, x_1],\ [x_1, x_2] \ldots [x_{i-1}, x_i] \ldots [x_{N-1}, x_N]$$

where $x_0 = a$ and $x_N = b$, as shown in Figure 9.1 This is called a **partition** of the interval $[a,b]$. The length of $[a,b]$ is $b-a$ and therefore the length of each subinterval is

$$\Delta x = \frac{b-a}{N}.$$

c) The partition points are constructed as follows:

$$x_i = a + \Delta x \times i \quad \text{for} \quad i = 0, 1, 2 \ldots N.$$

d) Then take the *minimum* value m_i of $f(x)$ on each subinterval, and draw rectangles of this height based on the subintervals. The result for $N = 8$ subintervals is shown in Figure 9.2.

e) Repeat the construction, this time using the *maximum* value M_i of $f(x)$ on each subinterval. This is shown in Figure 9.3.

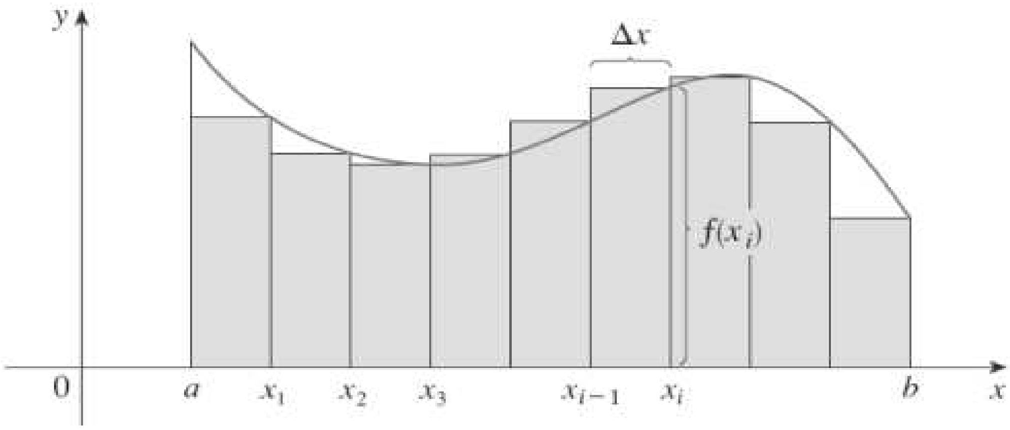


Figure 9.1: **Partition of the interval** $[a,b]$

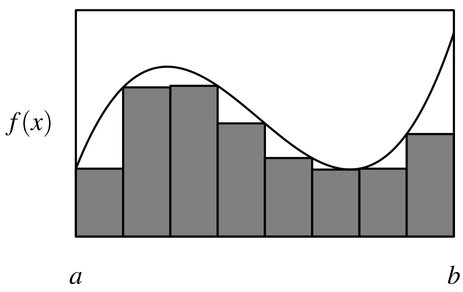


Figure 9.2: Rectangle height = m_i.

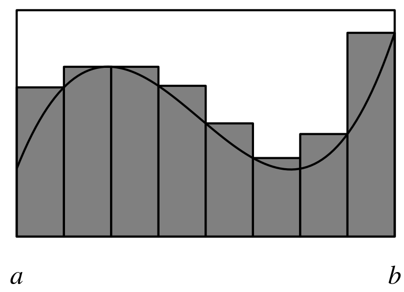


Figure 9.3: Rectangle height = M_i.

Properties

a) In the case of Figure 9.2 the total area of the rectangles is clearly a *lower estimate* for the area under the graph of $f(x)$.

b) Similarly the total shaded area in Figure 9.3 is an *upper estimate* for this area.

c) When the function increases, the minimum value m_i occurs on the left point of the subinterval $[x_{i-1}, x_i]$, that is $m_i = f(x_{i-1})$ and the maximum value M_i occurs on the right point of the subinterval $[x_{i-1}, x_i]$, that is $M_i = f(x_i)$.

d) Viceversa, when the function decreases, the minimum value m_i occurs on the right point of the subinterval $[x_{i-1}, x_i]$, that is $m_i = f(x_i)$ and the maximum value M_i occurs on the left point of the subinterval $[x_{i-1}, x_i]$, that is $M_i = f(x_{i-1})$.

Lower and Upper Riemann sums

In Figure 9.2 the area of the ith rectangle is (height$\times$base) $= m_i \times \Delta x$. Let L_N be the total area of the smaller rectangles. Then

(9.1a)
$$L_N = (m_1 \times \Delta x) + (m_2 \times \Delta x) + \cdots + (m_N \times \Delta x) = \sum_{i=1}^{N} m_i \times \Delta x\,.$$

The number L_N is called a **Lower Riemann Sum** for the function f on the interval $[a,b]$. It depends not only on N, but also on f and the interval $[a,b]$.

In Figure 9.3 the area of the ith rectangle is (height$\times$base) $= M_i \times \Delta x$. Let U_N be the total area of the larger rectangles. Then

(9.1b)
$$U_N = (M_1 \times \Delta x) + (M_2 \times \Delta x) + \cdots + (M_N \times \Delta x) = \sum_{i=1}^{N} M_i \times \Delta x\,.$$

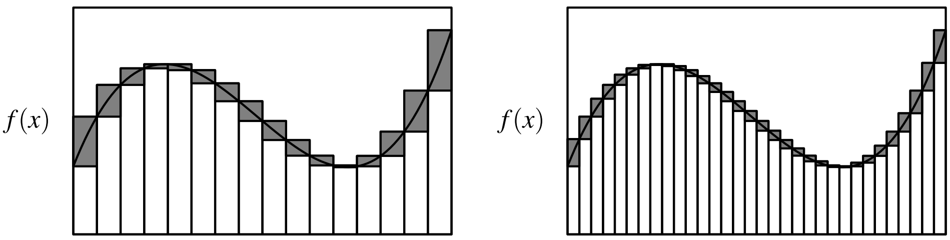
This is called a **Upper Riemann Sum** for f on $[a,b]$.

9.2 The Riemann integral

The Riemann lower and upper sums give us lower and upper estimates for the area under the graph of $f(x)$:

$$L_N \leq \text{Area under the Graph} \leq U_N\,.$$

Reall that Figures 9.2 and 9.3 correspond to $N = 8$ subintervals. Figure 9.4 shows the effect of increasing the number of partition points N, taking first 16 and then 32 subintervals. The



$n = 16$ $n = 32$

Figure 9.4:

shaded area represents the difference between the upper and lower rectangles. The pictures clearly suggest that the difference approaches zero as the number of intervals is increased. In fact, careful analysis (using more sophisticated mathematical ideas than we have available at this stage) confirms this intuition:

As the size of the subintervals is decreased to zero, the upper and lower Riemann sums approach a common value.

This is the number we call the **definite integral or Riemann integral** of $f(x)$ over the interval $[a,b]$.

We can summarize this as follows:

The Riemann Integral

Suppose that $f(x)$ is a continuous function defined on the interval $[a,b]$. For each integer $N \geq 1$ we can divide $[a,b]$ into N equal subintervals and form the associated upper and lower Riemann sums. As $N \to \infty$ both the upper and lower sums approach the same value. This value is called the **Riemann integral** (also called the **definite integral**) of f over the interval $[a,b]$, and is written as

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

It is the unique number which satisfies

$$L_N \leq \int_a^b f(x)\, dx \leq U_N$$

for all $N \geq 1$. Since the area under the graph of $f(x)$ satisfies the same inequalities, it must be equal to the integral of f over the interval. See below for how to interpret the area in the case where f takes negative values.

Since both the upper and lower sums have the same limit as N increases, we can also write

$$\lim_{N\to\infty} L_N = \lim_{N\to\infty} U_N = \int_a^b f(x)\, dx\,.$$

Non-positive Functions

If $f(x)$ takes negative values we must modify the above argument slightly. As before we divide $[a,b]$ into N equal subintervals and let m_i and M_i be the minimum and maximum values of f on the ith subinterval. Now one or both of these numbers may be *negative*.

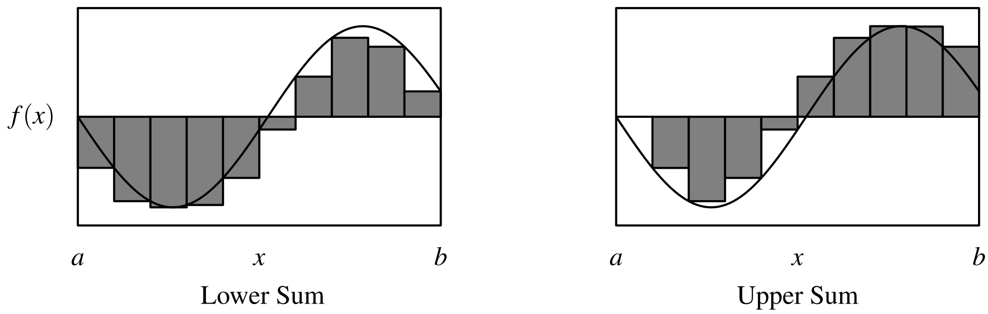


Figure 9.5:

If m_i (or M_i) is *negative* the rectangle appears *below* the axis. The Riemann sum is therefore equal to the sum of the areas of all the rectangles *above* the axis, minus the sum of the areas of all the rectangles below the axis.

It is still true that the lower and upper Riemann sums converge to a common value. The relation between the definite integral and area continues to hold, except that areas below the axis count as negative.

NOTE – In this discussion we have always used a subdivision of the interval of integration into subintervals of equal length. This is sufficient for many practical applications, but from a theoretical point of view there are some advantages in relaxing this condition. If we subdivide into (finitely many) subintervals of possibly different lengths the upper and lower Riemann sums are still defined. It can be proved that both sums converge to the value of the definite integral, as defined above, as the length of the *longest* subinterval is decreased towards zero.

9.3 Calculating Riemann sums

The Riemann sum provides a basic method to calculate approximate values of definite integrals from which other, more accurate techniques are derived. Therefore it is important that we become proficient at calculating Riemann sums.

Recall that if a function $f(x)$ *increases* with x then the minimum value on an interval is always at the left endpoint, and the maximum value at the right. If the function is *decreasing* on the interval then the minimum value on an interval is always at the right endpoint, and the maximum value at the left.

To handle the general situation we have to introduce a more general idea of Riemann sum. The upper and lower Riemann sums will then appear as special cases of this construction. As usual $f(x)$ is a continuous function defined on the interval $[a,b]$.

Suppose $[a,b]$ is divided into N equal subintervals. For $1 \le i \le N$, let c_i be *any* point in the ith subinterval and form the sum

(9.3a)
$$\boxed{(f(c_1)\times\Delta x)+(f(c_2)\times\Delta x)+\cdots+(f(c_N)\times\Delta x)=\sum_{i=1}^{N} f(c_i)\times\Delta x\,.}$$

This is the sum of areas of rectangles of width Δx and height $f(c_i)$. Whatever the choice of the c_i, we certainly have

(9.3b)
$$m_i \le f(c_i) \le M_i$$

since m_i and M_i are the minimum and maximum values of f on this subinterval. Multiplying all sides of 9.3b by Δx, gives

$$m_i\,\Delta x \le f(c_i)\,\Delta x \le M_i\,\Delta x$$

and adding up over all subintervals we obtain

$$\sum_{i=1}^{N} m_i \,\Delta x \le \sum_{i=1}^{N} f(c_i)\,\Delta x \le \sum_{i=1}^{N} M_i \,\Delta x$$

But the sums

$$\sum_{i=1}^{N} m_i \,\Delta x = L_N \quad \text{and} \quad \sum_{i=1}^{N} M_i \,\Delta x = U_N$$

and therefore we get the inequality:

(9.3c)
$$L_N \le \sum_{i=1}^{N} f(c_i)\Delta x \le U_N \,.$$

The middle expression here is an example of a (general) **Riemann Sum** for $f(x)$ on the interval $[a,b]$. Its value obviously depends on the choice of the c_i. By taking N large enough, we can make L_N and U_N as close as we like to the value of the definite integral. The inequalities (9.3c) then imply that *any* Riemann sum must be at least as close to the actual value of the integral. We illustrate this with an example.

Example 9.3d Use Riemann sums to estimate the integral

$$\int_1^2 \sin x \, dx \,.$$

Use a partition of $[a,b] = [1,2]$ into $N = 20$ subintervals and calculate the Riemann sum for each of the three cases:

a) c_i is the left endpoint of the ith subinterval,
b) c_i is the right endpoint of the ith subinterval,
c) c_i is the midpoint of the ith subinterval.

Solution:
Each subinterval has length

$$\Delta x = \frac{b-a}{N} = \frac{2-1}{20} = 1/20 = 0.05.$$

The partition points are given by

(9.3e)
$$x_i = a + \Delta x \times i = 1.0 + 0.05 \times i \quad \text{for} \quad 0 \le i \le 20.$$

The ith subinterval will then be the interval $[x_{i-1}, x_i]$. The three Riemann sums correspond to the choices

(9.3f)
$$c_i = x_{i-1}, \quad c_i = x_i, \quad c_i = x_{i-1} + 0.025 \,.$$

In each case we have to work out the sum

(9.3g)
$$\sum_{i=1}^{20} \sin(c_i)\,\Delta x = \left(\sum_{i=1}^{20} \sin(c_i)\right) \times 0.05 \,.$$

This is quite easy with a programmable calculator or by writing a simple computer program.

i	x_{i-1}	(1)	(2)	(3)
1	1.000000	0.841471	0.867423	0.854714
2	1.050000	0.867423	0.891207	0.879590
3	1.100000	0.891207	0.912764	0.902268
4	1.150000	0.912764	0.932039	0.922690
5	1.200000	0.932039	0.948985	0.940806
6	1.250000	0.948985	0.963558	0.956570
7	1.300000	0.963558	0.975723	0.969944
8	1.350000	0.975723	0.985450	0.980893
9	1.400000	0.985450	0.992713	0.989391
10	1.450000	0.992713	0.997495	0.995415
11	1.500000	0.997495	0.999784	0.998952
12	1.550000	0.999784	0.999574	0.999991
13	1.600000	0.999574	0.996865	0.998531
14	1.650000	0.996865	0.991665	0.994576
15	1.700000	0.991665	0.983986	0.988134
16	1.750000	0.983986	0.973848	0.979223
17	1.800000	0.973848	0.961275	0.967864
18	1.850000	0.961275	0.946300	0.954086
19	1.900000	0.946300	0.928960	0.937923
20	1.950000	0.928960	0.909297	0.919416
		0.954554	0.957946	0.956549

We calculate the Riemann sums by summing the last three columns and using the formula (9.3g). This gives the values shown in the bottom line of the table.

In this example it is easy to calculate the integral exactly:

$$\int_1^2 \sin x \, dx = \int_1^2 \frac{d}{dx}(-\cos x)\ dx = (-\cos 2) - (-\cos 1) \approx 0.956449 .$$

Note that in this example none of the three Riemann sums give the lower or upper Riemann sum. The maximum value of $\sin x$ occurs at $x = \pi/2$, which is inside the range of integration. Up to this point the function is increasing. After this point it is decreasing. In some cases the maximum value of the function occurs at the right of the subinterval, other times on the left, and in one case (at $\pi/2$) inside the subinterval. $\Diamond$

Example 9.3h Let the function $f(x) = x^2$ be given on the interval $[0,1]$. Partition $[0,1]$ into four equal subintervals and calculate:

a) The lower Riemann sum L_4.
b) The upper Riemann sum U_4.
c) The exact value of

$$A = \int_0^1 x^2 \, dx$$

and check that

$$L_4 \leq A \leq U_4 .$$

Solution:
In this problem, $a = 0$, $b = 1$, $N = 4$ therefore

$$\Delta x = \frac{b-a}{N} = 1/4 = 0.25.$$

The partition points are

$$x_i = a + \Delta x \times i = 0.25 \times i \quad \text{for} \quad i = 0, 1, 2, 3, 4$$

that is, $x_0 = 0,\ x_1 = 0.25,\ x_2 = 0.50,\ x_3 = 0.75,\ x_4 = 1$ and therefore the subintervals are

$$[0, 0.25],\ [0.25, 0.50],\ [0.50, 0.75],\ [0.75, 1].$$

a) Since $f(x) = x^2$ is increasing in $[0, 1]$, the minimum values of f occur at the left point of the subintervals, therefore the lower sum is

$$\begin{aligned} L_4 &= \big[f(0) + f(0.25) + f(0.50) + f(0.75)\big]\ \Delta x \\ &= \big[0^2 + 0.25^2 + 0.50^2 + 0.75^2\big] \times 0.25 = 0.21875. \end{aligned}$$

b) Since $f(x) = x^2$ is increasing in $[0, 1]$, the maximum values of f occur at the right point of the subintervals, therefore the upper sum is

$$\begin{aligned} U_4 &= \big[f(0.25) + f(0.50) + f(0.75) + f(1)\big]\ \Delta x \\ &= \big[0.25^2 + 0.50^2 + 0.75^2 + 1^2\big] \times 0.25 = 0.46875. \end{aligned}$$

c) The exact value of the integral is

$$A = \int_0^1 x^2\, dx = \frac{x^3}{3}\bigg|_0^1 = 1/3 \approx 0.33333.$$

Since

$$0.21875 \le 0.33333 \le 0.46875 \quad \Longrightarrow \quad L_4 \le A \le U_4.$$

$\Diamond$

Example 9.3i Let the function $f(x) = e^x$ be given on the interval $[a, b] = [0, 1]$ and partition $[0, 1]$ into N equal subintervals. Find the smallest value of N such that the difference

$$U_N - L_N \le 10^{-3}.$$

Solution:
Divide the interval $[a, b] = [0, 1]$ into N subintervals of equal length

$$[x_0, x_1],\ [x_1, x_2] \ldots [x_{i-1}, x_i] \ldots [x_{N-1}, x_N]$$

where $x_0 = a = 0$ and $x_N = b = 1$.

Since $f(x) = e^x$ is increasing in $[0,1]$, the minimum values of f occur at the left point of the subintervals and the maximum values occur at the right point of the subinterval.

Hence the lower Riemann sum is

$$L_N = \big[f(x_0) + f(x_1) + \cdots + f(x_{N-2}) + f(x_{N-1})\big]\ \Delta x$$

and the upper Riemann sum is

$$U_N = \big[f(x_1) + f(x_2) + \cdots + f(x_{N-1}) + f(x_N)\big]\ \Delta x.$$

Subtracting L_N from the expression for U_N gives

$$U_N - L_N = \big[f(x_N) - f(x_0)\big]\Delta x = \big[f(b) - f(a)\big]\frac{(b-a)}{N}.$$

In this problem $a = 0$, $b = 1$ and $f(x) = e^x$ therefore

$$U_N - L_N = \big[e^1 - e^0\big]\frac{(1-0)}{N} \le 10^{-3}$$

or

$$\frac{(e-1)}{N} \le 10^{-3} \qquad \Longrightarrow \qquad N \ge 10^3(e-1) \approx 1718.3.$$

Hence the smallest integer value of N that satisfies $U_N - L_N \le 10^{-3}$ is $N = 1719$. ◊

9.4 Properties of the Riemann integral

a) If m and M are the minimum and maximum values of f on the interval $[a,b]$, then

(9.4a) $$m \times (b-a) \le \int_a^b f(x)\, dx \le M \times (b-a)\,.$$

b) If c is a constant, then

(9.4b) $$\int_a^b c f(x)\, dx = c \int_a^b f(x)\, dx\,.$$

c) For functions f and g defined on the interval $[a,b]$,

(9.4c) $$\int_a^b \big(f(x) + g(x)\big)\, dx = \int_a^b f(x)\, dx + \int_a^b g(x)\, dx\,.$$

d) If f is defined on the interval $[a,c]$, and b is a point between a and b, then

(9.4d) $$\int_a^c f(x)\, dx = \int_a^b f(x)\, dx + \int_b^c f(x)\, dx\,.$$

The first formula (9.4a) is just the relation between the definite integral and the upper and lower Riemann sums in the case $N = 1$, so there is a single interval equal to all of $[a,b]$.

To see where the next two equations come from, we subdivide $[a,b]$ into N subintervals. For each i with $1 \leq i \leq N$ we choose a point c_i in the ith subinterval. Let $S_N(f)$ and $S_N(cf)$ be the corresponding Riemann sums for the two functions $f(x)$ and $cf(x)$. It follows immediately from the definition of the Riemann sum that $S_N(cf) = cS_N(f)$. Then, from the standard properties of limits,

$$\int_a^b cf(x)\,dx = \lim_{N\to\infty} S_N(cf) = \lim_{N\to\infty} cS_N(f) = c\lim_{N\to\infty} S_N(f) = c\int_a^b f(x)\,dx\,.$$

This proves (9.4b). For (9.4c), take N and the c_i as before and let $S_N(f)$, $S_N(g)$ and $S_N(f+g)$ be the corresponding Riemann sums for f, g and $f+g$. Then

$$\begin{aligned} S_N(f+g) &= \sum_{i=1}^{N} [f(c_i)+g(c_i)] \times \Delta x \\ &= \left(\sum_{i=1}^{N} f(c_i) \times \Delta x\right) + \left(\sum_{i=1}^{N} g(c_i) \times \Delta x\right) \\ &= S_N(f) + S_N(g)\,. \end{aligned}$$

Then

$$\begin{aligned} \int_a^b (f(x)+g(x))\;dx &= \lim_{N\to\infty} S_N(f+g) \\ &= \lim_{N\to\infty} (S_N(f)+S_N(g)) \\ &= \lim_{N\to\infty} S_N(f) + \lim_{N\to\infty} S_N(g) \\ &= \int_a^b f(x)\,dx + \int_a^b g(x)\,dx\,. \end{aligned}$$

If we interpret the definite integral as an area, the final formula (9.4d) is just the fact that the area over the interval $[a,c]$ is the sum of the areas over the two intervals $[a,b]$ and $[b,c]$. A formal mathematical proof of this depends on the more general type of Riemann sum (with subintervals of different lengths) mentioned at the end of the previous chapter.

Reversing the Direction of Integration

So far we have only defined $\displaystyle\int_a^b f(x)\,dx$ in the case where $a \leq b$. We can easily extend the definition to the case $a > b$ in a way which is consistent with the existing definition and properties. Note that in the definition of the Riemann sum as

(9.4e) $$\sum_{i=1}^{N} f(c_i)\,\Delta x$$

we have $\Delta x = (b-a)/N$. If $a > b$ we can use the same formula. The only new feature is that now Δx is negative. From an algebraic point of view this has no effect on our formulas. Geometrically it means that in the Riemann sum areas of rectangles *above* the axis now count as *negative*, and areas *below* the axis count as positive.

The easiest way to see the implication of all this is to go back to the Riemann sum 9.4e and look at the effect of interchanging a and b. The only difference which this makes to the formula is to change the sign of Δx. Hence the Riemann sum also simply changes sign. In the limit as $\Delta x \to 0$ the same is true of the definite integral, and we conclude that

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

One result of this is that the formula (9.4d) now holds whatever the order of the numbers a, b, c. For example, starting with $a \le b \le c$ and the relation

$$\int_a^b + \int_b^c = \int_a^c$$

we can rearrange to get

$$\int_a^b = \int_a^c - \int_b^c = \int_a^c + \int_c^b\,.$$

This is essentially just (9.4d) again, except the point c no longer lies between a and b.

Summary of Chapter 9

- **Riemann sums** were introduced using the calculation of the area under a curve as a motivation. We assumed a function $f(x)$ that is continuous on an interval $[a,b]$ and introduced a partition of $[a,b]$ into N equal subintervals.
- **The upper and lower Riemann sums** U_N and L_N provide upper and lower estimates of the exact area, so that

$$L_N \leq \text{Area under the Graph} \leq U_N\,.$$

- **The Riemann integral** is defined as

$$\lim_{N\to\infty} L_N = \lim_{N\to\infty} U_N = \int_a^b f(x)\,dx\,.$$

provided the limit exists.

Exercises

9.1 Partition the interval $[1,2]$ into N subintervals of equal length,

$$[x_0,x_1],\ [x_1,x_2],\ \ldots,\ [x_{N-1},\ x_N]\,.$$

a) Show that a general point in the partition is

$$x_i = 1+\left(\frac{1}{N}\right)i.$$

b) Show that the maximum value of $f(x)=\dfrac{1}{x}$ on $[x_{i-1},x_i]$ is

$$M_i = N/(N+i-1)$$

and the minimum value is

$$m_i = N/(N+i).$$

c) Show that the lower and upper Riemann sums for $f(x)=1/x$ on $[1,2]$ with N subintervals are

$$L_N = \frac{1}{N+1}+\frac{1}{N+2}+\cdots+\frac{1}{2N}$$

and

$$U_N = \frac{1}{N} + \frac{1}{N+1} + \cdots + \frac{1}{2N-1}.$$

d) Find a value of N such that the difference between the upper and lower sums is less than 10^{-6}, that is,

$$U_N - L_N < 10^{-6}.$$

9.2 Given that

$$\int_{-3}^{1} f(x)\,dx = -2, \quad \int_{1}^{2} f(x)\,dx = 5, \quad \int_{-3}^{2} g(x)\,dx = 8,$$

evaluate, where possible:

$$\int_{-3}^{2} (f(x)+g(x))\,dx, \qquad \int_{2}^{-3} \frac{g(x)}{2}\,dx, \qquad \int_{-3}^{2} f(x)g(x)\,dx.$$

CHAPTER 10

Fundamental Theorem of Calculus

In the previous chapter we used the calculation of the area under a curve to motivate the introduction of the definite integral. However, we have not so far discussed the exact mathematical relation between differentiation and integration that allows us to simplify the calculations. There are two parts to this relation. One part involves the derivative of an integral and the other the integral of a derivative. Together these parts make up the **Fundamental Theorem of Calculus**.

10.1 Integrals as functions

In the previous chapter we introduced the definite integral $\int_a^b f(x)\,dx$ over a fixed interval $[a,b]$ as a real *number* representing the area under the curve $f(x)$. In this chapter we will allow the endpoints to vary so we can look at the definite integral as a *function*.

Dependence on the endpoint

Let f be a continuous function defined on an interval $[a,b]$. For any point x in the interval we have the definite integral of f over the smaller interval $[a,x]$. Of course, the value of this integral depends on x, so we can think of it as a *function* $F(x)$ of x. Formally,

$$\boxed{F(x) = \int_a^x f(t)\,dt\,.} \tag{10.1a}$$

Notice that we have used t as the variable of integration to avoid confusion with the use of x as an endpoint of the interval of integration. The function $F(x)$ is defined at least for all x in the interval $[a,b]$. If f happens to be defined on a larger interval (or perhaps on the whole real line), we can extend the definition of F accordingly via the formula (10.1a). In such cases we can even define $F(x)$ for $x < a$ by the same formula. Some properties of F follow immediately from the definition:

- Changing the starting point a changes F by the addition of a constant. For if we use a_1 instead of a, then by (9.4d):

$$\int_{a_1}^x f(t)\,dt = \int_{a_1}^a f(t)\,dt + \int_a^x f(t)\,dt = C + F(x)\,,$$

where C is the definite integral of f from a_1 to a.

- If f is positive at a point x, then F is an *increasing* function at that point. For, if $f(x)$ is positive, increasing the range of integration adds a positive quantity to the area under the graph. Similarly F is a decreasing function where f is negative.

- Intuitively, $F(x)$ ought to be a *continuous* function. For if we change x to $x+\Delta x$, then $F(x)$ changes by the area under the graph between these two points. This area approaches zero as $\Delta x \to 0$. We will shortly prove the stronger result that $F(x)$ is a *differentiable* function of x.

Example 10.1b Sketch the graph of the function

$$F(x) = \int_0^x \sin t \, dt$$

on the interval $[0, 2\pi]$.

The graph of $\sin x$ on this interval is shown as the solid line in Figure 10.1.

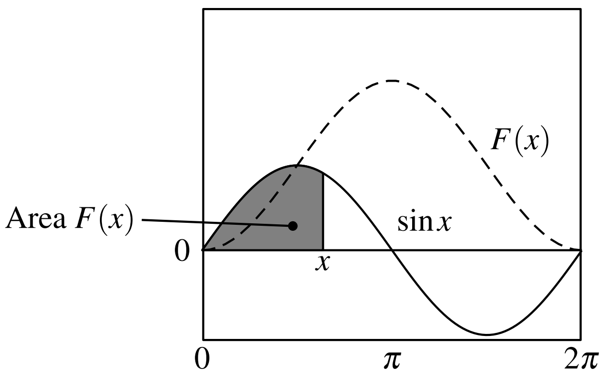


Figure 10.1:

At $x = 0$ the interval of integration has zero length, so $F(0) = 0$. Since $\sin x$ is positive on $[0, \pi]$ the function $F(x)$ is increasing over this interval. Similarly $F(x)$ is decreasing on the interval $[\pi, 2\pi]$. Together these facts show that $F(x)$ attains a *maximum* at $x = \pi$. Over the interval $[0, 2\pi]$ the areas above and below the graph of $\sin x$ are exactly equal, so we must have $F(2\pi) = 0$.

We conclude that the graph of $F(x)$ has the general shape shown by the broken line in Figure 10.1. Note that this curve has been drawn to show points of inflection where $\sin x$ has stationary points. Can you see any justification for this? $\Diamond$

10.2 The Fundamental Theorem of Calculus I

If we think about the integral itself as a function, we can look at what happens when we try to differentiate this function. This is the idea behind the first part of the Fundamental Theorem of Calculus (FTC).

The Fundamental Theorem of Calculus I

Let $f(x)$ be a continuous function defined on an interval $[a,b]$ of the real line and let $F(x)$ be defined by

$$F(x) = \int_a^x f(t)\,dt\,. \tag{10.2a}$$

Then $F(x)$ is a differentiable function of x and

$$F'(x) = f(x) \tag{10.2b}$$

for all x in the interval.

The significance of the result is that it confirms that *every* continuous function has an antiderivative—given by the formula (10.2a). An alternative way to denote the Fundamental Theorem of Calculus I is

$$\frac{d}{dx}\left(\int_a^x f(t)\,dt\right) = f(x). \tag{10.2c}$$

Proof of the FTC Part 1

To prove the first part of the Fundamental Theorem we go back to the definition of the derivative of a function as a limit. Recall that a function F is *differentiable* at a point x if the limit of the differential quotient

$$\frac{F(x+h)-F(x)}{h}$$

exists as $h \to 0$. Then the value of the limit is the *derivative* $F'(x)$ of F at x. If F is defined by (10.2a),

$$F(x+h) = \int_a^{x+h} f(t)\,dt = \int_a^x f(t)\,dt + \int_x^{x+h} f(t)\,dt = F(x) + \int_x^{x+h} f(t)\,dt\,,$$

so

$$F(x+h) - F(x) = \int_x^{x+h} f(t)\,dt\,. \tag{10.2d}$$

We now have to see what happens as $h \to 0$. Let M and m be the maximum and minimum of $f(t)$ for t between x and $x+h$. Assume $h > 0$. Then from (9.4a) we have

$$m \times h \leq \int_x^{x+h} f(t)\,dt \leq M \times h\,.$$

Dividing by h and using (10.2d), we get

(10.2e) $$m \leq \frac{F(x+h) - F(x)}{h} \leq M\,.$$

This inequality can be derived similarly in the case $h < 0$. By the continuity of f both m and M must converge to $f(x)$, as $h \to 0$. But the middle term in the inequality (10.2e) is sandwiched between m and M. So this must converge to $f(x)$ also. This shows that the limit exists, and $F(x)$ is differentiable with derivative $f(x)$.
End of proof

The first part of the Fundamental Theorem confirms that differentiation and integration are 'inverse processes': we now see that differentiating an integral gives back the original function. The standard notation for the antiderivative of a function $f(x)$ is

(10.2f) $$\boxed{\int f(x)\,dx\,.}$$

In this form the antiderivative is also called the **indefinite integral** of f. Note that we can always add an arbitrary constant without changing the fact that the derivative is $f(x)$. For example, we write

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

to indicate that *any* function of the form $\sin x + C$ is an antiderivative for $\cos x$.

10.3 The Fundamental Theorem of Calculus II

The Fundamental Theorem of Calculus II

Let $F(x)$ be a function defined on an interval $[a,b]$ of the real line. Suppose that the derivative of F is defined at each point x of the interval, and that the resulting function $F'(x)$ is continuous. Then

(10.3a) $$\int_a^b F'(x)\,dx = F(b) - F(a)\,.$$

Recall that the definite integral of $F'(x)$ is the unique number which lies between the upper and lower Riemann sums of $F'(x)$ for all subdivisions of the interval $[a,b]$. If we can show that the number $F(b)-F(a)$ has the same property, then (10.3a) is proved.

For this we need the **Mean Value Theorem** of differential calculus introduced in Section 5.6.

This theorem is very important in calculus, so here we rewrite the statement:

Mean Value Theorem
For a function $F(x)$ defined on an interval $[a,b]$ with continuous derivative $F'(x)$, there exists a point c somewhere in the interval with the property that

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

Essentially, this theorem says that the average rate of change over the interval (the ratio of the change in $F(x)$ to the change in x) is equal to the derivative of F at some point in the interval.

To prove formula (10.3a) we introduce a partition of the interval $[a,b]$ into N equal subintervals and then apply the Mean Value Theorem to each subinterval of the partition. Label the partition points as x_i, with $0 \le i \le N$, so that

$$a = x_0 \le x_1 \le \cdots \le x_{N-1} \le x_N = b.$$

and the ith subinterval is $[x_{i-1}, x_i]$. As usual we let m_i, M_i be the minimum and maximum values of $F(x)$ on this subinterval, and Δx the length of the subintervals. According to the Mean Value Theorem there exists a point c_i in the ith subinterval where

(10.3b)
$$F'(c_i) = \frac{F(x_i)-F(x_{i-1})}{x_i - x_{i-1}}.$$

Rearranging this equation and using the fact that $x_i - x_{i-1} = \Delta x$, we obtain

(10.3c)
$$F(x_i)-F(x_{i-1}) = F'(c_i)\,\Delta x.$$

Adding up over the range $1 \le i \le N$, the left side is just

$$[F(x_1)-F(x_0)] + [F(x_2)-F(x_1)] + \cdots$$
$$\cdots + [F(x_{N-1}-F(x_{N-2})] + [F(x_N)-F(x_{N-1})].$$

All the terms except $F(x_0)$ and $F(x_N)$ appear twice, with opposite signs. Therefore they cancel out, leaving only

$$F(x_N)-F(x_0) = F(b)-F(a),$$

where we have used the fact that $x_N = b$ and $x_0 = a$. Now, summing the right side gives the Riemann sum

$$\sum_{i=1}^{N} F'(c_i)\,\Delta x$$

for $F'(x)$ over the interval $[a,b]$. Finally, equating the left and right sides, we conclude that

$$\sum_{i=1}^{N} F'(c_i)\,\Delta x = F(b) - F(a)\,.$$

But we know that any Riemann sum for the specified partition lies between the upper and lower Riemann sums, that is,

$$L_N \le \sum_{i=1}^{N} F'(c_i)\,\Delta x \le U_N,$$

or equivalently,

$$L_N \le \int_a^b F'(x)\,dx \le U_N,$$

and therefore

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

This completes the proof of this part of the Fundamental Theorem.

Notation The change $F(b) - F(a)$ of a function F over an interval is often denoted by

$$\boxed{F(b) - F(a) = [F(x)]_a^b\,.}$$

The theorem then appears in the form

$$\int_a^b F'(x)\,dx = [F(x)]_a^b\,.$$

The important thing about this result is that it gives us a potential shortcut to working out a definite integral. In order to evaluate the integral

$$\int_a^b f(x)\,dx$$

we can look for a function $F(x)$ with the property that $F'(x) = f(x)$ on the interval $[a,b]$. According to the fundamental theorem we then have

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

The function $F(x)$ is called an **antiderivative** of $f(x)$. In this way the Fundamental Theorem of Calculus gives us a link between 'the area under the curve' (in terms of Riemann sums) and 'reverse differentiation', that is, finding an antiderivative.

10.4 Leibniz Integral Rule

We have seen in Section 10.2, equation (10.2c), that the Fundamental Theorem of Calculus I can be expressed in the form

$$\frac{d}{dx}\left(\int_a^x f(t)\,dt\right) = f(x).$$

For example, the derivative of the function defined by the integral

$$F(x) = \int_0^x \sin^2(t)\,dt$$

is

$$\frac{dF}{dx} = \frac{d}{dx}\left(\int_0^x \sin^2(t)\,dt\right) = \sin^2(x).$$

If the upper limit is x^3 instead of x, then we have to use the chain rule to differentiate a function of a function,

$$\frac{dF}{dx} = \frac{d}{dx}\left(\int_0^{x^3} \sin^2(t)\,dt\right) = \sin^2(x^3) \times (3x^2).$$

Leibniz integral rule is a generalization of this problem when both upper and lower limits are functions of x.

Leibniz integral rule

$$\frac{d}{dx}\left(\int_{a(x)}^{b(x)} f(t)\,dt\right) = f\big[b(x)\big].\frac{d}{dx}b(x) - f\big[a(x)\big].\frac{d}{dx}a(x).$$

Example 10.4a Calculate the derivative of the function F defined by

$$F(x) = \int_{x^2}^{x^5} \cos^3(t)\,dt.$$

Applying Leibniz integral rule gives

$$\frac{d}{dx}\left(\int_{x^2}^{x^5} \cos^3(t)\,dt\right) = \cos^3(x^5).(5x^4) - \cos^3(x^2).(2x).$$

◊

10.5 The natural logarithm and exponential functions

Exponentials and natural logarithms have already made several appearances in this course, but we have not given any justification of their main properties. We start with a precise definition of these functions.

The natural logarithm

Note first of all that the function $f(x) = 1/x$ is defined and continuous on the interval $0 < x < \infty$. According to the Fundamental Theorem of Calculus (and formula (10.2b) in particular) we can define an antiderivative $\ln x$ of $f(x)$ by the formula

$$\boxed{\ln x = \int_1^x \frac{1}{t}\, dt\,.}$$

This formula is valid for all x in the range $0 < x < \infty$. Figure 10.2 shows the graph of the function $1/x$ and its relation to $\ln x$. Then $\ln x$ is the unique antiderivative of $1/x$ taking the

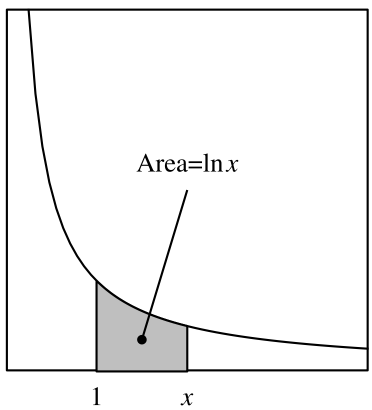


Figure 10.2:

value 0 at $x = 1$. We take this formula as the *definition* of the natural logarithm, and use it to prove the main properties of this function. Some of these properties are immediately obvious:

- $\ln 1 = 0$,
- $\ln x > 0$ if $x > 1$,
- $\ln x < 0$ if $0 < x < 1$.

NOTE We know that $\ln x$ is an antiderivative for $1/x$ for $x > 0$. If $x < 0$ then $\ln(-x)$ is defined, and the chain rule gives

$$\frac{d}{dx}\ln(-x) = \left(\frac{1}{-x}\right)\left(\frac{d(-x)}{dx}\right) = \frac{-1}{-x} = \frac{1}{x}.$$

We can combine the cases $x < 0$ and $x > 0$ into a single formula

(10.5a) $$\frac{d}{dx}\ln|x| = \frac{1}{x}$$

valid for all $x \neq 0$. Here we use the function $|x|$, where as usual

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

The formula (10.5a) should be used with caution. It is really two separate formulas, one for $x > 0$ and the other for $x < 0$. It does not, for example, allow us to assign a value to a definite integral over any interval including the point $x = 0$.

For any fixed positive number a we can define a function $g(x)$ by the formula $g(x) = \ln(ax)$. Then (using the chain rule)

$$g'(x) = \left(\frac{1}{ax}\right)\left(\frac{d(ax)}{dx}\right) = \frac{a}{ax} = \frac{1}{x}.$$

This shows that $g(x)$ is also an antiderivative of $1/x$, and hence differs from the function $\ln x$ itself by a constant:

(10.5b) $$\ln(ax) = \ln(x) + C$$

for all x. We can evaluate the constant C by putting $x = 1$ in equation (10.5b). We find that

$$\ln(a) = \ln(1) + C,$$

and hence $C = \ln(a)$ (since $\ln(1) = 0$). We have proved the formula

(10.5c) $$\boxed{\ln(ax) = \ln(a) + \ln(x)\,,}$$

for all $a, x > 0$. For any $a > 0$ and integer $n \geq 1$ we also have the useful formula

$$\ln(a^n) = \ln(a \cdot a \cdot \cdots \cdot a) = \ln a + \ln a + \cdots + \ln a = n\ln a\,.$$

As usual we define $a^{-n} = 1/a^n$, and we also have

$$\ln(a^n) + \ln(a^{-n}) = \ln(a^n \times a^{-n}) = \ln 1 = 0\,.$$

This gives us the formula $\ln(a^{-n}) = -n\ln a$.

The exponential function

From the definition we have $d(\ln x)/dx = 1/x > 0$ for all $x > 0$, so $\ln x$ is an *increasing* function of x. It must therefore be *one-to-one* on its domain of $(0,\infty)$. It can also be shown that the *range* of the function $\ln x$ is the whole real line, so $\ln x$ can take any real value.

In the language of set theory this implies that the function $x \to \ln x$ is a *bijection* from $(0,\infty)$ to the whole real line $\mathbb{R}$. It follows that ln has an *inverse function*. This inverse is known as the **exponential function**, and denoted by $\exp(x)$. The domain of exp is the range of ln, which is $\mathbb{R}$. The range of exp is the domain of ln. This is $(0,\infty)$, the set of all positive real numbers. Figures 10.3 and 10.4 show the graphs of the two functions:

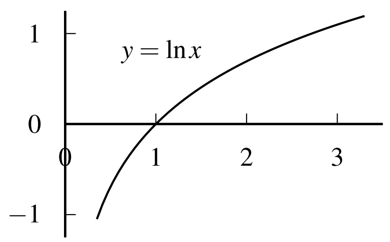


Figure 10.3:

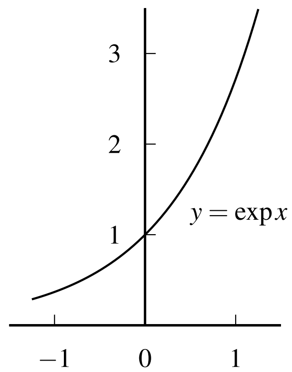


Figure 10.4:

Since exp and ln are inverse to each other we have that

$$\ln(\exp(x)) = x \qquad \text{for all } x \in \mathbb{R}$$

and

$$\exp(\ln(x)) = x \qquad \text{for all } x \in (0,\infty).$$

Equivalently, $y = \ln x$ if and only if $x = \exp(y)$. The fundamental property of the logarithm, contained in (10.5c) above, then shows that

$$ax = \exp(\ln(ax)) = \exp\big(\ln(a) + \ln(x)\big) \qquad \text{for all } a, x \in (0,\infty).$$

If we let $r = \ln(a)$ and $s = \ln(x)$ then $a = \exp(r)$ and $x = \exp(s)$. It follows that

(10.5d) $$\boxed{\exp(r)\exp(s) = \exp(r+s)}$$

for all $r, s \in \mathbb{R}$.

We can calculate the derivative of $\exp(x)$ using standard properties of the derivatives of inverse functions. Alternatively we can get the result more directly by using the chain rule to differentiate the identity

$$x = \ln(\exp(x))\,.$$

This gives

$$1 = \frac{1}{\exp(x)}\,\frac{d}{dx}\exp(x)\,,$$

which is easily rearranged into the formula

$$\boxed{\frac{d}{dx}\exp(x) = \exp(x)\,.}$$

NOTE This shows that the exponential function is a solution to the differential equation

$$\frac{df}{dx} = f(x)\,.$$

It can be shown that if a function f is defined and differentiable on the real line $\mathbb{R}$ and satisfies $f(0) = 1$ and $f'(x) = f(x)$, then $f(x) = \exp(x)$ for all x. The property that the derivative is the same as the original function is therefore said to be *characteristic* of the exponential function.

The General Exponential Function

We have already used the formula

$$\ln(a^n) = n\ln(a)\,,$$

valid for any positive integer n and real number $a > 0$. Applying the exponential function gives

$$a^n = \exp(n\ln(a))\,.$$

In this formula we have assumed that n is a positive integer, but the expression on the right is defined without any such restriction: n can be any real number. This suggests that we use this formula to *define* a^x for any pair of real numbers a, x (with $a > 0$), so

$$\boxed{a^x = \exp(x\ln(a))\,.}$$

For fixed a this is clearly a continuous (and even differentiable) function of x, and coincides with the usual definition of a^n when x is a positive integer.

The number e. One feature of the above definition is that it immediately leads us to consider the unique value of a with the property that $\ln(a) = 1$. The existence and uniqueness of such a number follows from our observation that the logarithm is a one-to-one map from $(0,\infty)$ onto the whole real line. Since the exponential is inverse to the logarithm we see that $a = \exp(1)$. This number is important enough to have its own name—it is called e and has the approximate value

$$e \approx 2.7182818284590452354\ldots .$$

In particular, we see that

$$\boxed{e^x = \exp(x\ln(e)) = \exp(x)\,,}$$

showing that the function exp is just the extension of the exponential powers e^n of this number to real values. In fact e^x is standard notation for the function $\exp(x)$.

Example 10.5e Show that the definition of the general exponential function still satisfies

$$(ab)^c = a^c b^c$$

for a, $b > 0$ and all c in $\mathbb{R}$. From the definition, we have

$$\begin{aligned}
(ab)^c &= \exp(c\ln(ab)) && \text{(definition)}\\
&= \exp(c(\ln a + \ln b)) && \text{(by (10.5c)}\\
&= \exp(c\ln a + c\ln b) &&\\
&= \exp(c(\ln a))\exp(c(\ln b)) && \text{(by (10.5d))}\\
&= a^c b^c && \text{(definition).}
\end{aligned}$$

◊

Summary of Chapter 10

- **The Fundamental Theorem of Calculus I** tells us that

$$F(x) = \int_a^x f(t)\,dt \qquad \Longrightarrow \qquad F'(x) = f(x).$$

- **The Fundamental Theorem of Calculus 2** shows that

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

- **The natural logarithm** is defined in terms of an integral as

$$\ln x = \int_1^x \frac{1}{t}\,dt\,.$$

The standard properties of $\ln x$ follow from this definition.

- **The exponential function** is defined as the inverse function of the logarithm. The standard properties of the exponential follow from this definition also.

Exercises

10.1 When applying the Fundamental Theorem it is important to check that the conditions for the theorem are satisfied. In particular, discontinuities in the function or its derivative can invalidate the formula. Consider the function $f(x) = 1/x^2$. This is not defined at $x = 0$ but everywhere else it has antiderivative $-1/x$. Since $f(x) > 0$ the integral over any interval should certainly be positive.

Show that an attempt to apply the Fundamental Theorem over the interval $[-1, 1]$ (ignoring the difficulties at $x = 0$) leads to the contradictory result

$$\int_{-1}^{+1} \frac{dx}{x^2} = -2\,.$$

10.2 Find the derivative with respect to x of the following integrals:

a) $F(x) = \displaystyle\int_0^x \sin^3 t\,dt$,

b) $F(x) = \displaystyle\int_0^{x^3} \sin^3 t\,dt$.

Hint: Think of the integral as a function of a function and use the chain rule.

10.3 Find the derivative of $\int_0^x u^2 e^{u^2}\,du$. Hence find the derivative of $\int_{\cos x}^{x^2} u^2 e^{u^2}\,du$.

10.4 Use the definition of the general exponential function to verify the following formulas (for any $a > 0$).

a) $a^x a^y = a^{x+y}$,

b) $a^0 = 1$,

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

d) Show that for $x > 0$, $\frac{d}{dx}x^a = ax^{a-1}$.

CHAPTER 11

Integration Techniques

One way to evaluate a definite integral is to find an *antiderivative* of the function being integrated. From a practical point of view there is a big difference between differentiation and integration. Differentiation is a fairly mechanical process, and follows a simple set of rules. Finding formulas for antiderivatives is usually much less straightforward, and often depends on recognising certain patterns in the function to be integrated.

In this chapter we look at the basic rules and then introduce the important methods of integration by substitution, integration by parts and partial fractions to integrate rational functions.

11.1 Basic rules of integration

In Section 9.4 we looked at some properties of the Riemann integral. The two relevant properties in this chapter are the so called *linearity expressions*:

Linearity properties

a) If k is a constant, then

(11.1a) $$\int_a^b k\,f(x)\,dx = k\int_a^b f(x)\,dx\,.$$

b) For functions f and g defined on the interval $[a,b]$,

(11.1b) $$\int_a^b \Big(f(x) \pm g(x)\Big)\,dx = \int_a^b f(x)\,dx \pm \int_a^b g(x)\,dx\,.$$

The expression (11.1a) is easily extended to sums of more than two terms, so we can integrate a sum of terms by integrating each term individually and adding up the results.

Rules (11.1a) and (11.1b) combined with the results from the *Table of Standard Integrals* given in Appendix F make it possible to find antiderivatives in a large number of cases.

Example 11.1c Use the linearity properties and the table in Appendix F to calculate the integral:

$$\int (e^x + 3x^2 + \cos x + 4)\, dx = \int e^x\, dx + 3\int x^2\, dx \int \cos x\, dx + 4\int dx$$

$$= e^x + x^3 + \sin x + 4x + C.$$

$\Diamond$

Example 11.1d Find the area between the graphs of $f(x) = x^2 + 3$ and $g(x) = 2\sin x$ from $x = -1$ to $x = 2$.

Observe first that $f(x) \geq g(x)$ for all x in the given range. According to the argument given in Example 12.3a the area we want is given by the definite integral

$$\int_{-1}^{2} (f(x) - g(x))\, dx = \int_{-1}^{2} \left(x^2 + 3 - 2\sin x\right)\, dx.$$

Using term-by-term integration and the standard integrals, we find that

$$\int \left(x^2 + 3 - 2\sin x\right)\, dx = \int x^2\, dx + 3\int 1\, dx - 2\int \sin x\, dx$$

$$= \frac{x^3}{3} + 3x + 2\cos x + C.$$

Therefore the area is

$$\left[\frac{x^3}{3} + 3x + 2\cos x\right]_{-1}^{2} = \left(\frac{2^3}{3} + 6 + 2\cos(2)\right) - \left(\frac{(-1)^3}{3} + 3(-1) + 2\cos(-1)\right)$$

$$= 12 + 2\cos(2) - 2\cos(-1).$$

$\Diamond$

11.2 Integration by substitution

Integration by substitution is based on the chain rule of differentiation that we studied in Section 5.4. Recall that the chain rule is used to differentiate a *composite function* or *function of a function* and says that if F and u are differentiable functions then

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

From the viewpoint of integration, this shows that $F(u(x))$ is an antiderivative of the function on the right, so

$$\int F'\left[u(x)\right] u'(x)\, dx = F\left[u(x)\right] + C.$$

The problem is clearer if we set $f(u) = F'(u)$, so that F itself is an antiderivative of f. The resulting formula is called the **change of variable** formula. It is an extremely useful tool in integration problems.

The Change of Variable Formula

$$\int f\big[u(x)\big]\,u'(x)\,dx = \int f(u)\,du. \tag{11.2a}$$

How do we find f and u? Usually the first step is to try to identify the function $u(x)$. The requirements on u are:

- The expression $u'(x)\,dx$ should appear as a factor in the integrand,
- The remaining factor should depend on x only via the function $u(x)$.

The following examples illustrate the technique.

Example 11.2b Calculate the indefinite integral $\displaystyle\int 3x^2\,\cos(x^3)\,dx$.

Note that the factor $3x^2$ is the derivative of x^3. This means that the integrand comes from using the chain rule on the function we are looking for. We therefore try the substitution $u = x^3$.

Differentiating gives $du = 3x^2\,dx$ and the change of variable formula then becomes

$$\begin{aligned}\int 3x^2\,\cos(x^3)\,dx &= \int \cos(x^3)\;3x^2\;dx\\ &= \int \cos(u)\;du = \sin(u) + C = \sin(x^3) + C.\end{aligned}$$

$\Diamond$

Example 11.2c Find a formula for the indefinite integral $\displaystyle\int \frac{2x\,dx}{\sqrt{x^2+5}}$.

In this case the presence of the factor $2x\,dx$ suggests we try the substitution $u = x^2+5$. Differentiating gives $du = 2x\,dx$. The change of variable formula then becomes

$$\begin{aligned}\int \frac{2x\,dx}{\sqrt{x^2+5}} &= \int (x^2+5)^{-1/2}\;2x\;dx\\ &= \int u^{-1/2}\;du = \frac{u^{1/2}}{1/2} + C \;=\; 2\sqrt{x^2+5} + C.\end{aligned}$$

Note that the variable of integration switches from x to u, and we conclude by replacing u by the function $u(x)$ of x. $\Diamond$

Example 11.2d Find an antiderivative for the function

$$\int (\tan x)\big(\ln(\cos x)\big)^4\, dx\,.$$

The choice of u is not so clear in this case. However we can write $\tan x = \dfrac{\sin x}{\cos x}$ and so the integral becomes

$$\int \left(\frac{\sin x}{\cos x}\right)\big(\ln(\cos x)\big)^4\, dx\,,$$

suggesting the substitution

$$u = \cos x\,, \quad du = -\sin x\, dx\,.$$

Then

$$\begin{aligned}\int \left(\frac{\sin x}{\cos x}\right)\big(\ln(\cos x)\big)^4\, dx &= \int (\ln(\cos x))^4 \left(\frac{\sin x}{\cos x}\right) dx \\ &= \int (\ln u)^4 \frac{(-1)}{u}\, du\,.\end{aligned}$$

This looks a bit better, but still needs work. Note the presence of the factor $1/u = d(\ln u)/du$, along with the fact that the remaining factors depend on u via $\ln u$. This leads to a second substitution

$$v = \ln u\,, \quad dv = \frac{1}{u}\, du\,,$$

giving

$$\int (\ln u)^4 \frac{(-1)}{u}\, du = \int -v^4\, dv\,.$$

Finally we get the result in terms of x by substituting $v(u)$ for v and then $u(x)$ for u. We can do both substitutions in one step, replacing v by $v(u(x)) = \ln(\cos x)$. The final result is

$$\int (\tan x)\big(\ln(\cos x)\big)^4\, dx = \frac{-v^5}{5} + C = \frac{-(\ln(\cos x))^5}{5} + C\,.$$

◊

Trigonometric substitutions

The change of variables formula (11.2a) can also be applied the other way around. Then we start with an integral with respect to u and turn it into an integral in x by the substitutions

$$u \to u(x)\,, \quad du \to \frac{du}{dx} dx\,,$$

where $u(x)$ is a suitable function of x. It is very common in this situation to use a trigonometric function for $u(x)$. Then it may be possible to simplify the integral using the standard trigonometric identities. Several 'standard integrals' can be evaluated in this way. For example, starting with the integral

$$\int \frac{du}{\sqrt{1-u^2}}$$

we observe that the substitution $u \to \sin x$ and the identity

$$\cos^2 x + \sin^2 x = 1$$

can be used to eliminate the square root. With $u \to \sin x$, we have

$$\sqrt{1-u^2} \to \sqrt{1-\sin^2 x} = \cos x\,, \quad du \to \frac{du}{dx} = \cos x\, dx\,.$$

The integral is transformed as follows:

$$\int \frac{du}{\sqrt{1-u^2}} = \int \frac{\cos x\, dx}{\cos x} = \int dx = x + C\,.$$

To recover the result as a function of u, we need to invert the relation $u = \sin x$ to get x as a function of u. But if $u = \sin x$ then $x = \sin^{-1} u$ (sometimes also called $\arcsin u$). We therefore recover the useful formula

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

It is sometimes useful to combine two applications of the change of variables formula into a single substitution. In this case we need to identify two functions $u(x)$ and $v(y)$ and make substitutions

$$u(x) \to v(y)\,, \quad \frac{du}{dx}dx \to \frac{dv}{dy}dy\,.$$

For example, the integral

$$\int \frac{dx}{(2x+1)^2+1}$$

is transformed by the substitutions

$$2x+1 \to \tan\theta\,, \quad 2dx \to \sec^2\theta\, d\theta\,.$$

into

$$\int \frac{\frac{1}{2}\sec^2\theta\, d\theta}{\tan^2\theta + 1} = \frac{1}{2}\int d\theta = \frac{\theta}{2} + C = \frac{1}{2}\tan^{-1}(2x+1) + C\,.$$

Substitution in definite integrals

Any method which applies to finding antiderivatives also applies to the evaluation of definite integrals. In the case of the substitution method we can avoid changing back the original variable by a suitable change in the limits of integration. To see how this happens we return the change of variables formula (11.2a):

$$\int f\left[u(x)\right] u'(x)\, dx = \int f(u)\, du\,.$$

The variable of integration in the formula on the left is x. By contrast, the integral on the right side of the formula is carried out with respect to u, so we have to change the limits to match. Therefore the change of variable formula takes the form

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

Example 11.2e Evaluate the definite integral

$$\int_0^{\sqrt{\pi}} x\sin(x^2)\,dx\,.$$

Setting $u(x) = x^2$ gives the substitutions

$$x^2 \to u\,, \quad 2x\,dx \to du\,.$$

Also $u(0) = 0$ and $u(\sqrt{\pi}) = \pi$, so the change of variable (and some juggling with the factor of 2) gives

$$\frac{1}{2}\int_0^{\sqrt{\pi}} 2x\sin(x^2)\,dx = \frac{1}{2}\int_0^{\pi} \sin u\,du = \frac{1}{2}\big[-\cos u\big]_0^{\pi} = \frac{1-(-1)}{2} = 1\,.$$

◊

11.3 Integration by parts

Integration by parts transforms the problem into another integration problem, which may be easier. The method is based on the product rule for differentiation:

(11.3a) $$\boxed{(uv)' = u'v + uv',}$$

where $u = u(x)$ and $v = v(x)$ are functions of x. Integrating both sides of (11.3a) yields the formula

$$uv = \int \big(u'v + uv'\big)\,dx \quad \text{or} \quad uv = \int \big(u'v\big)\,dx + \int \big(uv'\big)\,dx.$$

Rearranging the last expression we obtain the **integration by parts** formula:

Integration by parts

$$\int \big(uv'\big)\,dx = uv - \int \big(u'v\big)\,dx.$$

The formula is easier to use and to remember if it is written in terms of differentials as follows (see Appendix C),

(11.3b)
$$\int u\,dv = uv - \int v\,du.$$

Example 11.3c Use integration by parts to calculate $\int x \sin x\,dx$.

If we differentiate x and integrate $\sin x$ we may be able to simplify the integral on the right hand side. Therefore we try

$$u = x \quad\Longrightarrow\quad du = dx \qquad \text{and} \qquad dv = \sin x\,dx \quad\Longrightarrow\quad v = -\cos x.$$

Substituting into equation (11.3b), we obtain

$$\begin{aligned}\int x\sin x\,dx = uv - \int v\,du &= (-\cos x)x - \int(-\cos x)\,dx \\ &= -x\cos x + \int \cos x\,dx = -x\cos x + \sin x + C.\end{aligned}$$

◊

Example 11.3d Use integration by parts to calculate $\int \ln x\,dx$.

In this example the choice $u = \ln x$ would eliminate the logarithm because $du = \dfrac{1}{x}\,dx$. Next, we take $dv = 1\,dx$ and therefore $v = x$. Substituting into equation (11.3b) yields

$$\int \ln x\,dx = x\ln x - \int x \times \frac{1}{x}\,dx = x\ln x - x + C.$$

◊

Example 11.3e Use integration by parts to calculate $\int x\,\ln(x+1)\,dx$.

In this case we try

$$u = \ln(x+1) \quad\Longrightarrow\quad du = \frac{1}{x+1}\,dx \qquad \text{and} \qquad dv = dx \quad\Longrightarrow\quad v = x.$$

Substituting into equation (11.3b), we obtain

$$\begin{aligned}\int x\,\ln(x+1)\,dx = uv - \int v\,du.& \\ &= \frac{x^2}{2}\ln(x+1) - \int\left(\frac{x^2}{2}\right)\frac{1}{x+1}\,dx \\ &= \frac{x^2}{2}\ln(x+1) - \frac{1}{2}\int\frac{x^2}{x+1}\,dx.\end{aligned}$$

Next, to calculate $\displaystyle\int \frac{x^2}{x+1}\,dx$ we divide x^2 by $x+1$ using polynomial long division,

$$
\begin{array}{r|lll}
\multicolumn{1}{r}{} & x & -\ 1 & \\
\cline{2-3}
x+1 & x^2 & & \\
 & \underline{x^2} & \underline{+\ x} & \\
 & & -\ x & \\
 & & \underline{-\ x} & \underline{-\ 1} \\
 & & & 1
\end{array}
$$

Therefore we can write $\dfrac{x^2}{x+1} = \dfrac{(x-1)(x+1)+1}{x+1} = (x-1) + \dfrac{1}{x+1}$ and so

$$
\begin{aligned}
\int \frac{x^2}{x+1}\,dx &= \int (x-1)\,dx + \int \frac{dx}{x+1} \\
&= \frac{x^2}{2} - x + \ln(x+1).
\end{aligned}
$$

Collecting all the pieces together, the final answer becomes

$$\int x\,\ln(x+1)\,dx = \frac{x^2}{2}\ln(x+1) - \frac{1}{2}\Big(x^2 - x + \ln(x+1)\Big) + C.$$

$\Diamond$

Example 11.3f Use integration by parts to calculate $\displaystyle\int x^3 \sin x\,dx$.

Here we also let

$$u = x \quad\Longrightarrow\quad du = dx \qquad \text{and} \qquad dv = \sin x\,dx \quad\Longrightarrow\quad v = -\cos x.$$

Substituting into equation (11.3b), we obtain

$$
\begin{aligned}
\int x^3 \sin x\,dx &= uv - \int v\,du \\
&= (-\cos x)x^3 - \int (-\cos x)3x^2\,dx \\
&= -x^3\cos x + 3\int x^2\cos x\,dx.
\end{aligned}
$$

We have reduced the original problem to a simpler one. Repeating the method gives

$$
\begin{aligned}
\int x^2\cos x\,dx &= (\sin x)x^2 - \int (\sin x)2x\,dx \\
&= x^2\sin x - \left((-\cos x)2x - \int (-\cos x)2\,dx\right) \\
&= x^2\sin x + 2x\cos x - 2\sin x + C.
\end{aligned}
$$

So we have shown that

$$\int x^3\sin x\,dx = -x^3\cos x + 3x^2\sin x + 6x\cos x - 6\sin x + C.$$

$\Diamond$

Example 11.3g Use integration by parts to calculate $I = \int e^x \sin x \, dx$.

$$u = \sin x \quad \Longrightarrow \quad du = \cos x \, dx \qquad \text{and} \qquad dv = e^x \, dx \quad \Longrightarrow \quad v = e^x.$$

Substituting into equation (11.3b), gives

$$I = e^x \sin x - \int e^x \cos x \, dx.$$

This is no simpler than the original problem, but applying the same method again gives

$$\begin{aligned} I &= e^x \sin x - \left(e^x \cos x - \int e^x (-\sin x) \, dx \right) \\ &= e^x (\sin x - \cos x) - \int e^x \sin x \, dx \\ &= e^x (\sin x - \cos x) - I. \end{aligned}$$

We now have an equation for I. Adding I to both sides and dividing by 2 yields

$$I = \frac{1}{2} e^x (\sin x - \cos x) + C.$$

Differentiate this expression and check that you really do get $e^x \sin x$. ♢

Definite integrals

In the case of definite integrals we can express the formula for integration by parts in the form

$$\boxed{\int_a^b u \, dv = \big[uv \big]_a^b - \int_a^b v \, du.}$$

Example 11.3h Use integration by parts to calculate $\displaystyle\int_0^{\pi/2} x \sin x \, dx$.

The indefinite integral was found in Example 11.3c,

$$\begin{aligned} \int x \sin x \, dx = uv - \int v \, du = (-\cos x)x - \int (-\cos x) \, dx \\ = -x \cos x + \int \cos x \, dx = -x \cos x + \sin x + C. \end{aligned}$$

Therefore

$$\begin{aligned} \int_0^{\pi/2} x \sin x \, dx = \Big[uv \Big]_0^{\pi/2} - \int_0^{\pi/2} v \, du = [(-\cos x)x]_0^{\pi/2} - \int_0^{\pi/2} (-\cos x) \, dx \\ = \Big[-x \cos x + \sin x \Big]_0^{\pi/2} = 1. \end{aligned}$$

♢

11.4 Partial fractions

We are often led in applications (population growth, chemical reactions, etc.) to look for integrals of the form

$$\int \frac{2x+1}{(x-1)(x-2)}\,dx. \tag{11.4a}$$

The function $r(x) = \dfrac{2x+1}{(x-1)(x-2)}$ is an example of so-called **rational functions**. Rational functions are expressions of the form $f(x)/g(x)$, where $f(x)$ and $g(x)$ are polynomials.

In this example, $f(x) = 2x+1$, a polynomial of degree one and $g(x) = (x-1)(x-2)$, a polynomial of degree 2.

Example 11.4b As it stands the integral (11.4a) is difficult to calculate. However, if we split the rational function $r(x)$ as a sum of simpler functions in the form

$$\frac{2x+1}{(x-1)(x-2)} = \frac{A}{x-1} + \frac{B}{x-2},$$

then, it can easily be integrated. This process of spliting a rational function into sums of simpler functions is called **partial fractions**.

The way we calculate the constants A and B is as follows. First find the common denominator by cross-multiplying the right hand side,

$$\frac{2x+1}{(x-1)(x-2)} = \frac{A}{x-1} + \frac{B}{x-2} = \frac{A(x-2)+B(x-1)}{(x-1)(x-2)}.$$

Now, the denominators on the left and right hand side terms are the same, therefore the numerators must also be the same,

$$2x+1 = A(x-2)+B(x-1)$$

for all values of x. Collecting terms on the right hand side, gives

$$2x+1 = (A+B)x + (-2A-B).$$

Equating coefficients of like powers of x gives the system of equations

$$A+B = 2, \quad -2A-B = 1,$$

and solving simultaneously gives $a = -3$ and $b = 5$. Therefore

$$\frac{2x+1}{(x-1)(x-2)} = \frac{-3}{x-1} + \frac{5}{x-2}.$$

The integral (11.4a) can now be calculated,

$$\begin{aligned}\int \frac{2x+1}{(x-1)(x-2)}\,dx &= \int \frac{-3}{x-1}\,dx + \int \frac{5}{x-2}\,dx \\ &= -3\ln(x-1) + 5\ln(x-2) + C.\end{aligned}$$

◊

The same method works if there are three or more linear factors in the denominator.

Example 11.4c Find the partial fraction form for the function

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

Note that the degree of the denominator is 3, while the degree of the numerator is 2. We need to find the a, b and c such that

$$\frac{2x^2-1}{x(x-1)(x+1)} = \frac{a}{x} + \frac{b}{x-1} + \frac{c}{x+1}.$$

There is a handy shortcut method, as follows. First find the common denominator and equate numerators as before. But instead of equating coefficients of like powers of x and then solving a simultaneous system, we let $x = 0$, $x = 1$ and $x = -1$ succesively as follows.

$$2x^2 - 1 = a(x-1)(x+1) + bx(x+1) + cx(x-1).$$

To find a, put $x = 0$. Then two of the summands on the righthand side vanish. We find that $a = 1$.

To find b, put $x = 1$. Two summands vanish and we find that $b = \frac{1}{2}$.

Finally, to find c we put $x = -1$, the summands involving $x+1$ vanish and we get $c = \frac{1}{2}$. So we have

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

$\diamondsuit$

Important Rule
Partial Fractions applies to rational functions where **the degree of the numerator is strictly less than the degree of the denominator**. If this condition does not hold (as in the example 11.4d below) we have to divide the numerator by the denominator using polynomial (long) division.

Example 11.4d

$$\frac{x^4+2x+2}{x^3-x^2+x-1}. \tag{11.4e}$$

In this example the numerator (the upper term) has degree 4 and the denominator (the lower term) has degree 3, therefore we have to use polynomial division first. The division leaves a remainder term which has degree less than the denominator:

$$\begin{array}{r|rrrrrrr}
\multicolumn{1}{r}{} & x & + & 1 & & & & \\ \cline{2-8}
x^3-x^2+x-1 & x^4 & & & & +\ 2x & + & 2 \\
 & x^4 & - & x^3 & + & x^2 & - & x \\ \cline{2-7}
 & & & x^3 & - & x^2 & + & 3x\ +\ 2 \\
 & & & x^3 & - & x^2 & + & x\ -\ 1 \\ \cline{4-8}
 & & & & & & & 2x\ +\ 3
\end{array}$$

In each step we subtract off a multiple of the denominator, with the aim of eliminating the highest power of x remaining. The process stops when the remainder has degree less than the degree of the denominator. In this case the quotient is $x+1$ and the remainder $2x+3$. That is, we have shown that

$$x^4+2x+2=(x+1)\times(x^3-x^2+x-1)+(2x+3)$$

or equivalently,

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

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

The polynomial term ($x+1$ in our example) is easy to integrate directly, and so the remaining task is the integration of a rational function in which the degree of the numerator is less than the degree of the denominator. This is where the partial fraction method comes in. First factorise the denominator. In our example,

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

Now the task is to break the given rational function up into several rational functions, one for each of the factors of the denominator. In each the degree of the numerator should be less than the degree of the denominator. So in our case we need to find constants a, b and c such that

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

The basic method for doing this is to clear denominators and equate coefficients. Thus, multiplying through by $(x^2+1)(x-1)$ gives

$$2x+3=(ax+b)(x-1)+c(x^2+1)=(a+c)x^2+(b-a)x+(c-b), \tag{11.4f}$$

and we deduce that $a+c=0$, $b-a=2$ and $c-b=3$. We find that $a=-\frac{5}{2}$, $b=-\frac{1}{2}$ and $c=\frac{5}{2}$. By now we have found that

$$\int\frac{x^4+2x+2}{x^3-x^2+x-1}\,dx=\int(x+1)\,dx-\frac{5}{2}\int\frac{x}{x^2+1}\,dx-\frac{1}{2}\int\frac{1}{x^2+1}\,dx+\frac{5}{2}\int\frac{dx}{x-1}.$$

The second integral on the right hand side requires the substitution $v = x^2 + 1$, after which the answer is found to be $\frac{1}{2}\ln(x^2+1)$. The third integral is a standard integral, $\tan^{-1}(x)$. The fourth is easily found (after the straightforward substitution $u = x - 1$ to be $\ln|x-1|$. The final answer is

$$\int \frac{x^4+2x+2}{x^3-x^2+x-1}\,dx = \int (x+1)\,dx - \frac{5}{2}\int \frac{x}{x^2+1}\,dx - \frac{1}{2}\int \frac{1}{x^2+1}\,dx + \frac{5}{2}\int \frac{dx}{x-1}.$$

$$= \frac{x^2}{2} + x - \frac{5}{4}\ln(x^2+1) - \frac{1}{2}\tan^{-1}x + \frac{5}{2}\ln|x-1| + C.$$

$\Diamond$

We will not attempt to give a full account of the theory of partial fractions here. There are several different cases, depending on whether there are repeated factors in the denominator, and whether the factors have degree one or two. The most important case is definitely the case of distinct factors of degree one that we looked at.

Summary of Chapter 11

- **Integration by substitution** is a useful method of integration obtained from **the change of variable formula** which is a consequence of the chain rule for differentiation.

  The change of variables formula can also be applied to **definite integrals**. In this case it is necessary to change the limits of integration as well as the variable of integration.

- **Integration by parts** is a powerful integration technique obtained from the product rule for differentiation. The method involves separating the function to be integrated into two factors and transforming the problem into a different integral. In this process one factor is differentiated and the other is integrated.

- **Partial fraction expansions** is a method for integrating *rational functions*, that is, functions of the form $P(x)/Q(x)$ where $P(x)$ and $Q(x)$ are polynomials.

- **Reduction formulas** may be obtained when the integral involves an integer parameter n. In this case, it may be possible to relate the integral to other instances of the same integral with smaller values of the parameter.

Exercises

11.1 Use substitutions to evaluate the following integrals:

a) $\displaystyle\int x^{-1} \log x \, dx$,

b) $\displaystyle\int e^{\sqrt{x}} \, dx$,

c) $\displaystyle\int \frac{e^{2x}}{\sqrt{e^x+1}} \, dx$,

d) $\displaystyle\int \frac{\sqrt{1+x^2}}{x^4} \, dx$.

11.2 Find $\displaystyle\int 2 \sin x \cos x \, dx$

a) by using the substitution $u = \sin x$,

b) by using the substitution $v = \cos x$,

c) by using the formula from trigonometry for $\sin 2x$.

Check that your three answers are consistent with each other.

11.3 Show that the substitution $u = \tan x$ leads to the formula

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

where $\tan^{-1} u$ (or $\arctan u$) is the inverse function to the tangent.

11.4 Use partial fraction expansions to find the following integrals:

a) $\displaystyle\int \frac{dx}{(x-1)(x-3)} \, dx$

b) $\displaystyle\int \frac{x^2 \, dx}{(x-1)(x-3)}$.

11.5 Find a reduction formula for the indefinite integral

$$I_n = \int \frac{dx}{(1+x^2)^n}.$$

HINT: Take $dv/dx = 1$.

11.6 For a continuous function $f(x)$ defined for $x \geq a$ the integral from a to ∞ is defined to be the limit

$$\int_a^{\infty} f(x) \, dx = \lim_{N \to \infty} \int_a^N f(x) \, dx$$

(assuming the limit exists). By obtaining a suitable reduction formula, show that

$$\int_0^{\infty} x^n e^{-x} \, dx = n!$$

for all integers $n \geq 0$. This suggests the possibility of extending the definition of the factorial function to non-integer values of the argument. This is one motivation for the definition of the *Gamma Function* $\Gamma(s)$ by the formula

$$\Gamma(s) = \int_0^\infty x^{s-1} e^{-x}\, dx\,,$$

where s is any real number > 1. Although x^{s-1} is not defined at $x = 0$ when $s < 1$ we can still set

$$\int_0^\infty x^{s-1} e^{-x}\, dx = \lim_{a \to 0} \int_a^\infty x^{s-1} e^{-x}\, dx$$

to extend the definition of $\Gamma(s)$ to all $s > 0$. Then it is possible to check that $\Gamma(n+1) = n!$ (for integer n) and $\Gamma(s+1) = s\Gamma(s)$ for all $s > 0$.

11.7 Verify the formula

$$\int_0^1 x^m (1-x)^n\, dx = \frac{m!\,n!}{(m+n+1)!}\,,$$

where m, n are integers ≥ 0.

CHAPTER 12

Applications of Integration

In this chapter we apply a similar method to the calculation of area under a curve that we used in Chapter 10 to calculate

a) The length of a curve,
b) The area between two curves,
c) Volumes of solids of revolution.

For example, for solids of revolution, we divide the volume into smaller sections whose volumes are easy to calculate and then add them all up to obtain Riemann sums. By taking the limit as the number of sections tends to infinity, we obtain a definite integral that can be evaluated using the Fundamental Theorem of Calculus. Before that, however, we look at more advanced techniques of integration.

12.1 Further integration techniques

Powers of trigonometric functions

Since some of the applications of $\sin\theta$ and $\cos\theta$ in terms of complex exponentials usually involve the binomial theorem, we now revise it briefly.

Binomial theorem

The expression $x+y$ is called a *binomial expression*, and the binomial theorem is a generalisation of the familiar formula $(x+y)^2 = x^2+2xy+y^2$. It states that for all x, y and for all integers $n \geq 0$,

$$(x+y)^n = \binom{n}{0}x^n + \binom{n}{1}x^{n-1}y + \cdots + \binom{n}{r}x^{n-r}y^r + \cdots + \binom{n}{n}y^n.$$

Defining

$$^nC_r = \binom{n}{r} = \frac{n!}{r!(n-r)!},$$

the binomial theorem becomes

$$(x+y)^n =^n C_0x^n +^n C_1x^{n-1}y + \cdots +^n C_rx^{n-r}y^r + \cdots +^n C_ny^n.$$

The numbers $\binom{n}{0}, \binom{n}{1}, \ldots, \binom{n}{n}$ are called the *binomial coefficients*.

Thus, for example, with $n = 4$ we have

$$\text{If } n=4 \quad \Longrightarrow \quad (x+y)^4 = x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4,$$

and

$$\text{if } n=5 \quad \Longrightarrow \quad (x+y)^5 = x^5 + 5x^4y + 10x^3y^2 + 10x^2y^3 + 5xy^4 + y^5.$$

The binomial coefficients can also be found in the rows of *Pascal's Triangle*, part of which is shown below:

$$\begin{array}{ccccccccccccccc}
 & & & & & & & 1 & & & & & & & \\
 & & & & & & 1 & & 1 & & & & & & \\
 & & & & & 1 & & 2 & & 1 & & & & & \\
 & & & & 1 & & 3 & & 3 & & 1 & & & & \\
 & & & 1 & & 4 & & 6 & & 4 & & 1 & & & \\
 & & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & & \\
 & 1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1 & \\
1 & & 7 & & 21 & & 35 & & 35 & & 21 & & 7 & & 1
\end{array}$$

Note that each row gives the binomial coefficients for a particular value of n in the expansion of $(x+y)^n$. For example, in the expansion of $(x+y)^3$ the coefficients are $1,3,3,1$.

Another interesting feature of Pascal's triangle is that each entry is the sum of the two entries in the row above it. In the last row shown above, $21 = 15+6$, $35 = 20+15$, and so on. So you can add on as many additional rows as you require, remembering that each row begins and ends with a 1, to enlarge the triangle to any n value you please.

Example 12.1a Calculate $\cos^3\theta$ using the binomial theorem with $n = 3$ (binomial coefficients 1, 3, 3, 1) and the fact that $\cos\theta = \dfrac{e^{i\theta} + e^{-i\theta}}{2}$.

$$\begin{aligned}
\cos^3\theta &= \Big(\frac{e^{i\theta}+e^{-i\theta}}{2}\Big)^3 \\
&= \frac{1}{8}(e^{3i\theta} + 3e^{2i\theta}\,e^{-i\theta} + 3e^{i\theta}\,e^{-2i\theta} + e^{-3i\theta}) && \text{(binomial theorem)} \\
&= \frac{1}{8}(e^{3i\theta} + 3e^{i\theta} + 3e^{-i\theta} + e^{-3i\theta}) && \text{(simplify exponents)} \\
&= \frac{1}{4}\Big(\frac{1}{2}(e^{3i\theta} + e^{-3i\theta}) + \frac{3}{2}(e^{i\theta} + e^{-i\theta})\Big) && \text{(collect conjugate terms)} \\
&= \frac{1}{4}(\cos 3\theta + 3\cos\theta)
\end{aligned}$$

$\Diamond$

This formula for $\cos^3\theta$ may also be calculated simply using standard trigonometric identities, but the working is not as elegant as this approach. The same technique can be used to find expressions for any positive power of $\cos\theta$ and $\sin\theta$, which can then be used in integration problems.

Example 12.1b Find $\int \cos^3\theta\, d\theta$.

We have

$$\begin{aligned}\int \cos^3\theta\, d\theta &= \int \frac{1}{4}(\cos 3\theta + 3\cos\theta)\, d\theta\\ &= \frac{1}{4}\int \cos 3\theta\, d\theta + \frac{3}{4}\int \cos\theta\, d\theta\\ &= \frac{1}{12}\sin 3\theta + \frac{3}{4}\sin\theta + C,\end{aligned}$$

where C is an arbitrary constant. $\Diamond$

Example 12.1c Find a formula for $\sin^4\theta$ in terms of $\cos 4\theta$ and $\cos 2\theta$.

Using the binomial theorem with $n = 4$ and the expression for $\sin\theta$ in terms of exponentials, we obtain

$$\begin{aligned}\sin^4\theta &= \Big(\frac{e^{i\theta} - e^{-i\theta}}{2i}\Big)^4\\ &= \frac{1}{16i^4}\Big(e^{4i\theta} - 4e^{3i\theta}e^{-i\theta} + 6e^{2i\theta}e^{-2i\theta} - 4e^{i\theta}e^{-3i\theta} + e^{-4i\theta}\Big)\\ &= \frac{1}{16}\Big(e^{4i\theta} + e^{-4i\theta} - 4(e^{2i\theta} + e^{-2i\theta}) + 6\Big)\\ &= \frac{1}{8}\Big(\frac{1}{2}(e^{4i\theta} + e^{-4i\theta}) - 4\times\frac{1}{2}(e^{2i\theta} + e^{-2i\theta}) + 3\Big)\\ &= \frac{1}{8}\Big(\cos 4\theta - 4\cos 2\theta + 3\Big).\end{aligned}$$

$\Diamond$

So far we have shown how to obtain formulas for powers of $\sin\theta$ and $\cos\theta$ in terms of cosines of multiples of θ, but we can also reverse the process to find formulas for expressions like $\cos(n\theta)$ and $\sin(n\theta)$. In these cases we revert to the non-exponential polar form and apply both de Moivre's theorem and the binomial theorem to $(\cos\theta + i\sin\theta)^n$.

Example 12.1d Find formulas for $\cos 4\theta$ and $\sin 4\theta$ in terms of powers of $\sin\theta$ and $\cos\theta$.

$$\begin{aligned}\cos 4\theta + i\sin 4\theta &= \big(\cos\theta + i\sin\theta\big)^4\\ &= \cos^4\theta + 4\cos^3\theta\big(i\sin\theta\big) + 6\cos^2\theta\big(i^2\sin^2\theta\big) + 4\cos\theta\big(i^3\sin^3\theta\big) + i^4\sin^4\theta\\ &= \cos^4\theta - 6\cos^2\theta\sin^2\theta + \sin^4\theta + i\big(4\cos^3\theta\sin\theta - 4\cos\theta\sin^3\theta\big)\end{aligned}$$

Equating the real parts on both sides of the equation gives

$$\cos 4\theta = \cos^4\theta - 6\cos^2\theta\sin^2\theta + \sin^4\theta$$

and equating the imaginary parts gives

$$\sin 4\theta = 4\cos^3\theta\sin\theta - 4\cos\theta\sin^3\theta.$$

$\diamondsuit$

Reduction formulas

1) Indefinite integrals

In example 11.3f we found an antiderivative for the function $x^3\sin x$ by repeated application of the integration by parts formula. It is sometimes useful to make this process more systematic. For example, if we want to evaluate

$$\int x^{10}e^x\,dx$$

we can use integration by parts (with $u = x^{10}$) to reduce to a similar problem involving x^9. Repeating the process another nine times eventually eliminates the power of x altogether. This promises to be very tedious. We can reduce the effort by performing the integration for a general power x^n of x. For $n \geq 1$ set

$$I_n = \int x^n e^x\,dx.$$

Integration by parts with $u = x^n$ and $dv/dx = e^x$ gives

$$I_n = x^n e^x - \int nx^{n-1}e^x\,dx = x^n e^x - nI_{n-1}.$$

The formula $I_n = x^n e^x - nI_{n-1}$ relating I_n and I_{n-1} is an example of a **reduction formula**. Starting with I_0 we can apply this formula with $n = 1, 2, 3, 4\ldots$ to generate successive I_n directly. For example

$$\begin{aligned}
I_0 &= e^x,\\
I_1 &= xe^x - e^x,\\
I_2 &= x^2e^x - 2(xe^x - e^x)\\
&= (x^2 - 2x + 2)e^x,\\
I_3 &= x^3e^x - 3(x^2 - 2x + 2)e^x\\
&= (x^3 - 3x^2 + 6x - 6)e^x.
\end{aligned}$$

In each case we get the general solution by adding on the usual arbitrary constant.

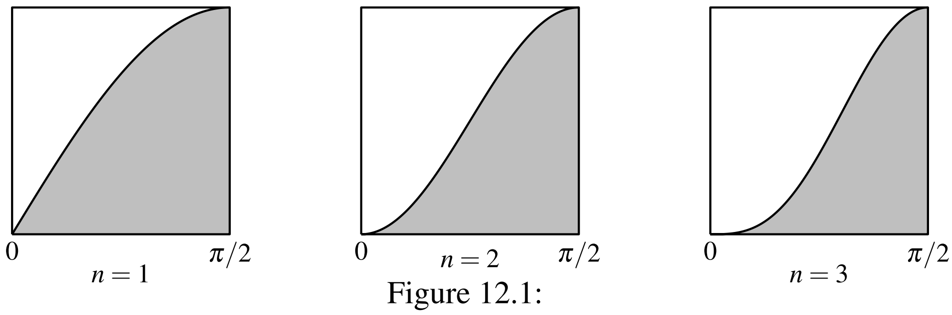


Figure 12.1:

2) Definite integrals

We can apply the same technique to *definite integrals*. An interesting example involves the integral

$$a_n = \int_0^{\pi/2} (\sin x)^n \, dx$$

for a general positive integer n. Since $0 \le \sin x \le 1$ for x in the range $0 \le x \le \pi/2$ we have

$$0 \le (\sin x)^{n+1} \le (\sin x)^n \le 1$$

for these values of x. Therefore

$$\int_0^{\pi/2} (\sin x)^{n+1} \, dx \le \int_0^{\pi/2} (\sin x)^n \, dx$$

also, showing that the numbers a_n form a *decreasing* sequence. Figure 12.1 shows the cases $1 \le n \le 3$; here a_n is the shaded area under the curve.

For small values of n the integral is easy to evaluate. For example,

$$a_0 = \int_0^{\pi/2} dx = \frac{\pi}{2},$$

while

$$a_1 = \int_0^{\pi/2} \sin x \, dx = \big[-\cos x\big]_0^{\pi/2} = 1.$$

For $n > 1$ we can try to use integration by parts to get a reduction formula. For this we first have to identify the factors u and dv/dx in $(\sin x)^n$. If $dv/dx = \sin x$ then $v = -\cos x$ and $u = (\sin x)^{n-1}$. This gives

$$a_n = \big[-(\sin x)^{n-1} \cos x\big]_0^{\pi/2} + \int_0^{\pi/2} (n-1)(\sin x)^{n-2} (\cos x)^2 \, dx.$$

Since $\sin x = 0$ when $x = 0$ and $\cos x = 0$ when $x = \pi/2$, the first term on the right is zero. We can use the familiar formula $\cos^2 x = 1 - \sin^2 x$ to rewrite the remaining term:

$$\begin{aligned} a_n &= \int_0^{\pi/2} (n-1)(\sin x)^{n-2} \big(1 - (\sin x)^2\big) \, dx \\ &= \int_0^{\pi/2} (n-1)(\sin x)^{n-2} - \int_0^{\pi/2} (n-1)(\sin x)^n \, dx \\ &= (n-1)a_{n-2} - (n-1)a_n. \end{aligned}$$

This immediately rearranges to give

$$na_n = (n-1)a_{n-2} \quad \text{or} \quad a_n = \frac{n-1}{n}a_{n-2}$$

for all $n > 1$. Note that this reduction formula gives *two* sequences of numbers, one for n even and the other for n odd. For example

$$\begin{aligned}
a_0 &= \frac{\pi}{2} \\
a_2 &= \frac{\pi}{2} \times \frac{1}{2} \\
a_4 &= \frac{\pi}{2} \times \frac{1}{2} \times \frac{3}{4} \\
a_6 &= \frac{\pi}{2} \times \frac{1}{2} \times \frac{3}{4} \times \frac{5}{6} \\
&\vdots \\
a_{2n} &= \frac{\pi}{2} \times \frac{1}{2} \times \frac{3}{4} \times \frac{5}{6} \times \cdots \times \frac{2n-1}{2n},
\end{aligned}$$

while the odd sequence looks like

$$\begin{aligned}
a_1 &= 1 \\
a_3 &= \frac{2}{3} \\
a_5 &= \frac{2}{3} \times \frac{4}{5} \\
a_7 &= \frac{2}{3} \times \frac{4}{5} \times \frac{6}{7} \\
&\vdots \\
a_{2n+1} &= \frac{2}{3} \times \frac{4}{5} \times \frac{6}{7} \times \cdots \times \frac{2n}{2n+1}.
\end{aligned}$$

12.2 Length of a curve

Consider a function $f(x)$ that is continuous on an interval $[a,b]$. In order to calculate the length of the curve from $x = a$ to $x = b$ we partition the interval $[a,b]$ into N equal subintervals. Label the division points as x_i, with $0 \le i \le N$, so that the typical ith subinterval is $[x_{i-1}, x_i]$ and

$$a = x_0 \le x_1 \le \cdots \le x_{N-1} \le x_N = b.$$

Figure 12.2 shows a section of the curve on the typical subinterval $[x_{i-1}, x_i]$ of length $\Delta x = x_i - x_{i-1}$. If Δx is small, the length of the curve between x_{i-1} and x_i, denoted by Δs, can be approximated by the length of the straight line segment, as shown in the figure. By the Pythagorean theorem the length of this segment is given by

(12.2a) $$\Delta s = \sqrt{\Delta x^2 + \Delta y^2} = \Delta x \sqrt{1 + \left(\frac{\Delta y}{\Delta x}\right)^2}.$$

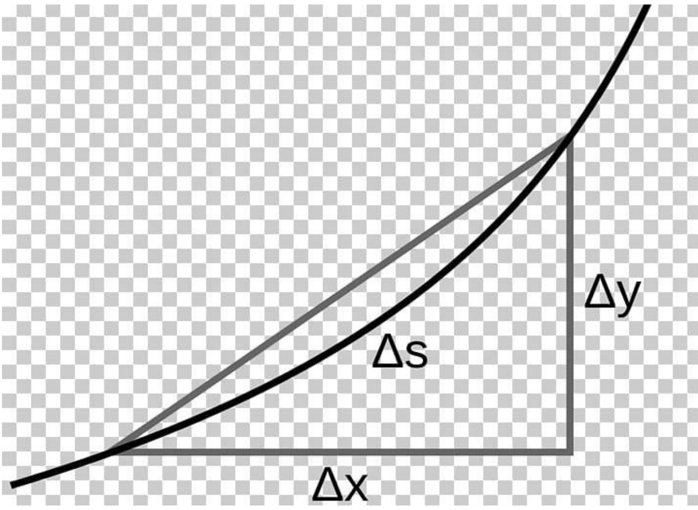


Figure 12.2:

Also, if Δx is small,

$$\frac{\Delta y}{\Delta x} \approx f'(x_{i-1}). \tag{12.2b}$$

Inserting (12.2b) into (12.2a) and adding the subscript i, we have

$$\Delta s_i = \sqrt{1 + \left[f'(x_{i-1})\right]^2}\,\Delta x_i,$$

and therefore, the length of the curve from a to b is given approximatelly by the Riemann sum

$$s \approx \sum_{i=1}^{N} \Delta s_i = \sum_{i=1}^{N} \sqrt{1 + \left[f'(x_{i-1})\right]^2}\,\Delta x_i.$$

Taking the limit as $N \to \infty$ the Riemann sum becomes the integral

$$s = \int_a^b \sqrt{1 + \left[f'(x)\right]^2}\,dx. \tag{12.2c}$$

This integral gives the length of the curve $f(x)$ between a and b and it is often written as

$$s = \int_0^s ds,$$

where

$$ds = \sqrt{1 + \left[f'(x)\right]^2}\,dx$$

is called the *arc length differential*.

Example 12.2d Use the integral expression (12.2c) to calculate the length of the straight line $y = x$ between $x = 0$ and $x = 1$.

We easily see, using elementary geometry, that the length of this segment of straight line is $\sqrt{2}$. However, we will use formula (12.2c) to illustrate the method.

In this case $f(x) = x$ and therefore $f'(x) = 1$. The arc length differential is

$$ds = \sqrt{1 + \left[f'(x)\right]^2}\,dx = \sqrt{1 + (1)^2}\,dx = \sqrt{2}\,dx$$

and the total length is

$$s = \int_0^s ds = \int_0^1 \sqrt{2}\,dx = \sqrt{2}.$$

$\diamond$

Example 12.2e Calculate the circumference of a circle of unit radius $x^2 + y^2 = 1$ using the formula for the arc length (12.2c).

We calculate the length of the upper semicircle $y = \sqrt{1 - x^2}$ and multiply by 2.

In this case $f'(x) = -\dfrac{x}{\sqrt{1 - x^2}}$, therefore

$$s = \int_a^b \sqrt{1 + \left[f'(x)\right]^2}\,dx = \int_{-1}^1 \sqrt{1 + \frac{x^2}{1 - x^2}}\,dx = \int_{-1}^1 \frac{dx}{\sqrt{1 - x^2}}\,dx.$$

Now make the substitution $x = \sin\theta$.

When $x = -1 \implies \theta = -\pi/2$ and when $x = 1 \implies \theta = \pi/2$. Therefore

$$s = \int_{-\pi/2}^{\pi/2} \frac{\cos\theta\,d\theta}{\cos\theta} = \int_{-\pi/2}^{\pi/2} d\theta = \pi/2 + \pi/2 = \pi.$$

Therefore, the circumference of the unit circle is equal to 2π as expected. $\diamond$

12.3 Area between two curves

Here we illustrate the method of Riemann sums to the calculation of area between two curves.

Example 12.3a Find a formula for the area between two curves $y = f(x)$ and $y = g(x)$ over the interval $a \leq x \leq b$.

We suppose $g(x) \leq f(x)$ in this interval. For a particular value of x we consider a thin strip of width Δx at the given value of x and lying between the two graphs. The length of the strip is approximately $f(x) - g(x)$. See figure 12.3. This gives an estimate of $[f(x) - g(x)]\Delta x$ for the area of the strip. The Riemann sum for the total area is a sum of such terms. To indicate that

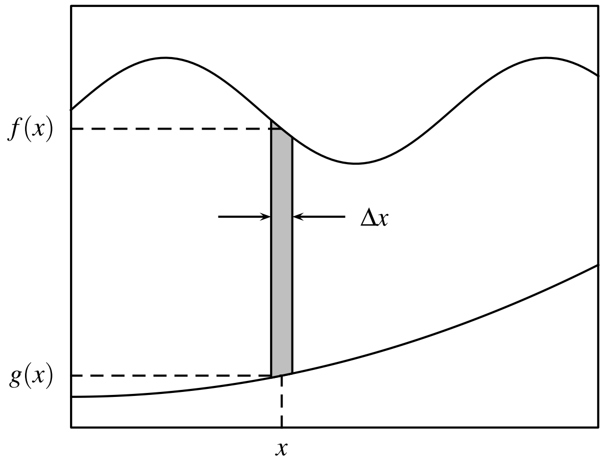


Figure 12.3:

we shrink the length of the intervals towards zero we replace Δx by dx and the summation sign Σ by the sign $\int_a^b$ for the definite integral. Then the formula

$$\sum [f(x) - g(x)] \Delta x$$

for the Riemann sum becomes the definite integral

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

$\diamondsuit$

12.4 Solids of revolution

The disk method

Suppose we have a continuous function $f(x) > 0$ defined on the interval $a \le x \le b$. Consider the region bounded by the x-axis, the lines $x = a$ and $x = b$, and the graph of f. Now imagine that this region is rotated around the x-axis, sweeping out a 3-dimensional *solid of revolution*. We look at the problem of working out the *volume* of this solid. We have already seen that we can approximate the *area* of the indicated region by a Riemann sum. Figure 12.4 shows the area in question, as well as a typical rectangle for a Riemann lower sum on the interval.

If $[a, b]$ is partitioned into N equal subintervals, then each subinterval has length

$$\Delta x = \frac{b - a}{N}.$$

As usual we let m_i and M_i be the minimum and maximum values of $f(x)$ on the ith subinterval. If we take the rectangle of height m_i based on this subinterval and rotate it around the x-axis

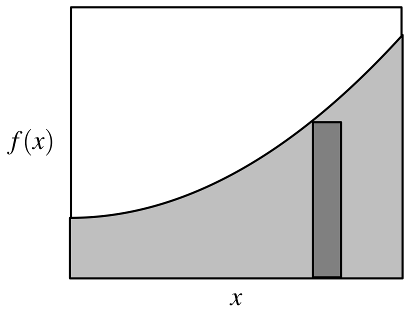


Figure 12.4:

it sweeps out a *disk* of thickness Δx and radius m_i. If we do the same for all rectangles contributing to the Riemann lower sum we get a solid of revolution made up of coaxial disks with a total volume of

$$\sum_{i=1}^{N} \pi m_i^2 \times \Delta x\,.$$

Figure 12.5 shows the disk generated by rotation on a single rectangle.

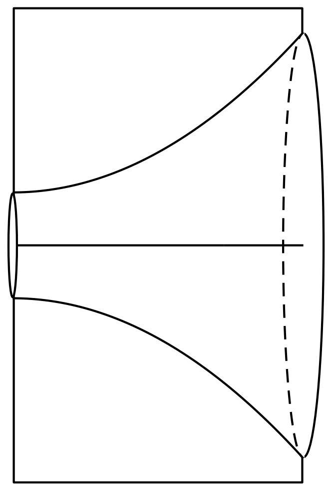


Figure 12.5:

Since each disk is contained in the original solid of revolution, this gives us a lower estimate for the volume of this solid. Replacing m_i by M_i, the maximum value of f on the ith subinterval, gives an upper estimate. If V is the volume we are trying to calculate, then

$$\sum_{i=1}^{N} \pi m_i^2 \times \Delta x \leq V \leq \sum_{i=1}^{N} \pi M_i^2 \times \Delta x\,.$$

But the first and last expressions here are just lower and upper Riemann sums for the function

$$F(x) = \pi f(x)^2.$$

Since the definite integral is the only number that satisfies all these inequalities, we conclude that

$$V = \int_a^b \pi f(x)^2 \, dx. \tag{12.4a}$$

If we can find an antiderivative $G(x)$ for the function $g(x) = \pi f(x)^2$ then we can calculate the volume directly as

$$V = G(b) - G(a).$$

For obvious reasons, this method of calculating volumes is sometimes called the **disk method**.

A Simpler way to obtain the formula

For a solid of revolution we can get the formula for volume more directly by finding the relation between a small change Δx in x and the corresponding small change ΔV in volume. Between points x and $x + \Delta x$ the volume of the solid of rotation is approximately that of a disk of radius $f(x)$ and thickness Δx. We write this as the formula

$$\Delta V \approx \pi f(x)^2 \Delta x,$$

where ΔV is the contribution to the total volume coming from the interval between x and $x + \Delta x$ and '$\approx$' means 'approximately equal'. Figure 12.6 shows the dimensions of the disk in question.

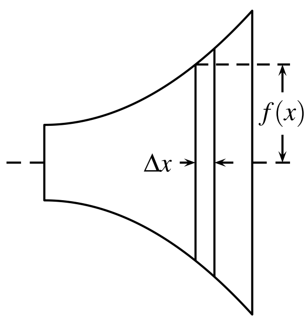


Figure 12.6:

Imagine the whole solid sliced up into thin disks of thickness Δx. The total volume V is then the sum of volumes of the individual disks, so

$$\text{TOTAL VOLUME } V = \sum \Delta V \approx \sum \pi f(x)^2 \Delta x.$$

As the number of subintervals increases and the thickness of the disks is decreased towards zero the error in the approximation also goes towards zero. The terms on the right are Riemann sums converging to the definite integral of 12.4a. Therefore

$$V = \lim_{\Delta x \to 0} \sum \pi f(x)^2 \times \Delta x = \int_a^b \pi f(x)^2 \, dx,$$

as before. We will use this type of argument again in the examples which follow.

Example 12.4b Find a formula for the volume of the solid generated by rotating the graph of the function

$$y = \sin x, \qquad 0 \le x \le 3$$

about the x-axis. Do not attempt to get a numerical answer; just express the answer as a definite integral.

We give an abbreviated version of the preceding argument, which may help you reconstruct the formula.

Slice the area under the graph of $\sin x$ on the interval $[0,3]$ into thin vertical strips. A typical such strip has width Δx and approximate height $\sin x$. Rotation of this strip about the x-axis generates a disk of approximate radius $\sin x$ and thickness Δx. So

$$\text{Volume of disk} \approx \pi(\sin x)^2 \Delta x\,.$$

As the strips are made narrower and narrower the sum of their volumes (over the interval $0 \le x \le 3$) converges towards the integral

$$\int_0^3 \pi(\sin x)^2 \, dx.$$

$\Diamond$

The shell method

We have already looked at one method for calculation of the volumes of solids of revolution using approximation by coaxial disks. Here we derive another method using approximation by *cylindrical shells*.

Consider the region below the graph of the (positive-valued) function $f(x)$ and above the x-axis between $x = a$ and $x = b$. We now look at the problem of finding the volume swept out when this region is rotated around the vertical axis $x = 0$ instead of the horizontal x-axis.

This time, rotating a thin rectangle as shown in Figure 12.7 about the y-axis generates a cylindrical shell of approximate height $f(x)$ and thickness Δx. In order to work out the total volume, we need an approximate formula for the volume of the solid lying between radii x and $x + \Delta x$. Since Δx is small, the cylinder can be thought of a thin shell. We can imagine cutting through it and flattening it out into a rectangle of width $2\pi x$ (the circumference of

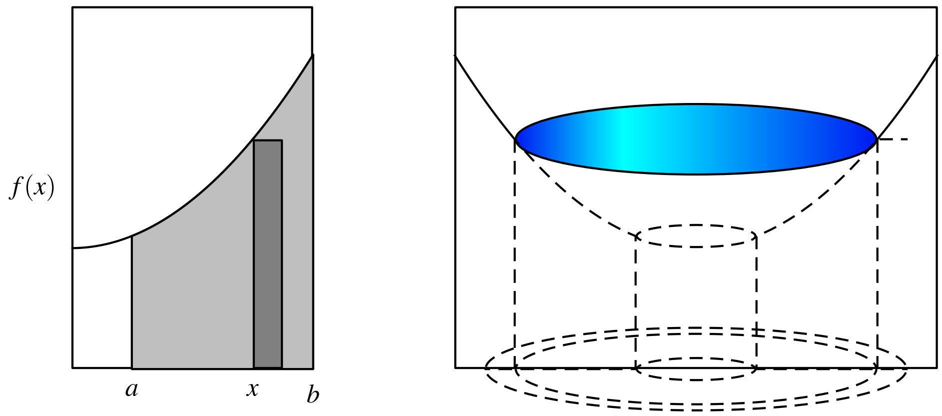


Figure 12.7:

the cylinder) and height $f(x)$ (the height of the cylinder). Since the thickness is Δx this contributes an amount

$$\Delta V \approx 2\pi x \times f(x) \times \Delta x$$

to the total volume. Summing up over the subintervals of a partition of $[a,b]$ gives

$$V \approx \sum 2\pi x f(x)\, \Delta x.$$

Letting $\Delta x \to 0$ gives the formula

$$\boxed{\text{VOLUME}\quad V = 2\pi \int_a^b x f(x)\, dx.}$$

In contrast to the disk method used previously, this is called the **shell method**.

It should be noted that the disk and shell methods have been used here to find the volume of *different* solids. Sometimes, though, either method can be used to find the volume of the same solid. Both methods will give the same result if correctly applied, but is often the case that one method will lead to easier calculations than the other.

Example 12.4c Find a definite integral for the volume of a donut with circular cross-section, inner hole of radius r and an outer radius of R.

In order to simplify the calculations we introduce the constants

$$a = \frac{R-r}{2} \quad \text{and} \quad b = \frac{R+r}{2}.$$

These correspond to the dimensions shown in Figure 12.8, which shows a cross-section through the donut. In order to apply the shell method we take a cylindrical shell of radius x

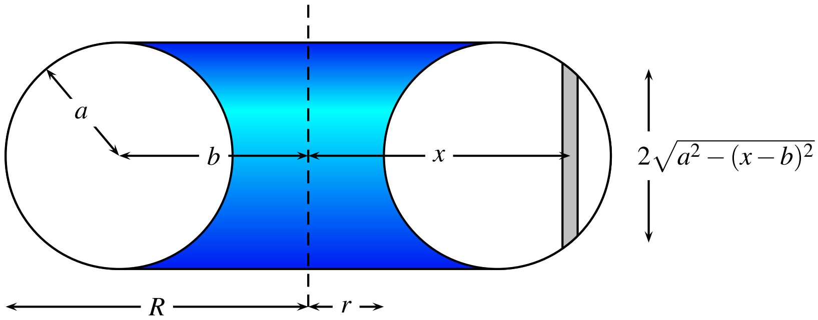


Figure 12.8:

and thickness Δx around the donut's axis of symmetry. Simple trigonometry shows that the height of this shell is $2\sqrt{a^2-(x-b)^2}$. To estimate the volume of this cylinder imagine that it is cut and rolled out into a rectangle of dimensions $2\sqrt{a^2-(x-b)^2}$ by $2\pi x$. Since the thickness is Δx, this contributes approximately

$$4\pi x\sqrt{a^2-(x-b)^2}\,\Delta x$$

to the total volume. We need to do this for $r \leq x \leq R$ to get the total volume. Converting to an integral we end up with the formula

$$\int_r^R 4\pi x\sqrt{a^2-(x-b)^2}\,dx$$

for the total volume. $\Diamond$

Summary of Chapter 12

- **Applications of integration** are usually based on the idea of decomposing some quantity into small pieces, finding a simple formula for each of the pieces, and summing the results into a Riemann sum approximating the total quantity. Taking the limit as the size of the pieces shrinks to zero gives the required quantity as a definite integral.
- **The length of a curve** was calculated by subdividing it into small sections called *the arc length differential* and then adding up all the small-sections to construct a Riemman sum. Finally, taking the limit as the number of sections tend to infinity, we obtained a definite integral that gives the total length of the curve.
- **Area between two curves** and **volumes of solids of revolution** were also calculated by subdividing them into thin disks or shells and then taking the limit too obtain a definite integral.

Exercises

12.1 Consider the part of the hyperbola $x^2 - y^2 = a^2$ in the first quadrant and between the lines $x = a$ and $x = b$, where $0 < a < b$. Find the volume obtained by rotating this curve

a) About the x-axis,
b) about the y-axis.

12.2 Let R be the region in the first quadrant bounded by the coordinate axes and the parabola $y = 4 - x^2$. Use the disc method to calculate the volume of the solid formed by revolving R about the y-axis. Use the shell method to check your answer.

Appendix A

Formal Definition of Limits

The informal definition of limit given in Chapter 4 is quite clear in an intuitive sense; however, it is not very precise. Example 4.1c shows that it may lead to the wrong guess and therefore we need a more rigourous approach.

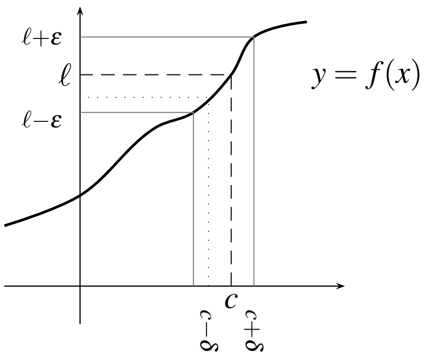
A mathematical way of saying that we can make $f(x)$ "as close as we like to ℓ" is to say that whenever $\varepsilon > 0$ is a "small" positive number then we can always ensure that $f(x)$ is between $\ell+\varepsilon$ and $\ell-\varepsilon$; that is $\ell-\varepsilon < f(x) < \ell+\varepsilon$, or in concise mathematical notation,

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

The informal definition says that this should be true whenever x is "sufficiently close" to ℓ; in other words, whenever $c-\delta < x < c+\delta$ for another "small" positive number δ; that is, $|x-c| < \delta$. In fact, since the informal definition states that x is close to c but **not equal** to c, the condition we actually require is $0 < |x-c| < \delta$. Putting all of this together gives the following mathematically precise definition of a limit.

> **Formal (rigorous) definition of limit**
> Suppose that ℓ is a real number. Then the limit of $f(x)$ as x approaches c is equal to ℓ if for each number $\varepsilon > 0$ there exists a number $\delta > 0$ such that $|f(x)-\ell| < \varepsilon$ whenever $0 < |x-c| < \delta$.

Since the ε, δ notation has become standard, this more precise version of the definition of limit is also known as the $\varepsilon-\delta$ definition. The roles of ε and δ are illustrated in the following diagram:



This picture tells us what we need to do in order to test whether or not ℓ is the limit of $f(x)$ as x approaches c. For each $\varepsilon > 0$ we consider $\ell \pm \varepsilon$, on the y–axis, and then look at the graph to see how small we need to make δ, on the x–axis, so that $\ell-\varepsilon < f(x) < \ell+\varepsilon$ whenever $c-\delta < x < c+\delta$ and $x \neq c$.

Another way of stating the formal definition is that given a number $\varepsilon > 0$, there exists a number $\delta > 0$ such that if $0 < |x-c| < \delta$ then $|f(x)-\ell| < \varepsilon$. It is important to understand that the condition $|f(x)-\ell| < \varepsilon$ must be satisfied for values of x to the right of c on the real number line (that is, when $c < x < c+\delta$) and also to the left of c (that is, when $c-\delta < x < c$).

Thus the definition of "limit" is a two-sided one and a consequence of this is that if $f(x)$ has a limit as x approaches c, this limit is unique; that is, there is *only one limit* as x approaches c regardless of whether it approaches from the left or the right.

Notice also that the value of δ will depend on ε and also on $f(x)$; in general, there will be a different value of δ for each value of ε and sometimes we emphazise this fact by writing $\delta = \delta(\varepsilon)$.

Example A.0a Use the formal definition to prove that, if $f(x) = 4x-6$ then $\lim\limits_{x\to 2} f(x) = 2$ (hence, in this example $c = 2$ and $\ell = 2$).

This example is quite straight forward if we think about it intuitively. The purpose, however, is to show how a formal proof is constructed. A simple problem is the best place to start.

The proof is done in two stages: 1) Guess a value for δ and 2) Show that the chosen δ works:

1) *Guessing a value for δ*

   Let ε be a given positive number. We want to find a number δ such that

   $$\text{if} \quad 0 < |x-c| < \delta \quad \text{then} \quad |f(x)-\ell| < \varepsilon$$

   or, in our case,

   $$\text{if} \quad 0 < |x-2| < \delta \quad \text{then} \quad |(4x-6)-2| < \varepsilon$$

   Now we have that $|f(x)-\ell| = |(4x-6)-2| = |4x-8| = 4|x-2|$. Therefore we want:

   $$\text{if} \quad 0 < |x-2| < \delta \quad \text{then} \quad 4|x-2| < \varepsilon$$

   or, equivalently,

   $$\text{if} \quad 0 < |x-2| < \delta \quad \text{then} \quad |x-2| < \frac{\varepsilon}{4}.$$

   This suggests that we should choose $\delta = \dfrac{\varepsilon}{4}$.

2) *Show that δ works*

   Given $\varepsilon > 0$, choose $\delta = \dfrac{\varepsilon}{4}$. If $0 < |x-2| < \delta$, then

   $$|f(x)-\ell| = |(4x-6)-2| = |4x-8| = 4|x-2| \le 4\,\delta = 4\left(\frac{\varepsilon}{4}\right) = \varepsilon.$$

   That is, we have found a $\delta = \dfrac{\varepsilon}{4}$ so that for any ε,

   $$\text{if} \quad 0 < |x-2| < \delta \quad \text{then} \quad |(4x-6)-2| < \varepsilon.$$

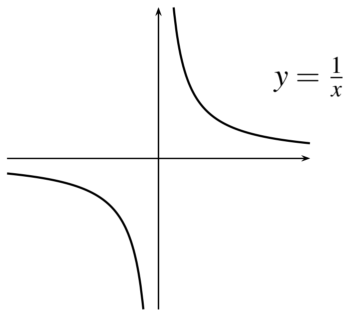
   Therefore by the formal definition of limit, we have $\lim_{x\to 2}\big(4x-6\big) = 2$.

$\Diamond$

Limits at infinity – Horizontal asymptotes

The definition of $\lim_{x\to c} f(x)$ given in Section 4.1 tells us about the expected behaviour of $f(x)$ as x approaches the **finite** number c. We can also ask how $f(x)$ behaves as x becomes arbitrarily **large and positive** ($x \to \infty$) and arbitrarily **large and negative** ($x \to -\infty$).

Looking at the graph of the function $f(x) = \dfrac{1}{x}$ below, it is intuitively clear that $\lim_{x\to\infty} \dfrac{1}{x} = \lim\limits_{x\to-\infty} \dfrac{1}{x} = 0$.



Based on these observations, the precise definition of a limit as $x \to \pm\infty$ can be formulated as follows:

Limits at infinity
Suppose that ℓ is a real number. Then the limit of $f(x)$, as x approaches ∞, is equal to ℓ if for each $\varepsilon > 0$ there exists a number $N > 0$ such that

$$|f(x) - \ell| < \varepsilon \quad \text{whenever} \quad x > N.$$

In this case we write $\lim_{x\to\infty} f(x) = \ell$.

Similarly, the limit of $f(x)$, as x approaches $-\infty$, is equal to ℓ if for each $\varepsilon > 0$ there exists a number $N > 0$ such that

$$|f(x) - \ell| < \varepsilon \quad \text{whenever} \quad x < -N.$$

In this case we write $\lim_{x\to-\infty} f(x) = \ell$.

Referring back to the graph of the function $f(x) = \dfrac{1}{x}$ we observe that the curve gets closer to the x-axis (the line $y = 0$) when $x \to \pm\infty$. In this case we say that the line $y = 0$ is a horizontal asymptote of the curve $y = \dfrac{1}{x}$. In general, we have the following definition:

Horizontal asymptotes
The line $y = L$ is called a horizontal asymptote of the curve $y = f(x)$ if either

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

As with limits as $x \to c$, we do not really want to use the formal definitions given above in order to calculate limits at infinity; luckily, *the limit laws we have already discussed also hold for limits at infinity*.

Appendix B

Geometric proof that $\lim_{x\to 0}\sin x/x = 1$

Here we use a geometric argument to show that $\lim\limits_{x\to 0}\dfrac{\sin x}{x} = 1$. This is the limit we calculated in Example 4.1b using a more heuristic approach.

Intuitively, this seems unlikely because $\dfrac{1}{x}$ becomes very large as x approaches 0, so we might guess that $\dfrac{\sin x}{x}$ must also get very large. However, this limit is in fact equal to 1. The problem with our intuition is that even though $\dfrac{1}{x}$ does become very large as x approaches 0, $\sin x$ simultaneously becomes very small; what happens is that these two effects exactly cancel out.

Just in case you are not convinced, Figure B.1 below shows again the graph of $y = \dfrac{\sin x}{x}$ with the gap at $x = 0$.

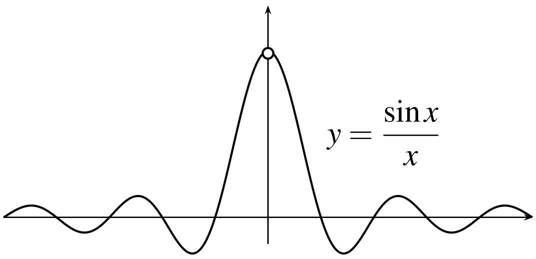
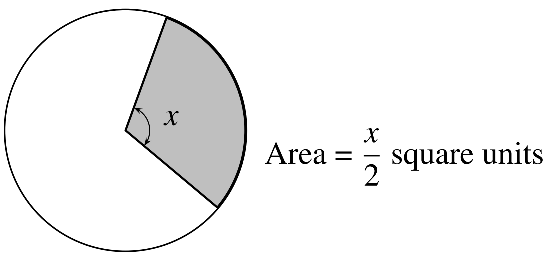


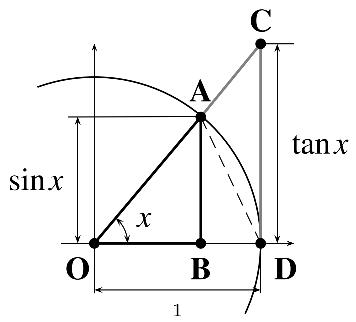
Figure B.1:

To show that $\lim\limits_{x\to 0}\dfrac{\sin x}{x} = 1$ we are going to use the squeeze law.

Before we begin we need to recall some facts about angles measured in radians. From school you know that the area of a unit circle is π units squared. You should also know that the area of a sector of the unit circle is $\dfrac{x}{2}$ square units if its inner angle is x radians — the area of the entire circle corresponds to the case $x = 2\pi$.



In order to calculate $\lim_{x\to 0} \frac{\sin x}{x}$ suppose that x is a non–zero angle *measured in radians* and consider the following picture of a circle of radius 1 in the Cartesian plane.



Now, the area of the *triangle AOD* is less than or equal to the area of the *sector AOD*; in turn, this sector has area less than or equal to the area of the *triangle COD*. We now compute these areas. First consider the triangle AOD. Its base, the line OD, has length 1, and its height is $\sin x$ (the length of the line AB). Therefore, AOD has area $\frac{1}{2}\sin x$. Next, as noted above, the area of the sector AOD is $\frac{x}{2}$. Finally, the length of the line CD is $\tan x$, and hence the area of COD is $\frac{1}{2}\tan x$. Combining these equations we see that

$$\begin{aligned} \frac{1}{2}\sin x \le \frac{1}{2}x \le \frac{1}{2}\tan x &\implies \sin x \le x \le \tan x \\ &\implies 1 \le \frac{x}{\sin x} \le \frac{1}{\cos x} \quad (\text{if } \sin x > 0), \\ &\implies 1 \ge \frac{\sin x}{x} \ge \cos x, \end{aligned}$$

where the last line follows by taking reciprocals. Remember when taking reciprocals of a set of inequalities where all terms have the same sign, we **reverse the inequalities**. We have assumed that $x \ge 0$ here so $\sin x \ge 0$; you might like to write out the argument for $x < 0$.

Rewriting the last inequality, we have

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

Now $\lim_{x\to 0} 1 = 1$ and $\lim_{x\to 0}\cos x = \cos 0 = 1$ because $\cos x$ is continuous. Therefore by the squeeze law, $\lim_{x\to 0}\frac{\sin x}{x}$ exists and equals 1, as claimed.

Appendix C

Linear approximations and differentials

Linear approximations

A differentiable function $f(x)$ can be approximated, close to the point $x = a$, by the tangent to the curve at that point. The equation of this tangent line is $y = f(a) + f'(a)(x-a)$ which can be written as the linear function

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

called the **linearisation** of the function f at a. The idea is to approximate the function f near $x = a$ by its linearisation, that is,

$$f(x) \approx L(x) \quad \text{near} \quad x = a.$$

Example C.0a Find the linearisation of the function $f(x) = \sqrt{x+3}$ at $x = 1$ and use it to calculate approximations of the numbers $\sqrt{3.98}$ and $\sqrt{4.05}$.

To calculate approximations of the given numbers, we first need to choose a point near $x = 1$ to evaluate the function. Writing $3.98 = 0.98 + 3$ gives us the clue that we need to choose $x = 0.98$. Similarly, writing $4.05 = 1.05 + 3$ means that $x = 1.05$.

Now, the linearisation of f at $x = 1$ is

$$\begin{aligned} L(x) &= f(1) + f'(1)(x-1) \\ &= 2 + \frac{1}{4}(x-1) \\ &= \frac{7}{4} + \frac{x}{4}. \end{aligned}$$

So we can approximate $f(x)$ as follows:

$$f(x) = \sqrt{x+3} \approx \frac{7}{4} + \frac{x}{4} \quad \text{when} \quad x \quad \text{is near } 1.$$

Therefore,

$$\sqrt{3.98} = \sqrt{0.98+3} \approx \frac{7}{4} + \frac{0.98}{4} = 1.995$$

and

$$\sqrt{4.05} = \sqrt{1.05+3} \approx \frac{7}{4} + \frac{1.05}{4} = 2.0125.$$

Because $x = 0.98$ and $x = 1.05$ are near $x = 1$, the approximations are reasonably good. If fact, using a calculator we get the exact values $\sqrt{3.98} = 1.99499$ and $\sqrt{4.05} = 2.01246$ which, after rounding to four decimal places, coincide with the approximate values. $\Diamond$

Differentials

In this section we introduce the concept of **differentials** and use them to find linear approximations to differentiable functions. Recall the alternative notation of the definition of derivative as a function given in Section 5.2,

$$f'(x) = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} \quad \text{where} \quad \Delta y = f(x+\Delta x) - f(x).$$

From the meaning of limit we see that for Δx "small", the ratio $\dfrac{\Delta y}{\Delta x}$ is close to the derivative $f'(x)$, that is,

$$\frac{\Delta y}{\Delta x} \approx f'(x) \quad \text{and therefore} \quad \Delta y \approx f'(x)\, \Delta x.$$

The last equation tells us that if x changes by a *small* amount Δx, then y will change by approximately the amount $\Delta y \approx f'(x)\, \Delta x$. This motivates the introduction of the concept of differential.

Differential
Let $y = f(x)$ where f is a differentiable function and let Δx be any nonzero real number. Then

a) The differential dx is a variable given by $dx = \Delta x$.
b) The differential dy of the function f is another function given by

$$dy = f'(x)\, dx \quad \text{with alternative notation} \quad df = f'(x)\, dx.$$

The differential of a function of one variable is itself a function of *two variables*; both x and dx are needed to evaluate the differential.

Note that if the definitions seem somewhat artificial it is because they have been introduced so that we can manipulate the symbols dx and dy and treat the notation for the derivative $\dfrac{dy}{dx}$ as a *"ratio"*, since in terms of differentials we have

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

Example C.0b Calculate the differential df of the function $f(x) = x^3 + 5x^2$.

Since $f'(x) = (3x^2 + 10x)$ then $df = f'(x)dx = (3x^2 + 10x)dx$. $\Diamond$

Example C.0c Write down the differential df of $f(x) = 2x\cos x$ and find an expression for the differential in terms of dx when $x = 0$.

Since $f'(x) = 2\cos x - 2x\sin x$ then $df = f'(x)dx = \big(2\cos x - 2x\sin x\big)dx$ and when $x = 0$ the differential becomes $df = 2dx$. $\Diamond$

Relationship between the increment Δy and the differential dy

Refering to the figure below, consider the function $y = f(x)$ and let x_0 be a fixed number where $f'(x_0)$ exists. For any value of Δx, we have $dx = \Delta x$. The equation

$$dy = f'(x_0)\,dx$$

is the equation of the tangent to the curve with slope $f'(x_0)$ in the coordinates (dx, dy).

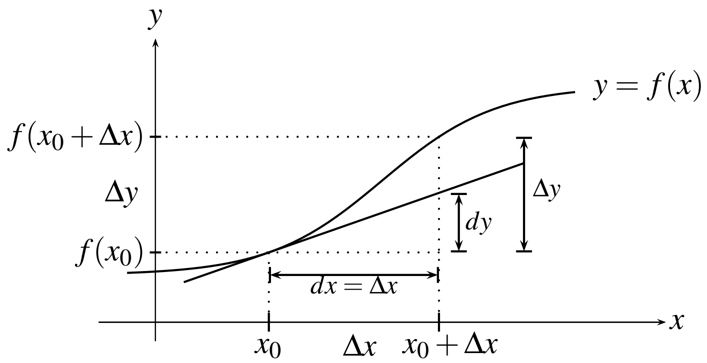
Observe carefully in the diagram that $\Delta y = f(x_0 + \Delta x) - f(x_0)$ changes along the curve $y = f(x)$ while dy changes along the tangent line. Therefore, assuming small values of $dx = \Delta x$, the values of dy and Δy become closer together, and we can say that

$$\Delta y \approx dy.$$

From the diagram we also note that $f(x_0 + \Delta x) = f(x_0) + \Delta y$ and using the approximation $\Delta y \approx dy$ we obtain $f(x_0 + \Delta x) = f(x_0) + \Delta y \approx f(x_0) + dy$. That is,

$$f(x_0 + \Delta x) \approx f(x_0) + dy. \tag{C.0d}$$

What this relation is telling us is that, if we have the value of a function $f(x_0)$ at a point x_0, to find an approximate value at a nearby point $x_0 + \Delta x$ we only have to find $dy = f'(x_0)\Delta x$ using the simpler equation of the tangent line.



Example C.0e

Given the function $f(x) = \sqrt{3x + 4}$, use differentials to find an approximate value for $f(7.1)$.

We first need to choose a point to evaluate the differential. We can see that $f(7)$ is easy to evaluate since $f(7) = 5$, therefore a sensible choice is $x_0 = 7$. Then the differential in x is $dx = \Delta x = 7.1 - 7 = 0.1$ and the derivative

$$f'(x) = \frac{3}{2\sqrt{3x+4}} \qquad \text{and so} \qquad f'(7) = \frac{3}{10} = 0.3.$$

Hence

$$dy = f'(x_0)dx = 0.3 \times 0.1 = 0.03.$$

This differential is the approximate *change* in $f(x)$ between $x = 7$ and $x = 7.1$ and so an approximation for $f(7.1)$, as given by equation (C.0d), is

$$f(7.1) = f(x_0 + \Delta x) = f(7 + 0.1) \approx f(7) + dy = 5 + 0.03 = 5.03.$$

Using a calculator gives $f(7.1) = 5.02991$ to six significant figures and so the approximation using differentials is accurate to three significant figures. $\Diamond$

Example C.0f The volume of a sphere is given by $V = \frac{4}{3}\pi r^3$ where r is the radius of the sphere. By approximately how much is the volume of a sphere of 6cm radius reduced if 0.25 cm is shaved off the radius of the sphere?

Here we seek to find the differential of V at $r = 6$. The differential in r is $dr = -0.25$, since we reduce the radius from 6 to 5.75. The differential in V is $dV = \dfrac{dV}{dr}dr$ where the derivative is evaluated at $r = 6$.

$$\frac{dV}{dr} = 4\pi r^2 = 144\pi \text{ when } r = 6.$$

So $dV = 144\pi \times -0.25 = -36\pi \approx -113.1$. Therefore the volume of the sphere is reduced by approximately 113.1 cm^3 if 0.25 cm is shaved off its radius. $\Diamond$

Example C.0g Find the approximate error in the area of a circle of radius 30, if there is an error of ± 0.2 cm in the measurement of its radius.

Let $A = \pi r^2$ be the area of the circle, where r is its radius. The error is approximated by the differential of A: $dA = \dfrac{dA}{dr}dr = 2\pi r dr$.

Here $r = 30$ and $|dr| = 0.2$. We use the absolute value of dr here because it is possible that the measurement is underestimated ($dr < 0$) or overestimated ($dr > 0$). We need only find the magnitude of the error in A and so we seek $|dA|$.

$$|dA| = \left|\frac{dA}{dr}dr\right| = \left|\frac{dA}{dr}\right||dr| = 60\pi \times 0.2 = 12\pi \approx 37.7$$

Therefore the estimated maximum error in the area of the circle is 38 cm^2. $\Diamond$

Example C.0h Relative error – Sometimes errors are given as a percentage or a fraction of the measurement. This is known as the *"relative error"*. These can be calculated using differentials with only a little more work.

For example: if the relative error in the measurement of the radius of a circle is 5%, find the relative error in the circle's area.

If the relative error of the radius is 5%, or 0.05 as a decimal, we have $\dfrac{|dr|}{r} = 0.05$. We seek to find $\dfrac{|dA|}{A}$:

$$\frac{|dA|}{A} = \frac{\left|\frac{dA}{dr}dr\right|}{A} = \frac{|2\pi r dr|}{\pi r^2} = \left|\frac{2\pi r dr}{\pi r^2}\right| = \left|\frac{2dr}{r}\right| = 2\left|\frac{dr}{r}\right| = 2 \times 0.05 = 0.1.$$

Therefore, the error in the area of the circle is 0.1, as a decimal, or 10% as a percentage. As you can see from the example, the aim is to rearrange the expression for $\dfrac{|dA|}{A}$ until the term $\dfrac{dr}{r}$ appears on the right hand side. $\Diamond$

Appendix D

The Distance Problem

Imagine a car accelerating along the road over a period of 10 seconds. Suppose the car starts with a speed of 5m/sec and ends up with a speed of 32.5 m/sec. What can we say about the distance travelled? We can make a rough estimate as follows. Since the car is accelerating the speed is always increasing. In particular, the speed is always between 5 m/sec and 32.5 m/sec. Over the period of 10 seconds the car therefore travels at least $5 \times 10 = 50$ metres, but no more than $32.5 \times 10 = 325$ metres. We can write these two inequalities as:

$$50\,\mathrm{m} \leq \textsc{Distance travelled} \leq 375\,\mathrm{m}\,.$$

This is a very rough estimate indeed. We can do much better if we know more about the velocity at intervening points of time. For example, suppose that we measure the velocity every two seconds. We can present the results as a table, which might look like the following:

Time (sec)	0	2	4	6	8	10
Velocity (m/sec)	5	14.5	22	27.5	31	32.5

The minimum and maximum velocities over the first two seconds are 5 m/sec and 14.5 m/sec. Therefore the distance travelled in this period is between $5 \times 2 = 10$ m and $14.5 \times 2 = 29$ m. Applying this to each interval in turn and adding up over all five intervals, we get a lower estimate of

(D.0a) $$(5 \times 2) + (14.5 \times 2) + (22 \times 2) + (27.5 \times 2) + (31.0 \times 2) = 200\,\mathrm{m},$$

and an upper estimate of

(D.0b) $$(14.5 \times 2) + (22 \times 2) + (27.5 \times 2) + (31 \times 2) + (32.5 \times 2) = 255\,\mathrm{m}.$$

The gap between the two estimates is now much smaller, with a maximum possible error of $255 - 200 = 55$ m. It is very instructive to draw a graph of velocity against time and use it to interpret these calculations.

This is done in Figure D.1 below, where the curved line shows the actual velocity of the car plotted against the time t. On each 2 second interval along the t-axis the height of the dark rectangle is equal to the *minimum* velocity on that interval. Since the velocity is increasing, this always occurs on the left endpoint of each subinterval. Thus the first rectangle has height 5, the second height 14.5, and so on. The total height of the dark and light rectangles together is equal to the *maximum* velocity on each interval. For an increasing function this will occur at the right endpoint.

We relate this geometrical construction to the distance travelled by introducing the idea of *area*. The width of each rectangle is a time interval and the height corresponds to our estimate of velocity over the same interval. Therefore the product WIDTH $\times$ HEIGHT gives the distance travelled during the interval, assuming the velocity is constant and equal to the height of the rectangle. Of course this product is also just the *area* of the rectangle. Therefore the lower estimate for the velocity given by (D.0a) is just the sum of the areas of the dark rectangles. Similarly the expression (D.0b) is the total area of the dark and light rectangles taken together.

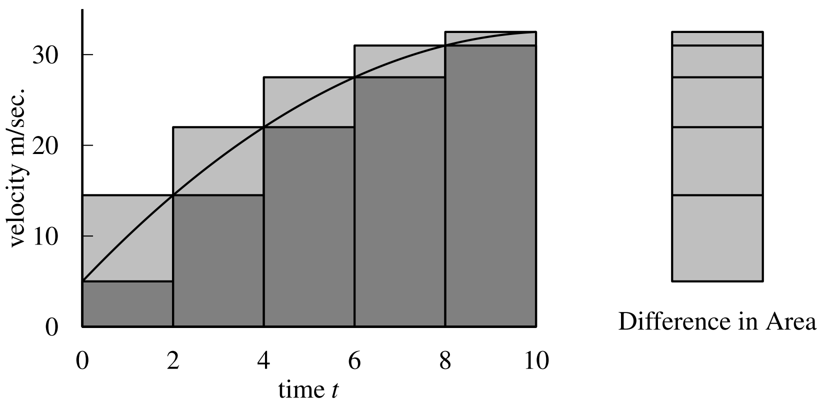


Figure D.1:

The difference between these upper and lower estimates of the distance is then equal to the sum of the areas of the light rectangles. In Figure D.1 the light rectangles have been copied over to the right of the diagram and stacked together, in order to better visualize their total area. In fact it is easy to see that the composite rectangle has dimensions $(32.5-5)\times 2$, with a total area of 55, in agreement with our earlier calculation.

With more data on the car's speed we can improve accuracy further still. Suppose we record the speed twice as often, so every second:

Time (sec)	0	1	2	3	4	5	6	7	8	9	10
Velocity (m/sec)	5	10	14.5	18.5	22	25	27.5	29.5	31	32	32.5

This gives us 10 intervals instead of 5, and we can again use the lowest and highest speeds on each interval to estimate the distance travelled. This is shown graphically in Figure D.2.

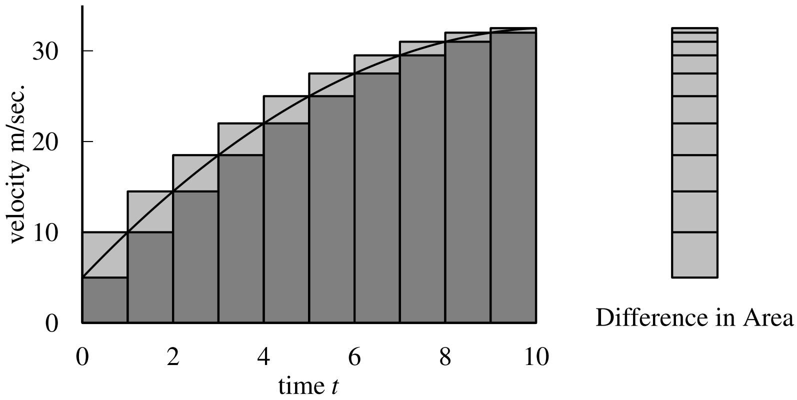


Figure D.2:

As before the areas of the shaded rectangles give lower and upper estimates for the distance travelled, and difference between these two estimates is equal to the total area of the light rectangles. Comparison with Figure D.1 shows that this difference is now much smaller (in fact it is equal to half its previous value). Adding up the upper and lower estimates of distance over each 1 second interval gives us inequalities:

$$215\,\mathrm{m} \leq \text{DISTANCE TRAVELLED} \leq 242.5\,\mathrm{m}\,.$$

The maximum possible error is now $242.5 - 215 = 27.5$ m. Of course, we can continue in the same way, using shorter and shorter subintervals. Figure D.3 shows the result of measuring the velocity every 0.5 seconds.

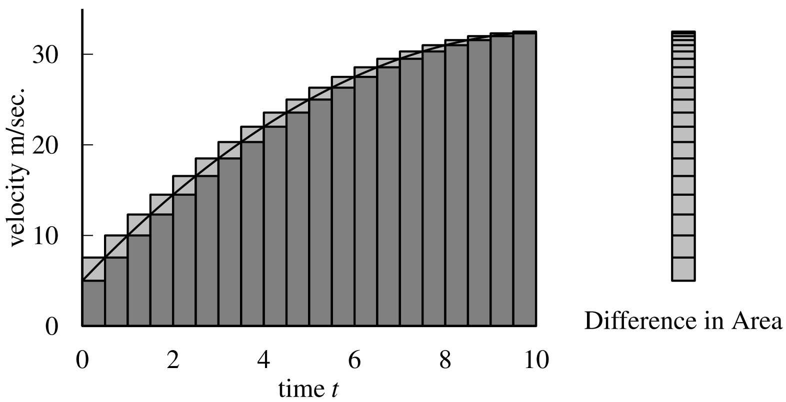


Figure D.3:

The difference between the upper and lower estimates is again equal to the area of the rectangle drawn at the side of the figure. It should be clear by now that by taking small enough steps, *we can make this area as small as we like*.

In mathematical terms this means that both the upper and lower estimates approach a common *limit* as the size of the steps shrinks towards zero. There is an obvious relation between this limit and area under the curved line. In each case the lower estimate is the sum of areas of rectangles which lie *inside* this curve. The upper estimate is a sum of areas of rectangles which *enclose* the curve. The *area under the curve*, like the total distance, therefore also lies between these upper and lower estimates. Since the upper and lower estimates have a common limit, there is only one number with this property. We conclude that

$$\boxed{\text{TOTAL DISTANCE} = \text{AREA UNDER THE CURVE.}}$$

There are two new concepts here. First, we have a way of estimating total distance travelled from a knowledge of velocity. By 'sampling' the velocity sufficiently frequently, we can make this estimate as accurate as we wish. Second, we see that the total distance travelled is the same as the area under the graph of velocity plotted against time. Note that the argument does not depend on having an algebraic formula for the velocity, and it does not use any differential calculus. The only place where we use the fact that velocity is rate of change of distance with time is in the formula DISTANCE $=$ VELOCITY $\times$ TIME for motion with *constant* velocity. In the next section we generalize this argument into a purely mathematical construction which we can apply to any continuous function.

Appendix E

Growth Rates

One of the important properties of a function $f(x)$ is its **growth rate**—how does $f(x)$ behave as $x \to \infty$. According to the definitions of the previous section we have $x^a = \exp(a \ln x)$. From the behaviour of the exponential and logarithm functions we deduce that

$$\lim_{x\to\infty} x^a = \begin{cases} 0, & \text{if } a < 0, \\ 1, & \text{if } a = 0, \\ \infty, & \text{if } a > 0. \end{cases}$$

But the limit of a function $f(x)$ as $x \to \infty$ is not the only interesting feature. Also significant are the *relative growth rates* of different functions. We say that a function $f(x)$ *grows faster* than $g(x)$ as $x \to \infty$ if

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

This is equivalent to the condition $g(x)/f(x) \to 0$ as $x \to \infty$. For example, if $a > b$ then x^a grows faster than x^b, since

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

(since $a - b > 0$).

How does the function $\ln x$ fit into this picture? In fact $\ln x$ grows more slowly than *any* positive power of x. This is not obvious from the graph of $\ln x$. Figure E.1 shows the graph of $\ln x$ along with the graphs of x^a for the cases $a = 0.5$ and $a = 0.2$. At least up to $x = 10$ the logarithm seems to be grower faster than $x^{0.2}$. But the power function eventually overtakes it, as the following argument shows.

Lemma 1 *For all $x > 0$ and $b > 0$ we have the inequality*

$$\ln(x^b) < x^b, \quad \text{or} \quad \ln x < \frac{x^b}{b}.$$

Proof. This is certainly true for $0 < x \le 1$, since $\ln x \le 0$ and $x^b > 0$ for these values of x. Define

$$F(x) = \frac{x^b}{b} - \ln x .$$

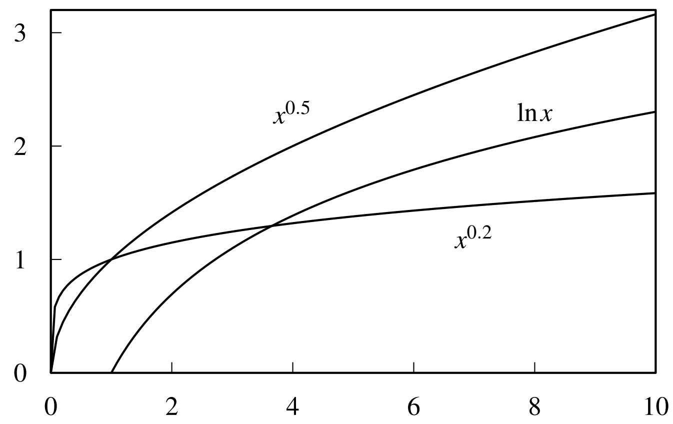


Figure E.1:

Then $F(1) = 1/b > 0$ and

$$F'(x) = x^{b-1} - \frac{1}{x} = \frac{x^b - 1}{x}.$$

Since $b > 0$ this shows that $F'(x) \geq 0$ for all $x \geq 1$. Since F is positive at $x = 1$ and non-decreasing for $x \geq 1$, we conclude that $F(x) > 0$ for all such x.

Theorem 1 *The function* $\ln x$ *grows more slowly than any positive power of x: for any* $a > 0$ *we have*

$$\lim_{x\to\infty} \frac{\ln x}{x^a} = 0.$$

Proof. We use Lemma 1 for a particular value of b. Given $a > 0$ we can choose b with $0 < b < a$. We could take $b = a/2$, for example. Then, for all $x > 0$, the previous lemma gives

$$\frac{\ln x}{x^a} < \frac{x^b}{bx^a} = \frac{x^{b-a}}{b}.$$

Since $b - a < 0$ we see that $x^{b-a} \to 0$ as $x \to \infty$.

Appendix F

Table of Standard Integrals

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

2. $\int \dfrac{dx}{x} = \log|x| + C$

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

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

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

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

7. $\int \operatorname{cosec}^2 x\,dx = -\cot x + C$

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

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

10. $\int \dfrac{dx}{\sqrt{a^2 - x^2}} = \sin^{-1}\dfrac{x}{a} + C$

11. $\int \dfrac{dx}{a^2 + x^2} = \dfrac{1}{a}\tan^{-1}\dfrac{x}{a} + C$

12. $\int \dfrac{dx}{\sqrt{x^2 + a^2}} = \sinh^{-1}\dfrac{x}{a} + C$

13. $\int \dfrac{dx}{\sqrt{x^2 - a^2}} = \cosh^{-1}\dfrac{x}{a} + C \qquad (x > a > 0)$

Appendix G

Answers to Selected Exercises

Chapter 1

1. (a) $7+i$ (c) $5-12i$ (e) $-i$

(g) $-(30/61)+(36/61)i$ (i) $(3/13)-2i/13$

(k) i (m) $(3-5i)/2$

2. $w+z=8+16i$, $z-w=2+8i$, $zw=-33+56i$, $z/w=(63+16i)/25$

3. (a) $(x^2-y^2)/(x^2+y^2)$ (c) $3x^2y-y^3$ (e) $(x^2+y^2)^3$ (g) $(x^2-y^2)/(x^2+y^2)^2$

4. (a) $-1/2$ (c) 1 (e) $8/17$

5. (a) $y=-1\pm 2i$ (c) $t=\dfrac{-1\pm\sqrt{5}}{2}$

8. z must be real.

Chapter 2

1. (a) $1, -\sqrt{3}$, $2, -\pi/3$ (c) $2, 2\sqrt{3}, 4, \pi/3$ (e) $-1, 0, 1, \pi$

2. $-4i=4\left(\cos(-\pi/2)+i\sin(-\pi/2)\right)$, $-2+2i=2\sqrt{2}\left(\cos(3\pi/4)+i\sin(3\pi/4)\right)$,

$1-i=\sqrt{2}\left(\cos(-\pi/4)+i\sin(-\pi/4)\right)$. (a) $4i$ (c) $8i$

3. (a) 1 (c) 1

4. (a) $\sqrt{2}e^{\frac{\pi}{4}i}$ (c) $2e^{-\frac{\pi}{2}i}$

5. 4096

6. $(3\pi-8)/32$

7. $\cos 6\theta=\cos^6\theta-15\cos^4\theta\sin^2\theta+15\cos^2\theta\sin^4\theta-\sin^6\theta$.

8. $z=i(\dfrac{\pi}{2}+2k\pi)$ for all $k\in\mathbb{Z}$.

9. The roots are $z=2i$ and $z=3+i$. Note that $p(z)$ is not a polynomial with real coefficients, so we do not expect the non–real roots to occur in complex conjugate pairs.

10. $2\pm 3i$, 1, -2

11. $1\pm i$, -1, 4

Chapter 3

1. (b) and (c)

2. (a) only

4. (a) $\mathbb{R}$ (c) $\mathbb{R}$ (e) $\{x \in \mathbb{R} \mid x \geq 2\}$ (g) $\mathbb{R}$

5. (a) $[0, \infty)$ (c) $[\ln 2, \infty)$ (e) $[0, 1]$ (g) $[\cos 1, 1]$

6. (a) $f \circ g$ is defined for all real x; $g \circ f$ is defined for all real x such that $x \geq 0$.

Chapter 4

1. (a) 7 (c) 0

2. (a) 3 (c) 2

3. (a) 0 (c) 0

4. (a) $3/4$ (c) -1 (e) does not exist

5. (a) 0

6. (a) Limit is 0.

7. (a) -4 (c) 3

Chapter 5

1. (a) e^{x+5} (c) $e^x(x+1)$ (e) $99(x+1)^{98}$

(g) $-\sin t e^{\cos t}$ (i) $2t \tan(1-t^2)$ (k) $\cos(\sin(\sin x) \cos(\sin x) \cos x$

2. (a) $f \circ f' = -x^2$, $f' \circ f = -x^2$ (c) $f \circ f' = 2$, $f' \circ f = 0$

3. (a) $-x^2/y^2$ (c) $\tan x \tan y$ (e) $-2\sqrt{y/x}$

4. (a) $y = 2 - x$

Chapter 6

1. (a) $21x^2 - 2 - x^{-2}$

(c) $-2x \sin(x^2) \cos(\cos(x^2))$

(d) $3x^2 \sin x + x^3 \cos x + e^{x \cos x}(\cos x - x \sin x)$

2. $y = 4x + 1$

3. (a) $9e^{-3}, 9e^3$

(c) $2, 6$

5. e^{-1}

6. 1

7. (a) 1 (c) 1 (e) 1 (g) -1

(i) $-\frac{1}{2}$

Chapter 7

1. (a) The constant approximation coinciding with $\sin x$ at $x = 0$ is identically zero. The 'error' is therefore equal to $\sin x$ itself.

(b) The tangent line at $x = 0$ is just the graph of x itself, so the error is $x - \sin x$.

2. $x + \frac{x^3}{3} + \frac{2x^5}{15} + \frac{17x^7}{315} + \frac{62x^9}{2835}$

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

4. You get the same polynomial back again, at least if you take the Taylor polynomial of degree equal to the degree of the given polynomial.

5. (a) $x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^3$ (c) $x + \frac{x^3}{3!} + \frac{x^5}{5!} + \frac{x^7}{7!}$ (e) $1 + \frac{x}{2} - \frac{x^2}{8} + \frac{x^3}{16}$

6. $(x-1)^3$

Chapter 8

2. $\frac{5}{9}$

3. $\frac{52}{99}$

4. (a) $w = \sqrt{2}e^{\frac{\pi}{4}i}$

5. (a) $zw = 2\sqrt{2}e^{-i\frac{\pi}{12}}$, $(zw)^{12} = (-2)^{18}$ (c) $e^z \approx 3.05 + 4.76i$

6. (a) $1 + \frac{1}{2}x + \frac{3\times 1}{4\times 2}x^2 + \frac{5\times 3\times 1}{6\times 4\times 2}x^3 + \frac{7\times 5\times 3\times 1}{8\times 6\times 4\times 2}x^4 + \cdots$.

(b) A series for $\sin^{-1} x$ is $x + \frac{1}{2}\frac{x^3}{3} + \frac{1.3}{2^2.2!}\frac{x^5}{5!} + \frac{1.3.5}{2^3 3!}\frac{x^7}{7} + \cdots$

Chapter 9

1. (d) Any $N > 5\times 10^5$ (e.g. $N = 500001$) works.

2. $\int_{-3}^{2}(f(x) + g(x))dx = 11$, $\int_{2}^{-3}\frac{g(x)}{2}dx = -4$.

$\int_{-3}^{2} f(x)g(x)dx$ cannot be evaluated with the given information.

Chapter 10

2. (b) $3x^2\sin^3(x^3)$

3. $\dfrac{d}{dx}\displaystyle\int_{\cos x}^{x^2} u^2 e^{u^2}\,du = 2x^5 e^{x^4} + \sin x \cos^2(x) e^{\cos^2(x)}$

Chapter 11

1. (a) $\dfrac{(\log x)^2}{2} + C$ (c) $\dfrac{2}{3}(e^x - 2)\sqrt{e^x + 1}$

2. $-\dfrac{\cos(2x)}{2}$

4. $\dfrac{1}{2}(\log(3-x) - \log(1-x))$

5. $2n I_{n+1} = (2n-1) I_n + x(1+x^2)^{-n}$

Chapter 12

1. (a) $\dfrac{\pi}{3}(2a^3 - 3a^2 b + b^3)$ (b) $\dfrac{2\pi}{3}(b^2 - a^2)^{\frac{3}{2}}$

2. 8π